5-21-2021

# Applications of Nonstandard Analysis in Probability and Measure Theory

Irfan Alam
*Louisiana State University and Agricultural and Mechanical College*

# APPLICATIONS OF NONSTANDARD ANALYSIS IN PROBABILITY AND MEASURE THEORY

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Mathematics

by
Irfan Alam
B.Math. (Hons.), Indian Statistical Institute, 2012
M.Math., Indian Statistical Institute, 2014
M.S., Louisiana State University, 2016
August 2021

This dissertation is dedicated to all those who consciously or unconsciously motivated me to be a mathematician.

There is no knowledge that is not power.

—Ralph Waldo Emerson
"Old Age" in *Society and Solitude*

## Acknowledgments

Everyone who I have had interactions with has helped shape me into the person I am today, and I wish to thank all who had a positive impact. While it would be impossible to acknowledge everyone's contributions, I would be remiss if I did not pay my dues to at least some people without whom my life would be unimaginably different.

Some of my most memorable experiences have been during my college days, both in Bangalore and in Kolkata. I spent countless hours doing all sorts of things with Ankit Rai, Anurag Rao, Ashirvad Kumar, Bhargobjyoti Saikia, Bibekananda Mishra, Chandranandan Gangopadhyay, Gautam Aishwarya, Nikhilesh Dasgupta, Priyam Biswas, Promit Kundu, Shariq Mohammed, Somnath Maity, Soumyashant Nayak, Subhadip Chowdhury, etc. It would require several pages to appropriately credit these memories and how they impacted me, so I will just say that I am grateful to all of these friends for being there for/with me over the years. I especially thank Shariq Mohammed for selflessly hosting me many times during my long research visits when he was in Connecticut.

In the US, new friends from LSU helped make sure I felt at home in this country. In particular, my transition in the first semester was helped immensely by Joshua Finkelstein with whom I would discuss math and play chess till late hours. I cherish the long-standing company of Tara Fife with whom I share too many memories to single something out. I am grateful to Erika Gerhold, Kaju Sarkar, Liem Nguyen, and Ryan Leigon, who have all provided moral support at times when I needed it the most. Mathematical discussions with Amy Peterson were helpful when we were both trying to get into research. Many senior graduate students were also incredibly warm and welcoming, and I especially

benefitted learning from the experiences of Bach Nguyen, Karthik Adimurthi, Lokendra Singh Thakur, Neal Livesay, and Paul Sinz. With Andrew Alaniz, using each other as sounding boards for ideas in the later parts of our PhD certainly helped ease some of the stress of that period. Outside LSU, I would like to thank Jacob Schirmer for helping me see things in new lights.

. I want to thank several professors from India who had a positive influence on me. First on that list is R.K. Choudhary, who was a teacher at my school. He recognized that I enjoyed mathematics and participated in discussions beyond the school curriculum after his classes. My deepest gratitude goes to B. Sury for believing in me when he was head of the B.Math. selection committee. Shashi Mohan Srivastava taught me logic, without which I might not have had the courage to take up my eventual research topic. Alok Goswami taught me perseverance and rigor that was instrumental in my growth as a mathematician. Manjunath Krishnapur inspired me to explore the beauty in probability theory.

At LSU, my PhD advisors Karl Mahlburg and Ambar Sengupta deserve a special acknowledgment for the numerous times they have supported me in all sorts of situations. They patiently listened to my unrefined ideas and always motivated me to aim high. I have discussed a lot of mathematics with Arnab Ganguly, Charles Delzell, Padmanabhan Sundar and Stephen Shipman, all of whom have been nothing but helpful. I would also like to thank Frank Neubrander and James Oxley—they have supported me immensely throughout my graduate school career, with both academic and non-academic matters. I am grateful to have found a second home in the philosophy department at LSU,

and would especially like to thank Marco Altamirano for helping me gather the confidence to study philosophy while being a full-time mathematics student. Jeffrey Roland and Jon Cogburn have been excellent philosophy mentors in these years.

Among external professors, I thank Isaac Goldbring and Steven Leth for welcoming me into the nonstandard community and helping with my job applications. I also thank Renling Jin, David Ross, and Robert Anderson for discussions that influenced my writing.

Last but not least, I am grateful to my family with whom I grew up. My pre-college life was enriched by the playful company of my sister. I will always miss the home-cooked meals of my mother whose tireless efforts ensured that I was nourished well at home. My father was very supportive of my decision to pursue mathematics and informed me of opportunities that I might not have found otherwise before I even went to college. My brother also had a big influence on my mathematical career before it was a career—a combinatorics book he gave me while I was in high school was one of the motivating factors that initially got me interested in mathematics for its own sake. I want to thank all of them for continuing to be a source of support and encouragement through various stages of my career.

## Table of Contents

## List of Figures

## Abstract

This dissertation broadly deals with two areas of probability theory and investigates how methods from nonstandard analysis may provide new perspectives in these topics. In particular, we use nonstandard analysis to prove new results in the topics of limiting spherical integrals and of exchangeability.

In the former area, our methods allow us to represent finite dimensional Gaussian measures in terms of marginals of measures on hyperfinite-dimensional spheres in a certain strong sense, thus generalizing some previously known results on Gaussian Radon transforms as limits of spherical integrals. This first area has roots in the kinetic theory of gases, which is also described.

In the latter area, we prove a new generalization of de Finetti's theorem for exchangeable random variables, a theorem important for the foundations of Bayesian statistics. In particular, we extend the de Finetti–Hewitt–Savage theorem to certain general sequences of exchangeable random variables taking values in any Hausdorff space. Under mild distributional conditions, our work expresses a sequence of exchangeable random variables taking values in any Hausdorff space as a mixture of sequences of iid random variables. Prior to this work, this result was known for random variables taking values in a Polish space. Hence, the current work has removed the need to have any assumptions on the state space, and shown that it is the underlying distribution of the random variables that is important. We prove several preparatory results in nonstandard and topological measure theory along the way, a highlight being a new generalization of Prokhorov's theorem.

# Chapter 1. Introduction

## 1.1. Brief overview of the dissertation

This dissertation is focused on two applications of nonstandard analysis in probability and measure theory—the first application concerns the connections between high-dimensional spherical integrals and Gaussian measures (which was originally covered in Alam [4, 2]), while the second application is a new generalization of a theorem of de Finetti–Hewitt–Savage on exchangeable random variables (which was originally covered in Alam [5, 3]).

In very broad terms, nonstandard analysis is a powerful framework in which for any structure on a set $S$ (consisting of atoms or urelements; that is, we view each element of $S$ as an "individual" without any structure, set-theoretic or otherwise), we have a superset ${}^*S$ that has the same first-order structural properties (this is called the *transfer principle*, or just *transfer* for brevity), but has more "ideal" elements. The set ${}^*S$ is called a nonstandard extension of $S$. Focusing on the nonstandard extension of $\mathbb{R}$, this allows a rigorous way to work with infinitesimals and infinite numbers, which provides firm foundations to several intuitive arguments in analysis that can be understood in such a language.

To give an example from combinatorial number theory, one can take a *hyperfinite* natural number $N$ (that is, $N$ is an element of the nonstandard extension ${}^*\mathbb{N}$ of $\mathbb{N}$ but not of $\mathbb{N}$ itself; we denote this by writing $N > \mathbb{N}$), and work with the so-called *Loeb measure* corresponding to the natural counting measure on the hyperfinite set $\{1, 2, \ldots, N\}$. This set contains $\mathbb{N}$ but still behaves like a finite set in certain ways (hence the term "hyperfinite"). This allows one to use measure-theoretic arguments to infer number theoretic state-

ments in the nonstandard universe, which are logically equivalent to corresponding statements in the standard universe.

Historically, nonstandard methods have been especially fruitful in probability theory where many results in discrete and continuous probability can be understood using the same intuition via a hyperfinite construction. The following excerpt from Albeverio et al. succinctly highlights this general philosophy:

> In this field hyperfinite structures play a particularly interesting and important role, combining in the same model the combinatorial aspects of the discrete theory and the analytic character of the continuous one. (Albeverio et al. [6, p. 107])

The aforementioned synthesis of combinatorial and analytic ideas is an important feature of the applications of nonstandard analysis that are presented in this dissertation. We briefly give an overview of these applications next. In our overview, we will only describe some key concepts from each chapter; refer to the introduction sections within each individual chapter to get more precise references to the results in that chapter. As stated earlier, the applications being presented in this dissertation can be roughly classified into two parts.

The first part uses properties of hyperfinite-dimensional spheres to generalize certain results on limits of spherical integrals and their connections with Gaussian measures. Aside from obtaining new standard results, this provides a new perspective in this field.

More concretely, it is well-known that for any $k \in \mathbb{N}$, the joint distribution of the first $k$ coordinates of the sphere $S^{n-1}(\sqrt{n})$ in $\mathbb{R}^n$ with center at origin and radius $\sqrt{n}$, converges to the standard Gaussian on $\mathbb{R}^k$ as $n \to \infty$. This result has roots in physics and goes back to Poincaré [80], Maxwell [74], and Boltzmann [21]. In Chapter 2, we provide a

new proof by working with hyperfinite dimensional spheres.

The main machinery in Chapter 2 is a nonstandard theory for the asymptotic behavior of integrals over varying domains in general (which we also use to give a new proof of the Riemann–Lebesgue lemma as a by-product). In nonstandard terminology, we then show that for any function $f\colon \mathbb{R}^k \to \mathbb{R}$ with finite Gaussian moment of an order larger than one, its expectation is given by a Loeb integral over a hyperfinite dimensional sphere. Some useful inequalities between high-dimensional spherical means of $f$ and its Gaussian mean are obtained in order to complete the above proof.

In Chapter 3, we use the theory developed in the previous chapter to tackle a generalization of the above asymptotic result that has origins in the works of Holmes–Sengupta [52], Sengupta [90], and Peterson–Sengupta [79, 78] on Gaussian Radon transforms. More concretely, Peterson and Sengupta proved that if a Gaussian measure has full support on a finite-dimensional Euclidean space, then the expected value of a bounded measurable function on that domain can be expressed as a limit of integrals over spheres $S^{n-1}(\sqrt{n})$ intersected with certain affine subspaces of $\mathbb{R}^n$. This allows one to realize the Gaussian Radon transform of such functions as a limit of spherical integrals. We study such limits in terms of Loeb integrals over a single hyperfinite dimensional sphere. This nonstandard geometric approach generalizes the known limiting result for bounded continuous functions to the case when the Gaussian measure is not necessarily fully supported.

The other part of this dissertation relates to exchangeability and de Finetti's theorem, a topic important for the foundations of Bayesian statistics (see, for example, Savage [88, Section 3.7], and Orbanz–Roy [76]). The original formulation of de Finetti's theorem

says that an exchangeable sequence of Bernoulli random variables is a mixture of independent and identically distributed (iid) sequences of random variables. In Chapter 4, we use combinatorial arguments to show that this probability distribution is induced by a hyperfinite sample mean. In Chapter 5, we provide a historical discussion on how de Finetti's theorem has been generalized in the literature and we set up the idea of our generalization. In Chapter 6, we prove various preparatory results from nonstandard and topological measure theory, before finally fitting together all the pieces of our generalization in Chapter 7.

Very briefly, following the work of Hewitt and Savage, de Finetti's theorem was previously known for several classes of exchangeable random variables (for instance, for Baire measurable random variables taking values in a compact Hausdorff space, and for Borel measurable random variables taking values in a Polish space). Under an assumption of the underlying common distribution being Radon, we show in Chapter 7 that de Finetti's theorem holds for a sequence of Borel measurable exchangeable random variables taking values in any Hausdorff space. This includes and generalizes the currently known versions of de Finetti's theorem. Indeed, we are now able to remove the need to have any assumptions on the state space, and show that it is the underlying distribution of the random variables that is important. We use nonstandard analysis to first study the empirical measures induced by hyperfinitely many identically distributed random variables, which leads to a proof of de Finetti's theorem in great generality while retaining the combinatorial intuition of the proof for Bernoulli random variables from Chapter 4.

The tools required in the above proof lie at the intersection of topological measure

theory and nonstandard analysis, which we present in Chapter 5. While some of this material can be viewed as a review of known results in topological measure theory (for which, Topsøe [93] is our main reference), we provide a self-contained exposition that is aided by perspectives provided from nonstandard analysis. This leads to both new proofs of known results as well as some new results. An example of a classical technique benefitting from this joint perspective is the technique of pushing down Loeb measures, which we are able to interpret as the topological operation of finding a standard measure that an internal measure is nearstandard to (with respect to the $A$-topology on the space of all Borel probability measures on a given topological space). This generalizes similar results obtained in the context of the topology of weak convergence by Anderson [13, Proposition 8.4(ii), p. 684], and by Anderson–Rashid [15, Lemma 2, p. 329] (see also Loeb [69]). This theory is useful in proving a generalization of Prokhorov's theorem as an intermediate consequence. Our generalization of Prokhorov's theorem postulates the sufficiency of uniform tightness for relative compactness of a subset of the space of Borel probability measures on *any* topological space (such a result was previously known for the space of Radon probability measures on any Hausdorff space). This version of Prokhorov's theorem is used in Chapter 6 as a key tool that allows pushing down certain internal measures on the space of all Radon probability measures on a Hausdorff space.

At the heart of our argument for the generalization of de Finetti–Hewitt–Savage theorem is a combinatorial result analogous to the approximate, finite version of de Finetti's theorem obtained by Diaconis and Freedman [32]. We establish a hyperfinite version of such a result as a part of our proof. This hyperfinite version of the result of

Diaconis and Freedman has a salient interpretation in terms of Bayes' theorem (this interpretation being fully and rigorously developed in Appendix E), which ties in nicely with the relevance of de Finetti's theorem in Bayesian statistics .

In the next section, we record some general topology and measure theory notation and conventions that will be used in the rest of the dissertation. The remainder of the present chapter provides a self-contained review of basic nonstandard methods (a significant part of this discussion is borrowed from a similar introduction in the arXiv version of Alam [4] and in Alam [5]) that will be used and referenced in the sequel.

### 1.2. General topology and measure theory notation

All measures considered in this dissertation are countably additive, and unless otherwise specified, probability measures. We will usually work with probability measures on the Borel sigma algebra $\mathcal{B}(T)$ of a topological space $T$ (thus $\mathcal{B}(T)$ is the smallest sigma algebra that contains all open subsets of $T$).

**Definition 1.2.1.** A subset of a topological space is called a $G_\delta$ *set* if it is a countable intersection of open sets. A topological space is called a $G_\delta$ *space* if all of its closed subsets are $G_\delta$ sets.

Let us recall the various notions of separation in topological spaces (for further topological background, we refer the interested reader to Kelley [60]):

($T_1$) A space $T$ is called *Fréchet* if any singleton subset of $T$ is closed.

($T_2$) A space $T$ is called *Hausdorff* if any two points in it can be separated via open sets. That is, given any two distinct points $x$ and $y$ in $T$, there exist disjoint open sets $G_1$ and $G_2$ such that $x \in G_1$ and $y \in G_2$.

($T_3$) A space $T$ is called *regular* if any closed set and a point outside that closed set can

be separated via open sets. That is, given a closed set $F \subseteq T$ and given $x \in T \backslash F$, there exist disjoint open sets $G_1$ and $G_2$ such that $x \in G_1$ and $F \subseteq G_2$.

($T_{3\frac{1}{2}}$) A space $T$ is called *completely regular* if any closed set and a point outside that closed set can be separated via some bounded real-valued function. That is, given a closed set $F \subseteq T$ and $x \in T \backslash F$, there is a continuous function $f\colon T \to [0,1]$ such that $f(x) = 0$ and $f(y) = 1$ for all $y \in F$.

($T_4$) A space $T$ is called *normal* if any two disjoint subsets of $T$ can be separated by open sets. That is, given closed sets $F_1, F_2 \subseteq T$ such that $F_1 \cap F_2 = \emptyset$, there exist disjoint open sets $G_1$ and $G_2$ such that $F_1 \subseteq G_1$ and $F_2 \subseteq G_2$.

($T_5$) A space $T$ is called *hereditarily normal* if all subsets of $T$ (under the subspace topology) are normal.

($T_6$) A space $T$ is called *perfectly normal* if it is a normal $G_\delta$ space.

We now recall the definitions of some important classes of probability measures.

**Definition 1.2.2.** For a Hausdorff space $T$, a Borel probability measure $\mu$ is called *tight* if given any $\epsilon \in \mathbb{R}_{>0}$, there is a compact subset $K_\epsilon$ such that the following holds:

$$\mu(K_\epsilon) > 1 - \epsilon. \tag{1.1}$$

An alternative way to write the above condition for tightness is the following:

$$\mu(T) = \sup\{\mu(K) : K \text{ is a compact subset of } T\}. \tag{1.2}$$

If a measure $\mu$ satisfies (1.2) with the occurrence of $T$ replaced by any Borel subset of $T$, then we call it a Radon measure. More formally we make the following definition (the second line in the equality following from the fact that we are only considering probability, and in particular finite, measures).

**Definition 1.2.3.** For a Hausdorff space $T$, a Borel probability measure $\mu$ is called *Radon*

if for each Borel set $B \in \mathcal{B}(T)$, the following holds:

$$\begin{aligned}\mu(B) &= \sup\{\mu(K) : K \subseteq B \text{ and } K \text{ is compact}\} \\ &= \inf\{\mu(G) : B \subseteq G \text{ and } G \text{ is open}\}.\end{aligned}$$

Note that the Hausdorffness of the topological space $T$ was assumed in the previous definitions so as to ensure that the compact sets appearing in them were Borel measurable (as a compact subset of any Hausdorff space is automatically closed). While not typically done (as many results do not generalize to those settings), these definitions can be made for arbitrary topological spaces if we replace the word "compact" by "closed and compact". See Schwarz [89, pp. 82-88] for more details on this generalization (Schwarz uses the phrase 'quasi-compact' instead of 'compact' in this discussion). In this dissertation, we will always have an underlying assumption of Hausdorffness of $T$ during any discussions involving tight or Radon measures.

**Remark 1.2.4.** It is clear that all Radon measures are tight. Note that any Borel probability measure on a $\sigma$-compact Hausdorff space (that is, a Hausdorff space that can be written as a countable union of compact spaces) is tight. Vakhania–Tarladze–Chobanyan [96, Proposition 3.5, p. 32] constructs a non-Radon Borel probability measure on a particular compact Hausdorff space (the construction being attributed to Dieudonné). Thus, not all tight measures are Radon.

**Definition 1.2.5.** Let $T$ be a topological space and let $\mathcal{K} \subseteq \mathcal{B}(T)$. We say that a Borel probability measure $\mu$ is *outer regular on* $\mathcal{K}$ if we have the following:

$$\mu(B) = \inf\{\mu(G) : B \subseteq G \text{ and } G \text{ is open}\} \text{ for all } B \in \mathcal{K}.$$

## 1.3. Nonstandard background

### 1.3.1. Basics of nonstandard extensions

There are many approaches to nonstandard analysis, eight of which were described in Benci–Forti–di Nasso [17]. We follow the superstructure approach, as done in Albeverio et al. [6].

In very broad terms, nonstandard analysis is a powerful framework in which for any structure on a set $S$, we have a superset ${}^*S$ that has the same first-order structural properties (this is called the *transfer principle*), but has more "ideal" elements. For instance, the set ${}^*\mathbb{R}$ has the same first-order theory as $\mathbb{R}$ in a certain model-theoretic sense, but it contains elements that are larger than all real numbers and positive elements that are smaller than all positive real numbers. More generally, a property which is expressible using finitely many symbols without quantifying over any collections of subsets of $S$ is true if and only if the same property is true of ${}^*S$. This is called the **transfer principle** (or just *transfer* for brevity). The set ${}^*S$ should contain, as a subset, ${}^*T$ for each $T \subseteq S$.

Like subsets, other mathematical objects defined on $S$ also have extensions. So, a function $f\colon S \to T$ extends to a map ${}^*f\colon {}^*S \to {}^*T$, and relations on $S$ extend to relations on ${}^*S$. Hence, there is a binary relation ${}^* <$ on ${}^*\mathbb{R}$, which we still denote by $<$ (an abuse of notation that we frequently make), and which is the same as the usual order when restricted to $\mathbb{R}$. Thus, ${}^*\mathbb{R}$ is an ordered field. Indeed all the axioms for ordered fields hold for it by transfer. The symbols in a sentence such as "$\forall x > 0\ \exists y(x = y^2)$" (which is expressing the proposition that each positive number has a square root) have new meanings in the nonstandard universe: by " $<$ ", we are now interpreting the extension of the order

on $\mathbb{R}$. Yet the sentence is true in ${}^*\mathbb{R}$ by transfer!

We shall soon see (cf. Proposition 1.3.2) that any "non-trivial" extension of $\mathbb{R}$ contains *infinite* elements (that is, those that are larger than all real numbers in absolute value), as well as *infinitesimal* elements (that is, those that are smaller than all positive real numbers in absolute value). Thus, ${}^*\mathbb{R}$ is not Archimedean. The set of finite nonstandard real numbers, denoted by ${}^*\mathbb{R}_{\text{fin}}$, is a subring of the non-Archimedean field ${}^*\mathbb{R}$. To see what went wrong, note that the following sentences formally express the Archimedean property for $\mathbb{R}$ and its transfer, respectively:

$$\forall x \in \mathbb{R} \qquad \exists\, n \in \mathbb{N}\ (n > x). \tag{1.3}$$

$$\forall x \in {}^*\mathbb{R} \qquad \exists\, n \in {}^*\mathbb{N}\ (n > x). \tag{1.4}$$

The transferred sentence (1.4) no longer expresses the Archimedean property (though it still expresses an interesting fact about ${}^*\mathbb{R}$). The issue is that we are only able to quantify over ${}^*\mathbb{N}$ (and not over $\mathbb{N}$) after transfer. To keep quantifying over $\mathbb{N}$, we would have to transfer an "infinite statement" (saying that for every $x$, either $1 > x$, or $2 > x$, or $3 > x$, or $\ldots$), which is not a valid first-order sentence.

A non-first-order property may not transfer. An example is the least upper bound principle—the set $\mathbb{N}$, viewed as a subset of ${}^*\mathbb{R}$, is bounded (by any positive infinite element), yet has no least upper bound (as any upper bound minus one is also an upper bound). The issue here is that the least upper bound property for $\mathbb{R}$ is expressed via the

second-order statement:

$$\forall A \subseteq \mathbb{R}$$

$$\langle\ [\exists x \in \mathbb{R}(\forall y \in \mathbb{R}\{(y \in A) \rightarrow (y \leq x)\})] \rightarrow$$

$$\exists z \in \mathbb{R}$$

$$\{(\forall y \in \mathbb{R}\ [(y \in A) \rightarrow (y \leq z)])$$

$$\wedge\ [\forall w \in \mathbb{R}(\forall y \in \mathbb{R}\{[(y \in A) \rightarrow (y \leq w)] \rightarrow (z \leq w)\})]\}\ \rangle.$$

One way to express this as a first-order statement is to quantify over the power-set, $\mathcal{P}(\mathbb{R})$, of $\mathbb{R}$. If our nonstandard map $^*$ was able to extend sets of subsets of $X$ as well, then the above would transfer to the following **-least upper bound property*:

$$\forall A \in {}^*\mathcal{P}(\mathbb{R})$$

$$\langle\ [\exists x \in {}^*\mathbb{R}(\forall y \in {}^*\mathbb{R}\{(y \in A) \rightarrow (y \leq x)\})] \rightarrow$$

$$\exists z \in {}^*\mathbb{R}$$

$$\{(\forall y \in {}^*\mathbb{R}\ [(y \in A) \rightarrow (y \leq z)])$$

$$\wedge\ [\forall w \in {}^*\mathbb{R}(\forall y \in {}^*\mathbb{R}\{[(y \in A) \rightarrow (y \leq w)] \rightarrow (z \leq w)\})]\}\ \rangle.$$

Notice that any quantification over a standard set was "transferred" to a quantification over the corresponding nonstandard extension of that set. The non-quantified occurrences of $\in$ in the original sentence were as relation symbols (that is, '$a \in b$' is true just in case $a$ is an element of $b$). Strictly speaking, an occurence of a relation (or function) symbol must be transferred to the nonstandard extension of that relation (function) symbol.

Thus, the second line of the transferred sentence must technically be

$$[\exists x \in {}^*\mathbb{R}(\forall y \in {}^*\mathbb{R}\{(y \,{}^*\!\in A) \to (y \,{}^*\!\leq x)\})].$$

However, as before, we suppress the $^*$ on the transferred relation symbols for better readability.

In practice, we often write informal logic sentences as long as it is clear that they can be made formal. For instance, instead of writing

$$(\forall y \in \mathbb{R}\{(y \in A) \to (y \leq x)\}),$$

one would often write '$\forall y \in A(y \leq x)$'.

The above discussion implies that $\mathbb{N}$ is not an element of ${}^*\mathcal{P}(\mathbb{R})$ (as it does not satisfy the $^*$-least upper bound property), whatever the latter object is. By the transfer of the sentence "$\forall A \in \mathcal{P}(\mathbb{R})\ \forall x \in A\ (x \in \mathbb{R})$", the object ${}^*\mathcal{P}(\mathbb{R})$ would in fact be a subset of $\mathcal{P}({}^*\mathbb{R})$. The previous example leads to the observation that ${}^*\mathcal{P}(\mathbb{R})$ is not a superset of $\mathcal{P}(\mathbb{R})$ in the literal sense. It does, however, contain as an element the extension ${}^*A$ for any $A \in \mathcal{P}(\mathbb{R})$.

In general, we fix a set $S$ consisting of atoms (that is, we view each element of $S$ as an "individual" without any structure, set-theoretic or otherwise), and extend what is called the superstructure $V(S)$ of $S$, which is defined inductively as follows (here, for any set $A$, the set $\mathcal{P}(A)$ denotes the power set of $A$):

$$
\begin{aligned}
V_0(S) &:= S, \\
V_{n+1}(S) &:= \mathcal{P}(V_n(S)) \text{ for all } n \in \mathbb{N}, \\
V(S) &:= \bigcup_{n\in\mathbb{N}} V_n(S).
\end{aligned} \tag{1.5}
$$

Choosing $S$ suitably, the superstructure $V(S)$ can be made to contain all mathematical objects relevant for a given theory. For example, if $\mathbb{R} \subseteq S$, then all collections of subsets of $\mathbb{R}$ live as objects in $V_2(S) \subseteq V(S)$. For a finite subset consisting of $k$ objects from $V_m(S)$, the ordered $k$-tuple of those objects is an element of $V_n(S)$ for some larger $n$; and hence the set of all $k$-tuples of objects in $V_m(S)$ lies as an object in $V_{n+1}(S)$. For example, if $x, y \in V_m(S)$, then the ordered pair $(x, y)$ is just the set $\{\{x\}, \{x, y\}\} \in V_{m+2}(S)$. Identifying functions and relations with their graphs, $V(S)$ also contains, if $\mathbb{R} \subseteq S$, all functions from $\mathbb{R}^n$ to $\mathbb{R}$, all relations on $\mathbb{R}^n$, etc., for all $n \in \mathbb{N}$.

We extend the superstructure $V(S)$ via a *nonstandard map*,

$$
{}^*\colon V(S) \to V({}^*S),
$$

which, by definition, is any map satisfying the following axioms:

(NS1) The transfer principle holds.

(NS2) ${}^*\alpha = \alpha$ for all $\alpha \in S$.

(NS3) $\{{}^*a : a \in A\} \subsetneq {}^*A$ for any infinite set $A \in V(S)$.

A nonstandard map may not be unique. In practice, however, we fix a standard universe $V(S)$ and a nonstandard map ${}^*$. The reader is referred to Chang–Keisler [24, Theorem 4.4.5, p. 268] or Albeverio et al. [6, Chapter 1] for a proof of the existence of a nonstandard map.

An object that belongs to ${}^*A$ for some $A \in V(S)$ is called *internal*. A useful way to understand this concept is to think that internal objects are those that inherit properties from their standard counterparts by transfer. For instance, the internal subsets of ${}^*S$ are precisely the elements of ${}^*\mathcal{P}(S)$—a (reasonable) property satisfied by all elements of $\mathcal{P}(S)$ (that is, by all subsets of $S$) will thus transfer to all internal sets. As a consequence, the class of internal sets is closed under Boolean operations such as finite unions, finite intersections, complements, etc.

**Definition 1.3.1.** For a cardinal number $\kappa$, a nonstandard extension is called *$\kappa$-saturated* if any collection of internal sets that has cardinality less than $\kappa$ and that has the finite intersection property has a non-empty intersection.

We will henceforth assume that the nonstandard extension we work with is sufficiently saturated (cf. Chang and Keisler [24, Lemma 5.1.4, p. 294 and Exercise 5.1.21, p. 305]). The next proposition shows that infinite (and infinitesimal) elements do exist in any sufficiently saturated nonstandard extension.

**Proposition 1.3.2.** *${}^*\mathbb{R}$ contains infinite as well as infinitesimal elements.*

*Proof.* By saturation, the set $\cap_{n\in\mathbb{N}}\{x \in {}^*\mathbb{R} : x > n\}$ is non-empty. It is clear that any element in this set must be infinite. The multiplicative inverse of any infinite element is infinitesimal. □

The following consequence of saturation will be useful in the sequel (see also [6, Lemma 3.1.1, p. 64]).

**Proposition 1.3.3.** *A countable union of disjoint internal sets is internal if and only if all but finitely many of them are empty.*

*Proof.* Suppose $\{A_i\}_{i\in\mathbb{N}}$ is a countable collection of disjoint internal sets. Let $A = \cup_{i\in\mathbb{N}} A_i$. If all but finitely many of the $A_i$ are empty, then $A$ being a finite union of internal sets is also internal due to transfer.

Conversely, if $A$ is internal, then $A\backslash A_i$ is internal for each $i \in \mathbb{N}$ by transfer. In that case, if all but finitely many of the $A_i$ are not empty, then the collection $\{A\backslash A_i\}_{i\in\mathbb{N}}$ would satisfy the finite intersection property. By saturation, this would lead to $\cap_{i\in\mathbb{N}}(A\backslash A_i) \neq \emptyset$, which is absurd. This completes the proof by contradiction. □

The next result says that all legitimately nonstandard natural numbers (that is, those elements of ${}^*\mathbb{N}$ that are not elements of $\mathbb{N}$) are infinite (this gives an alternative proof of Proposition 1.3.2 as well).

**Proposition 1.3.4.** *Any $N \in {}^*\mathbb{N}\backslash\mathbb{N}$ is infinite. We express this by writing $N > \mathbb{N}$.*

*Proof.* Let $N \in {}^*\mathbb{N}\backslash\mathbb{N}$. Suppose, if possible, that $N$ is finite. In particular, there exist elements of $\mathbb{N}$ that are larger than $N$. Thus the set $\{n \in \mathbb{N} : n > N\}$ is non-empty and hence has a smallest element, say $n_0$. By transfer of the fact that elements in $\mathbb{N}$ are at least one unit apart, we know that $n_0 - N \geq 1$. If $n_0 - N = 1$, then $N = n_0 - 1 \in \mathbb{N}$, a contradiction. Hence, we must have $n_0 - N \geq 2$ (by transfer of the fact that if the distance between two natural numbers is larger than one, then it is at least two). But then $n_0 - 1 \geq N + 1$ and $n_0 - 1 \in \mathbb{N}$, contradicting the minimality of $n_0$. □

As discussed earlier, the existence of an infinite element in ${}^*\mathbb{R}$ implies that the set $\mathbb{N}$ has an upper bound in ${}^*\mathbb{R}$, but it does not have a least upper bound. Since all bounded internal subsets of ${}^*\mathbb{R}$ have a least upper bound in ${}^*\mathbb{R}$ by transfer, it follows that the set $\mathbb{N}$

is not internal. The following useful result is a consequence of this fact. See also Albeverio et al. [6, Proposition 1.2.7, p.21].

**Proposition 1.3.5.** *Let $A$ be an internal set.*

*(i) [**Overflow**] If $\mathbb{N} \subseteq A$, then there is an $N > \mathbb{N}$ such that*

$$\{n \in {}^*\mathbb{N} : n \leq N\} \subseteq A.$$

*(ii) [**Underflow**] If $A$ contains all hyperfinite natural numbers, then there is an $n_0 \in \mathbb{N}$ such that ${}^*\mathbb{N}_{\geq n_0} := \{n \in {}^*\mathbb{N} : n \geq n_0\} \subseteq A$.*

*Proof.* To see (i), note that if $\mathbb{N} \subseteq A$, then the internal set $B := \{m \in {}^*\mathbb{N} : \forall k \in {}^*\mathbb{N}((k \leq m) \to (k \in A))\}$ contains $\mathbb{N}$. Since $\mathbb{N}$ is not internal, there must exist an $N \in ({}^*\mathbb{N} \backslash \mathbb{N}) \cap B$, which completes the proof.

The proof of (ii) follows similarly, using the fact that ${}^*\mathbb{N}\backslash\mathbb{N}$ is not internal. Indeed if ${}^*\mathbb{N}\backslash\mathbb{N} \subseteq A$, then the internal set $C := \{m \in {}^*\mathbb{N} : \forall k \in {}^*\mathbb{N}((k \geq m) \to (k \in A))\}$ contains ${}^*\mathbb{N}\backslash\mathbb{N}$. Since ${}^*\mathbb{N}\backslash\mathbb{N}$ is not internal, there must exist an $n_0 \in [{}^*\mathbb{N}\backslash({}^*\mathbb{N}\backslash\mathbb{N})] \cap C = \mathbb{N} \cap C$, which completes the proof. □

We have seen several examples of internal sets and functions—${}^*\mathbb{N}$, ${}^*\mathbb{R}$, ${}^*f$ (for any standard function $f$), etc. Unlike these examples, (NS3) guarantees the existence of internal objects that are not ${}^*\alpha$ for any $\alpha \in V(S)$. For instance, for any $N > \mathbb{N}$, the set $\{1, \ldots, N\}$ of the "first $N$ nonstandard natural numbers" is internal, yet it does not equal the nonstandard extension of any standard set. This set is rigorously defined as the initial segment of $N$ in ${}^*\mathbb{N}$. The fact that it is internal follows from the transfer of the following sentence:

$$\forall n \in \mathbb{N}\ \exists! A \in \mathcal{P}(\mathbb{N})\ [\forall x \in \mathbb{N}(x \in A \leftrightarrow x \leq n)].$$

For a standard set $A$, let $\mathcal{P}_{\text{fin}}(A)$ denote the collection of finite subsets of $A$. There is a function $\#\colon \mathcal{P}_{\text{fin}}(A) \to \mathbb{N} \cup \{0\}$ that counts the number of elements in each finite subset. By transfer, we have a corresponding counting function ${}^*\#\colon {}^*\mathcal{P}_{\text{fin}}(A) \to {}^*\mathbb{N} \cup \{0\}$ (which we often still denote by $\#$ by an abuse of notation) that satisfies the same first order properties as the usual counting function (for example, it satisfies the inclusion-exclusion principle). The elements of ${}^*\mathcal{P}_{\text{fin}}(A)$ are called the hyperfinite subsets of ${}^*A$. Hyperfinite sets behave like finite sets even though they are not finite in the standard sense. For instance, an internal set $H$ is hyperfinite if and only if there is an $N \in {}^*\mathbb{N}$ and an internal bijection $f\colon H \to \{1, \ldots, N\}$.

There is a "sum function" that takes any finite set of real numbers as an input and produces the sum of those real numbers. By transfer, we can thus abstractly make sense of "hyperfinite sums" (that is, the sum of hyperfinitely many nonstandard real numbers). For nonstandard real numbers $a_i$, this is the sense in which we interpret objects such as $\displaystyle\sum_{i=1}^{N} a_i$ where $N \in {}^*\mathbb{N}$ (or in general, $\displaystyle\sum_{i \in H} a_i$, where $H$ is a hyperfinite set).

The next result says that one can think of a finite nonstandard real number $z$ as having a real part, and an infinitesimal part (in fact, this real part is just $\sup\{y \in \mathbb{R} : y \leq z\}$). See Cutland [26, Theorem 2.10, p. 55] for a proof.

**Proposition 1.3.6.** *For all $z \in {}^*\mathbb{R}_{fin}$, there is a unique $x \in \mathbb{R}$ (called the standard part of $z$) such that $(z - x)$ is infinitesimal. We write $\mathbf{st}(z) = x$ or $z \approx x$.*

The next result gives a nice characterization of limit points of sequences in terms of standard parts of terms with hyperfinite indices (see Cutland [26, Theorems 3.1 and 3.3] for proofs of the two statements):

**Proposition 1.3.7.** *For a sequence of real numbers $\{a_n\}_{n\in\mathbb{N}}$, there is an extended sequence $\{a_n\}_{n\in{}^*\mathbb{N}}$ (by viewing the original sequence as a function on $\mathbb{N}$). A real number $L$ is an accumulation point of the sequence $\{a_n\}_{n\in\mathbb{N}} \iff$ there is an $N > \mathbb{N}$ such that $\mathbf{st}(a_N) = L$. Thus $\lim a_n = L \iff \mathbf{st}(a_N) = L$ for all $N > \mathbb{N}$.*

### 1.3.2. Nonstandard extensions of topological spaces

Note that, more generally, one can define the notion of standard parts for elements in the nonstandard extension of any Hausdorff space. In general, we will need a point to be *nearstandard*, instead of finite, for it to have a standard part. We develop this idea next.

Let $T$ be a topological space. For a point $y \in T$, we can think of points *infinitesimally close* to $y$ in ${}^*T$ as the set of points that lie in the nonstandard extensions of all open neighborhoods of $y$. More formally, we define:

$$\mathbf{st}^{-1}(y) := \{x \in {}^*T : x \in {}^*G \text{ for any open set } G \text{ containing } y\}. \tag{1.6}$$

The notation in (1.6) is suggestive—given a point $x \in {}^*T$, we may be interested in knowing if it is infinitesimally close to any standard point $y \in T$, in which case it would be nice to call $y$ as the standard part of $x$ (written $y = \mathbf{st}(x)$). The issue with this is that for a general topological space $T$, there is no guarantee that if a nonstandard point $x$ is *nearstandard* (that is, if there is a $y \in T$ for which $x \in \mathbf{st}^{-1}(y)$) then it is also uniquely

nearstandard to only one point of $T$. This pathological situation is remedied in Hausdorff spaces. Indeed, given two standard points $x_1$ and $x_2$ in a Hausdorff space $T$, one may separate them by open sets (say) $G_1$ and $G_2$ respectively, so that ${}^*G_1$ and ${}^*G_2$ are disjoint, thus making $\mathbf{st}^{-1}(x_1)$ and $\mathbf{st}^{-1}(x_2)$ also disjoint.

Conversely, thinking along the same lines, if the standard inverses of any two distinct points are disjoint, then those points can be separated by disjoint open sets. Thus, we have the following nonstandard characterization of Haudorffness (see also [6, Proposition 2.1.6 (i), p. 48]):

**Lemma 1.3.8.** *A topological space $T$ is Hausdorff if and only if for any distinct elements $x, y \in T$, we have $\mathbf{st}^{-1}(x) \cap \mathbf{st}^{-1}(y) = \emptyset$.*

Regardless of whether $T$ is Hausdorff or not, (1.6) allows us to naturally talk about $\mathbf{st}^{-1}(A)$ for subsets $A \subseteq T$. That is, we define:

$$\mathbf{st}^{-1}(A) := \{y \in {}^*T : y \in \mathbf{st}^{-1}(x) \text{ for some } x \in A\}. \tag{1.7}$$

Using this notation, Lemma 1.3.8 can be immediately modified to obtain the following nonstandard characterization of Hausdorffness, which will be useful in the sequel.

**Lemma 1.3.9.** *A topological space $T$ is Hausdorff if and only if for any disjoint collection $(A_i)_{i\in I}$ of subsets of $T$ (indexed by some set $I$), we have*

$$\mathbf{st}^{-1}\left(\bigsqcup_{i\in I} A_i\right) = \bigsqcup_{i\in I} \mathbf{st}^{-1}(A_i), \tag{1.8}$$

*where $\sqcup$ denotes a disjoint union.*

We define the set of *nearstandard points* of ${}^*T$ as follows:

$$\mathbf{Ns}({}^*T) := \mathbf{st}^{-1}(T).$$

Thus, by Lemma 1.3.8, if $T$ is Hausdorff then $\mathbf{st}\colon \mathbf{Ns}({}^*T) \to T$ is a well-defined map.

Using the notation in (1.7), there are succinct nonstandard characterizations of open, closed, and compact sets, which we note next (see [6, Proposition 2.1.6, p. 48], with the understanding that Albeverio et al. only use the set function $\mathbf{st}^{-1}$ when the underlying space is Hausdorff, but that is not needed for these characterizations).

**Theorem 1.3.10.** *Let $T$ be a topological space.*

*(i) A set $G \subseteq T$ is open if and only if $\mathbf{st}^{-1}(G) \subseteq {}^*G$.*

*(ii) A set $F \subseteq T$ is closed if and only if for all $x \in {}^*F \cap \mathbf{Ns}({}^*T)$, the condition $x \in \mathbf{st}^{-1}(y)$ implies that $y \in F$.*

*(iii) A set $K \subseteq T$ is compact if and only if ${}^*K \subseteq \mathbf{st}^{-1}(K)$.*

Using the standard inverse notation, we also have the following useful characterization of continuity (see, for example, Albeverio et al. [6, Proposition 1.3.3, p. 27]):

**Proposition 1.3.11.** *Let $S$ and $T$ be topological spaces, and let $f\colon S \to T$ be a function. Then $f$ is continuous at $x \in S$ if and only if ${}^*f(\mathbf{st}^{-1}(x)) \subseteq \mathbf{st}^{-1}(f(x))$.*

*Proof.* First suppose that $f$ is continuous at $x \in S$. Let $V$ be any open neighborhood of $f(x)$ in $T$. By continuity, there exists an open neighborhood $U$ of $x$ in $S$ such that $f(U) \subseteq V$. If $y \in \mathbf{st}^{-1}(x)$, then we have, by definition, $y \in {}^*U$, and hence by transfer (of the sentence '$\forall z \in U(f(z) \in V)$') we have ${}^*f(y) \in {}^*V$. Since $V$ was an arbitrary open neighborhood of $f(x)$, this shows that ${}^*f(y) \in \mathbf{st}^{-1}(f(x))$ for all $y \in \mathbf{st}^{-1}(x)$, completing the proof of the "only if" part.

Conversely, suppose that ${}^*f(\mathbf{st}^{-1}(x)) \subseteq \mathbf{st}^{-1}(f(x))$. Let $V$ be any open neighborhood of $f(x)$ in $T$. If $\tau(S)$ denotes the topology on $S$ and if $\tau(x)$ denotes the collection of

open neighborhoods of $x$, then the for each $U \in \tau(x)$, we define $\mathcal{G}_U := \{\mathfrak{U} \in {}^*\tau(S) : \mathfrak{U} \subseteq {}^*U\}$. This is nonempty as we have ${}^*U \in \mathcal{G}_U$, and by the same argument we also have that the collection $\{\mathcal{G}_U : U \in \tau(x)\}$ satisfies the finite intersection property. By saturation, we find some $\mathfrak{U} \in \cap_{U\in\tau(x)}\mathcal{G}_U$. Then, by construction, we have $\mathfrak{U} \subseteq \mathbf{st}^{-1}(x)$. As a consequence, ${}^*f(\mathfrak{U}) \subseteq \mathbf{st}^{-1}(f(x)) \subseteq {}^*V$. Thus, given this open neighborhood $V$ of $f(x)$ in $T$, the following sentence is true in the nonstandard universe:

$$\exists \mathfrak{U} \in {}^*\tau(S)\ ((x \in \mathfrak{U}) \wedge (\forall y \in \mathfrak{U}({}^*f(y) \in {}^*V))).$$

By transfer, we find an open neighborhood $U$ of $x$ such that $f(U) \subseteq V$, thus showing that $f$ is continuous at $x$. This completes the proof. □

The following technical consequence of Theorem 1.3.10 will be useful in Section 7.2.

**Lemma 1.3.12.** *Suppose $(F_i)_{i\in I}$ is a collection of closed subsets of a Hausdorff space $T$ (where $I$ is an index set). Suppose that $K := \cap_{i\in I}F_i$ is compact. Then for any open set $G$ with $K \subseteq G$, we have:*

$$ {}^*K \subseteq \left[\left(\bigcap_{i\in I} {}^*F_i\right) \cap \mathbf{Ns}({}^*T)\right] \subseteq {}^*G. \tag{1.9}$$

*Proof.* The first inclusion in (1.9) is true since ${}^*K \subseteq {}^*F_i$ for all $i \in I$ (which follows because $K \subseteq F_i$ for all $i \in I$), and since $K$ is compact (so that all elements of ${}^*K$ are nearstandard by Theorem 1.3.10(iii)). To see the second inclusion in (1.9), suppose we take $x \in \cap_{i\in I}\,({}^*F_i \cap \mathbf{Ns}({}^*T))$. Since $T$ is Hausdorff, $x \in \mathbf{Ns}({}^*T)$ has a unique standard part, say $\mathbf{st}(x) = y \in T$. Since $F_i$ is closed for each $i \in I$, it follows from the nonstandard characterization of closed sets (Theorem 1.3.10(ii)) that $y \in F_i$ for all $i \in I$. As a consequence, $y \in K \subseteq G$. Thus by the nonstandard characterization of open sets (see Theorem

1.3.10(i)), it follows that $x \in {}^*G$, thus completing the proof. $\square$

If $T$ is a topological space and $T' \subseteq T$ is viewed as a topological space under the subspace topology (thus a subset $G' \subseteq T'$ is open in $T'$ if and only if $G' = T' \cap G$ for some open subset $G$ of $T$), then there are multiple ways to interpret (1.7). There is a similar issue in general when we have two topological spaces in which we could be taking standard inverses. We will generally use '$\mathbf{st}$' and '$\mathbf{st}^{-1}$' for all such usages when the underlying topological space is clear from context. If it is not clear from context, then we mention the space in a subscript. Thus in the above situation where $T' \subseteq T$, we denote by $\mathbf{st}_T^{-1}$ and $\mathbf{st}_{T'}^{-1}$ the corresponding set functions on subsets of $T$ and $T'$ respectively. Thus, for subsets $A \subseteq T$ and $A' \subseteq T'$, we have:

$$\mathbf{st}_T^{-1}(A) = \{x \in {}^*T : \exists y \in A \text{ such that } x \in {}^*G \text{ for all open neighborhoods } G \text{ of } y \text{ in } T\},$$

and

$$\mathbf{st}_{T'}^{-1}(A') = \{x \in {}^*T : \exists y \in A' \text{ such that } x \in {}^*G' \text{ for all open neighborhoods } G' \text{ of } y \text{ in } T'\}.$$

The following useful relation is immediate from the fact that the nonstandard extension of a finite intersection of sets is the same as the intersection of the nonstandard extensions.

**Lemma 1.3.13.** *Let $T$ be a topological space and let $T' \subseteq T$ be viewed as a topological space under the subspace topology. For a subset $A \subseteq T' \subseteq T$, we have:*

$$^*T' \cap \mathbf{st}_T^{-1}(A) \subseteq \mathbf{st}_{T'}^{-1}(A).$$

### 1.3.3. Loeb measures

Let $(\mathfrak{T}, \mathcal{A}, \nu)$ be an internal probability space (that is, $\mathfrak{T}$ is an internal set, $\mathcal{A}$ is an internal algebra of subsets of $\mathfrak{T}$, and $\nu\colon \mathcal{A} \to {}^*[0,1]$ is an internal finitely additive function with $\nu(\mathfrak{T}) = 1$). There are multiple equivalent ways to define the so-called *Loeb measure* (which is a standard measure on a sigma algebra containing $\mathcal{A}$) induced by $\nu$; see Loeb [67] for the original exposition. We adopt the definition using inner and outer measures (see Albeverio et. al. [6, Remark 3.1.5, p. 66]). Formally, we define, for any $A \subseteq \mathfrak{T}$,

$$\begin{aligned} \underline{\nu}(A) &:= \sup\{\mathbf{st}(\nu(B)) : B \in \mathcal{A} \text{ and } B \subseteq A\}, \text{ and} \\ \overline{\nu}(A) &:= \inf\{\mathbf{st}(\nu(B)) : B \in \mathcal{A} \text{ and } A \subseteq B\}. \end{aligned} \tag{1.10}$$

The collection of sets for which the inner and outer measures agree form a sigma algebra called the *Loeb sigma algebra* $L(\mathcal{A})$. The common value $\underline{\nu}(A) = \overline{\nu}(A)$ in that case is defined as the Loeb measure of $A$, written $L\nu(A)$. We call $(\mathfrak{T}, L(\mathcal{A}), L\nu)$ the *Loeb space* of $(\mathfrak{T}, \mathcal{A}, \nu)$. More formally, we have:

$$L(\mathcal{A}) := \{A \subseteq \mathfrak{T} : \underline{\nu}(A) = \overline{\nu}(A)\}, \tag{1.11}$$

and

$$L\nu(A) := \underline{\nu}(A) = \overline{\nu}(A) \text{ for all } A \in L(\mathcal{A}). \tag{1.12}$$

When the internal measure $\nu$ is clear from context, we will frequently write 'Loeb measurable' (in the contexts of both sets and functions) to mean measurable with respect to the corresponding Loeb space $(\mathfrak{T}, L(\mathcal{A}), L\nu)$. Note that the Loeb sigma algebra $L(\mathcal{A})$, as defined above, depends on the original internal measure $\nu$ on $(\mathfrak{T}, \mathcal{A})$—we will use

appropriate notation such as $L_\nu(\mathcal{A})$ to indicate this dependence if there is any chance of confusion regarding the original measure inducing the Loeb sigma algebra. If we use the notation $L(\mathcal{A})$, then it is understood that a specific internal measure $\nu$ has been fixed on $(\mathfrak{T}, \mathcal{A})$ during that discussion.

There is a more abstract way of defining the Loeb measure $L\nu$ from an internal probability space $(\mathfrak{T}, \mathcal{A}, \nu)$ which is sometimes useful to think in terms of as well. We first note that $\mathbf{st}(\nu)\colon \mathcal{A} \to [0,1]$ is a finitely additive probability measure on an algebra. By Proposition 1.3.3, it follows that $\mathbf{st}(\nu)$ satisfies the premises of Carathéodory Extension Theorem. By that theorem, it extends to a standard probability measure on the smallest sigma algebra containing $\mathcal{A}$ (this is denoted by $\sigma(\mathcal{A})$). Then the Loeb measure $L\nu$ happens to be the completion of this standard measure on $(\mathfrak{T}, \sigma(\mathcal{A}))$, and $L(\mathcal{A})$ is a sigma algebra containing $\sigma(\mathcal{A})$ that arises out of this completion. Note that this construction could have been done with any finite internal measure $\nu$.

In the Loeb integration theory that will be explained next, we will use the following simplification of Ross [84, Theorem 5.1, p. 105] extensively:

**Proposition 1.3.14.** *Let $(\mathfrak{T}, L(\mathcal{A}), L\nu)$ be the Loeb probability space of $(\mathfrak{T}, \mathcal{A}, \nu)$. Suppose $F\colon \mathfrak{T} \to {}^*\mathbb{R}$ is an internal function that is measurable in the sense that $F^{-1}(B) \in \mathcal{A}$ for all $B \in {}^*\mathcal{B}(\mathbb{R})$ (where $\mathcal{B}(\mathbb{R})$ is the Borel $\sigma$-algebra on $\mathbb{R}$). If $F(x) \in {}^*\mathbb{R}_{fin}$ for $L\nu$-almost all $x \in \mathfrak{T}$, then $\mathbf{st}(F)$ is Loeb measurable (i.e., measurable as a map from $(\mathfrak{T}, L(\mathcal{A}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$).*

For any probability measure $\nu$, there is an *integral operator* $\displaystyle\int$ that takes certain functions (those in the space $L^1(\nu)$ of integrable real-valued functions on the underlying

sample space of $\nu$) to their integrals with respect to $\nu$. By transfer, for any internal probability space $(\mathfrak{T}, \mathcal{A}, \nu)$, we also have the associated space ${}^*L^1(\mathfrak{T}, \nu)$ of ${}^*$-integrable functions, and a corresponding ${}^*$-integral operator ${}^*\!\!\int$. For any ${}^*$-integrable $F \colon \mathfrak{T} \to {}^*\mathbb{R}$, one then has ${}^*\!\!\int_{\mathfrak{T}} F d\nu \in {}^*\mathbb{R}$, which we call the ${}^*$-*integral* of $F$ over $(\mathfrak{T}, \mathcal{A}, \nu)$.

The ${}^*$-integral on ${}^*L^1(\mathfrak{T})$ inherits many properties (an important one being linearity) from the ordinary integral by transfer. If $F$ is finite almost surely with respect to the corresponding Loeb measure, then $\mathbf{st}(F)$ is Loeb measurable by Proposition 1.3.14. In that case, it is interesting to study the relation between the ${}^*$-integral of $F$ and the Loeb integral of $\mathbf{st}(F)$. The following result covers this for a useful class of functions (see Ross [84, Theorem 6.2, p.110] for a proof):

**Theorem 1.3.15.** *Suppose $(\mathfrak{T}, \mathcal{A}, \nu)$ is an internal probability space and $F \in {}^*L^1(\mathfrak{T}, \nu)$ is such that $L\nu(F \in {}^*\mathbb{R}_{fin}) = 1$. Then the following are equivalent:*

*(1)* ${}^*\!\!\int_{\mathfrak{T}} |F| \, d\nu \in {}^*\mathbb{R}_{fin}$, *and*

$$\mathbf{st}\left({}^*\!\!\int_{\mathfrak{T}} |F| \, d\nu\right) = \lim_{m \to \infty} \mathbf{st}\left({}^*\!\!\int_{\mathfrak{T}} |F| \, \mathbb{1}_{\{|F| \leq m\}} d\nu\right).$$

*(2) For every $M > \mathbb{N}$, we have* $\mathbf{st}\left({}^*\!\!\int_{\mathfrak{T}} |F| \, \mathbb{1}_{\{|F| > M\}} d\nu\right) = 0.$

*(3)* ${}^*\!\!\int_{\mathfrak{T}} |F| \, d\nu \in {}^*\mathbb{R}_{fin}$; *and for any $A \in \mathcal{A}$ we have:*

$$\nu(A) \approx 0 \Rightarrow {}^*\!\!\int_{\mathfrak{T}} |F| \, \mathbb{1}_A d\nu \approx 0.$$

*(4)* $\mathbf{st}(F)$ *is Loeb integrable, and* $\mathbf{st}\left({}^*\!\!\int_{\mathfrak{T}} |F| \, d\nu\right) = \int_{\mathfrak{T}} |\mathbf{st}(F)| \, dL\nu.$

A function satisfying the conditions in Theorem 1.3.15 is called *$S$-integrable* on

$(\mathfrak{T}, \mathcal{A}, \nu)$. The notion of $S$-integrability, first developed by Anderson [11], is one of the most ubiquitous concepts in nonstandard measure theory.

Given a Loeb measurable $f\colon \mathfrak{T} \to \mathbb{R}$, a natural question to ask is whether or not it occurs as the standard part of an internal function. An internal measurable function $F : \mathfrak{T} \to {}^*\mathbb{R}$ is called a *lifting* of a Loeb measurable function $f$ if $L\nu(\mathbf{st}(F) = f) = 1$.

The following theorem shows that $^*$-integrable functions can be characterized as those possessing $S$-integrable liftings (see Ross [84, Theorem 6.4, p.111] for a proof).

**Theorem 1.3.16.** *Let $(\mathfrak{T}, \mathcal{A}, \nu)$ be an internal probability space and let $(\mathfrak{T}, L(\mathcal{A}), L\nu)$ be the associated Loeb space. Suppose $f : \mathfrak{T} \to \mathbb{R}$ is Loeb measurable. Then $f$ is Loeb integrable if and only if it has an $S$-integrable lifting.*

Using $S$-integrability, we now obtain a result that we will later use in our proof of de Finetti's theorem for Bernoulli random variables. The following result is applicable to more general situations (refer to the settings in Sections 3.4 and 3.5 of Albeverio et al. [6]). However, we restrict to compact Hausdorff spaces and real-valued functions on them for convenience.

**Theorem 1.3.17.** *Let $S$ be a compact Hausdorff space. Suppose ${}^*\mathcal{B}(S)$ is the internal algebra of $^*$-Borel subsets of $S$. Let $\nu$ be an internal (finitely additive) probability measure on $({}^*S, {}^*\mathcal{B}(S))$. Let $L\nu$ be the associated Loeb measure. Define a map $\mu\colon \mathcal{B}(S) \to [0,1]$ by:*

$$\mu(B) := L\nu(\mathbf{st}^{-1}(B)) \text{ for all } B \in \mathcal{B}(S). \tag{1.13}$$

*Then, we have:*

*(i) $\mu$ is a Radon probability measure.*

*(ii) For any nonnegative continuous function $f\colon S \to \mathbb{R}_{\geq 0}$, we have:*

$$ {}^{*}\!\!\int_{{}^*S} {}^*f d\nu \approx \int_S f d\mu. \tag{1.14}$$

*Proof.* Note that since $S$ is a compact space, we have $\mathbf{st}^{-1}(S) = {}^*S$. That $\mu$ is well-defined (that is, $\mathbf{st}^{-1}(B)$ is Loeb measurable for each $B \in \mathcal{B}(S)$) and is a Radon measure then follow from Proposition 3.4.5 and Corollary 3.4.3 in Albeverio et al. [6, pp. 88-89].

To see (ii), let $f\colon S \to \mathbb{R}_{\geq 0}$ be a nonnegative function (which is automatically bounded, as the domain is a compact space). Since $f$ is bounded, it follows that $\mathbf{st}({}^*f)$ is Loeb measurable, satisfying the following (see Proposition 1.3.14 and (2) $\Rightarrow$ (4) of Theorem 1.3.15):

$$ {}^{*}\!\!\int_{{}^*S} {}^*f d\nu \approx \int_{{}^*S} \mathbf{st}({}^*f) dL\nu. \tag{1.15}$$

Also, with $\lambda$ denoting the one-dimensional Lebesgue measure, we have (since $\mathbf{st}({}^*f)$ is nonnegative):

$$\begin{aligned} \int_{{}^*S} \mathbf{st}({}^*f) dL\nu &= \int_{(0,\infty)} L\nu \left\{x \in {}^*S : \mathbf{st}({}^*f(x)) > y\right\} d\lambda(y) \\ &= \int_{(0,\infty)} L\nu \left\{x \in {}^*S : f(\mathbf{st}(x)) > y\right\} d\lambda(y). \end{aligned} \tag{1.16}$$

We used the nonstandard characterization of continuity (i.e., that $\mathbf{st}({}^*f(x)) = {}^*f(\mathbf{st}(x))$ for all nearstandard points $x \in {}^*S$, which in our case includes all $x \in {}^*S$ since $S$ is compact) to obtain (1.16) in the above.

For $y \in (0, \infty)$, let

$$A_y := \{x \in {}^*S : f(\mathbf{st}(x)) > y\}$$

$$\text{and } B_y := \{x \in S : f(x) > y\}.$$

It is routine to verify that

$$A_y = \mathbf{st}^{-1}(B_y) \text{ for all } y \in (0, \infty). \tag{1.17}$$

Thus, (1.16) becomes:

$$\begin{aligned}
\int_{^*S} \mathbf{st}(^*f) dL\nu &= \int_{(0,\infty)} L\nu(A_y) d\lambda(y) \\
&= \int_{(0,\infty)} L\nu(\mathbf{st}^{-1}(B_y)) d\lambda(y) \\
&= \int_{^*S} \mathbf{st}(^*f) dL\nu \\
&= \int_{(0,\infty)} \mu(B_y) d\lambda(y) \\
&= \int_S f d\mu. \qquad (1.18)
\end{aligned}$$

Equations (1.15) and (1.18) complete the proof. □

We finish our review of basic nonstandard methods with the following remark about the nature of the standard universe we are extending in this dissertation.

**Remark 1.3.18.** Let $S$ be a set of urelements and let $V(S)$ be its superstructure. As discussed earlier, we fix a sufficiently saturated nonstandard extension of $V(S)$. In Chapters 2 and 3, we work with measures defined on a sequence of measure spaces, and want to construct a natural Loeb measure on any element in the nonstandard extension of such a sequence. One issue in doing so could be that the measure spaces might not all lie in a single iterated power set over $S$ (in which case, we cannot think of the sequence of measure spaces as an element of $V(S)$). In particular, this would be an issue if our measure spaces were the Borel spaces $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ and $S$ was the set of real numbers. To

get around this difficulty, we take a set $S$ that contains (copies of) $\mathbb{R}^n$ for each $n \in \mathbb{N}$ (or copies of any other standard sets that may be under consideration).

# Chapter 2. Limiting Probability Measures and Revisiting a Theorem of Boltzmann–Maxwell–Poincaré

## 2.1. Introduction

Gaussian measures have been mathematically connected with the uniform surface area measures on high-dimensional spheres since at least the time of Poincaré, who observed in [80] that if $n$ real numbers are randomly chosen under the constraint that their sum of squares equals $n$ (this is equivalent to choosing a random vector on $S^{n-1}(\sqrt{n})$, the sphere in $\mathbb{R}^n$ centered at the origin, of radius $\sqrt{n}$), then as $n \to \infty$, the probability distribution of the first number converges to that of a standard Gaussian random variable (that is, with zero mean and covariance equaling one). Considering works on the kinetic theory of gases in Physics, this connection goes back another century (we briefly outline this connection with Physics in Appendix A). We will attribute this result to Poincaré for having made the connection explicit.

For any sphere $S$ centered at the origin in a Euclidean space, there is a unique orthogonal transformation invariant probability measure $\bar{\sigma}_S$ (we will omit the subscript when the sphere under consideration is clear from context). For each $k \in \mathbb{N}$ and $n \in \mathbb{N}_{\geq k}$, let $\pi_k^{(n)} \colon \mathbb{R}^n \to \mathbb{R}^k$ denote the projection on to the first $k$ coordinates under the standard basis (we will omit the superscript when the dimension is clear from context). For a Borel set $B \subseteq \mathbb{R}^k$, we write:

$$\bar{\sigma}_{S^{n-1}(\sqrt{n})}(B) := \bar{\sigma}_{S^{n-1}(\sqrt{n})}[S^{n-1}(\sqrt{n}) \cap (\pi_k^{(n)})^{-1}(B)].$$

In the same spirit, we identify each measurable function $f \colon \mathbb{R}^k \to \mathbb{R}$ with a function on $\mathbb{R}^n$ by composing it with the projection $\pi_k^{(n)}$. This allows us to talk about integrals of

such an $f$ over domains in $\mathbb{R}^n$ for $n \in \mathbb{N}_{\geq k}$.

We let $\mu_{(k)}$ denote the standard Gaussian measure on $\mathbb{R}^k$ (again, omitting the subscript when the dimension is clear). With these conventions, we may write Poincaré's observation succinctly in terms of the following limit.

$$\lim_{n\to\infty} \bar{\sigma}_{S^{n-1}(\sqrt{n})}(B) = \mu(B) \text{ for all Borel sets } B \subseteq \mathbb{R}. \tag{2.1}$$

By standard measure theory, it is not difficult to see that the above can be rephrased in a more general form as follows. (As discussed above, the integral on the left side of (2.2) will be understood as that of the function $f \circ \pi_k^{(n)}$ for all $n \in \mathbb{N}_{>k}$.)

**Theorem 2.1.1** (Poincaré, [80])**.** *For all bounded measurable functions $f : \mathbb{R}^k \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu. \tag{2.2}$$

Similar ideas were later used by Lévy [64] to do infinite dimensional analysis, and then by Wiener [100] to construct Brownian motion. McKean [75] surveyed most of the relevant work from that period. Cutland and Ng explored these themes using nonstandard analysis (which provides the language of hyperfinite dimensional spheres) in [25]. They gave a new construction of the Wiener measure using the nonstandard machinery.

The current chapter may be considered a sequel to [25] in some sense. Indeed one of our aims is to view the above classical result (Theorem 2.1.1) as a statement about Loeb integrals on hyperfinite dimensional spheres, and obtain the same result for a larger class of functions. Toward that end, we give a new nonstandard proof of Poincaré's theorem in Section 2.2.3. A novel feature of this proof is that it does not require any explicit

integral calculations – it follows from straightforward applications of the weak law of large numbers and the definition of the uniform surface area measure on a sphere as a pushforward of a Gaussian measure. In Section 2.3, we also establish a nonstandard approach of extending such results from bounded measurable functions to other classes of functions. The general framework described in Sections 2.2 and 2.3 may be thought of as an invitation to apply nonstandard analysis to other asymptotic problems in probability and measure theory. One such application is carried out in [2] to generalize recent works of Sengupta [90] and Peterson–Sengupta [79] that connect Gaussian Radon transforms with limiting spherical integrals. This generalization is the topic of Chapter 3.

We also give a classical standard proof of Theorem 2.1.1 in Section 2.4.1 – it follows by dominated convergence theorem once the integral over the sphere is "disintegrated" properly (for example, using Sengupta [90, Proposition 4.1]). As pointed out in Remark 2.4.1, this proof of Theorem 2.1.1 does not immediately generalize to work for an arbitrary $\mu$–integrable function. The nonstandard framework of Section 3 allows one to get conditions (see Theorems 2.3.1 and 2.3.4) under which a result of the type of Theorem 2.1.1 for bounded measurable functions (over general domains) can be extended to unbounded functions. Though we do not use this terminology, the framework in Section 2.3 is similar to the framework of *graded probability spaces*, as in Hoover [54] and Keisler [59].

Aside from its application to spherical integrals, the approach of Section 2.4 is potentially useful in many other situations in which limits of integrals may be studied. A new proof of the Riemann-Lebesgue Lemma is provided (see Theorem 2.3.5) as an example of its use. Finally, in order to verify the sufficient conditions from Section 4 in the

case of spherical integrals, we also prove some inequalities between spherical means and $L^p(\mathbb{R}^k, \mu)$ norms of functions on $\mathbb{R}^k$ (see Theorem 2.4.6 and Corollary 2.4.7). Thus, the main results of this chapter can be divided into three types:

- Results viewing the limiting behavior of integrals over varying abstract domains as a single integral over a nonstandard domain.
- Inequalities between spherical integrals and Gaussian integrals.
- Applications of the results of the above types to systematically generalize Theorem 2.1.1 on limiting spherical integrals to a bigger class of functions.

### 2.1.1. Summary and motivation of our key results

Recall that for a Borel measurable function $f : \mathbb{R}^k \to \mathbb{R}$, we are interested in

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f(x_1, \ldots, x_k) d\bar{\sigma}(x_1, \ldots, x_n),$$

where we view $f$ as a function on $\mathbb{R}^n$ by first projecting the input into the first $k$ coordinates. Assuming Theorem 2.1.1, if $f$ is bounded, then we know from (2.2) that this limit is equal to the expected value of $f$ with respect to the standard Gaussian measure $\mu$ on $\mathbb{R}^k$. Since we are assuming the limiting result (2.2) for bounded functions, we have (using $\mathbb{1}_B$ to denote the indicator function of a set $B$) the following for a possibly unbounded Borel measurable function $f : \mathbb{R}^k \to \mathbb{R}$.

$$\lim_{m\to\infty} \lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f \mathbb{1}_{|f|\leq m} d\bar{\sigma} = \lim_{m\to\infty} \int_{\mathbb{R}^k} f \mathbb{1}_{|f|\leq m} d\mu = \int_{\mathbb{R}^k} f d\mu. \tag{2.3}$$

However, we wanted to find $\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma}$, which (assuming that $f$ is integrable over $S^{n-1}(\sqrt{n})$ for large $n \in \mathbb{N}$) is the same as the following:

$$\lim_{n\to\infty} \lim_{m\to\infty} \int_{S^{n-1}(\sqrt{n})} f \mathbb{1}_{|f|\leq m} d\bar{\sigma}.$$

Thus, in order to go from a result on bounded functions to a result on more general functions, we want to be able to switch the order of limits in (2.3). However, there is no general theory of switching double limits.

From the point of view of nonstandard analysis, the situation is simpler since the large–$n$ behavior of any sequence is captured in the values attained by the nonstandard extension of that sequence at hyperfinite indices. For a hyperfinite $N > \mathbb{N}$, the sphere $S^{N-1}(\sqrt{N})$ inherits a finitely additive internal probability measure from the sequence $(S^{n-1}(\sqrt{n}), \bar{\sigma}_{S^{n-1}(\sqrt{n})})_{n\in\mathbb{N}}$. The $N^{\text{th}}$ term in the nonstandard extension of the sequence $\left(\int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma}\right)_{n\in\mathbb{N}}$ is then the $^*$–integral of $^*f$ with respect to this internal measure. It turns out that the limiting integral for a general measurable function $f\colon \mathbb{R}^k \to \mathbb{R}$ exists (knowing that it exists and is equal to the Gaussian mean for bounded measurable functions) if $^*f$ is $S$–integrable over $S^{N-1}(\sqrt{N})$. In a more abstract setting, Theorem 2.3.1 essentially tells us that we can switch these limits if the tail double-limit $\lim_{m\to\infty}\lim_{n\to\infty}\int_{S^{n-1}(\sqrt{n})} |f|\, \mathbb{1}_{|f|>m} d\bar{\sigma}$ is zero. This condition of the tail double-limit being zero is just a standard reformulation of one of the equivalent conditions that ensure the $S$-integrability of $^*f$ over $S^{N-1}(\sqrt{N})$ (see (2) of Theorem 1.3.15).

A partial converse of the above result holds for nonnegative functions, which is covered in Theorem 2.3.4. Thus the set of all nonnegative functions for which the limit of spherical integrals is equal to the Gaussian integral is precisely the set of nonnegative functions for which the above tail double-limit is zero. While Theorems 2.3.1 and 2.3.4 come out of nonstandard measure theoretic considerations, we paraphrase a standard version for convenience as follows:

**Theorem 2.1.1.** *Let $(E, \mathcal{E})$ be a measure space. Let $k \in \mathbb{N}$ and for each $n \in \mathbb{N}_{>k}$, suppose $\Omega_n \subseteq E^{n'}$ for some $n' \in \mathbb{N}_{>k}$. Suppose that $\mathcal{F}_n$, the given sigma-algebra on $\Omega_n$, is induced by the product sigma-algebra $\mathcal{E}_{n'}$ on $E^{n'}$. Let $(\Omega_n, \mathcal{F}_n, \nu_n)$ be a sequence of Borel probability spaces. Let $\mathbb{P}$ be a probability measure on $(E^k, \mathcal{E}_k)$ such that $\lim_{n\to\infty} \nu_n(B) = \mathbb{P}(B)$ for any $B \in \mathcal{E}_k$. Then, for any function $f \colon E^k \to \mathbb{R}$, (1) implies (2) below.*

1. *The function $f$ is integrable on $(\Omega_n, \nu_n)$ for all large $n \in \mathbb{N}$, and furthermore:*

$$\lim_{m\to\infty} \lim_{n\to\infty} \int_{\Omega_n \cap \{|f| \geq m\}} |f| \, d\nu_n = 0.$$

2. *The function $f$ is $\mathbb{P}$-integrable and $\lim_{n\to\infty} \int_{\Omega_n} f d\nu_n = \int_{E^k} f d\mathbb{P}$.*

*Furthermore, if $f$ is assumed to be nonnegative, then the above conditions (1) and (2) are equivalent.*

The above theorem can also be interpreted more classically as a statement involving uniform integrability. While we do not focus on this aspect, it is interesting to emphasize that the nonstandard arguments using $S$–integrability thus encompass standard uniform integrability techniques.

In the case when $\Omega_n$ are the spheres $S^{n-1}(\sqrt{n})$, we verify the above double limit condition for all functions on $\mathbb{R}^k$ with a finite $(1+\epsilon)$–Gaussian moment, where $\epsilon$ is any positive real number. This allows us to extend the result in Theorem 2.1.1 to all such functions (see Theorem 2.4.8). The main step in this verification is an inequality (see Theorem 2.4.6 and Corollary 2.4.7) between sufficiently high-dimensional spherical means and $L^p(\mathbb{R}^k, \mu)$ norms of functions on $\mathbb{R}^k$, which we summarize as follows:

**Theorem 2.1.2.** *For each $p \in \mathbb{R}_{>1}$, there is a constant $C_p \in \mathbb{R}_{>0}$ such that the following*

*holds.*

$$\int_{S^{n-1}(\sqrt{n})} |g|\, d\bar{\sigma}_n \leq C_p[\mathbb{E}_\mu(|g|^p)]^{\frac{1}{p}} \textit{ for all } g \in L^p(\mathbb{R}^k, \mu) \textit{ and } n \in \mathbb{N}_{>4(k+1)q}, \tag{2.4}$$

*where* $q \in \mathbb{R}_{>0}$ *is such that* $\frac{1}{p} + \frac{1}{q} = 1$.

*Furthermore, we may replace the constant* $C_p$ *in the above inequality by a real number as close to* $1$ *as desired if* $n$ *is taken large enough (this large* $n$ *depends only on* $p \in \mathbb{R}_{>1}$ *and the desired distance of the constant from* $1$*).*

### 2.1.2. Structure of the chapter

Section 2.2 contains a nonstandard proof of Theorem 2.1.1 (carried out in Section 2.2.3), which is prefaced by some basic nonstandard measure theory that we will use and a discussion on spherical measures (alongwith their nonstandard counterparts).

Section 2.3 continues the theme by studying sequences of abstract measure spaces for which a result of the type of Poincaré is known. It gives conditions under which such results hold for more general functions, allowing us to express the limiting behavior of certain integrals by a Loeb integral on a single limiting measure space. An application that yields a new proof of the Riemann-Lebesgue lemma is carried out in Theorem 2.3.5.

In Section 2.4, we apply the results of Section 3 to the case of high-dimensional spheres, and obtain a generalization of the classical result on limits of spherical integrals to a large class of Gaussian integrable functions (see Theorem 2.4.8). Toward that end, we also obtain some useful inequalities between high-dimensional spherical means and Gaussian means (see Theorem 2.4.6 and Corollary 2.4.7).

## 2.2. A Quick nonstandard proof of Poincaré's theorem

Using the nonstandard characterization of limit points, Poincaré's theorem is essentially a statement about the Loeb measure of the fiber (in the hyperfinite-dimensional sphere $S^{N-1}(\sqrt{N})$ for $N > \mathbb{N}$) of a finite-dimensional set equaling its Gaussian measure. In a more general setting, we analyze this type of phenomenon in the next subsection. These results are routine but essential in setting up later proofs.

### 2.2.1. When a Loeb measure matches up with a standard measure on a subspace

In what follows, there will be a measure space $(E, \mathcal{E})$ such that we assume $X$ to contain copies of $E^n$ for all $n \in \mathbb{N}$. The corresponding product sigma-algebra on $E^n$ will be denoted by $\mathcal{E}_n$. Recall that we will be working with a sufficiently saturated nonstandard extension of the superstructure $V(X)$ over $X$. Let $k \in \mathbb{N}$. For $n \in \mathbb{N}_{\geq k}$, if $\Omega \in \mathcal{E}_n$ and $\nu$ is a measure on the induced sub-sigma-algebra on $\Omega$, then for any $B \in \mathcal{E}_k$, we denote $\nu(\Omega \cap (B \times E^{n-k}))$ by $\nu(B)$. Similarly, we can talk about integrating a measurable function $f \colon E^k \to \mathbb{R}$ over $\Omega$ by extending $f$ canonically to $E^n$.

**Proposition 2.2.1.** *Let $\Omega \in {}^*V(X)$ be such that $\Omega \subseteq {}^*E^N$ for some $N \in {}^*\mathbb{N}$. Let $\mathcal{E}$ be a sigma-algebra on $E$, and let $\mathcal{E}_k$ denote the corresponding product sigma-algebra on $E^k$ for each $k \in \mathbb{N}$. Let ${}^*\mathcal{E}_N$ denote the corresponding internal algebra on ${}^*E^N$ (defined by extension of the sequence $\{\mathcal{E}_k\}_{k\in\mathbb{N}}$, which is an element of $V(X)$ when viewed as a function on $\mathbb{N}$). Let $\mathcal{F}$ be the restriction of ${}^*\mathcal{E}_N$ to $\Omega$.*

*Fix $k \in \mathbb{N}$ and suppose $\mathbb{P} \in$ $\mathbf{Prob}(E^k, \mathcal{E}_k)$. Let $\nu \in {}^*\mathbf{Prob}(\Omega, \mathcal{F})$. If $L\nu$ is the*

*corresponding Loeb measure, and if $N \geq k$, then:*

$$\int_{\Omega} \mathbf{st}({}^*f) dL\nu = \int_{E^k} f d\mathbb{P} \text{ for all bounded measurable } f\colon E^k \to \mathbb{R} \tag{2.5}$$

$$\Updownarrow$$

$$L\nu({}^*B) = \mathbb{P}(B) \text{ for all } B \in \mathcal{E}_k. \tag{2.6}$$

*Proof.* If $f\colon E^k \to \mathbb{R}$ is bounded measurable, then $\mathbf{st}({}^*f)$ is Loeb measurable on $\Omega$ by Proposition 1.3.14. Hence the left side of equation (2.5) is well-defined.

The forward implication is immediate by taking $f = \mathbb{1}_B$, the indicator function of $B \in \mathcal{E}_k$. For the reverse implication, assume that $L\nu({}^*B) = \mathbb{P}(B)$ for all $B \in \mathcal{E}_k$ (that is, indicator functions of measurable sets satisfy (2.5)). The set of functions satisfying (2.5) is closed under taking finite $\mathbb{R}$-linear combinations, and hence all simple functions satisfy (2.5). Fix a bounded measurable function $f\colon E^k \to \mathbb{R}$. By standard measure theory (see, for example, Folland [40, Theorem 2.10]), there is a sequence $\{f_n\}_{n\in\mathbb{N}}$ of simple functions that converges to $f$ uniformly on $E^k$.

For $\epsilon \in \mathbb{R}_{>0}$, find $n_\epsilon \in \mathbb{N}$ such that we have the following inequality.

$$|f_n(x) - f(x)| < \epsilon \text{ for all } x \in E^k \text{ and } n \in \mathbb{N}_{\geq n_\epsilon}.$$

By transfer, for all $n \in \mathbb{N}_{\geq n_\epsilon}$, we get $|{}^*f_n(x) - {}^*f(x)| < \epsilon$ on ${}^*E^k$. Hence,

$$|\mathbf{st}({}^*f_n(x)) - \mathbf{st}({}^*f(x))| \leq \epsilon \text{ for all } n \in \mathbb{N}_{\geq n_\epsilon} \text{ and } x \in {}^*E^k.$$

As a consequence, we get:

$$\left|\int_{\Omega} \mathbf{st}({}^*f) dL\nu - \int_{\Omega} \mathbf{st}({}^*f_n) dL\nu\right| \leq \epsilon \text{ for all } n \in \mathbb{N}_{\geq n_\epsilon},$$

$$\text{that is, } \left|\int_{\Omega} \mathbf{st}({}^*f) dL\nu - \int_{E^k} f_n d\mathbb{P}\right| \leq \epsilon \text{ for all } n \in \mathbb{N}_{\geq n_\epsilon}.$$

But $\lim_{n\to\infty} \int_{E^k} f_n d\mathbb{P} = \int_{E^k} f d\mathbb{P}$, by dominated convergence theorem. Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary, this implies $\int_{\Omega} \mathbf{st}({}^*f) dL\nu = \int_{E^k} f d\mathbb{P}$, completing the proof. $\square$

The hypothesis in Proposition 2.2.1 is an abstract rendering of the premise of our central problem about limits of spherical measures. Indeed, we may think of $E$ as $\mathbb{R}$, the space $\Omega$ as the hyperfinite dimensional sphere $S^{N-1}(\sqrt{N})$ for some $N > \mathbb{N}$, and $\mathbb{P}$ as the standard Gaussian measure $\mu$. Then, (2.6) is the nonstandard characterization of (2.1), while (2.5) corresponds to (2.2). To strengthen this theme, in the next subsection, we will take a standard sequence of probability spaces and replace $\Omega$ by the $N^{\text{th}}$ term (for any $N > \mathbb{N}$) of the nonstandard extension of that sequence. We first record some useful implications of Proposition 2.2.1 below.

**Corollary 2.2.2.** *In the setting of Proposition 2.2.1, suppose* (2.5)*, and hence* (2.6)*, hold. Then*

$$L\nu(\{x \in \Omega : {}^*f(x) \in {}^*\mathbb{R}_{fin}\}) = 1 \textit{ for all measurable } f\colon E^k \to \mathbb{R}.$$

*Proof.* If $B_n := \{x \in E^k : |f(x)| < n\}$ for $n \in \mathbb{N}$, then the required probability is

$$L\nu\left(\cup_{n\in\mathbb{N}} {}^*B_n\right) = \lim_{n\to\infty} L\nu({}^*B_n) \overset{(2.6)}{=} \lim_{n\to\infty} \mathbb{P}(B_n) = 1,$$

thus completing the proof. $\square$

**Corollary 2.2.3.** *In the setting of Proposition 2.2.1, suppose* (2.5) *holds. Then, for any $\mathbb{P}$–integrable function $f\colon E^k \to \mathbb{R}$, we have that $\mathbf{st}({}^*f)$ is $L\nu$–integrable. Furthermore, we have:*

$$\int_{\Omega} |\mathbf{st}(^*f)| \, dL\nu = \int_{E^k} |f| \, d\mathbb{P},$$
$$\textit{and} \int_{\Omega} \mathbf{st}(^*f) dL\nu = \int_{E^k} f d\mathbb{P}.$$

*Proof.* We see that $\mathbf{st}(^*f)$ is Loeb measurable on $\Omega$ by Corollary 2.2.2 and Proposition 1.3.14. Also, by Corollary 2.2.2, $\mathbf{st}(^*f)\mathbb{1}_{\{|^*f|<n\}} \uparrow \mathbf{st}(^*f)$ $L\nu$–almost surely. Hence, we have:

$$\begin{aligned}
\int_{\Omega} |\mathbf{st}(^*f)| \, dL\nu &= \lim_{n\to\infty} \int_{\Omega} \mathbf{st}(^* |f|) \cdot \mathbb{1}_{\{^*|f|\leq n\}} dL\nu \\
&= \lim_{n\to\infty} \int_{E^k} |f| \cdot \mathbb{1}_{\{|f|\leq n\}} d\mathbb{P} \\
&= \int_{E^k} |f| \, d\mathbb{P} < \infty.
\end{aligned}$$

The first line follows from the monotone convergence theorem (applied on the Loeb space $(\Omega, L(\mathcal{F}), L\nu)$), the second line follows from (2.5), and the third line follows from the monotone convergence theorem (applied on the probability space $(E^k, \mathcal{E}_k, \mathbb{P})$).

Now, since $\lim_{n\to\infty} \left(\mathbf{st}(^*f) \cdot \mathbb{1}_{\{|^*f|<n\}}\right) = \mathbf{st}(^*f)$ $L\nu$-almost surely (using Corollary 2.2.2), and since $\left|\mathbf{st}(^*f) \cdot \mathbb{1}_{\{|^*f|<n\}}\right| \leq |\mathbf{st}(^*f)| \in L^1(\Omega, L\nu)$, it follows that:

$$\begin{aligned}
\int_{\Omega} \mathbf{st}(^*f) dL\nu &= \lim_{n\to\infty} \int_{\Omega} \mathbf{st}(^*f) \cdot \mathbb{1}_{\{|^*f|\leq n\}} dL\nu \\
&= \lim_{n\to\infty} \int_{E^k} f \cdot \mathbb{1}_{\{|f|\leq n\}} d\mathbb{P} \\
&= \int_{E^k} f d\mathbb{P}.
\end{aligned}$$

The first line follows from the dominated convergence theorem (applied on the Loeb space $(\Omega, L(\mathcal{F}), L\nu)$), the second line follows from (2.5), and the third line follows from the dominated convergence theorem (applied on the measure space $(E^k, \mathcal{E}_k, \mathbb{P})$). This completes the proof. □

**Corollary 2.2.4.** *In the setting of Proposition 2.2.1, the following are equivalent:*

(1) $\int_\Omega \mathbf{st}({}^*f)dL\nu = \int_{E^k} f d\mathbb{P}$ *for all bounded measurable* $f\colon E^k \to \mathbb{R}$.

(2) $L\nu({}^*B) = \mathbb{P}(B)$ *for all* $B \in \mathcal{E}_k$.

(3) $L\nu({}^*B) \leq \mathbb{P}(B)$ *for all* $B \in \mathcal{E}_k$.

(4) $L\nu({}^*B) \geq \mathbb{P}(B)$ *for all* $B \in \mathcal{E}_k$.

*Proof.* (1) $\Leftrightarrow$ (2) follows from Proposition 2.2.1. Also, (3) and (4) follow from (2) immediately. Conversely, assume (3). For any Borel set $B \subseteq E^k$, we have

$$L\nu({}^*B) \leq \mathbb{P}(B), \text{ and} \tag{2.7}$$

$$L\nu({}^*E^k \backslash {}^*B) \leq \mathbb{P}(E^k \backslash B) \Rightarrow L\nu({}^*B) \geq \mathbb{P}(B). \tag{2.8}$$

Combining (2.7) and (2.8) gives (2). The proof of (4) $\Rightarrow$ (2) is similar. $\square$

We end this subsection with the remark that if $E$ is a Hausdorff topological space equipped with its Borel sigma-algebra, and if the probability measure $\mathbb{P}$ is Radon, then (2.5) and (2.6) are both equivalent to the Loeb measure $L\nu$ agreeing with $\mathbb{P}$ on the nonstandard extensions of all open (or all compact) subsets of $E$.

**Proposition 2.2.5.** *In the setting of Proposition 2.2.1, suppose $E$ is a Hausdorff topological space and let $\mathcal{B}(E^k)$ be the Borel sigma-algebra on $E^k$. If $\mathbb{P}$ is a Radon probability measure on $E^k$, then the following are equivalent:*

(1) $\int_\Omega \mathbf{st}({}^*f)dL\nu = \int_{E^k} f d\mathbb{P}$ *for all bounded Borel measurable* $f\colon E^k \to \mathbb{R}$.

(2) $L\nu({}^*B) = \mathbb{P}(B)$ *for all* $B \in \mathcal{B}(E^k)$.

(3) $L\nu({}^*B) \leq \mathbb{P}(B)$ *for all* $B \in \mathcal{B}(E^k)$.

(4) $L\nu({}^*B) \geq \mathbb{P}(B)$ *for all* $B \in \mathcal{B}(E^k)$.

(5) $L\nu({}^*O) = \mathbb{P}(O)$ *for all open sets* $O \subseteq E^k$.

(6) $L\nu({}^*C) = \mathbb{P}(C)$ *for all compact sets* $C \subseteq E^k$.

*Proof.* The equivalence of (1), (2), (3), and (4) has been established without any conditions on $\mathbb{P}$ in the previous corollary. Also, (2) $\Rightarrow$ (5) is immediate. To complete the proof, we will show that (5) $\Rightarrow$ (6) and (6) $\Rightarrow$ (4).

To see (5) $\Rightarrow$ (6), note that if $C$ is a compact subset of the Hausdorff space $E^k$, then $C$ is closed, so that the subset $O := E^k \backslash C$ is open. By using the fact that ${}^*C = {}^*E^k \backslash {}^*O$, and then applying (5) to $O$, we obtain the following:

$$L\nu({}^*C) = 1 - L\nu({}^*O) = 1 - \mathbb{P}(O) = \mathbb{P}(C).$$

We now prove (6) $\Rightarrow$ (4). To that end, take any $B \in \mathcal{B}(E^k)$. For any compact subset $C \subseteq B$, we have ${}^*C \subseteq {}^*B$, so that (6) implies the following:

$$L\nu({}^*B) \geq L\nu({}^*C) = \mathbb{P}(C) \text{ for all compact subsets } C \text{ of } B.$$

Taking supremum over all compact subsets of $B$ and using the fact that the measure $\mathbb{P}$ is Radon, we thus obtain the desired inequality as follows:

$$L\nu({}^*B) \geq \sup\{\mathbb{P}(C) : C \text{ is a compact subset of } B\} = \mathbb{P}(B).$$

$\square$

### 2.2.2. Basic facts about surface area measures and their nonstandard counterparts

In this subsection, we review three different ways to think about the uniform surface area measure on spheres in Euclidean spaces. One aim of our review is to explain the

corresponding internal probability measures on hyperfinite dimensional spheres that we obtain by transfer. We refer to Matilla [73, Chapter 3] and Sengupta [90, Section 4] for basic properties of spherical surface area measures.

For each $n \in \mathbb{N}$, we let $\mathcal{B}_n = \mathcal{B}(\mathbb{R}^n)$, the Borel sigma-algebra on $\mathbb{R}^n$, and $O(n)$ be the set of all orthogonal linear transformations of $\mathbb{R}^n$. Let $\mathcal{S}_0$ be the set of all spheres centered at the origin (in any dimension $n \in \mathbb{N}$ and of any radius $r \in \mathbb{R}_{>0}$). Consider the function $\dim\colon \mathcal{S}_0 \to \mathbb{N}$ that takes each sphere $S$ to the smallest dimension $n \in \mathbb{N}$ such that $S \subseteq \mathbb{R}^n$. We are being pedantic about the "smallest dimension" since we have been identifying (during discussions on measures of sets) a subset $S$ of $\mathbb{R}^n$ with the subset $S \times \mathbb{R}^{n'-n} \subseteq \mathbb{R}^{n'}$ for $n' \in \mathbb{N}_{>n'}$.

It is known that there is a unique rotation-preserving probability measure on any sphere centered at origin equipped with its Borel sigma-algebra. More formally:

$$\forall S \in \mathcal{S}_0 \ \exists! \bar{\sigma} \in \mathbf{Prob}(S, \mathcal{B}(S)) \ \forall n \in \mathbb{N}$$
$$(n = \dim(S)) \to \ (\forall R \in O(n) \ \forall A \in \mathcal{B}(S) \ [\bar{\sigma}(R(A)) = \bar{\sigma}(A)]). \tag{2.9}$$

For any $S \in {}^*\mathcal{S}_0$ in the nonstandard universe, the transfer principle implies that the set ${}^*\mathrm{Prob}(S, {}^*\mathcal{B}(S))$ consists of a unique finitely additive internal function, say $\bar{\sigma}_S\colon {}^*\mathcal{B}(S) \to {}^*[0,1]$, that is ${}^*$–rotation preserving and satisfying $\bar{\sigma}_S(S) = 1$. By the usual Loeb measure construction, we get $L\bar{\sigma}_S$ on $\mathcal{L}({}^*\mathcal{B}(S))$ (a sigma-algebra containing $\sigma({}^*\mathcal{B}(\mathcal{S}))$, which we call the *uniform Loeb surface measure* on $S$. As before, we will often drop the superscript $S$ in $\bar{\sigma}_S$ when the sphere is clear from context.

In finite dimensions, we also have the notion of surface area. For the sphere $S :=$

$S^d(R)$ of radius $R \in \mathbb{R}_{>0}$, centered at the origin in $\mathbb{R}^{d+1}$, one can consider the surface area map $\sigma_S \colon \mathcal{B}(S) \to \mathbb{R}$, which satisfies the following volume-of-cone formula:

$$\lambda_{d+1}\left(\cup_{0\leq t\leq 1} tA\right) = \frac{1}{d+1} R\sigma_S(A),$$

where $\lambda_{d+1}$ is the Lebesgue measure on $\mathbb{R}^{d+1}$, and $A \in \mathcal{B}(S)$. This surface area function has the following properties:

- For any $d \in \mathbb{N}$ and any $R \in \mathbb{R}_{>0}$, we have $\sigma_{S^d(R)}(S^d(R)) = c_d \cdot R^d$, where $c_d = \sigma_{S^d(1)}(S^d(1)) = (d+1) \cdot \dfrac{\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}+1\right)} = 2\dfrac{\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)}$.
- For any $S \in \mathcal{S}$ and any $A \in \mathcal{B}(\mathcal{S})$, we have $\bar{\sigma}_S(A) = \dfrac{\sigma_S(A)}{\sigma_S(S)}$, where $\bar{\sigma}_S$ is the rotation preserving probability measure on $S$, as in (2.9).

By transfer, we have the notion of $^*$-surface area (that is applicable to hyperfinite-dimensional spheres as well) in the nonstandard universe. This could be used as an alternative way to define the uniform Loeb surface measure.

Yet another way to arrive at the uniform surface area measure on a sphere is by looking at an appropriate pushforward of a Gaussian measure. If $\mu$ is the standard Gaussian measure on $\mathbb{R}^n$ (here $n \in \mathbb{N}$), and $S^{n-1}$ is the unit sphere in $\mathbb{R}^n$, then the rotation invariance of $\mu$ implies that $\mu \circ g^{-1}$ is a rotation invariant probability measure on $S^{n-1}$ (and hence is the same as $\bar{\sigma}$), where

$$g : \mathbb{R}^n \backslash \{0\} \to S^{n-1} \text{ defined by } g(x) = \frac{x}{||x||}.$$

For spheres centered at origin but having radius $R \in \mathbb{R}_{>0}$, we can use the pushforward through the map $Rg$ (this is scalar multiple by $R$). For instance, for $N > \mathbb{N}$, if $\bar{\sigma}$ is the internal uniform surface area measure on $S^{N-1}(\sqrt{N})$ and $\mu_{(N)}$ is the internal Gaussian mea-

sure on ${}^*\mathbb{R}^N$ with mean $\mathbf{0}$ and covariance identity, then for any set $B \in {}^*\mathcal{B}(S^{N-1}(\sqrt{N}))$, we have:

$$\bar{\sigma}(B) = \mu_{(N)}\left(\left\{x \in {}^*\mathbb{R}^N : \frac{\sqrt{N}x}{||x||} \in B\right\}\right). \tag{2.10}$$

This characterization of the uniform surface area measure yields the classical result of Poincaré (Theorem 2.1.1) without doing any computations. We show that in the next subsection.

### 2.2.3. A nonstandard proof of Poincaré's theorem

Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, and $(X_n)_{n\in\mathbb{N}}$ is a sequence of iid $\mathcal{N}(0,1)$ random variables (that is, the $X_i$ are independent Gaussian random variables with mean 0 and variance 1). In that case, $({X_n}^2 - 1)_{n\in\mathbb{N}}$ is an iid sequence of random variables with mean zero and finite variance (in fact, the variance is equal to one). Hence the weak law of large numbers implies the following:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{({X_1}^2 - 1) + \ldots + ({X_n}^2 - 1)}{n}\right| > \epsilon\right) = 0 \text{ for all } \epsilon \in \mathbb{R}_{>0}. \tag{2.11}$$

Each $X_i$ (where $i \in \mathbb{N}$), as a function from $\Omega$ to $\mathbb{R}$, has a nonstandard extension ${}^*X_i$, which, by transfer, is a ${}^*\mathcal{N}(0,1)$ random variable, that is, ${}^*\mathbb{P} \circ {}^*{X_i}^{-1}$ is the same as the internal measure ${}^*\mu_{(1)}$ (the nonstandard extension of the standard Gaussian measure $\mu_{(1)}$ on $\mathbb{R}$).

Consider the function $X \colon \mathbb{N} \times \Omega \to \mathbb{R}$ defined by:

$$X(n,\omega) := X_n(\omega) \text{ for all } n \in \mathbb{N}, \text{ and } \omega \in \Omega. \tag{2.12}$$

Considering the nonstandard extension of $X$, we see that

$$^*X(n,\omega) = {}^*X_n(\omega) \text{ for all } n \in \mathbb{N}, \text{ and } \omega \in {}^*\Omega.$$

Since $^*X$ is a function from $^*\mathbb{N} \times {}^*\Omega$, this allows us to naturally talk about the $N^{\text{th}}$ element of the original sequence of random variables for any $N \in {}^*\mathbb{N}$ (and all those elements will be independent and internally Gaussian distributed with mean 0 and variance 1). In the sequel, we will often be loose with notation, and use $X_i$ as both a standard and a nonstandard random variable (when it is considered as a nonstandard random variable, it is understood to be given by the nonstandard extension of the map $X \colon \mathbb{N} \times \Omega \to \mathbb{R}$), with the usage being clear from context.

For the rest of this section, fix $N > \mathbb{N}$. Let $\bar{\sigma}$ be the internal uniform surface area measure on $S^{N-1}(\sqrt{N})$. Let $Y = (X_1)^2 + \ldots + (X_N)^2$.

**Lemma 2.2.6.** *There exists an infinitesimal $\xi > 0$ such that*

$$^*\mathbb{P}\left(\left|\frac{Y}{N} - 1\right| > \xi\right) \approx 0. \tag{2.13}$$

*Proof.* Consider $\epsilon \in \mathbb{R}$ such that $0 < \epsilon < 1$. Then we have the following:

$$^*\mathbb{P}\left(\left|\frac{Y}{N} - 1\right| > \epsilon\right) = {}^*\mathbb{P}\left(\left|\frac{({X_1}^2 - 1) + \ldots + ({X_N}^2 - 1)}{N}\right| > \epsilon\right),$$

where the right side is infinitesimal by the nonstandard characterization of limits applied to (2.11). The lemma now follows by underflow applied to the following internal set.

$$\left\{\epsilon \in {}^*\mathbb{R}_{>0} : {}^*\mathbb{P}\left(\left|\frac{Y}{N} - 1\right| > \epsilon\right) < \epsilon\right\}.$$

$\square$

For a set $S \subseteq \mathbb{R}^k$ and a real number $\alpha \in \mathbb{R}$, the set $\alpha S$ is the set of all scalar products (of elements of $S$) by $\alpha$. That is,

$$\alpha S := \{y \in \mathbb{R}^k : y = \alpha x \text{ for some } \alpha \in A\}.$$

For $S \subseteq \mathbb{R}^k$ and $A \subseteq \mathbb{R}$, the set $AS$ is defined as the set of all scalar products of elements of $S$ with elements in $A$. That is,

$$AS := \cup_{\alpha \in A} \alpha S. \tag{2.14}$$

Scalar products (with elements of ${}^*\mathbb{R}$ or with internal subsets of ${}^*\mathbb{R}$) are analogously defined in the nonstandard universe by transfer. We note the following elementary fact about small scalings of compact sets that will be useful in the sequel.

**Lemma 2.2.7.** *Let $C$ be a compact subset of $\mathbb{R}^k$. Then we have:*

$$\bigcap_{n \in \mathbb{N}_{>1}} \left[1 - \frac{1}{n}, 1 + \frac{1}{n}\right] C = C. \tag{2.15}$$

*Proof.* Let the left side of (2.15) be called $\tilde{C}$ for brevity. It is clear that $C \subseteq \tilde{C}$. To show the inclusion from the other side, consider $x \in \tilde{C}$. Thus for each $n \in \mathbb{N}_{>1}$, there exist $\alpha_n \in \mathbb{R}$ and $y_n \in C$ such that $x = \alpha_n y_n$. By the sequential compactness of $C$, find a subsequence $(n_k)_{k \in \mathbb{N}}$ such that $\lim_{k \to \infty} y_{n_k}$ exists as an element of $C$. Say, $\lim_{k \to \infty} y_{n_k} = y \in C$. Note that, by construction, we have $\lim_{k \to \infty} \alpha_{n_k} = 1$. By continuity of the scalar product map, we thus have the following:

$$x = \lim_{k \to \infty} \alpha_{n_k} y_{n_k} = \left(\lim_{k \to \infty} \alpha_{n_k}\right) \left(\lim_{k \to \infty} y_{n_k}\right) = y \in C, \tag{2.16}$$

completing the proof. $\square$

We now prove Poincaré's theorem that we restate here for convenience:

**Theorem 2.1.1** (Poincaré, [80]). *For all bounded measurable functions $f : \mathbb{R}^k \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu. \tag{2.2}$$

*Proof.* Let $B$ be a Borel subset of $\mathbb{R}^k$ and let $\mathbf{X} = (X_1, \ldots, X_N)$ be as defined in (2.12). For $k \in \mathbb{N}$, let $\mathbf{X}_{(k)}$ be the projection $(X_1, \ldots, X_k)$ onto ${}^*\mathbb{R}^k$. Using (2.10) and taking standard parts on both sides yields the following:

$$\begin{aligned} L\bar{\sigma}(\{(x_1 \ldots, x_N) \in S^{N-1}(\sqrt{N}) : (x_1, \ldots, x_k) \in {}^*B\}) &= L^*\mathbb{P}\left(\frac{\sqrt{N}\mathbf{X}_{(k)}}{\sqrt{Y}} \in {}^*B\right) \\ &= L^*\mathbb{P}\left(\mathbf{X}_{(k)} \in \sqrt{\frac{Y}{N}}\,{}^*B\right). \end{aligned} \tag{2.17}$$

Using Lemma 2.2.6, the last expression is less than or equal to

$$L^*\mathbb{P}\left(\mathbf{X}_{(k)} \in \bigcap_{n=2}^{m}{}^*\left[1-\frac{1}{n}, 1+\frac{1}{n}\right]B\right) = L^*\mathbb{P}\left(\mathbf{X}_{(k)} \in {}^*\left(\bigcap_{n=2}^{m}\left[1-\frac{1}{n}, 1+\frac{1}{n}\right]B\right)\right)$$

for all $m \in \mathbb{N}$.

Taking limits as $m \to \infty$, we obtain:

$$\begin{aligned} &L\bar{\sigma}(\{(x_1 \ldots, x_N) \in S^{N-1}(\sqrt{N}) : (x_1, \ldots, x_k) \in {}^*B\}) \\ &\le \lim_{m\to\infty} L^*\mathbb{P}\left(\mathbf{X}_{(k)} \in {}^*\left[\bigcap_{n=2}^{m}\left[1-\frac{1}{n}, 1+\frac{1}{n}\right]B\right)\right) \\ &= \lim_{m\to\infty} \mathbb{P}\left(\mathbf{X}_{(k)} \in \bigcap_{n=2}^{m}\left[1-\frac{1}{n}, 1+\frac{1}{n}\right]B\right) \\ &= \mathbb{P}\left(\mathbf{X}_{(k)} \in \bigcap_{n\in\mathbb{N}_{>1}}\left[1-\frac{1}{n}, 1+\frac{1}{n}\right]B\right). \end{aligned} \tag{2.18}$$

By (2.18) and Lemma 2.2.7, we have the following inequality:

$$L\bar{\sigma}(\{(x_1 \ldots, x_N) \in S^{N-1}(\sqrt{N}) : (x_1, \ldots, x_k) \in {}^*C\}) \leq \mathbb{P}\left(\mathbf{X}_{(k)} \in C\right) = \mu_{(k)}(C)$$

$$\text{for all compact subsets } C \subseteq \mathbb{R}^k.$$

Since $N > \mathbb{N}$ is arbitrary and $\mu_{(k)}$ is a Radon measure, Proposition 2.2.5 and the nonstandard characterization of limits complete the proof. □

## 2.3. On the limiting behavior of a sequence of probability spaces

Toward the proof of Poincaré's theorem in the previous section, we showed that for an arbitrary $N > \mathbb{N}$, the surface area measure over $S^{N-1}(\sqrt{N})$ (which may be thought of as the $N^{\text{th}}$ element of the sequence of of spheres $(S^{n-1}(\sqrt{n}))_{n\in\mathbb{N}}$) assigns the same measure (up to infinitesimals) to fibers of finite dimensional sets as the Gaussian measures of such sets (in their respective ambient Euclidean spaces). This idea is explored in more abstract settings in the current section in order to generalize to limiting results for integrals of unbounded functions.

### 2.3.1. Integrating finite dimensional functions along nice sequences of probability spaces

Let $\{(\Omega_n, \mathcal{F}_n, \nu_n)\}_{n\in\mathbb{N}}$ be a sequence of probability spaces. Viewing the sequence as a function on $\mathbb{N}$, we get an internal probability space $(\Omega_N, \mathcal{F}_N, \nu_N)$ for each $N > \mathbb{N}$. Note that we have been dropping the $^*$ when it is clear from context that the index $N$ is hyperfinite. Philosophically, the Loeb space $(\Omega_N, L(\mathcal{F}_N), L\nu_N)$ for $N > \mathbb{N}$ should capture the long-term behavior of the sequence $\{(\Omega_n, \mathcal{F}_n, \nu_n)\}_{n\in\mathbb{N}}$ of probability spaces. We will often omit the sigma-algebra when there is no chance of confusion. Drawing inspiration

from Theorem 1.3.15(4), we obtain the following theorem in this regard.

**Theorem 2.3.1.** *Let $(E, \mathcal{E})$ be a measure space. Let $k \in \mathbb{N}$ and for each $n \in \mathbb{N}_{>k}$, suppose $\Omega_n \subseteq E^{n'}$ for some $n' \in \mathbb{N}_{>k}$. Suppose that $\mathcal{F}_n$, the given sigma-algebra on $\Omega_n$, is induced by the product sigma-algebra $\mathcal{E}_{n'}$ on $E^{n'}$. Let $(\Omega_n, \mathcal{F}_n, \nu_n)$ be a sequence of Borel probability spaces. Let $f\colon E^k \to \mathbb{R}$ satisfy*

$$\lim_{m\to\infty} \lim_{n\to\infty} \int_{\Omega_n \cap \{|f| \geq m\}} |f| \, d\nu_n = 0. \tag{2.19}$$

*Then, $f$ is integrable over $\Omega_n$ for large $n$, so that the sequence $\alpha_{f,n} := \int_{\Omega_n} f d\nu_n$ is well-defined for large $n$. Furthermore, for any $N > \mathbb{N}$, the function $\mathbf{st}({}^*f)$ is Loeb integrable over $(\Omega_N, L(\mathcal{F}_N), L\nu_N)$ and satisfies*

$$\mathbf{st}(\alpha_{f,N}) = \int_{\Omega_N} \mathbf{st}({}^*f) dL\nu_N.$$

**Remark 2.3.2.** Bounded measurable functions trivially satisfy the hypothesis in (2.19).

*Proof.* For a fixed $\epsilon \in \mathbb{R}_{>0}$, there exists $\ell_\epsilon \in \mathbb{N}$ such that the following holds: for any $m \geq \ell_\epsilon$, there is an $n_{\epsilon,m} \in \mathbb{N}$ such that for all $n \geq n_{\epsilon,m}$, we have

$$\int_{\Omega_n \cap \{|f| \geq m\}} |f| \, d\nu_n < \epsilon. \tag{2.20}$$

In particular, $f$ is integrable on $\Omega_n$ for all $n > n_{\epsilon,\ell_\epsilon}$, with the integral of the absolute value being at most $(\ell_\epsilon + \epsilon)$. Further, for any $M, N > \mathbb{N}$, transfer yields

$${}^*\!\!\int_{\Omega_N} |{}^*f| \, \mathbb{1}_{\{|{}^*f| > M\}} d\nu_N \leq {}^*\!\!\int_{\Omega_N} |{}^*f| \, \mathbb{1}_{\{|{}^*f| > \ell_\epsilon\}} d\nu_N < \epsilon \text{ for all } \epsilon \in \mathbb{R}_{>0}.$$

Given $N > \mathbb{N}$, ${}^*f$ is $S$–integrable on $\Omega_N$ by Theorem 1.3.15(2).

Now, $\alpha_{f,N}$ is the ${}^*$–integral of ${}^*f$ over $(\Omega_N, \nu_N)$ by transfer. Note that

$$f = f_+ - f_-,$$

where $f_+ := \max\{f, 0\}$ and $f_- := \max\{-f, 0\}$. By transfer, we then have:

$$\alpha_{f,N} = \alpha_{f_+,N} - \alpha_{f_-,N}. \tag{2.21}$$

Since ${}^*f$ is $S$–integrable on $(\Omega_N, \nu_N)$, so are ${}^*f_+$ and ${}^*f_-$ (this is because $|{}^*f_+|$ and $|{}^*f_-|$ are at most equal to $|{}^*f|$). Since ${}^*f_+$ and ${}^*f_-$ are nonnegative functions, Theorem 1.3.15(4) implies:

$$\begin{aligned} \alpha_{f_+,N} &= \int_{\Omega_N} \mathbf{st}({}^*f_+) dL\nu_N, \\ \text{and } \alpha_{f_-,N} &= \int_{\Omega_N} \mathbf{st}({}^*f_-) dL\nu_N. \end{aligned} \tag{2.22}$$

Using this in (2.21) and then using the fact that $\mathbf{st}({}^*f)$ is Loeb integrable completes the proof. □

**Corollary 2.3.3.** *Let $(E, \mathcal{E})$ be a measure space. Let $k \in \mathbb{N}$ and for each $n \in \mathbb{N}_{>k}$, suppose $\Omega_n \subseteq E^{n'}$ for some $n' \in \mathbb{N}_{>k}$. Suppose that $\mathcal{F}_n$, the given sigma-algebra on $\Omega_n$, is induced by the product sigma-algebra $\mathcal{E}_{n'}$ on $E^{n'}$. Let $(\Omega_n, \mathcal{F}_n, \nu_n)$ be a sequence of Borel probability spaces. Let $\mathbb{P}$ be a probability measure on $(E^k, \mathcal{E}_k)$ such that $L\nu_N({}^*B) = \mathbb{P}(B)$ for any $B \in \mathcal{E}_k$ and $N > \mathbb{N}$.*

*(i) If $f \colon E^k \to \mathbb{R}$ is measurable, then*
$$L\nu_N(\{x \in \Omega_N : {}^*f(x) \in {}^*\mathbb{R}_{fin}\}) = 1 \textit{ for all } N > \mathbb{N}.$$

*(ii) If $f \colon E^k \to \mathbb{R}$ is bounded and measurable, then*
$$\lim_{n\to\infty} \int_{\Omega_n} f d\nu_n = \int_{E^k} f d\mathbb{P} = \int_{\Omega_N} \mathbf{st}({}^*f) dL\nu_N \textit{ for all } N > \mathbb{N}.$$

*(iii) If $f \colon E^k \to \mathbb{R}$ is $\mathbb{P}$–integrable, then we have that $\mathbf{st}({}^*f)$ is $L\nu_N$–integrable for all $N > \mathbb{N}$. Furthermore, for any $N > \mathbb{N}$, we have:*
$$\int_{E^k} f d\mathbb{P} = \int_{\Omega_N} \mathbf{st}({}^*f) dL\nu_N, \textit{ and } \int_{E^k} |f|\, d\mathbb{P} = \int_{\Omega_N} |\mathbf{st}({}^*f)|\, dL\nu_N$$

*Proof.* (i) follows from Corollary 2.2.2. (ii) follows from Theorem 2.3.1, Corollary 2.2.3 and the nonstandard characterization of limits. Finally, (iii) follows from Corollary 2.2.3, completing the proof. $\square$

Note that Corollary 2.3.3(iii) allows us to express the expected value of a $\mathbb{P}$–integrable function $f\colon E^k \to \mathbb{R}$ as the Loeb integral of $\mathbf{st}({}^*f)$ over $\Omega_N$ for all hyperfinite $N$. However, this does not necessarily imply that the sequence $\alpha_{f,n} := \int_{\Omega_n} f d\nu_n$ converges to $\int_{E^k} f d\mathbb{P}$, as $\alpha_{f,N}$ may not be infinitesimally close to the Loeb integral of $\mathbf{st}({}^*f)$ over $\Omega_N$ in general. To see a counterexample, consider $(E, \mathcal{E}) = (\mathbb{N}_0, \mathcal{P}(\mathbb{N}))$ (where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$), with $\Omega_n := \{0, n\}$ for each $n \in \mathbb{N}$. Define $\mathbb{P} := \mathbb{1}_{\{0\}}$, the probability measure concentrated at 0. Define $\nu_n(\{0\}) = 1 - \frac{1}{n}$ and $\nu_n(\{n\}) = \frac{1}{n}$. Then for any $N > \mathbb{N}$, the Loeb measure $L\nu_N$ assigns full mass to $\{0\}$. Thus the hypotheses of Corollary 2.3.3 are satisfied. Consider the measurable function $f\colon \mathbb{N}_0 \to \mathbb{R}$ defined by $f(n) := n$ for all $n \in \mathbb{N}$. It is clear that $\alpha_{f,N}$ equals 1 while the Loeb integral of $\mathbf{st}({}^*f)$ equals 0.

In view of Theorem 1.3.15, the correct criterion needed for $\alpha_{f,N}$ to be infinitesimally close to the Loeb integral of $\mathbf{st}({}^*f)$ over $\Omega_N$ for nonnegative functions $f$ is the $S$–integrability of ${}^*f$ over $\Omega_N$. This also means that the sufficient criterion (2.19) in Theorem 2.3.1 is necessary if we restrict to nonnegative functions. We record and prove these observations in the following theorem.

**Theorem 2.3.4.** *In the setting of Corollary 2.3.3, the following are equivalent for a nonnegative function $f\colon E^k \to \mathbb{R}_{\geq 0}$:*

1. *$f$ is $\mathbb{P}$–integrable and $\lim_{n\to\infty} \int_{\Omega_n} f d\nu_n = \int_{E^k} f d\mathbb{P}$.*

2. *The nonstandard extension* $^*f$ *is* $S$*–integrable on* $\Omega_N$ *for all* $N > \mathbb{N}$.

3. *The function* $f$ *is integrable on* $(\Omega_n, \nu_n)$ *for all large* $n \in \mathbb{N}$, *and furthermore:*

$$\lim_{m\to\infty}\lim_{n\to\infty}\int_{\Omega_n\cap\{f\geq m\}} f d\nu_n = 0.$$

*Proof.* $\boxed{(1) \Rightarrow (2)}$

Assume that $f$ is $\mathbb{P}$–integrable and $\lim_{n\to\infty}\int_{\Omega_n} f d\nu_n = \int_{E^k} f d\mathbb{P}$. Using the nonstandard characterization of limits, Corollary 2.3.3[(iii)], and Theorem 1.3.15(4) (making use of the fact that $f = |f|$ since $f$ is assumed to be nonnegative), it follows that $^*f$ is $S$–integrable on $\Omega_N$ for any $N > \mathbb{N}$.

$\boxed{(2) \Rightarrow (3)}$

Now assume that $^*f$ is $S$–integrable on $\Omega_N$ for all $N > \mathbb{N}$. As a consequence (using either Theorem 1.3.15(2) or Theorem 1.3.15(3)), we have that $^*f\mathbb{1}_{\{|^*f|\geq m\}}$ is $S$–integrable on $\Omega_N$ for any $N > \mathbb{N}$ and $m \in \mathbb{N}$. Fix $N_0 > \mathbb{N}$ such that the following is true (existence of such an $N_0$ is guaranteed by the nonstandard characterization of limit superiors):

$$\limsup_{n\to\infty}\int_{\Omega_n\cap\{|f|\geq m\}} |f|\, d\nu_n = \mathbf{st}\left({}^*\!\int_{\Omega_{N_0}} {}^*|f|\, \mathbb{1}_{\{^*|f|\geq m\}} d\nu_{N_0}\right).$$

By Theorem 1.3.15(4), we get:

$$\begin{aligned} &\limsup_{n\to\infty}\int_{\Omega_n\cap\{|f|\geq m\}} |f|\, d\nu_n = \int_{\Omega_{N_0}} \mathbf{st}\left({}^*|f|\, \mathbb{1}_{\{^*|f|\geq m\}}\right) dL\nu_{N_0} \\ \Rightarrow &\lim_{m\to\infty}\limsup_{n\to\infty}\int_{\Omega_n\cap\{|f|\geq m\}} |f|\, d\nu_n = \lim_{m\to\infty}\int_{\Omega_{N_0}} \mathbf{st}\left({}^*|f|\, \mathbb{1}_{\{^*|f|\geq m\}}\right) dL\nu_{N_0}. \end{aligned} \tag{2.23}$$

Since $^*f$ is $S$–integrable on $\Omega_{N_0}$, it follows that $\mathbf{st}(^*f)$ is Loeb integrable on $\Omega_{N_0}$. Hence the limit on the right side of (2.23) is zero, as desired.

$\boxed{\textit{(3)} \Rightarrow \textit{(1)}}$

This follows from Theorem 2.3.1, Corollary 2.3.3(iii), and Theorem 1.3.15(4). □

### 2.3.2. Application to a proof of the Riemann-Lebesgue Lemma

The theory of limiting integrals built over the last two subsections may theoretically be applied to a lot of situations in which the probability spaces are changing. While we will cover its application to spherical integrals in the next section, we include here a new proof of the famous Riemann–Lebesgue lemma as an illustration of the versatility of this theory. We paraphrase the Riemann–Lebesgue lemma below (see, for example, Rudin [87, 5.14, p. 103]).

**Theorem 2.3.5** (Riemann–Lebesgue Lemma)**.** *Let $\lambda$ be the Lebesgue measure on the interval $T := [-\pi, \pi]$. If $f \in L^1(T, \lambda)$, then we have:*

$$\lim_{n\to\infty} \int_T f(x)\cos(nx)d\lambda(x) = 0 \text{ and } \lim_{n\to\infty} \int_T f(x)\sin(nx)d\lambda(x) = 0.$$

*Proof.* For each $n \in \mathbb{N}$, define $g_n : T \to \mathbb{R}$ by $g_n(x) = \dfrac{1-\cos(nx)}{2\pi}$. The functions $g_n$ are probability densities on $[-\pi, \pi]$. For each $n \in \mathbb{N}$, let $\mathbb{P}_n$ denote the probability measure on $T$ with the density $g_n$. By integrating the densities for $n \in \mathbb{N}$, we find that the corresponding probability distribution functions are given by:

$$G_n(x) := \mathbb{P}_n\{(-\infty, x]\} = \frac{1}{2\pi}\left(x - \frac{\sin(x)}{n}\right) \text{ for all } x \in T.$$

As $n \to \infty$, the sequence $G_n$ converges pointwise to the distribution function of the uniform (normalized) Lebesgue measure $\mathbb{P}$ on $[-\pi, \pi]$. Thus $\mathbb{P}_n \xrightarrow{weak} \mathbb{P}$, that is,

$$\lim_{n\to\infty} \int_T f d\mathbb{P}_n = \int_T f d\mathbb{P} \text{ for all bounded continuous } f : T \to \mathbb{R}.$$

By an equivalent criterion for weak convergence, we obtain:

$$\liminf_{n\to\infty} \mathbb{P}_n(U) \geq \mathbb{P}(U) \text{ for all open subsets } U \subseteq T. \tag{2.24}$$

By the nonstandard characterization of limit inferiors, this is equivalent to:

$$\mathbb{P}_N({}^*U) \geq \mathbb{P}(U) \text{ for all open subsets } U \subseteq T \text{ and } N > \mathbb{N}. \tag{2.25}$$

Since the density function $g_n$ for $\mathbb{P}_n$ is pointwise bounded above by the density function for $\mathbb{P}$, by transfer we also obtain the other side of the above inequality. That is, we obtain:

$$\mathbb{P}_N({}^*U) \leq \mathbb{P}(U) \text{ for all open subsets } U \subseteq T \text{ and } N > \mathbb{N}. \tag{2.26}$$

Combining (2.25) and (2.26), we obtain:

$$\mathbb{P}_N({}^*U) = \mathbb{P}(U) \text{ for all open subsets } U \subseteq T \text{ and } N > \mathbb{N}. \tag{2.27}$$

By Proposition 2.2.5, we obtain:

$$\int\limits_{{}^*T} \mathbf{st}({}^*f) dL\mathbb{P}_N = \int\limits_{T} f d\mathbb{P} \text{ for all bounded measurable } f\colon T \to \mathbb{R} \text{ and } N > \mathbb{N}. \tag{2.28}$$

For any $f \in L^1(T,\lambda)$, we use the facts that $|g_n| \leq \dfrac{1}{\pi}$ and $f \in L^1(T,\lambda)$ to get:

$$\lim_{m\to\infty}\lim_{n\to\infty} \int_T |f|\, \mathbb{1}_{|f|>m} d\mathbb{P}_n(x) \leq \frac{1}{\pi} \lim_{m\to\infty} \int_T |f|\, \mathbb{1}_{|f|>m} d\lambda(x) = 0. \tag{2.29}$$

Using (2.28) and (2.29) in Theorem 2.3.4 (with $(T, \mathbb{P}_n)$ playing the role of $(\Omega_n, \nu_n)$ in that theorem), we obtain, for each $f \in L^1(T, \lambda) = L^1(T, \mathbb{P})$:

$$\begin{aligned}
&\lim_{n\to\infty} \int_T f(x)d\mathbb{P}_n(x) = \int_T f(x)d\mathbb{P}(x) \\
\Rightarrow &\lim_{n\to\infty} \int_T \left( \frac{f(x)}{2\pi} - \frac{f(x)\cos(nx)}{2\pi} \right) d\lambda(x) = \int_T \frac{f(x)}{2\pi} d\lambda(x) \\
\Rightarrow &\lim_{n\to\infty} \int_T f(x)\cos(nx)d\lambda(x) = 0.
\end{aligned}$$

The proof for $\sin(nx)$ goes exactly the same way if we replace the $f_n$ by the probability density functions $g_n(x) = \dfrac{1-\sin(nx)}{2\pi}$ for $x \in T$. □

### 2.3.3. What happens if the finite dimensional function is not nice in the limiting space?

In general, for a function $f\colon E^k \to \mathbb{R}$ (not necessarily satisfying the conditions in Theorem 2.3.4), the following result allows us to still approximate its integral by a suitably modified sequence of integrals over $(\Omega_n, \nu_n)$. Note that this result is in the spirit of Littlewood's three principles from measure theory (see [65, p. 26])—approximating a potentially ill-behaved integrable function by well-behaved bounded functions.

**Lemma 2.3.6.** *In the setting of Corollary 2.3.3, let $f\colon E^k \to \mathbb{R}$ be $\mathbb{P}$–integrable. Given any $\epsilon, \delta, \theta \in \mathbb{R}_{>0}$ there exist an $n_0 \in \mathbb{N}$ and functions $g_n\colon \Omega_n \to \mathbb{R}$ for all $n \in \mathbb{N}_{\geq n_0}$ such that the following hold:*

*(i)* $|g_n|$ *is bounded by* $n$ *for all* $n \in \mathbb{N}_{\geq n_0}$.

*(ii)* $\nu_n\left(|g_n - f| > \delta\right) < \epsilon$ *for all* $n \in \mathbb{N}_{\geq n_0}$.

*(iii)* $\left| \int_{\Omega_n} g_n d\nu_n - \int_{E^k} f d\mathbb{P} \right| < \theta$ *for all* $n \in \mathbb{N}_{\geq n_0}$.

*Proof.* By Corollary 2.3.3(iii), we know that

$$\int_{E^k} |f|\, d\mu = \int_{\Omega_N} \mathbf{st}({}^*|f|) dL\nu_N \text{ for all } N > \mathbb{N}.$$

Thus, for any $N > \mathbb{N}$, the map $\mathbf{st}({}^*f)$ is Loeb integrable on $\Omega_N$, and hence has an $S$–integrable lifting $G_N \colon \Omega_N \to {}^*\mathbb{R}$ by Theorem 1.3.16. In particular,

$$L\nu_N\left(\mathbf{st}(G_N) = \mathbf{st}({}^*f)\right) = 1, \text{ and} \tag{2.30}$$

$$\mathbf{st}\left({}^*\int_{\Omega_N} G_N d\nu_N\right) = \int_{\Omega_N} \mathbf{st}(G_N) dL\nu_N = \int_{\Omega_N} \mathbf{st}({}^*f) dL\nu_N = \int_{E^k} f d\mathbb{P}. \tag{2.31}$$

Equation (2.30) follows from the definition of lifting. The first equality in (2.31) follows from Theorem 1.3.15(4) applied to the nonnegative $S$–integrable functions $(G_N)_+ := \max\{G_N, 0\}$ and $(G_N)_- := \max\{-G_N, 0\}$. The second equality in (2.31) follows from equation (2.30), while the last equality in (2.31) follows from Corollary 2.3.3[(iii)].

Without loss of generality, we can assume that $|G_N| \leq N$ for all $N > \mathbb{N}$ (as we may replace $G_N$ by the function $G_N \mathbb{1}_{|G_N| \leq N}$, which still satisfies (2.30) and (2.31)). Thus, for the given $\epsilon, \delta, \theta \in \mathbb{R}_{>0}$, the following internal set contains ${}^*\mathbb{N}\backslash\mathbb{N}$.

$$\mathcal{G}_{\epsilon,\delta,\theta} := \Bigg\{ n \in {}^*\mathbb{N} : \ \exists G_n \in {}^*L^1(\Omega_n, \nu_n) \text{ such that } \ |G_n| \leq n,$$
$${}^*\nu_n\left(|G_n - {}^*f| > \delta\right) < \epsilon, \text{ and } \left| {}^*\int_{{}^*\Omega_n} G_n d^*\nu_n - \int_{E^k} f d\mathbb{P} \right| < \theta \Bigg\}.$$

By underflow, we find $n_0 \in \mathbb{N}$ such that $\mathbb{N}_{\geq n_0} \subseteq \mathcal{G}_{\epsilon,\delta,\theta}$. Now fix an $n \in \mathbb{N}_{\geq n_0}$. In the

nonstandard universe, the following statement is true:

$$\exists G_n \in {}^*L^1(\Omega_n, \nu_n)$$
$$\left( (|G_n| \le n) \wedge ({}^*\nu_n\left(|G_n - {}^*f| > \delta\right) < \epsilon)) \wedge \left(\left| {}^*\!\!\int_{{}^*\Omega_n} G_n d^*\nu_n - \int_{E^k} f d\mathbb{P} \right| < \theta \right)\right).$$

Transfer of this sentence yields a $g_n \in L^1(\Omega_n, \nu_n)$ with the desired properties. ☐

We can strengthen Lemma 2.3.6 as follows, by requiring the functions to have the same domain $E^k$.

**Theorem 2.3.7.** *In the setting of Corollary 2.3.3, let $f\colon E^k \to \mathbb{R}$ be $\mathbb{P}$–integrable. Given any $\epsilon, \delta, \theta \in \mathbb{R}_{>0}$ there exist an $n_0 \in \mathbb{N}$ and functions $g_n\colon E^k \to \mathbb{R}$ for all $n \in \mathbb{N}_{\ge n_0}$ such that the following hold:*

*(i)* $|g_n|$ *is bounded by* $n$ *for all* $n \in \mathbb{N}_{\ge n_0}$.

*(ii)* $\nu_n\left(|g_n - f| > \delta\right) < \epsilon$ *for all* $n \in \mathbb{N}_{\ge n_0}$.

*(iii)* $\left| \int_{\Omega_n} g_n d\nu_n - \int_{E^k} f d\mathbb{P} \right| < \theta$ *for all* $n \in \mathbb{N}_{\ge n_0}$.

*Proof.* For $n \in \mathbb{N}_{\ge k}$, define $\nu'_n\colon \mathcal{E}_k \to [0,1]$ by $\nu'_n(B) = \nu_n((B \times E^{n-k}) \cap \Omega_n)$. For any bounded measurable $g\colon E^k \to \mathbb{R}$, expressing $g$ as a uniform limit of simple functions yields

$$\int_{\Omega_n} g d\nu_n = \int_{E^k} g d\nu'_n. \tag{2.32}$$

Let $(g_n)_{n\in\mathbb{N}}$ be a sequence of functions obtained by applying Lemma 2.3.6 to the sequence $(E^k, \nu'_n)_{n\in\mathbb{N}}$ of probability spaces. Then (i), (ii) and (iii) follow from the corresponding results in Lemma 2.3.6 together with (2.32). ☐

## 2.4. Generalizing Poincaré's theorem

### 2.4.1. Revisiting a standard proof of Poincaré's theorem

For the rest of the chapter, we let $S_n$ denote the sphere $S^{n-1}(\sqrt{n})$ and $\bar{\sigma}_n$ denote $\bar{\sigma}_{S_n}$, for all $n \in \mathbb{N}$. Fix $k \in \mathbb{N}$ and let $\mu$ denote the standard k-dimensional Gaussian measure. Let $B_k(a)$ denote the open ball of radius $a$ in $\mathbb{R}^k$. For a set $B \in \mathcal{B}(\mathbb{R}^k)$ and any $n \in \mathbb{N}_{\geq k}$, we define $\bar{\sigma}_n(B)$ to be the value of $\bar{\sigma}_n(\{x \in S_n : \pi_k(x) \in B\}) = \bar{\sigma}_n\left((B \times \mathbb{R}^{n-k}) \cap S_n\right)$, where $\pi_k$ is the projection onto $\mathbb{R}^k$. Similarly, a function $f\colon \mathbb{R}^k \to \mathbb{R}$ is canonically extended to $\mathbb{R}^n$ by using '$f(x, y)$' to denote $f(x)$ for all $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^{n-k}$.

In an attempt to generalize Theorem 2.1.1, we first look at another proof of the same result using classical analysis. This proof requires directly evaluating the spherical integrals and using dominated convergence theorem (compare with the less computational proof of Theorem 2.1.1 in Section 2.2.3). We restate Theorem 2.1.1 below for convenience.

**Theorem 2.1.1** (Poincaré, [80]). *For all bounded measurable functions $f : \mathbb{R}^k \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu. \tag{2.2}$$

*Proof.* Let $\lambda$ denote the Lebesgue measure on $\mathbb{R}^k$. By Sengupta's disintegration formula (see [90, Proposition 4.1]), we have the following chain of equalities for any bounded measurable $f\colon \mathbb{R}^k \to \mathbb{R}$.

$$
\begin{aligned}
&\int\limits_{S^{n-1}(\sqrt{n})} f d\bar{\sigma}_n \\
&= \frac{1}{\sigma(S_n)} \int\limits_{x\in B_k(\sqrt{n})} \int_{y\in S^{n-k-1}(\sqrt{n-||x||^2})} f(x,y)d\sigma(y)\frac{\sqrt{n}}{\sqrt{n-||x||^2}}d\lambda(x) \\
&= \frac{1}{\sigma(S_n)} \int_{\mathbb{R}^k} \sigma\left(S^{n-k-1}\left(\sqrt{n-||x||^2}\right)\right)\cdot\frac{\mathbb{1}_{B_k(\sqrt{n})}(x)f(x)\sqrt{n}}{\sqrt{n-||x||^2}}d\lambda(x) \\
&= \frac{\Gamma\left(\frac{n}{2}\right)}{2\pi^{\frac{n}{2}}\cdot(\sqrt{n})^{n-1}} \int\limits_{\mathbb{R}^k} \frac{2\pi^{\frac{n-k}{2}}(n-||x||^2)^{\frac{n-k-1}{2}}}{\Gamma\left(\frac{n-k}{2}\right)}\cdot\frac{\mathbb{1}_{B_k(\sqrt{n})}(x)f(x)\sqrt{n}}{\sqrt{n-||x||^2}}d\lambda(x) \\
&= a_{n,k}b_{n,k}\int_{\mathbb{R}^k}\frac{1}{\left(\sqrt{2\pi}\right)^k}\left(1-\frac{||x||^2}{n}\right)^{\frac{n}{2}}\frac{\mathbb{1}_{B_k(\sqrt{n})}(x)f(x)}{\left(1-\frac{||x||^2}{n}\right)^{\frac{k+2}{2}}}d\lambda(x),
\end{aligned}
\tag{2.33}
$$

where $a_{n,k}=\dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right)\cdot\left(\frac{n-k}{2}\right)^{\frac{k}{2}}}$ and $b_{n,k}=\left(1-\dfrac{k}{n}\right)^{\frac{k}{2}}$.

Note that $\lim\limits_{n\to\infty} a_{n,k} = \lim\limits_{n\to\infty} b_{n,k} = 1$ for all $k\in\mathbb{N}$ (the first limit following from Stirling's formula, see Rudin [86, equation 103, p. 194]).

Modulo constants, for large values of $n$, the integrand in (2.33) is bounded by $|f(x)|\,e^{-\frac{||x||^2}{4}}$, which is integrable on $\mathbb{R}^k$ since $f$ is assumed to be bounded. Thus by the dominated convergence theorem, the integral in (2.33) converges to $\displaystyle\int_{\mathbb{R}^k} f d\mu$ as $n\to\infty$, as desired. $\square$

**Remark 2.4.1.** Due to the factor of $\left(1-\dfrac{||x||^2}{n}\right)^{\frac{k+2}{2}}$ in the denominator of (2.33), dominated convergence theorem does not directly work when we work with an unbounded function $f$, as there is no reason for $|f(x)|\,e^{-\frac{||x||^2}{4}}$ to be Lebesgue integrable in general. Indeed for a general Gaussian integrable $f$, we can bound $|f(x)|\left(1-\dfrac{||x||^2}{n}\right)^{\frac{n}{2}}$ by $|f(x)|\,e^{-\frac{||x||^2}{2}}$, but there is still no obvious way to bound the whole integrand in (2.33) by a Lebesgue integrable function due to that extra factor in the denominator.

**Corollary 2.4.2.** *For $k \in \mathbb{N}$ and $N > \mathbb{N}$, almost all points on $S_N$ have finite first $k$ coordinates. That is,*

$$L\bar{\sigma}_N(\{(x_1,\ldots,x_N) \in S^{N-1}(\sqrt{N}) : x_1,\ldots,x_k \in {}^*\mathbb{R}_{fin}\}) = 1.$$

*Proof.* Fix $k$ and $N$ as above. If $m \in \mathbb{N}$, we have $L\bar{\sigma}_N({}^*(-m,m)^k) = \mu((-m,m)^k)$ by Theorem 2.1.1. Letting $m \to \infty$ on both sides completes the proof. □

**Corollary 2.4.3.** *For any $t \in \mathbb{R}_{>1}$, we have*

$$\lim_{n\to\infty} \int_{\{x\in\mathbb{R}^k : \frac{n}{t} < ||x||^2 < n\}} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{4}} d\lambda(x) = 0.$$

*Proof.* Let $t \in \mathbb{R}_{>1}$ and $N > \mathbb{N}$. As a consequence of Corollary 2.4.2, we obtain:

$$\bar{\sigma}_N\left(\left\{x \in S^{N-1}(\sqrt{N}) : \frac{N}{t} < ||\pi_k(x)||^2 < N\right\}\right) \approx 0.$$

The nonstandard characterization of limits and equation (2.33) thus yield the following.

$$\lim_{n\to\infty} \int_{\{x\in\mathbb{R}^k : \frac{n}{t} < ||x||^2 < n\}} \frac{1}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{2}} \frac{1}{\left(1 - \frac{||x||^2}{n}\right)^{\frac{k+2}{2}}} d\lambda(x) = 0. \qquad (2.34)$$

For all $n \in \mathbb{N}_{\geq 2(k+2)}$, the sequence in the statement of the Corollary is bounded above by (a constant times) the sequence in (2.34), thus completing the proof. □

**Remark 2.4.4.** We can also prove Corollary 2.4.3 directly by noting that

$$\left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{4}} \mathbb{1}_{||x||^2 \leq n} \leq e^{-\frac{||x||^2}{4}},$$

where the right side is Lebesgue integrable over $\mathbb{R}^k$. The proof presented above is still valuable because it exposes a connection between these integrals and surface area measures.

### 2.4.2. A useful inequality between spherical and Gaussian measures

In this subsection, we derive an inequality comparing the $L^1$ norm (over the sphere $S^{n-1}(\sqrt{n})$) of a function defined on $\mathbb{R}^k$ and its $p^{\text{th}}$ moment (for any $p \in \mathbb{R}_{>1}$) with respect to the standard Gaussian measure on $\mathbb{R}^k$.

With the foresight provided by the philosophy of spherical integrals being close to a Gaussian integral, we expect these spherical integrals to be asymptotically bounded by the $L^p(\mathbb{R}^k, \mu)$–norms as the dimensions increase. Theorem 2.4.6 shows that depending on the value of $p \in \mathbb{R}_{>1}$, there is a dimension (namely $4(k+1)q$) beyond which this does happen. Before we prove that theorem, we need to generalize Sengupta's disintegration formula to work for any nonnegative function.

**Theorem 2.4.5.** *Let $N$ and $k$ be positive integers with $k < N$. Suppose $f$ is either a bounded measurable or a nonnegative measurable function on $S^{N-1}(a)$, the sphere in $\mathbb{R}^N = \mathbb{R}^k \times \mathbb{R}^{N-k}$ of radius $a$ and with center $0$. Then, with $\sigma$ denoting surface measure (non-normalized) on spheres,*

$$\int_{z \in S^{N-1}(a)} f(z)d\sigma(z) = \int_{x \in B_k(a)} \left( \int_{y \in S^{N-k-1}(a_x)} f(x,y)d\sigma(y) \right) \frac{a}{a_x} dx \tag{2.35}$$

*for any $a \in \mathbb{R}_{>0}$, where $a_x = \sqrt{a^2 - ||x||^2}$. The above equality means that either both sides are finite and equal, or both sides are infinite.*

*Proof.* If $f$ is bounded measurable, then this is just Sengupta's disintegration formula (see [90, Proposition 4.1]). Otherwise, if $f$ is nonnegative, then apply Sengupta's disintegration formula to the bounded measurable functions $f_m := f \cdot \mathbb{1}_{f \leq m}$ for each $m \in \mathbb{N}$, and then use monotone convergence theorem on both sides to obtain (2.35). $\square$

**Theorem 2.4.6.** *For each $p \in \mathbb{R}_{>1}$, there is a constant $C_p \in \mathbb{R}_{>0}$ such that*

$$\int_{S^{n-1}(\sqrt{n})} |g| \, d\bar{\sigma}_n \leq C_p [\mathbb{E}_\mu(|g|^p)]^{\frac{1}{p}} \text{ for all } g \in L^p(\mathbb{R}^k, \mu) \text{ and } n \in \mathbb{N}_{>4(k+2)q}, \tag{2.36}$$

*where $q \in \mathbb{R}_{>0}$ is such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$.*

*Proof.* Fix $g \in L^p(\mathbb{R}^k, \mu)$, where $p \in \mathbb{R}_{>1}$. Also, let $t \in \mathbb{N}_{>1}$. Using Theorem 2.4.5 instead of [90, Proposition 4.1], we can follow the same steps leading up to (2.33) to see that $\int_{S^{n-1}(\sqrt{n})} |g| \, d\bar{\sigma}_n$ is equal to

$$\begin{aligned}
&\int_{||x||^2 \leq \frac{n}{t}} \frac{a_{n,k} b_{n,k} \, |g(x)|}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n-k-2}{2}} d\lambda(x) \\
&+ \int_{\frac{n}{t} < ||x||^2 \leq n} \frac{a_{n,k} b_{n,k} \, |g(x)|}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{2}} \frac{1}{\left(1 - \frac{||x||^2}{n}\right)^{\frac{k+2}{2}}} d\lambda(x),
\end{aligned} \tag{2.37}$$

where $a_{n,k} = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right) \cdot \left(\frac{n-k}{2}\right)^{\frac{k}{2}}}$ and $b_{n,k} = \left(1 - \dfrac{k}{n}\right)^{\frac{k}{2}}$ are the same constants that appear in (2.33).

Note that

$$\left(1 - \frac{||x||^2}{n}\right)^{-\frac{k+2}{2}} \leq \left(\frac{t}{t-1}\right)^{\frac{k+2}{2}} \quad \text{whenever } ||x||^2 \leq \frac{n}{t}.$$

Also, $\left(1 - \dfrac{||x||^2}{n}\right)^{\frac{n}{2}} \mathbb{1}_{||x||^2 \leq n}$ is at most equal to $e^{-\frac{||x||^2}{2}}$ for all $x \in \mathbb{R}^k$. Noting that $b_{n,k} < 1$ for all $n \in \mathbb{N}_{>k}$, the first summand in (2.37) is at most

$$\left(\frac{t}{t-1}\right)^{\frac{k+2}{2}} \frac{a_{n,k}}{(2\pi)^{\frac{k}{2}}} \int_{\mathbb{R}^k} |g(x)| \, e^{-\frac{||x||^2}{2}} d\lambda(x)$$

for all $n \in \mathbb{N}_{>k}$. Writing this integral as a Gaussian expected value, and then using Jensen's inequality, we have:

$$I_1 \leq a_{n,k} \left(\frac{t}{t-1}\right)^{\frac{k+2}{2}} ||g||_{L^p(\mathbb{R}^k,\mu)} \text{ for all } n \in \mathbb{N}_{>k}, \tag{2.38}$$

where $I_1$ is the first summand in (2.37), and $||g||_{L^p(\mathbb{R}^k,\mu)} = (\mathbb{E}_\mu(|g|^p))^{\frac{1}{p}}$.

Let $q \in \mathbb{R}_{>1}$ be such that $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Then we can write the second summand in (2.37) as follows:

$$a_{n,k} b_{n,k} \int_{\frac{n}{t} < ||x||^2 \leq n} \frac{|g(x)|}{(\sqrt{2\pi})^{\frac{k}{p}}} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{2p}} \cdot \frac{1}{(\sqrt{2\pi})^{\frac{k}{q}}} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{2q} - \frac{k+2}{2}} d\lambda(x). \tag{2.39}$$

Note that $b_{n,k} < 1$ for all $n \in \mathbb{N}_{>k}$. By Hölder's inequality applied to the functions $x \mapsto \dfrac{|g(x)|}{(\sqrt{2\pi})^{\frac{k}{p}}} \left(1 - \dfrac{||x||^2}{n}\right)^{\frac{n}{2p}}$ and $x \mapsto \dfrac{1}{(\sqrt{2\pi})^{\frac{k}{q}}} \left(1 - \dfrac{||x||^2}{n}\right)^{\frac{n}{2q} - \frac{k+2}{2}}$ (on the domain $\{x \in \mathbb{R}^k : \dfrac{n}{t} < ||x||^2 < n\}$ equipped with its Lebesgue measure), the expression in (2.39) is at most equal to the following [1]:

$$a_{n,k} \left( \int_{\substack{x \in \mathbb{R}^k \\ \frac{n}{t} < ||x||^2 \leq n}} |g(x)|^p \cdot \frac{1}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{2}} d\lambda(x) \right)^{\frac{1}{p}}$$
$$\times \left( \int_{\substack{x \in \mathbb{R}^k \\ \frac{n}{t} < ||x||^2 \leq n}} \frac{1}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\left(\frac{n}{2q} - \frac{k+2}{2}\right) \cdot q} d\lambda(x) \right)^{\frac{1}{q}}.$$

The first term in this product is at most $a_{n,k}(\mathbb{E}_\mu(|g|^p))^{\frac{1}{p}}$. Also, the integrand in the second term in this product is at most $\left(1 - \dfrac{||x||^2}{n}\right)^{\frac{n}{4}}$ for all $n \in \mathbb{N}_{>2(k+2)q}$. To summarize, if $I_2$ is the second summand in (2.37), then we have:

[1] An anonymous referee has pointed out that one could also apply Hölder's inequality to the functions $x \mapsto |g(x)|$ and $x \mapsto \left(1 - \dfrac{||x||^2}{n}\right)^{-\frac{k+2}{2}}$, on the same domain but with the measure given by $d\nu(x) = \dfrac{1}{(\sqrt{2\pi})^k} \left(1 - \dfrac{||x||^2}{n}\right)^{\frac{n}{2}} d\lambda(x)$.

$$I_2 \leq a_{n,k}\, ||g||_{L^p(\mathbb{R}^k,\mu)} \cdot \theta_{n,t} \text{ for all } n \in \mathbb{N}_{>(k+2)q}, \tag{2.40}$$

where

$$\theta_{n,t} = \left( \int_{\substack{x\in\mathbb{R}^k \\ \frac{n}{t}<||x||^2\leq n}} \frac{1}{(\sqrt{2\pi})^k} \left(1 - \frac{||x||^2}{n}\right)^{\frac{n}{4}} d\lambda(x) \right)^{\frac{1}{q}}. \tag{2.41}$$

Combining (2.38) and (2.40), we get:

$$\int_{S^{n-1}(\sqrt{n})} |g|\, d\bar{\sigma}_n \leq a_{n,k} \left[ \left(\frac{t}{t-1}\right)^{\frac{k+2}{2}} + \theta_{n,t} \right] ||g||_{L^p(\mathbb{R}^k,\mu)} \tag{2.42}$$

for all $n \in \mathbb{N}_{>4(k+2)q}$ and $t \in \mathbb{N}_{>1}$.

Here $a_{n,k} = \dfrac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right) \cdot \left(\frac{n-k}{2}\right)^{\frac{k}{2}}}$ and $\theta_{n,t}$ is as in (2.41). Note that $\lim_{n\to\infty} a_{n,k} = 1$, and by Corollary 2.4.3, $\lim_{n\to\infty} \theta_{n,t} = 0$ for all $t \in \mathbb{N}$. Thus, for any $t \in \mathbb{N}$, the coefficient of $||g||_{L^p(\mathbb{R}^k,\mu)}$ in (2.42) is uniformly bounded above, by (say) $C_p$. This completes the proof of the theorem. □

Focusing on the coefficient in (2.42), we note that given $\epsilon \in \mathbb{R}_{>0}$ we can choose $t \in \mathbb{N}_{>1}$ large enough for which the following inequality holds.

$$\left(\frac{t}{t-1}\right)^{\frac{k+2}{2}} < 1 + \frac{\epsilon}{2}.$$

For this $t$, using Corollary 2.4.3, we can choose an $n_p \in \mathbb{N}$ large enough such that $\theta_{n,t} < \dfrac{\epsilon}{2}$ for all $n \in \mathbb{N}_{>n_p}$. Since $\lim_{n\to\infty} a_{n,k} = 1$, we can also ensure that the $n_p$ we choose is large enough such that $a_{n,k} < 1 + \epsilon$ for all $n \in \mathbb{N}_{>n_p}$. Combining all of this, (2.42) yields the following useful corollary: we are able to bound the ratio of the spherical integral and

the Gaussian $L^p$ norm by a constant as close to 1 as we want, with the price of having to go to a potentially higher dimension to observe this phenomenon.

**Corollary 2.4.7.** *For each $p \in \mathbb{R}_{>1}$ and $\epsilon \in \mathbb{R}_{>0}$, there is an $n_p \in \mathbb{N}$ such that*

$$\int_{S^{n-1}(\sqrt{n})} |g| \, d\bar{\sigma}_n \leq (1+\epsilon)[\mathbb{E}_\mu(|g|^p)]^{\frac{1}{p}} \text{ for all } g \in L^p(\mathbb{R}^k, \mu) \text{ and } n \in \mathbb{N}_{>n_p}. \tag{2.43}$$

Using Theorem 2.4.6, the condition (2.19) of Theorem 2.3.1 is easily verified for all functions in $L^p(\mathbb{R}^k, \mu)$, where $p \in \mathbb{R}^k$. Using that theorem and Theorem 2.1.1, we obtain our main limiting result for spherical integrals.

**Theorem 2.4.8.** *If $\mu$ is the standard Gaussian measure on $\mathbb{R}^k$ and $f \in L^p(\mathbb{R}^k, \mu)$ for some $p \in \mathbb{R}_{>1}$, then the nonstandard extension ${}^*f$ is $S$–integrable on $S^{N-1}(\sqrt{N})$ for all $N > \mathbb{N}$. As a consequence, the function $f$ is integrable on $(S^{n-1}(\sqrt{n}), \bar{\sigma}_n)$ for all large $n \in \mathbb{N}$, and*

$$\lim_{m\to\infty} \lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n}) \cap \{|f| \geq m\}} |f| \, d\bar{\sigma}_n = 0.$$

*Furthermore, the spherical integrals of $f$ satisfy the following limiting behavior:*

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n})} f d\bar{\sigma}_n = \int_{\mathbb{R}^k} f d\mu.$$

*This limit of spherical integrals can be written as a single spherical integral (over an infinite sphere) $\int_{S^{N-1}(\sqrt{N})} \mathbf{st}({}^*f) dL\bar{\sigma}_N$ for any hyperfinite $N$.*

# Chapter 3. Limiting Spherical Integrals of Bounded Continuous Functions

## 3.1. Introduction

The study of the connection between high-dimensional surface area measures on spheres and Gaussian measures, in essence, dates back to the works on the kinetic theory of gas by Boltzmann [21] and Maxwell [74]. We studied this phenomenon from a nonstandard analytic perspective in the previous chapter. A key idea in that work was to express the limiting behavior of spherical integrals through certain Loeb integrals over spheres of hyperfinite dimensions.

The aforementioned idea of using a nonstandard measure space as a limiting object of a sequence of measure spaces is applicable to many situations in which we are studying asymptotics of marginals along a given direction while the ambient spaces are changing. Sengupta [90] and Peterson–Sengupta [79] studied the Gaussian Radon transform of finite dimensional functions as a limit of spherical integrals over certain spheres of increasing dimension, which is an appropriate setting to work with nonstandard analysis in. This is the main theme of this chapter. We refer the reader to [48] and [49] for earlier standard approaches in this context.

In [90], Sengupta fixed a hyperplane $H$ in $\ell^2(\mathbb{R})$ and analyzed the limit of integrals over $S^{n-1}(\sqrt{n})$ intersected with an appropriate "truncation" of $H$ to the $n^{\text{th}}$ dimension. More precisely, let $H$ be the set of all square summable real sequences orthogonal to a unit vector $u \in \ell^2(\mathbb{R})$. The integral of a function $f\colon \ell^2(\mathbb{R}) \to \mathbb{R}$ with respect to the infinite dimensional Gaussian measure with mean $\vec{0} = (0, 0, \ldots)$ and covariance operator equaling

the projection $P_H$ onto $H$ is the Gaussian–Radon transform of $f$ evaluated at the hyperplane $H$ (see also Holmes–Sengupta [52]). In general, one could work with a codimension-1 affine subspace $A := pu + H$, and integrate $f : \ell^2(\mathbb{R}) \to \mathbb{R}$ with respect to the Gaussian measure with mean $pu$ and covariance $P_H$ in order to evaluate the Gaussian Radon transform at $A$.

In the case when $f\colon \mathbb{R}^k \to \mathbb{R}$ (identifying it as a function on $\ell^2(\mathbb{R})$ by composing it from the right with the projection to the first $k$ coordinates) is bounded measurable, Sengupta [90] showed that the corresponding Gaussian–Radon transform evaluated at many codimension-1 affine subspaces (more precisely, at those affine subspaces for which the marginal onto $\mathbb{R}^k$ of the Gaussian measure described above has full support) can be thought of as limits of spherical integrals of $f$ over the intersection of the spheres $S^{n-1}(\sqrt{n})$ with an appropriate finite dimensional approximation (in $\mathbb{R}^n \subseteq \ell^2(\mathbb{R})$) to $A$. This generalizes the earlier known results on limiting spherical integrals as we are not integrating over the full sphere $S^{n-1}(\sqrt{n})$, but rather on slices of this sphere. In [79], Peterson and Sengupta generalized the above result further to the case of affine subspaces of any finite codimension (see also [78]).

To more rigorously state the key results in this context, we first need to set up some notation and definitions that will be used throughout the rest of the chapter.

### 3.1.1. Notation and definitions

Let $\mathbb{R}^{\mathbb{N}}$ be the vector space of sequences of real numbers, with the standard basis $e_1 = (1, 0, 0, \ldots), e_2 = (0, 1, 0, \ldots)$, etc. As usual, $\ell^2(\mathbb{R})$ will denote the subspace consisting of all square summable real sequences. For $x = (x_1, x_2, \ldots) \in \mathbb{R}^{\mathbb{N}}$ and $n \in \mathbb{N}$, we define the

$n^{\text{th}}$ truncation/projection by

$$x_{(n)} := (x_1, \dots, x_n).$$

If $x := (x_1, \dots, x_m) \in \mathbb{R}^m$ for some $m \in \mathbb{N}$, then we will use the same symbol $x$ to denote $(x_1, \dots, x_m, 0, \dots, 0) \in \mathbb{R}^n$ for any $n \in \mathbb{N}_{>m}$, as well as to denote $(x_1, \dots, x_m, 0, 0, \dots) \in \ell^2(\mathbb{R})$, with the ambient space being clear from the context.

For $k \in \mathbb{N}$, we use $\pi_{(k)}$ to denote the projection from $\mathbb{R}^{\mathbb{N}}$ (or from some fixed $\mathbb{R}^n$ for $n \in \mathbb{N}_{\geq k}$ if the dimension $n$ is clear from context) onto the first $k$ coordinates:

$$\pi_{(k)}(x_1, x_2, \dots) = (x_1, \dots, x_k).$$

Let $u^{(1)}, \dots, u^{(\gamma)}$ be mutually orthonormal vectors in $\ell^2(\mathbb{R})$. For real numbers $p_1, \dots, p_\gamma$ (with $\vec{p} := (p_1, p_2, \dots, p_\gamma) \in \mathbb{R}^\gamma$), and $n \in \mathbb{N}$, define (see also Figure 3.1):

$$\begin{aligned}
A(\vec{p}) = A &:= \{x \in \ell^2(\mathbb{R}) : \langle x, u^{(i)} \rangle = p_i \text{ for all } i \in [\gamma]\}, \\
H_n &:= \{x \in \mathbb{R}^n : \langle x, (u^{(i)})_{(n)} \rangle = 0 \text{ for all } i \in [\gamma]\}, \\
A_n(\vec{p}) = A_n &:= \{x \in \mathbb{R}^n : \langle x, (u^{(i)})_{(n)} \rangle = p_i \text{ for all } i \in [\gamma]\}, \\
S_{A_n(\vec{p})} = S_{A_n} &:= S^{n-1}(\sqrt{n}) \cap A_n(\vec{p}), \text{ and} \\
S_{H_n} &:= S^{n-1}(\sqrt{n}) \cap H_n.
\end{aligned}$$

We also denote $S_{H_n}$ by $S_{n,u^{(1)},\dots,u^{(\gamma)}}$ when it is important to emphasize which vectors in $\ell^2(\mathbb{R})$ we are working with. When the sphere is clear from context, we will use $\bar{\sigma}$ to denote its uniform surface area measure.

The Borel sigma-algebra of a topological space $\Omega$ is denoted by $\mathcal{B}(\Omega)$. Let $\mathcal{S}_0$ be the set of all spheres that are centered at the origin in some real Euclidean space (in any

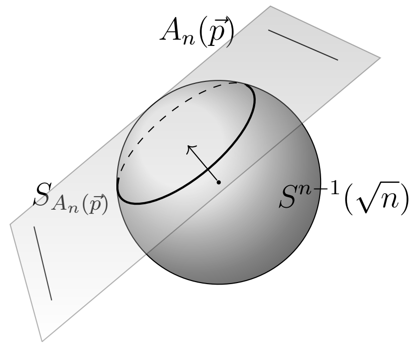


Figure 3.1. Intersecting $S^{n-1}(\sqrt{n})$ by the affine plane $A_n$

dimension $n \in \mathbb{N}$ and of any radius $r \in \mathbb{R}_{>0}$). For any $S \in \mathcal{S}_0$, we have an orthogonal transformation preserving map called the surface area $\sigma_S \colon \mathcal{B}(S) \to \mathbb{R}$, which satisfies the following (see [73, Chapter 3] for more background):

- For any $d \in \mathbb{N}$ and any $a \in \mathbb{R}_{>0}$, we have $\sigma_{S^d(a)}(S^d(a)) = c_d \cdot a^d$, where $c_d = \sigma_{S^d(1)}(S^d(1)) = (d+1) \cdot \dfrac{\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}+1\right)} = 2\dfrac{\pi^{\frac{d+1}{2}}}{\Gamma\left(\frac{d+1}{2}\right)}$.
- For any $S \in \mathcal{S}$ and any $A \in \mathcal{B}(\mathcal{S})$, we have $\bar{\sigma}_S(A) = \dfrac{\sigma_S(A)}{\sigma_S(S)}$.

Recall that we follow the superstructure approach to nonstandard extensions, as in Albeverio et al. [6]. In particular, we fix a sufficiently saturated nonstandard extension of a superstructure containing all standard mathematical objects under study. The nonstandard extension of a set $A$ is denoted by ${}^*A$. For $x, y \in {}^*X$ (where $X$ is a normed space), we will write $x \approx y$ to denote that $||x-y||$ is an infinitesimal. The set of finite nonstandard real numbers will be denoted by ${}^*\mathbb{R}_{\text{fin}}$ and the standard part map $\mathbf{st} \colon {}^*\mathbb{R}_{\text{fin}} \to \mathbb{R}$ takes a finite nonstandard real to its closest real number. We write $N > \mathbb{N}$ (and call such an $N$ *hyperfinite*) if $N \in {}^*\mathbb{N} \setminus \mathbb{N}$. Viewing other spheres as translations of spheres in $S_0$, the concept of the surface area measure canonically extends to all finite-dimensional spheres. By transfer, we have the notion of ${}^*$-surface area in the nonstandard universe. Taking stan-

dard parts of the uniform $^*$-surface area $\bar{\sigma}_S$ leads to the construction of the uniform Loeb surface measure $L\bar{\sigma}_S$ on any hyperfinite-dimensional sphere $S$. When the sphere is clear from context, we drop the subscript and use $\bar{\sigma}$ and $L\bar{\sigma}$ to denote these measures.

Fix $k \in \mathbb{N}$. For a set $B \in \mathcal{B}(\mathbb{R}^k)$ and any $n \in \mathbb{N}_{\geq k}$, if $S$ is a (possibly lower dimensional) sphere in $\mathbb{R}^n$, then we use $\bar{\sigma}_S(B)$ to denote $\bar{\sigma}_S\left((B \times \mathbb{R}^{n-k}) \cap S\right)$. Similarly, a function $f\colon \mathbb{R}^k \to \mathbb{R}$ is canonically extended to $\mathbb{R}^n$ by using "$f(x, y)$" to denote $f(x)$ for all $x \in \mathbb{R}^k$ and $y \in \mathbb{R}^{n-k}$.

For an element $x$ in a Hilbert space, $P_x$ is the projection operator onto the span of $x$. Let $\bar{\eta} = p_1(u^{(1)})_{(k)} + \ldots + p_\gamma(u^{(\gamma)})_{(k)}$. With $I_k$ being the identity operator on $\mathbb{R}^k$, let $\mu^{(k)}_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ be the Gaussian measure on $\mathbb{R}^k$ with mean $\bar{\eta}$ and covariance

$$I_k - \left|\left|(u^{(1)})_{(k)}\right|\right|^2 P_{(u^{(1)})_{(k)}} - \ldots - \left|\left|(u^{(\gamma)})_{(k)}\right|\right|^2 P_{(u^{(\gamma)})_{(k)}}. \tag{3.1}$$

We drop the superscript in $\mu^{(k)}_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ when the dimension $k$ is clear from context. Also, when the $u^{(i)}$ and $p_i$ are clear from context, we denote $\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ by just $\mu$. If the $p_i$ are all zero, then we denote the corresponding measure by $\mu_0$.

### 3.1.2. Description of key results

In the notation set up above, the result of Peterson–Sengupta on limits of integrals on slices of high-dimensional spheres can be summarized as follows (see [79, Theorem 2.1]).

**Theorem 3.1.1** (Peterson–Sengupta)**.** *Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be bounded and Borel measurable. If the Gaussian measure $\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ has full support on $\mathbb{R}^k$, then*

$$\lim_{n\to\infty} \int_{S_{A_n}} f\, d\bar{\sigma} = \int_{\mathbb{R}^k} f\, d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}.$$

Nonstandard analysis allows us to view the limit of spherical integrals as (the standard part of) another "spherical integral" in a hyperfinite dimension. We will use this idea to generalize Theorem 3.1.1 for bounded continuous functions in the case when $\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ does not necessarily have full support on $\mathbb{R}^k$. Our proof is done in several steps of increasing complexity:

(i) We first prove Theorem 3.1.1 in the case when the coordinates of the vectors $u^{(1)}, \ldots, u^{(\gamma)}$ are zero after a finite index, and the $p_i$ are zero(this is done in the next section—see Lemma 3.2.4 and Proposition 3.2.6).

(ii) In Section 3, we continue in the case when the $p_i$ are zero (we call this the *case of great circles*), and use overflow to obtain an approximation result for Loeb integrals over a hyperfinite dimensional sphere intersected with an internal affine subspace defined by a hyperfinite truncation of the $u^{(i)}$. Some continuity properties of our integrals then yield the limiting result for bounded uniformly continuous functions in the case of great circles.

(iii) We use the scaling and translation properties of the surface area measures to generalize the result for bounded uniformly continuous functions further to the case of non-great circles. See Theorem 3.4.1.

(iv) Using Theorem 3.4.1, it follows that almost all points of $S_{A_N}$ (where $N$ is hyperfinite) have finite coordinates along any given direction (this is Theorem 3.4.2). Using this and the notion of $S$-integrability, we are able to finally generalize to all bounded continuous functions (see Theorem 3.4.3).

In our proof, we also use a fact from asymptotic linear algebra with a nonstandard proof in an appendix (see Lemma B.1).

### 3.2. Integrating bounded functions on certain great circles

In this section, we prove Theorem 3.1.1 in the case when the following hold:

(i) The function $f \colon \mathbb{R}^k \to \mathbb{R}$ is bounded measurable.

(ii) All the $p_i$ are zero (thus $\bar{\eta}$ is the zero vector in $\mathbb{R}^k$ in this case).

(iii) The vectors $u^{(i)} \in \ell^2(\mathbb{R})$ are finite-dimensional (their sequence representations with

respect to the standard basis have finitely many nonzero terms).

We will make use of a disintegration formula from [90], which we quote below.

**Theorem 3.2.1.** *[90, Proposition 4.1, p. 19] Let $N$ and $k$ be positive integers with $k < N$, and $f$ any bounded measurable function on $S^{N-1}(a)$, the sphere in $\mathbb{R}^N = \mathbb{R}^k \times \mathbb{R}^{N-k}$ of radius $a$ and with center $0$. Then, with $\sigma$ denoting surface measure (non-normalized) on spheres, we have the following for all $a \in \mathbb{R}_{>0}$:*

$$\int_{z \in S^{N-1}(a)} f(z) d\sigma(z) = \int_{x \in B_k(a)} \left( \int_{y \in S^{N-k-1}(a_x)} f(x,y) d\sigma(y) \right) \frac{a}{a_x} dx, \tag{3.2}$$

*where $a_x = \sqrt{a^2 - ||x||^2}$ and $B_k(a)$ is the open ball of radius $a$ in $\mathbb{R}^k$.*

The following lemma ensures that the Gaussian measures appearing in this chapter are well-defined.

**Lemma 3.2.2.** *For orthonormal vectors $v^{(1)}, \ldots, v^{(\gamma)}$ in $\ell^2(\mathbb{R})$, the $(k \times k)$ matrix $I - ||(v^{(1)})_{(k)}||^2 P_{(v^{(1)})_{(k)}} - \ldots - ||(v^{(\gamma)})_{(k)}||^2 P_{(v^{(\gamma)})_{(k)}}$ is positive-semidefinite. In particular, it is the covariance matrix of a Gaussian measure on $\mathbb{R}^k$.*

*Proof.* Let $x \in \mathbb{R}^k$. Then,

$$\begin{aligned}
&\left\langle x, \left( I - ||(v^{(1)})_{(k)}||^2 P_{(v^{(1)})_{(k)}} - \ldots - ||(v^{(\gamma)})_{(k)}||^2 P_{(v^{(\gamma)})_{(k)}} \right) x \right\rangle \\
&= ||x||^2 - \sum_{i=1}^{\gamma} \left\langle x, ||(v^{(i)})_{(k)}||^2 \left\langle x, \frac{(v^{(i)})_{(k)}}{||(v^{(i)})_{(k)}||} \right\rangle \frac{(v^{(i)})_{(k)}}{||(v^{(i)})_{(k)}||} \right\rangle \\
&= ||x||^2 - \sum_{i=1}^{\gamma} \langle x, (v^{(i)})_{(k)} \rangle^2 \\
&\geq ||x||^2 - \sum_{i=1}^{\gamma} \left( \langle x, v^{(i)} \rangle_{\ell^2} \right)^2 \\
&\geq ||x||^2 - ||x||^2 = 0,
\end{aligned}$$

which completes the proof. ∎

**Notation 3.2.3.** The Gaussian measure on $\mathbb{R}^k$ with the above covariance and mean $\rho \in \mathbb{R}^k$ will be denoted by $\mu_{\rho;v^{(1)},\ldots,v^{(\gamma)}}$. In general, for a positive-semidefinite $(k \times k)$-matrix $L$, we will also use $\mu_{\rho,L}$ to denote the Gaussian measure on $\mathbb{R}^k$ with mean $\rho$ and covariance $L$.

We now study the simplest case when the $u^{(i)}$ are all in $\mathbb{R}^k$, the domain of $f$.

**Lemma 3.2.4.** *Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be a bounded measurable function. Let $u^{(1)}, \ldots, u^{(\gamma)}$ be mutually orthonormal vectors in $\mathbb{R}^k$ (hence, $\gamma \leq k$ necessarily). Then we have:*

$$\lim_{n\to\infty} \int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu_{0;u^{(1)},\ldots u^{(\gamma)}}.$$

*Proof.* Without loss of generality, let $\gamma < k$ (if $\gamma = k$, then $u^{(1)}, \ldots, u^{(\gamma)}$ span $\mathbb{R}^k$, and hence the above equality is trivial with both sides being identical to zero). Let $\{u^{(1)}, \ldots, u^{(\gamma)}, z^{(1)}, \ldots z^{(k-\gamma)}\}$ be an orthonormal basis of $\mathbb{R}^k$. Define $g\colon \mathbb{R}^{k-\gamma} \to \mathbb{R}$ by $g(y_1, \ldots, y_{k-\gamma}) = f(y_1 z^{(1)} + \ldots + y_{k-\gamma} z^{(k-\gamma)})$.

The map $T\colon S_{n,u^{(1)},\ldots,u^{(\gamma)}} \to S^{n-\gamma-1}(\sqrt{n})$ defined as follows is a measure isomorphism:

$$T\left(\sum_{i=1}^{k-\gamma} y_i z^{(i)} + \sum_{j=k+1}^{n} y_j e_j\right) := \sum_{i=1}^{n-\gamma} y_i e_i.$$

It thus follows that for any bounded measurable function $f\colon \mathbb{R}^k \to \mathbb{R}$, we have:

$$\begin{aligned}
&\int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f(x_1, \ldots, x_k) d\bar{\sigma}(x_1, \ldots, x_n) \\
=&\int_{S^{n-\gamma-1}(\sqrt{n})} g(y_1, \ldots, y_{k-\gamma}) d\bar{\sigma}(y_1, \ldots, y_{n-\gamma}) \\
=&\frac{1}{\sigma(S^{n-\gamma-1}(\sqrt{n}))} \int_{S^{n-\gamma-1}(\sqrt{n})} g(y_1, \ldots, y_{k-\gamma}) d\sigma(y_1, \ldots, y_{n-\gamma}) \\
=&\frac{1}{c_{n-\gamma-1} n^{\frac{n-\gamma-1}{2}}} \int_{B_{k-\gamma}(\sqrt{n})} \frac{g(\vec{y}) n^{\frac{1}{2}}}{(n - ||\vec{y}||^2)^{\frac{1}{2}}} \cdot \sigma\left(S^{n-\gamma-1-(k-\gamma)}\left(\sqrt{n - ||\vec{y}||^2}\right)\right) d\lambda(\vec{y}),
\end{aligned}$$

where $\vec{y} = (y_1, \ldots, y_{k-\gamma})$. We have used Theorem 3.2.1 in the last equality above. Simplifying further, we thus obtain the following:

$$
\begin{aligned}
&\int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f(x_1, \ldots, x_k) d\bar{\sigma}(x_1, \ldots, x_n) \\
=&\frac{\left(\frac{2\pi^{\frac{n-k}{2}}}{\Gamma\left(\frac{n-k}{2}\right)}\right)}{\left(\frac{2\pi^{\frac{n-\gamma}{2}}}{\Gamma\left(\frac{n-\gamma}{2}\right)}\right)} \cdot \frac{1}{n^{\frac{n-\gamma-2}{2}}} \cdot \int_{B_{k-\gamma}(\sqrt{n})} g(y_1, \ldots, y_{k-\gamma})(n - ||\vec{y}||^2)^{\frac{n-k-2}{2}} d\lambda(\vec{y}) \\
=&\frac{1}{(2\pi)^{\frac{k-\gamma}{2}}} \cdot \frac{\Gamma\left(\frac{n-k}{2} + \frac{k-\gamma}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right) \cdot \left(\frac{n-k}{2}\right)^{\frac{k-\gamma}{2}}} \\
&\quad \cdot \int_{B_{k-\gamma}(\sqrt{n})} g(y_1, \ldots, y_{k-\gamma}) \cdot \frac{\left((n - ||\vec{y}||^2)^{\frac{n-k-2}{2}}\right) \cdot (n-k)^{\frac{k-\gamma}{2}}}{n^{\frac{n-\gamma-2}{2}}} d\lambda(\vec{y}) \\
=&\frac{a_{n,k} b_{n,k}}{(2\pi)^{\frac{k-\gamma}{2}}} \cdot \int_{B_{k-\gamma}(\sqrt{n})} g(y_1, \ldots, y_{k-\gamma}) \cdot \left(1 - \frac{||\vec{y}||^2}{n}\right)^{\frac{n-k-2}{2}} d\lambda(\vec{y}),
\end{aligned}
\tag{3.3}
$$

where $a_{n,k} = \dfrac{\Gamma\left(\frac{n-k}{2} + \frac{k-\gamma}{2}\right)}{\Gamma\left(\frac{n-k}{2}\right) \cdot \left(\frac{n-k}{2}\right)^{\frac{k-\gamma}{2}}}$ and $b_{n,k} = \left(1 - \dfrac{k}{n}\right)^{\frac{k-\gamma}{2}}$. Note that

$$\lim_{n\to\infty} a_{n,k} = 1 = \lim_{n\to\infty} b_{n,k}.$$

Since $f$ is bounded, therefore for large values of $n$, the integrand in (3.3) is bounded by $||f||_\infty \cdot e^{-\frac{||y||^2}{4}}$ in absolute value, the latter being integrable on $\mathbb{R}^{k-\gamma}$. We thus obtain the following by dominated convergence theorem:

$$
\begin{aligned}
&\lim_{n\to\infty} \int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f(x_1, \ldots, x_k) d\bar{\sigma}(x_1, \ldots, x_n) \\
=&\frac{1}{(2\pi)^{\frac{k-\gamma}{2}}} \cdot \int_{\mathbb{R}^{k-\gamma}} g(y_1, \ldots, y_{k-\gamma}) \cdot e^{\frac{-||\vec{y}||^2}{2}} d\lambda(\vec{y}) \\
=&\frac{1}{(2\pi)^{\frac{k-\gamma}{2}}} \cdot \int_{\mathbb{R}^{k-\gamma}} f(y_1 z^{(1)} + \ldots + y_{k-\gamma} z^{(k-\gamma)}) \cdot e^{\frac{-||\vec{y}||^2}{2}} d\lambda(\vec{y}) \\
=&\frac{1}{(2\pi)^{\frac{k}{2}}} \cdot \int_{\mathbb{R}^k} f(y_1 z^{(1)} + \ldots + y_{k-\gamma} z^{(k-\gamma)}) \cdot e^{-\frac{{y_1}^2 + \ldots + {y_k}^2}{2}} d\lambda(\tilde{y}),
\end{aligned}
\tag{3.4}
$$

where $\tilde{y} = (y_1, \ldots, y_k)$. (3.4) follows from Fubini's theorem, using the fact that the integral over the last $\gamma$ coordinates is $(2\pi)^{\frac{\gamma}{2}}$. Rewriting (3.4), we have:

$$\lim_{n\to\infty} \int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f d\bar{\sigma} = \frac{1}{(2\pi)^{\frac{k}{2}}} \cdot \int_{\mathbb{R}^k} f(P_{u^{(1)},\ldots,u^{(\gamma)}}(\tilde{y})) \cdot e^{\frac{-||\tilde{y}||^2}{2}} d\lambda(\tilde{y})$$
$$= \int_{\mathbb{R}^k} f d\mu_{0;u^{(1)},\ldots,u^{(\gamma)}}, \tag{3.5}$$

completing the proof. $\square$

The following basic fact about Gaussian measures allows us to strengthen Lemma 3.2.4 to the case when the vectors $u^{(1)}, \ldots u^{(\gamma)}$ are vectors in $\ell^2(\mathbb{R})$ that are eventually zero (but not necessarily zero after the $k^{\text{th}}$ coordinate)

**Lemma 3.2.5.** *Let $u^{(1)}, \ldots, u^{(\gamma)}$ be orthonormal vectors in $\mathbb{R}^m$ where $m \in \mathbb{N}_{>k}$. Let $\mu'$ be the Gaussian measure on $\mathbb{R}^m$ with mean $\mathbf{0} \in \mathbb{R}^m$ and covariance $I - \sum_{i=1}^{\gamma} P_{u^{(i)}}$. Let $\mu_{0;u^{(1)},\ldots,u^{(\gamma)}}$ be the Gaussian measure on $\mathbb{R}^k$ with mean $0 \in \mathbb{R}^k$ and covariance as in (3.1). For any bounded measurable function $f\colon \mathbb{R}^k \to \mathbb{R}$, we have:*

$$\int_{\mathbb{R}^m} f(x_1, \ldots, x_m) d\mu' = \int_{\mathbb{R}^k} f(x_1, \ldots, x_k) d\mu_{0;u^{(1)},\ldots,u^{(\gamma)}}. \tag{3.6}$$

*Proof.* The collection of functions satisfying (3.6) is closed under taking $\mathbb{R}$-linear combinations and uniform limits. Hence, it is enough to show that indicator functions of Borel subsets of $\mathbb{R}^k$ satisfy (3.6). If $X \sim N(\rho, \Sigma)$ is an $m$-dimensional Gaussian random variable, then $X_{(k)} \sim N(\rho_{(k)}, \Sigma_{(k,k)})$. Let $\rho = \mathbf{0} \in \mathbb{R}^m$ and $\Sigma$ be the matrix of the operator on $\mathbb{R}^m$ given by $I_m - \sum_{i=1}^{\gamma} P_{u^{(i)}}$. Note that for any $i \in \{1, \ldots, \gamma\}$ and $j \in \{1, \ldots, k\}$, we have

$\langle e_j, u^{(i)} \rangle = \langle e_j, (u^{(i)})_{(k)} \rangle$, which implies

$$\begin{aligned}
(P_{u^{(i)}}(e_j))_{(k)} &= (\langle e_j, u^{(i)} \rangle u^{(i)})_{(k)} \\
&= \left|\left|(u^{(i)})_{(k)}\right|\right|^2 \left\langle e_j, \frac{(u^{(i)})_{(k)}}{\left|\left|(u^{(i)})_{(k)}\right|\right|} \right\rangle \frac{(u^{(i)})_{(k)}}{\left|\left|(u^{(i)})_{(k)}\right|\right|} \\
&= \left|\left|(u^{(i)})_{(k)}\right|\right|^2 P_{(u^{(i)})_{(k)}}(e_j).
\end{aligned}$$

Thus the operator represented by $\Sigma_{(k,k)}$ is

$$I_k - \left|\left|(u^{(1)})_{(k)}\right|\right|^2 P_{(u^{(1)})_{(k)}} - \ldots - \left|\left|(u^{(\gamma)})_{(k)}\right|\right|^2 P_{(u^{(\gamma)})_{(k)}},$$

which completes the proof. □

**Proposition 3.2.6.** *If $u^{(1)}, \ldots, u^{(\gamma)}$ are orthonormal vectors in $\mathbb{R}^m$ and $f\colon \mathbb{R}^k \to \mathbb{R}$ is a bounded measurable function, then*

$$\lim_{n\to\infty} \int_{S_{n,u^{(1)},\ldots,u^{(\gamma)}}} f(x_1, \ldots, x_k) d\bar{\sigma}(x_1, \ldots, x_n) = \int_{\mathbb{R}^k} f d\mu_{0;u^{(1)},\ldots,u^{(\gamma)}}. \tag{3.7}$$

*Proof.* In the case when $m \leq k$, this follows from Lemma 3.2.4. Now suppose $m > k$. By Lemma 3.2.4, the limit on the left side of (3.7) is equal to

$$\int_{\mathbb{R}^m} f(x_1, \ldots, x_m) d\mu',$$

where $\mu'$ is the Gaussian measure on $\mathbb{R}^m$ with mean 0 and covariance $I_m - \sum_{i=1}^{\gamma} P_{u^{(i)}}$. The proof is now completed by Lemma 3.2.5. □

### 3.3. A hyperfinite approximation and integrating on any great circle

Throughout this section, $N > \mathbb{N}$ will be a hyperfinite number. The goal of this section is to generalize Theorem 3.1.1 to the case when the function $f\colon \mathbb{R}^k \to \mathbb{R}$ is continuous with compact support (we will henceforth write $f \in C_c(\mathbb{R}^k)$), while the $p_i$ are zero

(i.e., $S_{A_N}$ is a great circle on $S^{N-1}(\sqrt{N})$), with no restriction on the orthonormal vectors $u^{(1)}, \dots, u^{(\gamma)} \in \ell^2(\mathbb{R})$. The key steps are as follows:

(i) Show that the ${}^*$-integral of ${}^*f$ does not change much when $S^{N-1}(\sqrt{N})$ is intersected by two different internal hyperplanes that are infinitesimally close enough to each other.

(ii) In view of (i) and certain continuity properties of the Gaussian integral (with varying covariance), prove some continuity results that show that the integrals in Theorem 3.1.1 do not change much when we work with two different hyperfinite truncations of the $u^{(i)}$.

(iii) Use overflow together with the results of Section 2 to get an approximation result when the vectors $u^{(i)}$ are zero after a small but hyperfinite index $M$. Then, use (ii) to complete the proof.

These three ideas will be pursued in the next three subsections respectively.

### 3.3.1. Effect of an infinitesimal rotation on the ${}^*$-integral of a bounded uniformly continuous function

**Proposition 3.3.1.** *Let $N > \mathbb{N}$ and $R \in {}^*\mathbb{R}_{>0}$ be such that $\dfrac{R}{\sqrt{N}} \in {}^*\mathbb{R}_{fin}$. Then, almost all points of $S^{N-1}(R)$ have finite coordinates in any given direction, i.e.,*

$$L\bar{\sigma}(\{(x_1, \dots, x_N) \in S^{N-1}(R) : x_i \in {}^*\mathbb{R}_{fin}\}) = 1 \textit{ for all } i \leq N.$$

*As a consequence, for any $v^{(1)}, \dots, v^{(\gamma)} \in {}^*\mathbb{R}^N$, almost all points on the sphere $S^{N-1}(R) \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp}$ have finite coordinates along a given direction (unit) vector $w \in {}^*\mathbb{R}^N$. That is, given a unit vector $w \in {}^*\mathbb{R}^N$, almost all points $x$ on this sphere have $\langle x, w \rangle \in {}^*\mathbb{R}_{fin}$.*

*Proof.* The first half of the proposition was proved in [4, Corollary 4.3] for the case $R = \sqrt{N}$ (i.e. for the sphere $S^{N-1}(\sqrt{N})$). The result for $S^{N-1}(R)$ in general then follows from the (transfer of) scaling property of uniform surface area measures. Indeed, for any $B \in$

${}^*\mathcal{B}(S^{N-1}(\sqrt{N}))$, we have:

$$\bar{\sigma}_{S^{N-1}(\sqrt{N})}(B) = \bar{\sigma}_{S^{N-1}(R)}\left(\frac{R}{\sqrt{N}}B\right).$$

Now, let

$$S' := S^{N-1}(R) \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp}.$$

Also, let $L$ be the internal span $\left({}^*\mathbb{R}v^{(1)} + \ldots + {}^*\mathbb{R}v^{(\gamma)}\right)$ of $v^{(1)}, \ldots, v^{(\gamma)}$ in ${}^*\mathbb{R}^N$. Consider an arbitrary unit vector $w \in {}^*\mathbb{R}^N$. We want to show that almost all points on $S'$ have finite coordinates along $w$, i.e., $\langle x, w\rangle \in {}^*\mathbb{R}_{\text{fin}}$ for $L\bar{\sigma}_{S'}$-almost all $x \in S'$. Let $w'$ and $w''$ be the orthogonal projections of $w$ onto $L$ and its orthogonal complement $L^{\perp}$ (in ${}^*\mathbb{R}^N$) respectively. Since $S' \subseteq L^{\perp}$, we have:

$$\langle x, w\rangle = \langle x, w'\rangle + \langle x, w''\rangle = \langle x, w''\rangle \text{ for all } x \in S'.$$

If $w'' = 0$, then clearly all points of $S$ have the coordinate 0 along $w$. Otherwise, if $w''$ is not zero, then define

$$w^{(1)} := \frac{w''}{||w''||}.$$

Let $c = {}^*\dim(L)$. It is clear that $c \leq \gamma$. Extend $w^{(1)}$ to an orthonormal basis $\{w^{(1)}, \ldots, w^{(N-c)}\}$ of $L^{\perp} = {}^*\mathbb{R}^N \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp}$. Consider the map $\phi\colon {}^*\mathbb{R}^N \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp} \to {}^*\mathbb{R}^{N-c}$ defined by

$$\phi(w^{(i)}) = \mathbf{e}_i \text{ for all } i \in [N-c].$$

The map $\phi$ restricted to $S'$ is a measure isomorphism onto $S^{N-1-c}(R)$. The first half of the proposition (applied to $S^{N-1-c}(R)$) now completes the proof. $\square$

In the following, we use the concept of Separation Property (SP) defined in Appendix C; see (C.2). Roughly speaking, a set of vectors satisfy SP if they are linearly independent in a non-infinitesimal way (this is made precise in Appendix C). The hypothesis of Theorem 3.3.2 is the same as that of Theorem C.2 (with ${}^*\mathbb{R}^N$ as the ambient internal vector space).

**Theorem 3.3.2.** *Fix $N > \mathbb{N}$. For each $i \in \{1, \ldots, \gamma\}$, let $v^{(i)}, v'^{(i)} \in ({}^*\mathbb{R})^N$ be such that the following conditions hold:*

*(i) The collections $\{v^{(1)}, \ldots, v^{(\gamma)}\}$ and $\{v'^{(1)}, \ldots, v'^{(\gamma)}\}$ both satisfy the Separation Property (see (C.2)).*

*(ii) $\left|\left|v^{(i)}\right|\right|, \left|\left|v^{(i)}\right|\right| \in {}^*\mathbb{R}_{fin}$.*

*(iii) $\left|\left|v^{(i)} - v'^{(i)}\right|\right| \approx 0$.*

*Then for any bounded and uniformly continuous $f \colon \mathbb{R}^k \to \mathbb{R}$, we have:.*

$$\int\limits_{S^{(1)}} \mathbf{st}({}^*f(x)) dL\bar{\sigma}(x) = \int\limits_{S^{(2)}} \mathbf{st}({}^*f(x)) dL\bar{\sigma}(x),$$

*where $S^{(1)} = S^{N-1}(\sqrt{N}) \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp}$ and $S^{(2)} = S^{N-1}(\sqrt{N}) \cap {v'^{(1)}}^{\perp} \cap \ldots \cap {v'^{(\gamma)}}^{\perp}$.*

*Proof.* The idea is to first show that $S^{(1)}$ and $S^{(2)}$ are spheres of the same topological dimension that are infinitesimally apart (to be made precise below). See also Figure 3.2. Note that the hypotheses on the sets $\{v^{(1)}, \ldots, v^{(\gamma)}\}$ and $\{v'^{(1)}, \ldots, v'^{(\gamma)}\}$ are the same as in Theorem C.2. Thus, using Theorem C.2, we obtain orthonormal sets of vectors $\{w^{(1)}, \ldots, w^{(\gamma)}\}$ and $\{z^{(1)}, \ldots, z^{(\gamma)}\}$ such that the following hold:

1. For any $i \in \{1, \ldots, \gamma\}$, we have $\text{span}(v^{(1)}, \ldots, v^{(i)}) = \text{span}(w^{(1)}, \ldots, w^{(i)})$ and $\text{span}(v'^{(1)}, \ldots, v'^{(i)}) = \text{span}(z^{(1)}, \ldots, z^{(i)})$.

2. For all $i \in \{1, \ldots, \gamma\}$, we have $\left|\left|w^{(i)} - z^{(i)}\right|\right| \approx 0$.

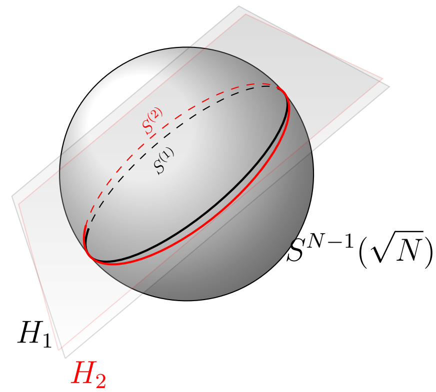


Figure 3.2. $S^{(1)}$ and $S^{(2)}$ are separated infinitesimally

Define the internal subspaces $H_1 := \{x \in {}^*\mathbb{R}^N : \langle x, w^{(i)} \rangle = 0$ for all $i \in \{1, \ldots, \gamma\}\}$ and $H_2 := \{x \in {}^*\mathbb{R}^N : \langle x, z^{(i)} \rangle = 0$ for all $i \in \{1, \ldots, \gamma\}\}$. Therefore,

$$S^{(1)} = S^{N-1}(\sqrt{N}) \cap {v^{(1)}}^{\perp} \cap \ldots \cap {v^{(\gamma)}}^{\perp} = S^{N-1}(\sqrt{N}) \cap H_1, \text{ and}$$

$$S^{(2)} = S^{N-1}(\sqrt{N}) \cap {v'^{(1)}}^{\perp} \cap \ldots \cap {v'^{(\gamma)}}^{\perp} = S^{N-1}(\sqrt{N}) \cap H_2.$$

Note that $\dim(H_1 \cap H_2) \geq N - 2\gamma$. Obtain an internally orthonormal set $\{c_1, \ldots, c_{N-2\gamma}\}$ in $H_1 \cap H_2$. For $i \in \{1, \ldots, N - 2\gamma\}$, define $w^{(\gamma+i)} = z^{(\gamma+i)} = c_i$. We thus have internally orthonormal sets $\{w^{(1)}, \ldots w^{(N-\gamma)}\}$ and $\{z^{(1)}, \ldots z^{(N-\gamma)}\}$ such that $||w^{(i)} - z^{(i)}|| \approx 0$ for all $i \in \{1, \ldots, N - \gamma\}$. Now extend to an internal orthonormal basis $\{w^{(1)}, \ldots, w^{(N-\gamma)}, \ldots, w^{(N)}\}$ of ${}^*\mathbb{R}^N$, and inductively define the following for $i \in \{1, \ldots, \gamma\}$:

$$z^{(N-\gamma+1)} := \frac{w^{(N-\gamma+1)} - \sum_{j=1}^{\gamma} \langle z^{(j)}, w^{(N-\gamma+1)} \rangle z^{(j)}}{\left|\left|w^{(N-\gamma+1)} - \sum_{j=1}^{\gamma} \langle z^{(j)}, w^{(N-\gamma+1)} \rangle z^{(j)}\right|\right|},$$

$$z^{(N-\gamma+i+1)} := \frac{z}{||z||}, \text{ where}$$

$$z = w^{(N-\gamma+i+1)} - \sum_{j=1}^{\gamma} \langle z^{(j)}, w^{(N-\gamma+i+1)} \rangle z^{(j)} - \sum_{l=1}^{i} \langle z^{(N-\gamma+l)}, w^{(N-\gamma+i+1)} \rangle z^{(N-\gamma+l)}.$$

It is straightforward to verify that $\{z^{(1)}, \ldots, z^{(N)}\}$ is also an internal orthonormal basis of ${}^*\mathbb{R}^N$ (orthonormality follows by construction and we then use the transfer of the standard fact about Euclidean spaces that says that all orthonormal sets containing as many elements as the dimension span the space). Furthermore, $\left|\left|w^{(i)} - z^{(i)}\right|\right| \approx 0$ for all $i \in \{1, \ldots, N\}$ by construction.

Define a ${}^*\mathbb{R}$-linear map $R\colon {}^*\mathbb{R}^N \to {}^*\mathbb{R}^N$ by $R(w^{(i)}) = z^{(i)}$ for all $i \in \{1, \ldots, N\}$. Since $R$ takes an internal orthonormal basis to an internal orthonormal basis, it is an internal orthogonal map. Also, $R(S^{(1)}) = S^{(2)}$. By transfer, it follows that for any $A \in {}^*\mathcal{B}(S^{(1)})$, we have $\bar{\sigma}_{S^{(2)}}(R(A)) = \bar{\sigma}_{S^{(1)}}(A)$. Hence, it follows that

$$L\bar{\sigma}_{S^{(2)}}(R(A)) = L\bar{\sigma}_{S^{(1)}}(A) \text{ for all } A \in {}^*\mathcal{B}(S^{(1)}).$$

By a change of variables argument, we conclude the following for any bounded measurable function $g\colon \mathbb{R}^k \to \mathbb{R}$:

$$\int_{S^{(2)}} \mathbf{st}({}^*g(x))dL\bar{\sigma}(x) = \int_{S^{(1)}} \mathbf{st}({}^*g(R(x)))dL\bar{\sigma}(x).$$

Thus it suffices to prove that the following holds for all bounded and uniformly continuous functions $f\colon \mathbb{R}^k \to \mathbb{R}$.

$$\int_{S^{(1)}} \mathbf{st}({}^*f(x))dL\bar{\sigma}(x) = \int_{S^{(1)}} \mathbf{st}({}^*f(R(x)))dL\bar{\sigma}(x) \tag{3.8}$$

In order to show (3.8), we first need the following claim.

**Claim 3.3.3.** *We have $||x - R(x)|| \approx 0$ for almost all $x \in S^{(1)}$.*

*Proof of Claim 3.3.3.* Note that $x = \sum_{i=1}^{N} \langle x, w^{(i)} \rangle w^{(i)}$ for any $x \in {}^*\mathbb{R}^N$. Since $w^{(i)} = z^{(i)}$ for all $i \in \{\gamma + 1, \ldots, N - \gamma\}$, the facts that $R(w^{(i)}) = z^{(i)}$ for all $i$ and that $\langle x, w^{(i)} \rangle = 0$

for all $i \in \{1, \ldots, \gamma\}$ imply that

$$||x - R(x)|| = \left|\left|\sum_{i=N-\gamma+1}^{N} \langle x, w^{(i)} \rangle (w^{(i)} - z^{(i)})\right|\right| \text{ for all } x \in S^{(1)}.$$

By Proposition 3.3.1, it follows that for each $i > \gamma$, almost surely $\langle x, w^{(i)} \rangle \in {}^*\mathbb{R}_{\text{fin}}$. Thus, being a maximum of finitely many elements of ${}^*\mathbb{R}_{\text{fin}}$, we have:

$$\max_{N-\gamma+1 \leq i \leq N} \left|\langle x, w^{(i)} \rangle\right| = s(x) \text{ (say)},$$

where $s(x) \in {}^*\mathbb{R}_{\text{fin}}$ for almost all $x \in S^{(1)}$. Hence, for almost all $x \in S^{(1)}$, we get:

$$||x - R(x)|| \leq s(x) \cdot \sum_{i=N-\gamma+1}^{N} \left|\left|w^{(i)} - z^{(i)}\right|\right| \approx 0, \text{ as desired.}$$

Using Claim 3.3.3, we have $||{}^*\pi_k(x) - {}^*\pi_k(R(x))|| \approx 0$ for almost all $x \in S^{(1)}$. Thus, by the nonstandard characterization of uniform continuity, we have that $\mathbf{st}({}^*f(x)) = \mathbf{st}({}^*f(R(x)))$ for almost all $x \in S^{(1)}$. This completes the proof. $\square$

### 3.3.2. Some integral continuity properties

**Definition 3.3.4.** Let PSD be the set of all positive-semidefinite $(k \times k)$-matrices with real entries. For a bounded measurable function $f \colon \mathbb{R}^k \to \mathbb{R}$, we define $G_f : \text{PSD} \to \mathbb{R}$ to be the function that maps $L$ to the expectation of (our fixed function) $f$ with respect to the Gaussian measure on $\mathbb{R}^k$ with mean 0 and covariance $L$, i.e.,

$$G_f(L) = \int_{\mathbb{R}^k} f d\mu_{0,L}.$$

Being a subset of the space of linear operators on $\mathbb{R}^k$, the space PSD inherits the metric induced by the operator norm.

**Lemma 3.3.5.** *With respect to the operator norm on PSD, the map $G_f$ is continuous for all bounded continuous $f\colon \mathbb{R}^k \to \mathbb{R}$.*

*Proof.* Let $L_n \to L$ in PSD. It suffices to prove that $\mu_{0,L_n} \to \mu_{0,L}$ weakly. Equivalently, we want to show that $Z_n \to Z$ in distribution, where $Z_n \sim N(0, L_n)$ and $Z \sim N(0, L)$. Since $Z_n$ and $Z$ are $\mathbb{R}^k$-valued Gaussian random variables, $Z_n \to Z$ in distribution if and only if for any $\vec{x} \in \mathbb{R}^k$, the Gaussian random variables $\langle \vec{x}, Z_n\rangle$ converge in distribution to $\langle \vec{x}, Z\rangle$ (see, for example, [56, Corollary 4.5, p. 64]). Toward that end, fix $\vec{x} \in \mathbb{R}^k$. We know that $\langle \vec{x}, Z_n\rangle \sim N(0, \langle \vec{x}, L_n\vec{x}\rangle)$, while $\langle \vec{x}, Z\rangle \sim N(0, \langle \vec{x}, L\vec{x}\rangle)$. For real-valued Gaussian random variables, convergence in distribution is equivalent to the convergence of means and variances. The proof is thus completed by the observation that $L_n\vec{x} \to L\vec{x}$ (which follows from the fact that $L_n \to L$ in operator norm). □

**Definition 3.3.6.** For an inner product space $V$ and $\gamma \in \mathbb{N}$, let $V^{[\gamma]}$ be the set of $\gamma$-tuples of orthonormal vectors from $V$.

**Definition 3.3.7.** For a bounded measurable function $f\colon \mathbb{R}^k \to \mathbb{R}$ and $m \in \mathbb{N}_{\geq k}$, let $\theta_{f,m}\colon (\mathbb{R}^m)^{[\gamma]} \to \mathbb{R}$ and $a_{f,m}\colon (\mathbb{R}^m)^\gamma \times \mathbb{N}_{\geq k} \to \mathbb{R}$ be defined by

$$\theta_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}) := \int_{\mathbb{R}^k} f d\mu_{0;v^{(1)},\ldots,v^{(\gamma)}},$$
$$a_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}, n) := \int_{S_{n,v^{(1)},\ldots,v^{(\gamma)}}} f d\bar{\sigma}.$$

Here, $S_{n,v^{(1)},\ldots,v^{(\gamma)}}$ is equal to the intersection of $S^{n-1}(\sqrt{n})$ with $\cap_{i\leq\gamma}(v^{(i)})^\perp$. Note that $\theta_{f,m}$ is defined only on $(\mathbb{R}^m)^{[\gamma]}$ (instead of on $(\mathbb{R}^m)^\gamma$) since $\mu_{0;v^{(1)},\ldots,v^{(\gamma)}}$ is well-defined for $(v^{(1)}, \ldots, v^{(\gamma)}) \in (\mathbb{R}^m)^{[\gamma]}$ by Lemma 3.2.2 while it may not be defined in general for an arbitrary set of vectors in $\mathbb{R}^m$. On the other hand, $a_{f,m}(\cdot, n)$ is defined on all of $(\mathbb{R}^m)^\gamma$,

though we will usually only be interested in the case when $v^{(1)}, \ldots, v^{(\gamma)}$ are truncations of orthonormal vectors in $\ell^2(\mathbb{R})$.

The space $(\mathbb{R}^m)^{[\gamma]}$ inherits a metric from $(\mathbb{R}^m)^\gamma$. The next two lemmas respectively prove that under the topology of that metric, the function $\theta_{f,m}$ is continuous and that $a_{f,m}(\cdot, N)\colon {}^*(\mathbb{R}^m)^{[\gamma]} \to {}^*\mathbb{R}$ is $S$-continuous (an internal function is called *$S$-continuous* if it maps points that are infinitesimally close in the domain to points that are infinitesimally close in the range).

**Lemma 3.3.8.** *For each bounded continuous function $f\colon \mathbb{R}^k \to \mathbb{R}$ and $m \in \mathbb{N}$, the map $\theta_{f,m}$ is continuous. In other words, if $(v^{(1)}, \ldots, v^{(\gamma)}) \in (\mathbb{R}^m)^{[\gamma]}$ and $(v'^{(1)}, \ldots, v'^{(\gamma)}) \in ({}^*\mathbb{R}^m)^{[\gamma]}$ are such that $\left|\left|v^{(i)} - v'^{(i)}\right|\right| \approx 0$ for each $i \in \{1, \ldots, \gamma\}$, then we have $\theta_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}) \approx {}^*\theta_{f,m}(v'^{(1)}, \ldots, v'^{(\gamma)})$.*

*Proof.* The two statements in the lemma are equivalent by the nonstandard characterization of continuity. We will prove the latter statement. Toward that end, fix $(v^{(1)}, \ldots, v^{(\gamma)}) \in (\mathbb{R}^m)^{[\gamma]}$ and $(v'^{(1)}, \ldots, v'^{(\gamma)}) \in ({}^*\mathbb{R}^m)^{[\gamma]}$ such that

$$\left|\left|v^{(i)} - v'^{(i)}\right|\right| \approx 0 \text{ for each } i \in \{1, \ldots, \gamma\}.$$

Let

$$L = I - \left|\left|(v^{(1)})_{(k)}\right|\right|^2 P_{(v^{(1)})_{(k)}} - \ldots - \left|\left|(v^{(\gamma)})_{(k)}\right|\right|^2 P_{(v^{(\gamma)})_{(k)}}, \text{ and}$$
$$L' = I - \left|\left|(v'^{(1)})_{(k)}\right|\right|^2 P_{(v'^{(1)})_{(k)}} - \ldots - \left|\left|(v^{(\gamma)})_{(k)}\right|\right|^2 P_{(v'^{(\gamma)})_{(k)}}.$$

By Definition 3.3.4 and transfer, we have

$$\theta_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}) = G_f(L), \text{ and } {}^*\theta_{f,m}(v'^{(1)}, \ldots, v'^{(\gamma)}) = {}^*G_f(L').$$

By Lemma 3.3.5, the map $G_f$ is continuous. Hence, by nonstandard characterization of continuity, it suffices to show that $||L - L'||_{op} \approx 0$. This is straightforward if one uses the representation of the projection operator as given by the inner product in the direction of the projection. □

**Lemma 3.3.9.** *Let $N > \mathbb{N}$. If $f\colon \mathbb{R}^k \to \mathbb{R}$ is bounded and uniformly continuous, then the map $a_{f,m}(\cdot, N)\colon {}^*(\mathbb{R}^m)^{[\gamma]} \to {}^*\mathbb{R}$ is S-continuous.*

*Equivalently, if $(v^{(1)}, \ldots, v^{(\gamma)})$ and $(v'^{(1)}, \ldots, v'^{(\gamma)}) \in ({}^*\mathbb{R}^m)^{[\gamma]}$ are such that $\left|\left|v^{(i)} - v'^{(i)}\right|\right| \approx 0$ for each $i \in \{1, \ldots, \gamma\}$, then*

$${}^*a_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}, N) \approx {}^*a_{f,m}(v'^{(1)}, \ldots, v'^{(\gamma)}, N) \text{ for all } N > \mathbb{N}.$$

*Proof.* This is immediate from Theorem 3.3.2 followed by applications of the transfer principle and the S-integrability of finitely bounded internal functions. □

### 3.3.3. A hyperfinite approximation via overflow

**Theorem 3.3.10.** *With $u^{(1)}, \ldots, u^{(\gamma)}$ orthonormal in $\ell^2(\mathbb{R})$, for any bounded and uniformly continuous $f\colon \mathbb{R}^k \to \mathbb{R}$, we have*

$$\lim_{n\to\infty} \int_{S^{n-1}(\sqrt{n}) \cap u^{(1)\perp} \cap \ldots \cap u^{(\gamma)\perp}} f(x) d\bar{\sigma}(x) = \int_{\mathbb{R}^k} f(x) d\mu_{0;u^{(1)},\ldots,u^{(\gamma)}}(x).$$

*Proof.* Fix $f$ as above. Since the limit of a sequence, if it exists, is the same as the standard part of any element with a hyperfinite index in the nonstandard extension of the sequence, it suffices to show that $\mathbf{st}\left({}^*a_{f,N}\left(u^{(1)}_{(N)}, \ldots, u^{(\gamma)}_{(N)}, N\right)\right)$ equals $\int_{\mathbb{R}^k} f(x) d\mu_0(x)$. Con-

sider the following internal set:

$$\mathcal{G} := \Bigg\{ m \in {}^*\mathbb{N} : m \leq N, \text{ and } \forall (v^{(1)}, \ldots, v^{(\gamma)}) \in ({}^*\mathbb{R}^m)^{[\gamma]}$$
$$\left( \left| {}^*a_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}, N) - {}^*\theta_{f,m}(v^{(1)}, \ldots, v^{(\gamma)}) \right| < \frac{1}{m} \right) \Bigg\}.$$

By Lemma 3.3.8, Lemma 3.3.9 and Corollary 3.2.4, it follows that $\mathbb{N} \subseteq \mathcal{G}$. By overflow, there exists $M > \mathbb{N}$ such that $\{1, \ldots, M\} \subseteq \mathcal{G}$. Fix this $M$. By Lemma B.1, the vectors $(u^{(1)})_{(M)}, \ldots, (u^{(\gamma)})_{(M)}$ are ${}^*\mathbb{R}$-linearly independent. Use the Gram-Schmidt algorithm to get orthonormal vectors with the same linear span:

$$\begin{aligned}
w^{(1)} &:= \frac{(u^{(1)})_{(M)}}{\left|\left|(u^{(1)})_{(M)}\right|\right|}, \qquad (3.9)\\
w^{(2)} &:= \frac{(u^{(2)})_{(M)} - \langle (u^{(2)})_{(M)}, w^{(1)} \rangle w^{(1)}}{\left|\left|(u^{(2)})_{(M)} - \langle (u^{(2)})_{(M)}, w^{(1)} \rangle w^{(1)}\right|\right|},\\
w^{(3)} &:= \frac{(u^{(3)})_{(M)} - \langle (u^{(3)})_{(M)}, w^{(1)} \rangle w^{(1)} - \langle (u^{(3)})_{(M)}, w^{(2)} \rangle w^{(2)}}{\left|\left|(u^{(3)})_{(M)} - \langle (u^{(3)})_{(M)}, w^{(1)} \rangle w^{(1)} - \langle (u^{(3)})_{(M)}, w^{(2)} \rangle w^{(2)}\right|\right|},\\
&\vdots\\
w^{(\gamma)} &:= \frac{(u^{(\gamma)})_{(M)} - \langle (u^{(\gamma)})_{(M)}, w^{(1)} \rangle w^{(1)} - \ldots - \langle (u^{(\gamma)})_{(M)}, w^{(\gamma-1)} \rangle w^{(\gamma-1)}}{\left|\left|(u^{(\gamma)})_{(M)} - \langle (u^{(\gamma)})_{(M)}, w^{(1)} \rangle w^{(1)} - \ldots - \langle (u^{(\gamma)})_{(M)}, w^{(\gamma-1)} \rangle w^{(\gamma-1)}\right|\right|}.
\end{aligned}$$

Since $M \in \mathcal{G}$, we have

$$\left| {}^*a_{f,M}(w^{(1)}, \ldots, w^{(\gamma)}, N) - {}^*\theta_{f,M}(w^{(1)}, \ldots, w^{(\gamma)}) \right| < \frac{1}{M} \approx 0. \qquad (3.10)$$

Since $\langle u^{(i)}, u^{(j)} \rangle_{\ell^2(\mathbb{R})} = \lim_{m\to\infty} \langle (u^{(i)})_{(m)}, (u^{(j)})_{(m)} \rangle = 0$ if $i \neq j$, the nonstandard characterization of limits implies that

$$\langle (u^{(i)})_{(M)}, (u^{(j)})_{(M)} \rangle \approx 0 \text{ for } i \neq j. \qquad (3.11)$$

Similarly, $\left|\left|(u^{(i)})_{(M)}\right|\right| \approx \lim_{m\to\infty} \left|\left|(u^{(i)})_{(M)}\right|\right| = 1$, and $\langle (u^{(i)})_{(M)}, w^{(j)} \rangle \approx 0$ for all $j \in \{1, \ldots, \gamma\} \backslash \{i\}$ (for a given $i \in \{1, \ldots, j\}$, this follows by induction on $j$ using (3.11)).

Truncating to the first $k$ coordinates, we thus obtain (by induction on $i$):

$$\left|\left|(w^{(i)})_k - (u^{(i)})_{(k)}\right|\right| \approx 0 \text{ for all } i. \tag{3.12}$$

Hence the covariance matrix defined by $(w^{(1)})_{(k)}, \ldots, (w^{(\gamma)})_{(k)}$ is infinitesimally close to that defined by $(u^{(1)})_{(k)}, \ldots, (u^{(\gamma)})_{(k)}$ in ${}^*$operator norm, i.e.,

$$\begin{aligned}\Bigg|\Bigg| &\left(I - \left|\left|(w^{(1)})_{(k)}\right|\right|^2 P_{(w^{(1)})_{(k)}} - \ldots - \left|\left|(w^{(\gamma)})_{(k)}\right|\right|^2 P_{(w^{(\gamma)})_{(k)}}\right) \\ &- \left(I - \left|\left|(u^{(1)})_{(k)}\right|\right|^2 P_{(u^{(1)})_{(k)}} - \ldots - \left|\left|(u^{(\gamma)})_{(k)}\right|\right|^2 P_{(u^{(\gamma)})_{(k)}}\right) \Bigg|\Bigg| \approx 0.\end{aligned} \tag{3.13}$$

The continuity of $G_f$ thus yields the following:

$${}^*G_f\left(I - \left|\left|(w^{(1)})_{(k)}\right|\right|^2 P_{(w^{(1)})_{(k)}} - \ldots - \left|\left|(w^{(\gamma)})_{(k)}\right|\right|^2 P_{(w^{(\gamma)})_{(k)}}\right) \approx \int_{\mathbb{R}^k} f d\mu_0$$

Also, by transfer we have:

$$\begin{aligned}{}^*G_f&\left(I - \left|\left|(w^{(1)})_{(k)}\right|\right|^2 P_{(w^{(1)})_{(k)}} - \ldots - \left|\left|(w^{(\gamma)})_{(k)}\right|\right|^2 P_{(w^{(\gamma)})_{(k)}}\right) \\ &= {}^*\theta_{f,M}(w^{(1)}, \ldots, w^{(\gamma)}).\end{aligned}$$

Hence, using (3.10), we get ${}^*a_{f,M}(w^{(1)}, \ldots, w^{(\gamma)}, N) \approx \int_{\mathbb{R}^k} f d\mu_0$. Thus, it suffices to show that ${}^*a_{f,M}(w^{(1)}, \ldots, w^{(\gamma)}, N) \approx a_{f,N}((u^{(1)})_{(N)}, \ldots, (u^{(\gamma)})_{(N)}, N)$. Since $f$ is bounded, ${}^*f$ is $S$-integrable on $S^{N-1}(\sqrt{N}) \cap u^{(1)}{}^{\perp}_{(N)} \cap \ldots \cap u^{(\gamma)}{}^{\perp}_{(N)}$, so that the above is equivalent to showing the following for any $f \in C_c(\mathbb{R}^k)$:

$$\begin{aligned}&\int_{S^{N-1}(\sqrt{N}) \cap w^{(1)\perp} \cap \ldots \cap w^{(\gamma)\perp}} \mathbf{st}({}^*f(x)) dL\bar{\sigma}(x) \\ &\qquad = \int_{S^{N-1}(\sqrt{N}) \cap u^{(1)}{}^{\perp}_{(N)} \cap \ldots \cap u^{(\gamma)}{}^{\perp}_{(N)}} \mathbf{st}({}^*f(x)) dL\bar{\sigma}(x).\end{aligned} \tag{3.14}$$

This follows from Proposition C.4 and Theorem 3.3.2, completing the proof. □

## 3.4. Integrating continuous functions over non-great circles

In this section, we prove Theorem 3.1.1 for all bounded continuous functions. We recall some notation here for convenience. We fix $p_1, \dots, p_\gamma \in \mathbb{R}$, and for any $n \in \mathbb{N}$, we consider the sets

$$\begin{aligned} A &:= \{x \in \ell^2(\mathbb{R}) : \langle x, u^{(i)} \rangle = p_i \text{ for all } i \in \{1, \dots, \gamma\}\}, \\ H_n &:= \{x \in \mathbb{R}^n : \langle x, (u^{(i)})_{(n)} \rangle = 0 \text{ for all } i \in \{1, \dots, \gamma\}\}, \\ A_n &:= \{x \in \mathbb{R}^n : \langle x, (u^{(i)})_{(n)} \rangle = p_i \text{ for all } i \in \{1, \dots, \gamma\}\}, \\ S_{A_n} &:= S^{n-1}(\sqrt{n}) \cap A_n, \text{ and} \\ S_{H_n} &:= S^{n-1}(\sqrt{n}) \cap H_n. \end{aligned}$$

Let $z^{(1)}, \dots, z^{(\gamma)}$ be the Gram-Schmidt orthonormalization of the ${}^*\mathbb{R}$-linearly independent vectors $(u^{(1)})_{(N)}, \dots, (u^{(\gamma)})_{(N)}$ (see Lemma B.1). Define

$$S := S_{H_N} + \left( \frac{p_1}{||(u^{(1)})_{(N)}||} \right) \frac{(u^{(1)})_{(N)}}{||(u^{(1)})_{(N)}||} + \ldots + \left( \frac{p_\gamma}{||(u^{(\gamma)})_{(N)}||} \right) \frac{(u^{(\gamma)})_{(N)}}{||(u^{(\gamma)})_{(N)}||}.$$

It is clear that $S_{A_N}$ and $S$ are $(N-\gamma-1)$-dimensional spheres contained in $A_N$, and that they have the same center $\theta_N$, where

$$\begin{aligned} \theta_N &:= \left( \frac{p_1}{||(u^{(1)})_{(N)}||} \right) \frac{(u^{(1)})_{(N)}}{||(u^{(1)})_{(N)}||} + s \ldots + \left( \frac{p_\gamma}{||(u^{(\gamma)})_{(N)}||} \right) \frac{(u^{(\gamma)})_{(N)}}{||(u^{(\gamma)})_{(N)}||} && (3.15) \\ &= q_1 z^{(1)} + \ldots q_\gamma z^{(\gamma)} \text{ for some } q_1, \dots, q_\gamma \in {}^*\mathbb{R}. && (3.16) \end{aligned}$$

Using the expressions for the $z^{(i)}$ (see (C.11)) and the fact that $||(u^{(i)})_{(N)}|| \approx 1$ for all $i \in \{1, \dots, \gamma\}$, it follows by induction on $i$ that $q_i \approx p_i$ for all $i \in \{1, \dots, \gamma\}$. By

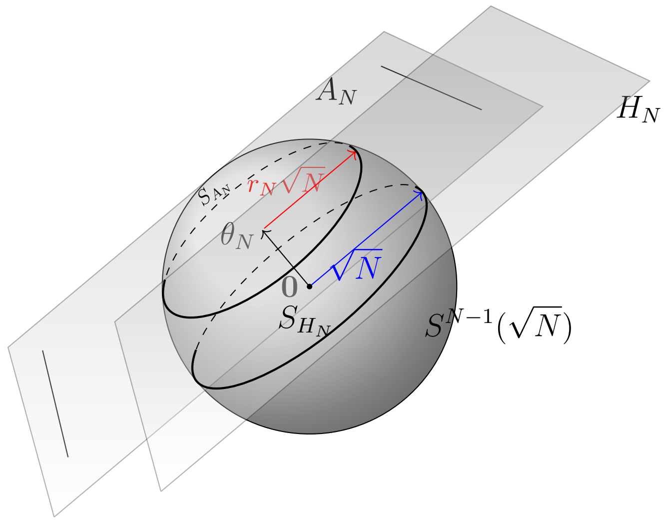


Figure 3.3. Visualizing $S_{A_N}$ in contrast with $S_{H_N}$

truncating onto the first $k$ coordinates in (3.15) and (3.16), we thus get:

$$q_1(z^{(1)})_{(k)} + \ldots q_\gamma(z^{(\gamma)})_{(k)} \approx p_1(u^{(1)})_{(k)} + \ldots p_\gamma(u^{(\gamma)})_{(k)}. \tag{3.17}$$

Let $r_N = \dfrac{\text{Radius}\,(S_{A_N})}{\text{Radius}\,(S)} = \dfrac{\sqrt{N - {q_1}^2 - \ldots - {q_\gamma}^2}}{\sqrt{N}} \approx 1$. Then we have

$$S_{A_N} = r_N \cdot S_{H_N} + \theta_N. \tag{3.18}$$

Since the $(u^{(i)})_{(n)}$ are $\mathbb{R}$-linearly independent in $\mathbb{R}^n$ for all large $n \in \mathbb{N}$, we can carry out the above construction to define $\theta_n$ for all $n \in \mathbb{N}_{\geq n'}$, where $n' \in \mathbb{N}$. By the formula corresponding to (3.18), we thus have:

$$\forall n \in \mathbb{N}_{\geq n'}\ \forall B \in \mathcal{B}(\mathbb{R}^k) \left[\bar{\sigma}_{S_{A_n}}(B) = \bar{\sigma}_{S_{H_n}}\left(\frac{1}{r_n}(B - \pi_k(\theta_n))\right)\right], \tag{3.19}$$

where $\pi_k$ denotes the projection onto the first $k$ coordinates under the standard orthonormal basis. We are now in a position to show that $\lim\limits_{n\to\infty} \displaystyle\int_{S_{A_n}} f d\bar{\sigma}$ equals the corresponding Gaussian expectation of $f$ for all $f \in C_c(\mathbb{R}^k)$.

**Theorem 3.4.1.** *Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be continuous with compact support. Then*

$$\lim_{n\to\infty} \int_{S_{A_n}} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}.$$

*Proof.* Define $h\colon \mathbb{R}^k \to \mathbb{R}$ by $h(y) = f\left(y + p_1(u^{(1)})_{(k)} + \ldots p_\gamma(u^{(\gamma)})_{(k)}\right)$ for all $y \in \mathbb{R}^k$. Note the following chain of equations (line 1 follows from transfer of the corresponding expressions for $a_{h,n}$ (as $n$ varies over $\mathbb{N}$) in Definition 3.3.7, line 2 follows from Theorem 3.3.10, line 3 follows from the definition of $h$, while line 4 follows from properties of Gaussian distributions):

$$\begin{aligned}
a_{h,n}(z^{(1)},\ldots,z^{(\gamma)},N) &= {}^*\!\!\int_{S^{N-1}(\sqrt{N})\cap(u^{(1)})_{(N)}{}^{\perp}\cap\ldots\cap(u^{(\gamma)})_{(N)}{}^{\perp}} {}^*h d^*\bar{\sigma} \\
&\approx \int_{\mathbb{R}^k} h d\mu_{0,u^{(1)},\ldots,u^{(\gamma)}} \\
&= \int_{\mathbb{R}^k} f\left(y + p_1(u^{(1)})_{(k)} + \ldots p_\gamma(u^{(\gamma)})_{(k)}\right) d\mu_{0,u^{(1)},\ldots,u^{(\gamma)}} \\
&= \int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}},
\end{aligned}$$

where $\bar{\eta} = p_1(u^{(1)})_{(k)} + \ldots p_\gamma(u^{(\gamma)})_{(k)}$.

The $S$-integrability of ${}^*h$ thus implies that

$$\int_{S^{N-1}(\sqrt{N})\cap z^{(1)\perp}\cap\ldots\cap z^{(\gamma)\perp}} \mathbf{st}({}^*h) dL\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}. \tag{3.20}$$

Using (3.20) and the nonstandard characterization of uniform continuity (which, in particular, implies that ${}^*h(x) \approx {}^*h(rx)$ for all $x \in {}^*\mathbb{R}^N$ and $r \approx 1$), we obtain:

$$\begin{aligned}
\int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}} &= \int_{S^{N-1}(\sqrt{N})\cap z^{(1)\perp}\cap\ldots\cap z^{(\gamma)\perp}} \mathbf{st}({}^*h(x)) dL\bar{\sigma}(x) \\
&= \int_{S_{N,z^{(1)},\ldots,z^{(\gamma)}}} \mathbf{st}({}^*h(rx)) dL\bar{\sigma}(x).
\end{aligned} \tag{3.21}$$

Note that the composition of ${}^*h$ with the scaling by $r$ is a finitely bounded, and hence $S$-integrable, function. This and transfer of the scaling properties of the uniform surface measures respectively imply the following:

$$\begin{aligned}\int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}} &\approx {}^*\!\!\int_{S_{N,z^{(1)},\ldots,z^{(\gamma)}}} {}^*h(rx) d^*\bar{\sigma} \\ &\approx {}^*\!\!\int_{r\cdot S_{N,z^{(1)},\ldots,z^{(\gamma)}}} {}^*h(x) d^*\bar{\sigma}(x). \end{aligned} \tag{3.22}$$

The proof is now contained in the following sequence of equations obtained by simplifying (3.22):

$$\begin{aligned}\int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}} &\approx {}^*\!\!\int_{r\cdot S_{N,z^{(1)},\ldots,z^{(\gamma)}}} {}^*f\left(x + q_1 z^{(1)} + \ldots + q_\gamma z^{(\gamma)}\right) d^*\bar{\sigma}(x) \\ &= {}^*\!\!\int_{r\cdot S_{N,z^{(1)},\ldots,z^{(\gamma)}} + q_1 z^{(1)} + \ldots + q_\gamma z^{(\gamma)}} {}^*f(x) d^*\bar{\sigma}(x) \\ &= {}^*\!\!\int_{S_{A_N}} {}^*f(x) d^*\bar{\sigma}(x). \end{aligned}$$

The first line follows from the fact that ${}^*h$ is bounded by a real number (and is hence $S$-integrable) and the following fact that is true for all $x \in {}^*\mathbb{R}^N$ (due to the nonstandard characterization of the uniform continuity of $f\colon \mathbb{R}^k \to \mathbb{R}$ and (3.17)):

$$\begin{aligned}{}^*h(x) &= {}^*f\left(x + p_1(u^{(1)})_{(k)} + \ldots p_\gamma(u^{(\gamma)})_{(k)}\right) \\ &\approx {}^*f\left(x + q_1(z^{(1)})_{(k)} + \ldots + q_\gamma(z^{(\gamma)})_{(k)}\right) \\ &= {}^*f\left(x + q_1 z^{(1)} + \ldots + q_\gamma z^{(\gamma)}\right). \end{aligned}$$

The second line follows by transfer of the translation properties of the uniform surface measures. The third line follows from (3.18). □

Using Theorem 3.4.1, we immediately deduce that the first $k$ coordinates of almost any point of $S_{A_N}$ are finite.

**Theorem 3.4.2.** *Almost all points of $S_{A_N}$ have finite projections to ${}^*\mathbb{R}^k$, i.e.,*

$$L\bar{\sigma}(\{x \in S_{A_N} : x_1, \ldots, x_k \in {}^*\mathbb{R}_{fin}\}) = 1.$$

*Proof.* We prove this for $k = 1$ (the general case follows from the fact that the intersection of finitely many almost sure events is almost sure). For each $m \in \mathbb{N}$, consider the function $f_m$ that is equal to 1 on $(-m+1, m-1)$, equal to zero on $\mathbb{R}\backslash(-m, m)$, and is linear in between. We thus have

$$\begin{aligned}
L\bar{\sigma}(x_1 \in {}^*(-m, m)) &= \mathbb{E}_{S_{A_N}}(\mathbf{st}(\mathbb{1}_{{}^*(-m,m)})) \\
&= \mathbb{E}_{S_{A_N}}(\mathbf{st}({}^*\mathbb{1}_{(-m,m)})) \\
&\geq \mathbb{E}_{S_{A_N}}(\mathbf{st}({}^*f_m)) \\
&\geq \int_{\mathbb{R}^k} f_m d\mu. && \text{[using Theorem 3.4.1]}
\end{aligned}$$

As a consequence, we obtain

$$\begin{aligned}
1 \geq L\bar{\sigma}(x_1 \in {}^*\mathbb{R}_{\text{fin}}) = L\bar{\sigma}(\cup_{m\in\mathbb{N}}\{x_1 \in {}^*(-m, m)\}) &= \lim_{m\to\infty} L\bar{\sigma}(x_1 \in {}^*(-m, m)) \\
\Rightarrow 1 \geq L\bar{\sigma}(x_1 \in {}^*\mathbb{R}_{\text{fin}}) \geq \lim_{m\to\infty} \int_{\mathbb{R}^k} f_m d\mu = 1 \\
\Rightarrow L\bar{\sigma}(x_1 \in {}^*\mathbb{R}_{\text{fin}}) = 1,
\end{aligned}$$

thus completing the proof. □

Using Theorem 3.4.1 and Theorem 3.4.2, we are now able to generalize the limiting spherical integral result to all bounded continuous functions on $\mathbb{R}^k$.

**Theorem 3.4.3.** *Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be a bounded continuous function. Then*

$$\lim_{n\to\infty} \int_{S_{A_n}} f d\bar{\sigma} = \int_{\mathbb{R}^k} f d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}.$$

*Proof.* Let $f\colon \mathbb{R}^k \to \mathbb{R}$ be bounded and continuous. For each $m \in \mathbb{N}$, let $f_m$ be the restriction of $f$ to $[-m,m]^k$, i.e., $f_m := f \cdot \mathbb{1}_{[-m,m]^k}$. Fix $N > \mathbb{N}$. Since $f$ is bounded, $\mathbf{st}(^*f)$ is $S$-integrable. This shows:

$$^*\!\!\int_{S_{A_N}} {}^*f d\bar{\sigma} \approx \int_{S_{A_N}} \mathbf{st}(^*f) dL\bar{\sigma}. \tag{3.23}$$

Using Theorem 3.4.2 and applying dominated convergence theorem, we obtain:

$$\int_{S_{A_N}} \mathbf{st}(^*f) dL\bar{\sigma} = \lim_{m\to\infty} \int_{S_{A_N}} \mathbf{st}(^*f_m) dL\bar{\sigma}. \tag{3.24}$$

The right side of (3.24) equals $\lim_{m\to\infty} \int_{\mathbb{R}^k} f_m(x) d\mu_{\bar{\eta},u^{(1)},\ldots,u^{(\gamma)}}$ using Theorem 3.4.1. Thus dominated convergence theorem and (3.23) now completes the proof. □

# Chapter 4. De Finetti's Theorem for Bernoulli Random Variables

## 4.1. Introduction

The rest of the dissertation focuses on de Finetti's theorem, which is a result for sequences of exchangeable random variables. This chapter presents a nonstandard analytic treatment of the original formulation of de Finetti's theorem, which holds for a sequence of exchangeable Bernoulli random variables. Throughout the chapter, all Bernoulli random variables take values in $\{0,1\}$. We begin with the definition of exchangeability.

**Definition 4.1.1.** A finite collection $X_1, \ldots, X_n$ of random variables is said to be *exchangeable* if for any permutation $\sigma \in S_n$, the random vectors $(X_1, \ldots, X_n)$ and $(X_{\sigma(1)}, \ldots, X_{\sigma(n)})$ have the same distribution. An infinite sequence $(X_n)_{n \in \mathbb{N}}$ of random variables is said to be *exchangeable* if any finite subcollection of the $X_i$ is exchangeable in the above sense.

See Feller [37, pp. 229-230] for some examples of exchangeable random variables. A well-known result of de Finetti says that an exchangeable sequence of Bernoulli random variables (that is, random variables taking values in $\{0,1\}$) is conditionally independent given the value of a random parameter in $[0,1]$ (the parameter being sampled through a unique probability measure on the Borel sigma algebra of the closed interval $[0,1]$). In a more technical language, we say that any exchangeable sequence of Bernoulli random variables is uniquely representable as a *mixture* of independent and identically distributed (iid) sequences of Bernoulli random variables. More precisely, we may write de Finetti's theorem in the following form.

**Theorem 4.1.2** (de Finetti)**.** *Let $X_1, X_2, \ldots$ be a sequence of exchangeable Bernoulli ran-*

*dom variables. There exists a unique measure $\mu$ on the interval $[0,1]$ such that the following holds:*

$$\mathbb{P}(X_1 = e_1, \dots, X_k = e_k) = \int_{[0,1]} p^{\sum_{j=1}^k e_j}(1-p)^{k-\sum_{j=1}^k e_j} d\mu(p) \tag{4.1}$$

*for any $k \in \mathbb{N}$ and $e_1, \dots, e_k \in \{0,1\}$.*

The integrand on the right side is the probability that $k$ iid Bernoulli($p$) random variables have the outcomes $e_1, \dots, e_k$. In this sense, de Finetti's theorem expresses an exchangeable sequence of Bernoulli random variables as a *mixture* of iid sequences of Bernoulli random variables.

See de Finetti [28, 29] for the original formulations of this theorem. Aldous [9] and Kingman [61] are good resources for an introduction to exchangeability and related topics. See Kirsch [62] for a recent elementary proof of de Finetti's theorem.

We will give a nonstandard proof of Theorem 4.1.2. In nonstandard analytic language, the idea is that the measure $\mu$ will be shown to be induced by a hyperfinite sample mean $\dfrac{X_1 + \ldots + X_N}{N}$.

For the rest of this chapter, we fix an exchangeable sequence $X_1, X_2, \dots$ of Bernoulli random variables. We also fix $k \in \mathbb{N}$ and $e_1, \dots, e_k \in \{0,1\}$. Taking $\alpha = \displaystyle\sum_{j=1}^{k} e_j$ and writing the integral in (4.1) as an expectation in terms of a random variable $Y \sim \mu$, de Finetti's theorem may be restated as follows:

$$\mathbb{P}(X_1 = e_1, \dots, X_k = e_k) = \mathbb{E}_\mu(Y^\alpha (1-Y)^{k-\alpha}). \tag{4.2}$$

Written this way, it is clear that any measure satisfying the conclusion of de Finetti's theorem must be unique. Indeed, taking $\alpha = k$ and varying $k$ through $\mathbb{N}$ in

(4.2) shows that such a measure has a unique sequence of moments, which implies that they agree on expected values of continuous functions on $[0, 1]$ (using the Weierstrass approximation theorem).

Hence, it is enough to prove the existence of a probability measure on $[0, 1]$ satisfying the conclusion of de Finetti's theorem. Toward that end, we will verify equation (4.2) for a standard measure $\mu$ that is naturally induced by an appropriate Loeb measure. Fix $N > \mathbb{N}$ and define:

$$Y_N = \frac{X_1 + \ldots + X_N}{N}. \tag{4.3}$$

Note that we are abusing notation by using $(X_i)$ to denote both the standard sequence $(X_i)_{i \in \mathbb{N}}$ of random variables and the nonstandard extension of this sequence, with the usage being clear from context. More precisely, if $\mathcal{X} \colon \Omega \times \mathbb{N} \to S$ is defined by $\mathcal{X}(\omega, i) := X_i(\omega)$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$, then for any $i \in {}^*\mathbb{N}$, the internal random variable $X_i \colon {}^*\Omega \to {}^*S$ is defined as follows:

$$X_i(\omega) = {}^*\mathcal{X}(\omega, i) \text{ for all } \omega \in {}^*\Omega \text{ and } i \in {}^*\mathbb{N}.$$

Let ${}^*\mathbb{P} \colon {}^*\mathcal{F} \to {}^*[0, 1]$ denote the nonstandard extension of $\mathbb{P}$. Then ${}^*\mathbb{P}$ is an internal probability measure. Note that $Y_N$ takes values in $\left\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, \frac{N}{N} = 1\right\}$. Naively conditioning on the value of $Y_N$, we obtain the following:

$$
\begin{aligned}
&\mathbb{P}(X_1 = e_1, \ldots,\ X_k = e_k) \\
&\qquad = \sum_{i=0}^{N} {}^*\mathbb{P}\left(X_1 = e_1, \ldots, X_k = e_k \middle| Y_N = \frac{i}{N}\right) {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right). \qquad (4.4)
\end{aligned}
$$

Note that we could have started the sum in (4.4) at $i = \alpha$ since the conditional probabilities in this sum are zero for all $i < \alpha$.

The random variable $Y_N$ induces an internal finitely additive internal probability measure $\mathbb{P}_N$ on ${}^*[0,1]$, which is supported on $\left\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, \frac{N}{N} = 1\right\}$, in the following way:

$$
\mathbb{P}_N(B) = {}^*\mathbb{P}(Y_N \in B) \text{ for all } {}^*\text{-Borel sets } B \subseteq {}^*[0,1]. \qquad (4.5)
$$

Consider the associated Loeb measure $L\mathbb{P}_N$. With $\mathcal{B}([0,1])$ denoting the Borel sigma algebra of $[0,1]$, define $\mu\colon \mathcal{B}([0,1]) \to [0,1]$ by:

$$
\mu(A) := L\mathbb{P}_N(\mathbf{st}^{-1}(A)) \text{ for all Borel subsets } A \subseteq [0,1]. \qquad (4.6)
$$

By Theorem 1.3.17, $\mu$ is a well-defined Radon probability measure on $[0,1]$ such that the following holds:

$$
{}^*\mathbb{E}_{\mathbb{P}_N}({}^*f) \approx \mathbb{E}_\mu(f) \text{ for all bounded nonnegative } f\colon [0,1] \to \mathbb{R}_{\geq 0}. \qquad (4.7)
$$

Consider the function $f\colon [0,1] \to \mathbb{R}_{\geq 0}$ defined by

$$f(p) = p^{\alpha}(1-p)^{k-\alpha} \text{ for all } p \in [0,1]. \tag{4.8}$$

Noting the form of the right side in (4.2), and using (4.4) and (4.7), it is clear that we need the following to be true:

**Theorem 4.1.3.** *We have*

$$\sum_{i=0}^{N} {}^*\mathbb{P}\left(X_1 = e_1, \ldots, X_k = e_k \middle| Y_N = \frac{i}{N}\right) {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right)$$
$$\approx \sum_{i=0}^{N} \left(\frac{i}{N}\right)^{\alpha} \left(1 - \frac{i}{N}\right)^{k-\alpha} {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right). \tag{4.9}$$

The rest of this chapter will build toward a proof of Theorem 4.1.3.

## 4.2. Proving Theorem 4.1.3

Our strategy is to use the following simple fact from nonstandard analysis:

**Lemma 4.2.1.** *If $\alpha_j, \beta_j \in {}^*\mathbb{R}_{\geq 0}$ (where $j \in H$ for some hyperfinite set $H$) and $\frac{\alpha_j}{\beta_j} \approx 1$ for all $j \in H$, then*

$$\frac{\sum_{j \in H} \alpha_j}{\sum_{j \in H} \beta_j} \approx 1. \tag{4.10}$$

*Proof.* Let $H$, $\alpha_j$, and $\beta_j$ be as in the statement of the lemma. Note that $\alpha_j, \beta_j$ must all be strictly positive. For any real number $\epsilon \in \mathbb{R}_{>0}$, the condition that $\frac{\alpha_j}{\beta_j} \approx 1$ for all $j \in H$ implies that

$$1 - \epsilon < \frac{\alpha_j}{\beta_j} < 1 + \epsilon \text{ for all } j \in H.$$

Multiplying all sides of the above inequality by $\beta_j$, we have:

$$\beta_j(1-\epsilon) < \alpha_j < \beta_j(1+\epsilon) \text{ for all } j \in H.$$

Summing as $j$ varies over the hyperfinite set (in this step, we are also using transfer of a similar inequality for finite sums), we get:

$$(1-\epsilon)\sum_{j\in H}\beta_j < \sum_{j\in H}\alpha_j < (1+\epsilon)\sum_{j\in H}\beta_j. \tag{4.11}$$

Dividing all sides of (4.11) by $\sum_{j\in H}\beta_j$ and noting that $\epsilon\in\mathbb{R}_{>0}$ was arbitrarily chosen completes the proof. $\square$

For brevity in future computations, we define

$$a_i = {}^*\mathbb{P}\left(X_1 = e_1,\ldots,X_k = e_k \middle| Y_N = \frac{i}{N}\right) \tag{4.12}$$

$$\text{and } b_i = \left(\frac{i}{N}\right)^{\alpha}\left(1-\frac{i}{N}\right)^{k-\alpha} \text{ for all } i\in\{0,1,2,\ldots,N\}. \tag{4.13}$$

Let us first try to understand the conditional probabilities $a_i$. As explained earlier, the $a_i$ are zero for $i<\alpha$. By summing over all possible cases, we have:

$$\begin{aligned} a_i &= {}^*\mathbb{P}\left(X_1 = e_1,\ldots,X_k = e_k \middle| Y_N = \frac{i}{N}\right)\\ &= \sum_{(u_1,\ldots,u_N)\in\mathcal{G}} {}^*\mathbb{P}\left(X_1 = u_1,\ldots,X_N = u_N \middle| X_1+\ldots+X_N = i\right), \end{aligned} \tag{4.14}$$

where

$$\mathcal{G} := \left\{(u_1,\ldots u_N)\in\{0,1\}^N : u_j = e_j \text{ for all } j\in\{1,\ldots,k\} \text{ and } \sum_{j=1}^{N}u_j = i\right\}.$$

It is clear that the internal cardinality of $\mathcal{G}$ is the number of ways of choosing $u_{k+1},\ldots,u_N\in\{0,1\}$ such that $\sum_{j=k+1}^{N}u_j = i-\alpha$. By a simple counting argument, this

yields:

$$\#(\mathcal{G}) = \binom{N-k}{i-\alpha}. \tag{4.15}$$

Also, by the transfer of exchangeability of the $X_i$, it is clear that:

$$\begin{aligned} &{}^*\mathbb{P}\left(X_1 = u_1, \ldots, X_N = u_N \middle| X_1 + \ldots + X_N = i\right) \\ = \; &\frac{1}{\text{Number of ways of writing } i \text{ as a sum of } N \text{ zeroes and ones}} \end{aligned} \tag{4.16}$$

for all $(u_1, \ldots, u_N) \in \mathcal{G}$.

To see (4.16), first define $\mathcal{G}'$ as the set of those $(u_1, \ldots, u_N)$ such that $\sum_{j=1}^{N} u_j = i$. Then exchangeability implies that

$$\begin{aligned} &{}^*\mathbb{P}((X_1, \ldots, X_N) = \vec{u} \mid X_1 + \ldots + X_N = i) \\ &\qquad = {}^*\mathbb{P}((X_1, \ldots, X_N) = \vec{u}' | X_1 + \ldots + X_N = i) \text{ for all } \vec{u}, \vec{u}' \in \mathcal{G}'. \end{aligned}$$

Since the sum of ${}^*\mathbb{P}((X_1, \ldots, X_N) = \vec{u} \mid X_1 + \ldots + X_N = i)$ as $\vec{u}$ varies over $\mathcal{G}'$ is equal to one, it must be the case that

$${}^*\mathbb{P}((X_1, \ldots, X_N) = \vec{u} \mid X_1 + \ldots + X_N = i) = \frac{1}{\#(\mathcal{G}')} \text{ for all } \vec{u} \in \mathcal{G}'. \tag{4.17}$$

In particular, since $\mathcal{G} \subseteq \mathcal{G}'$, equation (4.17) explains (4.16). Now, another simple counting argument shows that $\#(\mathcal{G}') = \binom{N}{i}$. Thus, (4.16) becomes:

$${}^*\mathbb{P}\left(X_1 = u_1, \ldots, X_N = u_N \middle| X_1 + \ldots + X_N = i\right) = \frac{1}{\binom{N}{i}} \tag{4.18}$$

for all $(u_1, \ldots, u_N) \in \mathcal{G}$.

Using (4.18) and (4.15) in (4.14), we obtain:

$$a_i = \frac{\binom{N-k}{i-\alpha}}{\binom{N}{i}} \text{ for all } i \in \{1, \ldots N\}, \tag{4.19}$$

where $\begin{pmatrix} N-k \\ i-\alpha \end{pmatrix}$ is understood to be zero when $i < \alpha$.

Using (4.19), we first prove Theorem 4.1.3 in a pathological case of zero probability (see Lemma 4.2.2) that we will avoid afterward. Note that the conclusion of de Finetti's theorem implies that this pathological case can never happen, unless all the random variables $X_i$ are zero almost surely. However, since we are proving de Finetti's theorem, we have to take care of this case in a non-circular way, without using de Finetti's theorem.

**Lemma 4.2.2.** *Suppose* $\mathbb{P}(X_1 = e_1, \ldots, X_k = e_k) = 0$. *Then,* (4.9) *holds.*

*Proof.* Suppose $\mathbb{P}(X_1 = e_1, \ldots, X_k = e_k) = 0$. Suppose $i \geq \alpha$ and consider the event $\left\{Y_N = \frac{i}{N}\right\}$, which is the same as the event $\{X_1 + \ldots + X_N = i\}$.

If the sum of $N$ zero-one random variables is $i \geq \alpha$ then some subcollection of $k$ such random variables must have had exactly $\alpha$ ones. Therefore, if $\mathcal{C}$ denotes the collection of all $k$ tuples of distinct indices from $\{1, \ldots, N\}$ (so that the internal cardinality $\#(\mathcal{C})$ is $\begin{pmatrix} N \\ k \end{pmatrix}$), then we have

$$\{X_1 + \ldots X_N = i\} \subseteq \bigcup_{(j_1, \ldots j_k) \in \mathcal{C}} \{X_{j_1} = e_1, \ldots, X_{j_k} = e_k\}.$$

By exchangeability, all events in the union on the right have the same probability as the event $\{X_1 = e_1, \ldots, X_k = e_k\}$, which is assumed to have probability zero. Since ${}^*\mathbb{P}$ is hyperfinitely subadditive, this implies that ${}^*\mathbb{P}(X_1 + \ldots + X_N = i) = 0$ whenever $i \geq \alpha$. Thus (using (4.19)), proving (4.9) is equivalent to proving the following:

$$
\begin{aligned}
&\sum_{i=0}^{\alpha-1} \frac{\binom{N-k}{i-\alpha}}{\binom{N}{i}} \mathbb{P}\left(Y_N = \frac{i}{N}\right) \\
\approx &\sum_{i=0}^{\alpha-1} \left(\frac{i}{N}\right)^{\alpha} \left(1 - \frac{i}{N}\right)^{k-\alpha} \mathbb{P}\left(Y_N = \frac{i}{N}\right).
\end{aligned} \tag{4.20}
$$

But the left side of (4.20) is zero (as $\binom{N-k}{i-\alpha} = 0$ for $i < \alpha$), while the right side is an infinitesimal (being a finite sum of infinitesimals). This completes the proof. □

Also using (4.19), we obtain the following result about the ratio of $a_i$ and $b_i$:

**Lemma 4.2.3.** *There exists a constant $r \approx 1$, such that for each $i \in {}^*\mathbb{N}_{>k}$, we have*

$$
\frac{a_i}{b_i} = \frac{i!}{(i-\alpha)! i^{\alpha}} \left(1 - \frac{1}{N-i}\right) \dots \left(1 - \frac{k-\alpha-1}{N-i}\right) r \leq r. \tag{4.21}
$$

*Proof.* From (4.19) and (4.13), we obtain:

$$
\begin{aligned}
\frac{a_i}{b_i} &= \frac{\frac{(N-k)(N-k-1)\dots(N-k-(i-\alpha-1))}{(i-\alpha)!}}{\frac{N(N-1)\dots(N-(i-1))}{i!}\left(\frac{i}{N}\right)^{\alpha}\left(1-\frac{i}{N}\right)^{k-\alpha}} \\
&= \frac{i!}{(i-\alpha)! i^{\alpha}} \frac{N^k}{(N-i)^{k-\alpha}} \frac{(N-k)(N-k-1)\dots(N-k-(i-\alpha-1))}{N(N-1)\dots(N-(i-1))} \\
&= \frac{i!}{(i-\alpha)! i^{\alpha}} \frac{N^k (N-i)(N-(i+1))\dots N-(i+k-\alpha-1)}{N(N-1)\dots(N-(k-1))(N-i)^{k-\alpha}}.
\end{aligned}
$$

Let

$$
r := \frac{N^k}{N(N-1)\dots(N-(k-1))} = \frac{1}{1\left(1-\frac{1}{N}\right)\dots\left(1-\frac{k-1}{N}\right)} \approx 1. \tag{4.22}
$$

Thus the proof is complete because:

$$
\begin{aligned}
&\frac{(N-i)(N-(i+1))\dots(N-(i+k-\alpha-1))}{(N-i)^{k-\alpha}} \\
= &1\left(1-\frac{1}{N-i}\right)\dots\left(1-\frac{k-\alpha-1}{N-i}\right).
\end{aligned}
$$

□

**Lemma 4.2.4.** *Suppose $\alpha \geq 1$. There is an $M_1 > \mathbb{N}$ such that $M_1 < N - \sqrt{N}$ and*

$$\sum_{i=0}^{M_1} a_i {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right) \approx 0 \text{ and } \sum_{i=0}^{M_1} b_i {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right) \approx 0.$$

*Proof.* Fix any $M_1 > \mathbb{N}$ such that $M_1 < \min\{N^{\frac{1}{3}}, N - \sqrt{N}\}$.

Note that $\sum_{i=0}^{k} a_i$ is an infinitesimal. Hence, by (4.21), it suffices to show that $\sum_{i=0}^{M_1} b_i$ is an infinitesimal. Now,

$$\sum_{i=0}^{M_1} b_i = \sum_{i=0}^{M_1} \left(\frac{i}{N}\right)^{\alpha} \left(1 - \frac{i}{N}\right)^{k-\alpha} \leq \frac{{M_1}^{1+\alpha}}{N^{\alpha}} < \frac{N^{\frac{1+\alpha}{3}}}{N^{\alpha}} = \frac{1}{N^{\frac{2\alpha-1}{3}}}.$$

But the right side is an infinitesimal because $2\alpha > 1$ (as $\alpha \geq 1$ is assumed in the statement of the lemma). This completes the proof. □

For the rest of this chapter, let

$$M_2 := [N - \sqrt{N}] + 1, \tag{4.23}$$

where $[\cdot]$ is the greatest integer function.

**Corollary 4.2.5.** *For $i \in {}^*\mathbb{N}$ with $\mathbb{N} < i \leq M_2$, we have $\frac{a_i}{b_i} \approx 1$.*

*Proof.* Note that $\frac{i!}{(i-\alpha)! i^{\alpha}} = 1$ when $\alpha = 0, 1$. And for $\alpha \geq 2$, we have

$$\frac{i!}{(i-\alpha)! i^{\alpha}} = \left(1 - \frac{1}{i}\right) \ldots \left(1 - \frac{\alpha - 1}{i}\right) \approx 1 \text{ if } i > \mathbb{N}.$$

Thus, we have:

$$\frac{i!}{(i-\alpha)! i^{\alpha}} \approx 1 \text{ for all } i > \mathbb{N}. \tag{4.24}$$

Now let $i$ be as in the statement of the corollary, i.e., $\mathbb{N} < i \leq M_2$. Then, $N - i \geq N - M_2 \geq \sqrt{N}$. Then,

$$\left(1 - \frac{1}{N-i}\right) \cdots \left(1 - \frac{k-\alpha-1}{N-i}\right) \approx 1 \text{ as well.} \tag{4.25}$$

Using (4.24) and (4.25) in (4.21) completes the proof. ◻

**Lemma 4.2.6.** *Suppose* $\alpha \leq (k-1)$. *Then*

$$\sum_{i=M_2+1}^{N} {a_i}^{*}\mathbb{P}\left(Y_N = \frac{i}{N}\right) \approx 0 \text{ and } \sum_{i=M_2+1}^{N} {b_i}^{*}\mathbb{P}\left(Y_N = \frac{i}{N}\right) \approx 0. \tag{4.26}$$

*Proof.* By (4.21), it suffices to show that the second sum is an infinitesimal. Since the $b_i$ are all positive, we have the following estimate for the second term:

$$\begin{aligned}
\sum_{i=M_2+1}^{N} {b_i}^{*}\mathbb{P}\left(Y_N = \frac{i}{N}\right) &\leq \left(\max_{M_2+1 \leq i \leq N} b_i\right) \sum_{i=M_2+1}^{N} {}^{*}\mathbb{P}\left(Y_N = \frac{i}{N}\right) \\
&\leq \max_{M_2+1 \leq i \leq N} \left(\frac{i}{N}\right)^{\alpha} \left(1 - \frac{i}{N}\right)^{k-\alpha} \\
&\leq 1 \cdot \left(1 - \frac{N - \sqrt{N}}{N}\right)^{k-\alpha} \\
&= \left(\frac{1}{\sqrt{N}}\right)^{k-\alpha},
\end{aligned}$$

where the last term is infinitesimal since $k - \alpha \geq 1$. ◻

We are now in a position to prove Theorem 4.1.3. We restate it here for convenience.

**Theorem 4.1.3.** *We have*

$$\sum_{i=0}^{N} {}^*\mathbb{P}\left(X_1 = e_1, \ldots, X_k = e_k \middle| Y_N = \frac{i}{N}\right) {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right)$$
$$\approx \sum_{i=0}^{N} \left(\frac{i}{N}\right)^{\alpha} \left(1 - \frac{i}{N}\right)^{k-\alpha} {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right). \tag{4.9}$$

*Proof.* The case when $\alpha = 0$ is verified directly by plugging in $\alpha = 0$ to the formulae for $a_i$ and $b_i$ and using Lemma 4.2.1.

In the case when $\alpha = k$, using (4.9) and (4.13), we get:

$$\frac{a_i}{b_i} = \frac{\binom{N-k}{i-k}}{\binom{N}{i}\frac{i^k}{N^k}} = \frac{i!}{(i-k)!i^k}\frac{(N-k)!N^k}{N!}.$$

This expression is infinitesimally close to 1 whenever $i > \mathbb{N}$. Thus, Lemma 4.2.4 and Lemma 4.2.1 complete the proof in this case.

By Lemma 4.2.2, we may also assume that

$$\mathbb{P}(X_1 = e_1, \ldots, X_k = e_k) \neq 0.$$

Then using (4.4), we obtain

$$\sum_{i=0}^{N} {}^*\mathbb{P}\left(X_1 = e_1, \ldots, X_k = e_k \middle| Y_N = \frac{i}{N}\right) {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right) \not\approx 0.$$

Thus, by Lemmas 4.2.4 and 4.2.6, we obtain:

$$\sum_{i=0}^{N} {}^*\mathbb{P}\left(X_1 = e_1, \ldots, X_k = e_k \middle| Y_N = \frac{i}{N}\right) {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right)$$
$$\approx \sum_{i=M_1+1}^{M_2} a_i {}^*\mathbb{P}\left(Y_N = \frac{i}{N}\right),$$

and

$$\sum_{i=0}^{N}\left(\frac{i}{N}\right)^{\alpha}\left(1-\frac{i}{N}\right)^{k-\alpha}{}^*\mathbb{P}\left(Y_N=\frac{i}{N}\right)\approx\sum_{i=M_1+1}^{M_2}b_i{}^*\mathbb{P}\left(Y_N=\frac{i}{N}\right).$$

Corollary 4.2.5 together with Lemma 4.2.1 now complete the proof in this case. $\square$

As $e_1,\ldots,e_k$ was an arbitrarily fixed finite sequence of zeros and ones, this proves de Finetti's Theorem 4.1.2 using Theorem 1.3.17.

We finish this section with a combinatorial-probabilistic interpretation of the proof. A main ingredient in the proof was Corollary 4.2.5. It shows that when $i$ is large (in the sense that it is hyperfinite) but not too large (in the sense that it is less than $M_2=[N-\sqrt{N}]+1$), then $\frac{a_i}{b_i}$ is infinitesimally close to 1. Looking at the expressions (4.19) and (4.13) for $a_i$ and $b_i$ respectively, we can express the ratio as follows:

$$\frac{a_i}{b_i}=\frac{\binom{N-k}{i-\alpha}\binom{k}{\alpha}}{\binom{N}{i}}\cdot\frac{1}{\binom{k}{\alpha}\left(\frac{i}{N}\right)^{\alpha}\left(1-\frac{i}{N}\right)^{k-\alpha}}.$$

The first term on the right is an expression related to a certain hypergeometric random variable, while the second term is related to a certain binomial random variable. We can thus interpret Corollary 4.2.5 as a statement about asymptotically approximating a hypergeometric random variable with a binomial random variable. More explicitly, Corollary 4.2.5 says that as long as $i$ is neither too small not too large, then the probabilities $P_1$ and $P_2$ described by the following are very close to each other in the sense that $\frac{P_1}{P_2}\approx 1$:

(1) Uniformly choose a random subset of size $i$ (here $i\geq\alpha$) from $\{1,\ldots,N\}$—thus all the $\binom{N}{i}$ subsets are equally likely to be chosen. Then $P_1$ is the probability that exactly $\alpha$ elements of $\{1,\ldots,k\}$ appear in this random subset of size $i$.

(2) Take a coin with a probability of Heads being $\dfrac{i}{N}$. Then $P_2$ is the probability that exactly $\alpha$ Heads appear in $k$ independent tosses of this coin.

# Chapter 5. Ideas for Generalizing De Finetti's Theorem

## 5.1. The form of possible generalizations

The previous chapter established de Finetti's theorem for $\{0,1\}$-valued exchangeable random variables (see Theorem 4.1.2). The work of generalizing de Finetti's theorem from $\{0,1\}$ to more general state spaces has been an enterprise spanning the better part of the twentieth century. This chapter provides the history of these generalizations and sets up an overview of our generalization, which is carried out over the next two chapters.

What counts as a generalization of Theorem 4.1.2? Notice that in equation (4.1), the variable of integration, $p$, can be identified with the measure induced on $\{0,1\}$ by a coin toss for which the chance of success (with success identified with the state 1) is $p$. Clearly, all probability measures on the discrete set $\{0,1\}$ are of this form. Thus, $\nu$ in (4.1) can be thought of as a measure on the set of all probability measures on $\{0,1\}$. The integrand in (4.1) then represents the probability of getting $\sum_{j=1}^{k} e_j$ successes in $k$ independent coin tosses, while the integral represents the expected value of this probability with respect to $\nu$.

With $S = \{0,1\}$, we can thus interpret (4.1) as saying that the probability that the random vector $(X_1, \ldots, X_k)$ is in the Cartesian product $B_1 \times \ldots \times B_k$ of measurable sets $B_1, \ldots, B_k \subseteq S$, is given by the expected value of $\mu(B_1) \cdot \ldots \cdot \mu(B_k)$ as $\mu$ is sampled (according to some distribution $\nu$) from the space of all Borel probability measures on $S$. Thus, one possible direction in which to generalize Theorem 4.1.2 is to look for a statement of the following type (although we now know this to be incorrect in such generality following the work of Dubins and Freedman [35], it is still illustrative to explore the kind

of statement that we are looking for).

**A natural guess for a generalization of de Finetti.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $(X_n)_{n\in\mathbb{N}}$ *be an exchangeable sequence of random variables taking values in some measurable space* $(S, \mathfrak{S})$ *(called the state space). If* $\mathfrak{P}(\mathrm{S})$ *denotes the set of all probability measures on* $(S, \mathfrak{S})$*, then there is a unique probability measure* $\mathscr{P}$ *on* $\mathfrak{P}(\mathrm{S})$ *such that the following holds for all* $k \in \mathbb{N}$*:*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) = \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{P}(\mu) \text{ for all } B_1, \ldots, B_k \in \mathfrak{S}. \tag{5.1}$$

The above statement is crude since we want a probability measure on the underlying set $\mathfrak{P}(\mathrm{S})$, yet we have not specified what sigma algebra on $\mathfrak{P}(\mathrm{S})$ we are working with. We shall soon see that there are multiple natural sigma algebras on $\mathfrak{P}(\mathrm{S})$. Since we want to integrate functions of the type $\mu \mapsto \mu(B)$ on $\mathfrak{P}(\mathrm{S})$ for all $B \in \mathfrak{S}$, the smallest sigma algebra ensuring the measurability of all such functions is appropriate for this discussion. That minimal sigma algebra, which we denote by $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$, is generated by cylinder sets. In other words, $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ is the smallest sigma algebra containing all sets of the type

$$\{\mu \in \mathfrak{P}(\mathrm{S}) : \mu(B_1) \in A_1, \ldots, \mu(B_k) \in A_k\},$$

where $k \in \mathbb{N}$; $B_1, \ldots, B_k \in \mathfrak{S}$; and $A_1, \ldots, A_k \in \mathcal{B}(\mathbb{R})$, the Borel sigma algebra on $\mathbb{R}$.

Hewitt and Savage [51, p. 472] called a measurable space $(S, \mathfrak{S})$ *presentable* (or in some usages, the sigma algebra $\mathfrak{S}$ itself is called *presentable*) if for any exchangeable sequence of random variables $(X_n)_{n\in\mathbb{N}}$ from $(\Omega, \mathcal{F}, \mathbb{P})$ to $(S, \mathfrak{S})$, the condition (5.1) holds for some probability measure $\mathscr{P}$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$. The *mixing measure* $\mathscr{P}$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ corresponding to an exchangeable sequence of random variables, if it

exists, is unique—this is shown in Hewitt–Savage [51, Theorem 9.4, p. 489].

**Remark 5.1.1.** In the situation when $S$ is a topological space, we will end up using the Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ induced by the so-called $A$-topology. This sigma algebra contains the aforementioned sigma algebra $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ generated by cylinder sets. While the integrand in (5.1) only "sees" $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$, using the larger Borel sigma algebra induced by the $A$-topology opens up the possibility to use tools from nonstandard topological measure theory. Thus our main result (Theorem 7.3.7) is stated in terms of measures on this larger sigma algebra, though it includes a corresponding statement in terms of measures on $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$. For the sake of historical consistency, we will continue using the sigma algebra $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ in the context of presentability during this introduction.

In this terminology, the original result of de Finetti [28] thus says that the state space $(\{0,1\},\mathcal{P}(\{0,1\}))$ is presentable (where by $\mathcal{P}(S)$ we denote the power set of a set $S$). In [29], de Finetti generalized the result to real-valued random variables and showed that the Borel sigma algebra on $\mathbb{R}$ is presentable. Dynkin [36] also solved the case of real-valued random variables independently.

Hewitt and Savage [51] observed that the methods used so far required some sense of separability of the state space $S$ in an essential way. They were able to overcome this requirement by using new ideas from convexity theory—they looked at the set of exchangeable distributions on the product space $S^\infty$ as a convex set, of which the (coordinate-wise) independent distributions (whose values at $B_1 \times \ldots \times B_k$ are being integrated on the right side of (5.1)) are the extreme points. Using the Krein–Milman–Choquet theorems, they were thus able to extend de Finetti's theorem to the case in

which the state space $S$ is a compact Hausdorff space with the sigma algebra $\mathfrak{S}$ being the collection of all Baire subsets of $S$ (see [51, Theorem 7.2, p. 483]). Thus in their terminology, Hewitt and Savage proved that all compact Hausdorff spaces equipped with their Baire sigma algebra are presentable:

**Theorem 5.1.2** (Hewitt–Savage)**.** *Let $S$ be a compact Hausdorff space and let $\mathcal{B}_{\mathrm{a}}(\mathrm{S})$ denote the Baire sigma algebra on $S$ (which is the smallest sigma algebra with respect to which any continuous function $f\colon S \to \mathbb{R}$ is measurable). Then $\mathcal{B}_{\mathrm{a}}(\mathrm{S})$ is presentable.*

What does the result of Hewitt and Savage say about the presentability of Borel sigma algebras, as opposed to Baire sigma algebras? As a consequence of their theorem, they were able to show that the Borel sigma algebra of an arbitrary Borel subset of the real numbers is presentable (see [51, p. 484]), generalizing the earlier works of de Finetti [29] and Dynkin [36] (both of whom independently showed the presentability of the Borel sigma algebra on the space of real numbers).

For a topological space $T$, we will denote its Borel sigma algebra (that is, the smallest sigma algebra containing all open subsets) by $\mathcal{B}(T)$. Recall that a Polish space is a separable topological space that is metrizable with a complete metric. A subset of a Polish space is called an *analytic set* if it is representable as a continuous image of a Borel subset of some (potentially different) Polish space. As pointed out by Varadarajan [98, p. 219], the result of Hewitt and Savage immediately implies that any state space $(S, \mathfrak{S})$ that is *analytic* is also presentable. Here an *analytic* space refers to a measurable space that is isomorphic to $(T, \mathcal{B}(T))$ where $T$ is an *analytic subset* of a Polish space, equipped with the subspace topology (see also, Mackey [72, Theorem 4.1, p. 140]). In particular, all Polish

spaces equipped with their Borel sigma algebras are presentable.

**Remark 5.1.3.** Note that both Mackey and Varadarajan use the standard conventions in descriptive set theory of referring to a measurable space as a Borel space (thus, the original conclusion of Varadarajan was stated for "Borel analytic spaces"). We will not use descriptive set theoretic considerations in this work, and hence we decided to not use the adjective 'Borel' in quoting Varadarajan above, so as to avoid confusion with Borel subsets of topological spaces that we will generally consider in this chapter.

The above observation of Varadarajan is the state of the art for modern treatments of de Finetti's theorem for Borel sigma algebras on topological state spaces. For example, Diaconis and Freedman [32, Theorem 14, p. 750] reproved the result of Hewitt and Savage using their approximate de Finetti's theorem for finite exchangeable sequences in any state space (wherein they needed a nice topological structure on the state space to be able to take the limit to go from their approximate de Finetti's theorem on finite exchangeable sequences to the exact de Finetti's theorem on infinite exchangeable sequences). They then concluded (see [32, p. 751]) that de Finetti's theorem holds for state spaces that are isomorphic to Borel subsets of a Polish space. Since any Borel subset of a Polish space is also analytic, this observation is a special case of Varadarajan's. In his monograph, Kallenberg [57, Theorem 1.1] has a proof of de Finetti's theorem for any state space that is isomorphic to a Borel subset of the closed interval $[0, 1]$, a formulation that is contained in the above.

As is justified from the above discussion, the generalization of de Finetti's theorem to more general state spaces is sometimes referred to in the literature as the de Finetti–

Hewitt–Savage theorem.

Due to a lack of counterexamples at the time, a natural question arising from the work of Hewitt and Savage [51] was whether de Finetti's theorem held without any topological assumptions on the state space $S$. This was answered in the negative by Dubins and Freedman [35] who constructed a separable metric space $S$ on which de Finetti's theorem does not hold for some exchangeable sequence of $S$-valued Borel measurable random variables. In terms of the (pushforward) measure induced by the sequence on the countable product $S^\infty$ of the state space, Dubins [34] further showed that the counterexample in [35] is singular to the measure induced by any presentable sequence. This counterexample suggests that some topological conditions are typically needed in order to avoid such pathological cases, though it may be difficult to identify the most general set of conditions that work.

Let us define the following related concept for individual sequences of exchangeable random variables.

**Definition 5.1.4.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let $(X_n)_{n\in\mathbb{N}}$ be an exchangeable sequence of random variables taking values in some state space $(S, \mathfrak{S})$. Then the sequence $(X_n)_{n\in\mathbb{N}}$ is said to be *presentable* if it satisfies (5.1) for some unique probability measure $\mathscr{P}$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$.

Thus a state space $(S, \mathfrak{S})$ is presentable if and only if *all* exchangeable sequences of $S$-valued random variables are presentable. It is interesting to note that any Borel probability measure on a Polish space (which is the setting for the modern treatments of de Finetti–Hewitt–Savage theorem) is automatically Radon (see Definition 1.2.3). Curiously

enough, the counterexample of Dubins and Freedman was for a state space on which non-Radon measures are theoretically possible. The main result of this chapter shows that the Radonness of the common distribution of the underlying exchangeable random variables is actually sufficient for de Finetti's theorem to hold for any Hausdorff state space (equipped with its Borel sigma algebra). In particular, this implies that the exchangeable random variables constructed in the counterexample of Dubins and Freedman do not have a Radon distribution. Restricting to random variables with Radon distributions (which is actually not that restrictive as many areas of probability theory work under that assumption in any case) shows that there does not exist a non-presentable exchangeable sequence of this type. For brevity of expression, let us make the following definitions.

**Definition 5.1.5.** An identically distributed sequence $(X_n)_{n\in\mathbb{N}}$ of random variables taking values in a Hausdorff space $S$ equipped with its Borel sigma algebra $\mathcal{B}(S)$ is said to be *Radon-distributed* if the pushforward probability measure induced on $(S, \mathcal{B}(S))$ by $X_1$ is Radon. It is said to be *tightly distributed* if this pushforward measure is tight (see also Definition 1.2.2).

Focusing on Hausdorff state spaces, while the answer to the original question of whether de Finetti's theorem holds without topological assumptions is indeed in the negative (as the counterexample of Dubins and Freedman shows), we are still able to show that the most commonly studied exchangeable sequences (that is, those that are Radon-distributed) taking values in *any* Hausdorff space are presentable, thus establishing an affirmative answer from a different perspective. Ignoring the various technicalities in the statement of our main result (Theorem 7.3.7), we can thus briefly summarize our contribu-

tion to the above question as follows.

**Theorem 5.1.6.** *Any Radon-distributed exchangeable sequence of random variables taking values in a Hausdorff space (equipped with its Borel sigma algebra) is presentable.*

A closer inspection of our proof shows that we will not use the full strength of the assumption of Radonness of the common distribution of exchangeable random variables—the theorem is still true for sequences of exchangeable random variables whose common distribution is tight and outer regular on compact sets (see the discussion following Theorem 7.3.7).

Before we give an overview of our methods, let us first describe a common practice in statistics that is intimately connected to the reasoning behind a statement like equation (5.1) that we are trying to generalize for sequences of Radon-distributed exchangeable random variables.

### 5.2. A heuristic strategy motivated by statistics

Let $\mathfrak{S}$ be a sigma algebra on a state space $S$. Suppose we devise an experiment to sample values from an identically distributed sequence $X_1, \dots, X_n$ (where $n \in \mathbb{N}$ can theoretically be as large as we please) of random variables from some underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(S, \mathfrak{S})$. Depending on the way the experiment is conducted, within each iteration of the experiment it might not be justified to assume that the sampled values are independent, but it might be reasonable to still believe that the distribution of $(X_1, \dots, X_n)$ is invariant under permutations of indices. Depending on the application, one might be interested in the joint distribution of two (or more) of the $X_i$, which is difficult to establish without an assumption of independence. However, only under an assump-

tion of exchangeability, it is not very difficult to show the following. (Theorem 7.3.1 is a nonstandard version of this statement, with the standard statement having a proof along the same lines—replace the step where we use the hyperfiniteness of $N$ in that proof by an argument about taking limits.)

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \lim_{n\to\infty} \mathbb{E}(\mu_{\cdot,n}(B_1) \cdot \dots \cdot \mu_{\cdot,n}(B_k)) \tag{5.2}$$

for all $k \in \mathbb{N}$ and $B_1, \dots, B_k \in \mathfrak{S}$, where

$$\mu_{\omega,n}(B) = \frac{\#\{i \in [n] \colon X_i(\omega) \in B\}}{n} \text{ for all } \omega \in \Omega \text{ and } B \in \mathfrak{S}. \tag{5.3}$$

Here $[n]$ denotes the initial segment $\{1, \dots, n\}$ of $n \in \mathbb{N}$. In (statistical) practice, for any $k \in \mathbb{N}$ and $B_1, \dots, B_k \in \mathfrak{S}$, we do multiple independent iterations of the experiment. For $j \in \mathbb{N}$, we calculate the product $\mu^{(j)}_{\cdot,n}(B_1) \cdot \dots \cdot \mu^{(j)}_{\cdot,n}(B_k)$ of the "empirical sample means" in the $j^{\text{th}}$ iteration of the experiment. The strong law of large numbers (which we can use because of the assumption that the experiments generating samples of $(X_1, \dots, X_n)$ are independent) thus implies the following:

$$\lim_{m\to\infty} \frac{\sum_{j\in[m]} \mu^{(j)}_{\cdot,n}(B_1) \cdot \dots \cdot \mu^{(j)}_{\cdot,n}(B_k)}{m} = \mathbb{E}\left(\mu_{\cdot,n}(B_1) \cdot \dots \cdot \mu_{\cdot,n}(B_k)\right) \text{ almost surely.} \tag{5.4}$$

By (5.4) and (5.2), we thus obtain the following for all $k \in \mathbb{N}$ and $B_1, \dots, B_k \in \mathfrak{S}$:

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \lim_{n\to\infty} \lim_{m\to\infty} \frac{\sum_{j\in[m]} \mu^{(j)}_{\cdot,n}(B_1) \cdot \dots \cdot \mu^{(j)}_{\cdot,n}(B_k)}{m}. \tag{5.5}$$

Thus, only under an assumption of exchangeability of the values sampled in each experiment, as long as we have a method to repeat the experiment independently, we have the following heuristic algorithm to statistically approximate the joint probability $\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k)$ for any $B_1, \dots, B_k \in \mathfrak{S}$:

(i) In each iteration of the experiment, sample a large number (this corresponds to $n$ in (5.5)) of values.

(ii) Conduct a large number (this corresponds to $m$ in (5.5)) of such independent experiments.

(iii) The average of the empirical sample means $\mu_{\cdot,n}^{(j)}(B_1)\cdot\ldots\cdot\mu_{\cdot,n}^{(j)}(B_k)$ (as $j$ varies in $[m]$) is then an approximation to $\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k)$.

As hinted earlier, the above heuristic idea is at the heart of the intuition behind de Finetti's theorem as well. How do we make this idea more precise to hopefully get a version of de Finetti theorem of the form (5.1)? Suppose for the moment that we have fixed some sigma algebra on $\mathfrak{P}(\mathrm{S})$ (we will come back to the issue of which sigma algebra to fix) such that the following natural conditions are met:

(i) For each $n \in \mathbb{N}$, the map $\omega \mapsto \mu_{\omega,n}$ is a $\mathfrak{P}(\mathrm{S})$-valued random variable on $\Omega$.

(ii) For each $B \in \mathfrak{S}$, the map $\mu \mapsto \mu(B)$ is a real-valued random variable on $\mathfrak{P}(\mathrm{S})$.

For each $n \in \mathbb{N}$, this would define a pushforward probability measure $\nu_n$ on $\mathfrak{P}(\mathrm{S})$ that is supported on $\{\mu_{\omega,n} \colon \omega \in \Omega\} \subseteq \mathfrak{P}(\mathrm{S})$, such that

$$\int_{\mathfrak{P}(\mathrm{S})} \mu(B_1)\ldots\mu(B_k)d\nu_n(\mu) = \int_{\Omega} \mu_{\omega,n}(B_1)\ldots\mu_{\omega,n}(B_k)d\mathbb{P}(\omega)$$
$$\text{for all } B_1, \ldots, B_k \in \mathfrak{S}. \tag{5.6}$$

Comparing (5.2) and (5.6), it is clear that we are looking for conditions that guarantee there to be a measure $\nu$ on $\mathfrak{P}(\mathrm{S})$ such that the following holds:

$$\lim_{n\to\infty} \int_{\mathfrak{P}(\mathrm{S})} \mu(B_1)\ldots\mu(B_k)d\nu_n(\mu) = \int_{\mathfrak{P}(\mathrm{S})} \mu(B_1)\ldots\mu(B_k)d\nu(\mu)$$
$$\text{for all } B_1, \ldots, B_k \in \mathfrak{S}. \tag{5.7}$$

Intuitively, equation (5.7) is a statement of convergence (in some sense) of $\nu_n$ to $\nu$.

A naive candidate for $\nu$ could come from (5.6) if the following are true:

1. There exists an almost sure set $\Omega' \subseteq \Omega$ such that for each $B \in \mathfrak{S}$, the limit $\lim_{n\to\infty} \mu_{\omega,n}(B)$ exists for all $\omega \in \Omega'$. Up to null sets in $\Omega$, this would thus define a map $\omega \mapsto \mu_\omega$ from $\Omega$ to the space of all real-valued functions on $\mathfrak{S}$, where $\mu_\omega(B) = \lim_{n\to\infty} \mu_{\omega,n}(B)$.

2. The function $\mu_\omega : \mathfrak{S} \to [0,1]$ is actually a probability measure on $(S, \mathfrak{S})$.

Indeed if these two conditions are true, then one may define $\nu$ to be the pushforward on $\mathfrak{P}(\mathrm{S})$ of the map $\omega \mapsto \mu_\omega$. A weaker version of (1) is often interpreted as a generalization of the strong law of large numbers for exchangeable random variables—see, for instance, Kingman [61, Equation (2.2), p. 185], which can be easily modified to work in the setting of an arbitrary $(S, \mathfrak{S})$ to conclude that $\lim_{n\to\infty} \mu_{\omega,n}(B)$ exists for all $\omega$ in an almost sure set that depends on $B$. Of course, an issue with this idea is that if we have too many (that is, uncountably many) different choices for $B \in \mathfrak{S}$, then there is no guarantee that an almost sure set would exist that works for all $B \in \mathfrak{S}$ simultaneously. The condition (2) is even more delicate, as showing countable additivity of $\mu_\omega$ would require some control on the rates at which the sequences $(\mu_{\omega,n}(B))_{n\in\mathbb{N}}$ converge for different $B \in \mathfrak{S}$.

Thus we seem to have reached a dead end in this heuristic strategy in the absence of having more information about the specific structure of our spaces and measures. We now describe a generalization of a slightly different type before explaining our method of proof.

### 5.3. Ressel's Radon presentability and the ideas behind our proof

As we describe next, our strategy (motivated by the statistical heuristics from Section 5.2) for proving de Finetti's theorem naturally leads to an investigation into a de

Finetti style theorem first proved by Ressel in [83]. Ressel studied de Finetti-type theorems using techniques from abstract harmonic analysis. His insight was to look for indirect generalizations of de Finetti's theorem; that is, those generalizations which do not prove (5.1) for a state space in a strict sense, but rather prove an analogous statement applicable to nicer classes of random variables, with the smaller space of Radon probability measures being considered (as opposed to the space of all Borel probability measures). Before we proceed, let us make some of these technicalities more precise.

**Definition 5.3.1.** Let $\mathfrak{P}(\mathrm{T})$ and $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ respectively denote the sets of all Borel probability measures and Radon probability measures on a Hausdorff space $T$. The *weak topology* (or *narrow topology*) on either of these sets is the smallest topology under which the maps $\mu \mapsto \mathbb{E}_\mu(f)$ are continuous for each real-valued bounded continuous function $f\colon S \to \mathbb{R}$.

**Definition 5.3.2.** Let a sequence of random variables $(X_n)_{n\in\mathbb{N}}$ taking values in a Hausdorff space $S$ be called *jointly Radon distributed* if the pushforward measure induced by the sequence on $(S^\infty, \mathcal{B}(S^\infty))$ (the product of countably many copies of $S$, equipped with its Borel sigma algebra) is Radon.

**Definition 5.3.3.** Let a jointly Radon distributed sequence of exchangeable random variables $(X_n)_{n\in\mathbb{N}}$ be called *Radon presentable* if there is a unique Radon measure $\mathscr{P}$ on the space $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ of all Radon measures on $S$ (equipped with the Borel sigma algebra induced by its weak topology) such that the following holds for all $k \in \mathbb{N}$:

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) = \int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{P}(\mu)$$
$$\text{for all } B_1, \ldots, B_k \in \mathcal{B}(S). \tag{5.8}$$

Note that (5.8) is an analog of (5.1). This terminology of Ressel is inspired from the similar terminology of presentable spaces introduced by Hewitt and Savage [51].

One of the results that Ressel proved (see [83, Theorem 3, p. 906]) says that all completely regular Hausdorff spaces are Radon presentable. Ressel's theorem, in particular, shows that all Polish spaces and all locally compact Hausdorff spaces are Radon presentable (see [83, p. 907]). In fact, as we show in Appendix D (see Theorem D.6), there is a standard measure theoretic argument by which Ressel's result on completely regular Hausdorff spaces implies the Hewitt–Savage generalization of de Finetti's theorem (Theorem 5.1.2). Thus, although it appears to be in a slightly different form, Ressel's result indeed is a generalization of the de Finetti–Hewitt–Savage theorem in a strict sense. Prior to the statement of his theorem, he remarked the following (see [83, p. 906]):

"It might be true that all Hausdorff spaces have this property."

This conjecture of Ressel was confirmed by Winkler [101] using ideas from convexity theory (similar in spirit to Hewitt–Savage [51]). Fremlin showed in his treatise [42] that a stronger statement is actually true. Replacing the requirement of being jointly Radon distributed with the weaker requirement of being jointly quasi-Radon distributed (this notion is defined in Fremlin [42, 411H, p. 5]) and marginally Radon distributed (that is, the individual common distribution of the random variables must be Radon), Fremlin [42, 459H, p. 166] showed that all such exchangeable sequences also satisfy (5.8). One of our main results generalizes this further to situations where no assumptions on the joint distribution of the sequence of exchangeable random variables are needed:

**Theorem 7.3.2.** *Let $S$ be a Hausdorff topological space, with $\mathcal{B}(S)$ denoting its Borel*

*sigma algebra. Let* $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ *be the space of all Radon probability measures on* $S$ *and* $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ *be the Borel sigma algebra on* $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ *with respect to the* $A$*-topology on* $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$*.*

*Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Let* $X_1, X_2, \dots$ *be a sequence of exchangeable* $S$*-valued random variables such that the common distribution of the* $X_i$ *is Radon on* $S$*. Then there exists a unique probability measure* $\mathscr{P}$ *on* $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ *such that the following holds for all* $k \in \mathbb{N}$*:*

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \int\limits_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{P}(\mu)$$

$$\textit{for all } B_1, \dots, B_k \in \mathcal{B}(S). \tag{7.63}$$

We have not yet described the concept of $A$-topology that appears in the above theorem. In general, if $S$ is a topological space and $\mathfrak{S} = \mathcal{B}(S)$ is the Borel sigma algebra on $S$, then there are natural ways to topologize the space $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$) of Borel probability measures (respectively Radon probability measures) on $S$, which would thus lead to natural (Borel) sigma algebras on $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$). Although we had already established that any such sigma algebra on $\mathfrak{P}(\mathrm{S})$ we work with under the aim of showing (5.1) should be at least as large as the cylinder sigma algebra $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$, a potentially larger Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ induced by some topology on $\mathfrak{P}(\mathrm{S})$ would be desirable in order to be able to use tools from topological measure theory (an analogous statement applies for $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ in the context of (5.8)).

For instance, perhaps the most common topology studied in probability theory is the topology of weak convergence (see Definition 5.3.1). The weak topology on $\mathfrak{P}(\mathrm{S})$, however, is interesting only when there are many real-valued continuous functions on $S$

to work with. If $S$ is completely regular (which is true of all the settings in the previous generalizations of de Finetti's theorem), for instance, then the weak topology on $\mathfrak{P}(\mathrm{S})$ is a natural topology to work with. However, if the state space $S$ is not completely regular then the weak topology may actually be too coarse to be of any interest.

Indeed, as extreme cases, there are regular Hausdorff spaces that do not have any nonconstant continuous real-valued functions. Identifying the most general conditions on the topological space $S$ that guarantee the existence of at least one nonconstant continuous real-valued function was part of Urysohn's research program (see [95] where he posed this question). Hewitt [50] and later Herrlich [47] both showed that regularity of the space $S$ is generally not sufficient. In fact, the result of Herrlich dramatically shows that given any Frechét space $F$ (see ($T_1$) on p. 15 for a definition of Frechét spaces) containing at least two points, there exists a regular Hausdorff space $S$ such that the only continuous functions from $S$ to $F$ are constants. If the topology on $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$) is too coarse, we might not be able to make sense of an equation such as (5.1) (respectively (5.8)), as we would want the induced sigma algebra on $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$) to be large enough such that the evaluation maps $\mu \mapsto \mu(B)$ are measurable for all $B \in \mathcal{B}(S)$.

Thus, we ideally want something finer than the weak topology when working with state spaces that are more general than completely regular spaces. A natural finer topology is the so-called $A$-topology (named after A.D. Alexandroff [10]) defined through bounded upper (or lower) semicontinuous functions from $S$ to $\mathbb{R}$, as opposed to through bounded continuous functions. Thus, the $A$-topology on $\mathfrak{P}(\mathrm{S})$ or $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ is the smallest topology such that the maps $\mu \mapsto \mathbb{E}_{\mu}(f)$ on either space are upper semicontinuous for

each bounded upper semicontinuous function $f: S \to \mathbb{R}$. With respect to the Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ or $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ induced by this topology, the evaluation maps $\mu \mapsto \mu(B)$ are indeed measurable for all $B \in \mathcal{B}(S)$ (see Theorem 6.2.7 and Theorem 6.3.2), which is something we necessarily need in order to even write an equation such as (5.1) or (5.8) meaningfully. The next section is devoted to a thorough study of this topology.

How is a generalization of Ressel's theorem in the form of Theorem 7.3.2 connected to our generalization of the classical de Finetti's theorem as stated in Theorem 5.1.6 (see Theorem 7.3.7 for a more precise statement)? The idea is that any sequence of exchangeable random variables satisfying (5.8) must also satisfy the more classical equation (5.1) of de Finetti–Hewitt–Savage (see Theorem 7.3.6). This follows from elementary topological measure theory arguments that exploit the specific structure of the subspace topology induced by the $A$-topology. Thus, extending Ressel's theorem to a wider class of exchangeable random variables also proves the classical de Finetti's theorem for that class of exchangeable random variables. Let us now describe the intuition behind our proof idea, which will complete the story by showing that such an idea naturally leads to an investigation into a generalization of Ressel's theorem in the form of Theorem 7.3.2.

The idea is to carry out the naive strategy from Section 5.2 using hyperfinite numbers from nonstandard analysis as tools to model large sample sizes. Fix a hyperfinite $N > \mathbb{N}$ and study the map $\omega \mapsto \mu_{\omega,N}$ from ${}^*\Omega$ to ${}^*\mathfrak{P}(\mathrm{S})$. This map induces an internal probability measure (through the pushforward) on the space ${}^*\mathfrak{P}(\mathrm{S})$ of all internal probability measures on ${}^*S$. That is, this pushforward measure $Q_N$ (say) lives in the space ${}^*\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$. In view of (5.7) (and the nonstandard characterization of limits), we want to

have a standard probability measure $\mathscr{Q}$ on $\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ that is close to $Q_N$ in the sense that the integral of the function $\mu \mapsto \mu({}^*B_1) \cdot \ldots \cdot \mu({}^*B_k)$ with respect to $Q_N$ is infinitesimally close to its integral with respect to ${}^*\mathscr{Q}$ for any $k \in \mathbb{N}$ and $B_1, \ldots, B_k \in \mathcal{B}(S)$.

As the space $\mathfrak{P}(\mathrm{S})$ (and hence the space $\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ has a topology on it (namely, the $A$-topology), a natural way to look for a standard element in $\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ close to a given element of ${}^*\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ is to try to see if this given element has a unique standard part (or if it is at least nearstandard). If $T$ is a Hausdorff space, then there are certain natural sufficient conditions for an element in ${}^*\mathfrak{P}(\mathrm{T})$ to be nearstandard (see Section 2.2, more specifically Theorem 6.2.15 and Theorem 6.1.2). However, in our case, the Hausdorffness of $\mathfrak{P}(\mathrm{S})$ is too much to ask for in general (see Corollary 6.2.18)! We remedy this situation by focusing on a nicer subspace of $\mathfrak{P}(\mathrm{S})$—it is known that if the underlying space $S$ is Hausdorff then the space $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ of all *Radon* probability measures on $S$ is also Hausdorff (see Topsøe [93], or Theorem 6.3.4 for our proof). The internal measures $\mu_{\omega,N}$ are internally Radon for all $\omega \in {}^*\Omega$ (as they are supported on the hyperfinite sets $\{X_1(\omega), \ldots, X_N(\omega)\}$). Hence, this move from $\mathfrak{P}(\mathrm{S})$ to $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ does not affect our strategy—the pushforward $P_N$ induced by the map $\omega \mapsto {}^*\mu_{\omega,N}$ from $\Omega$ to $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ lives in ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, in which we try to find its standard part $\mathscr{P}$ in order to complete our proof.

The main tool in finding a standard part of this pushforward is Theorem 6.2.15, which is used in conjunction with Theorem 6.1.2 (originally from Albeverio et al. [6, Proposition 3.4.6, p. 89]). This technique is called "pushing down Loeb measures" and is well-known in the nonstandard literature (see, for example, Albeverio et al. [6, Chapter 3.4] or Ross [84, Section 3]). It is often used to construct a standard measure that is close

in some sense to an internal (nonstandard) measure. The way we develop the theory of $A$-topology allows us to interpret this classical technique of pushing down Loeb measures as actually taking a standard part in a legitimate nonstandard space (of internal measures). See, for example, Theorem 6.2.15, Remark 6.2.16, and Theorem 6.3.5. Similar results were obtained in the context of the topology of weak convergence by Anderson [13, Proposition 8.4(ii), p. 684], and by Anderson–Rashid [15, Lemma 2, p. 329] (see also Loeb [69]).

Using Theorem 6.1.2 as described above requires us to first show the existence of large compact sets in $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ in some sense, which is shown to be the case in Theorem 7.2.11 using a version of Prokhorov's theorem in this setting (see Theorem 6.5.4). It is in this proof that we need the Radonnes of the underlying distribution of $X_1$, thus explaining how our statistical heuristic naturally leads to an investigation of a generalization of Ressel's theorem to sequences of Radon-distributed exchangeable random variables, rather than the classical presentability of Hewitt and Savage.

After setting up this abstract machinery for pushing down Loeb measures, the main computational result that is sufficient for Theorem 7.3.2 is Theorem 7.3.1, which, as mentioned earlier, is the nonstandard version of (5.2) from our statistical heuristic in Section 5.2. The fact that this is a sufficient condition follows naturally from the general topological measure theory of hyperfinitely many identically distributed random variables that is developed in Section 7.2. It should be pointed out that the proof of Theorem 7.3.1 uses a similar combinatorial construction as Diaconis–Freedman's proof of the finite, approximate version of de Finetti's theorem in [32]. In fact, the proof shows that the two results are different ways to express the same idea (see also the discussion following the

statement of Theorem 7.3.1). The form of the result presented here can be given an intuitive underpinning based on Bayes' theorem (this is made more precise in Appendix E, where an alternative proof of Theorem 7.3.1 is provided). This is noteworthy from the point of view that Theorem 7.3.1 is the key ingredient in our proof of the generalization of a result (namely de Finetti's theorem) usually considered foundational for Bayesian statistics (see Savage [88, Section 3.7], and Orbanz–Roy [76]).

In some sense, we prove a highly general de Finetti's theorem using the same underlying basic idea that works for the simplest versions of de Finetti's theorem (that being the idea of approximating using empirical sample means), the technical machinery from topological measure theory and nonstandard analysis notwithstanding. The next chapter is devoted to setting up this technical machinery, while our proof is finally fleshed out in Chapter 7.

For a more thorough introduction to exchangeability, see Aldous [9], Kingman [61], and Kallenberg [57]. Besides a recent paper of the author on a nonstandard proof of de Finetti's theorem for Bernoulli random variables (see Alam [5] which was covered in Chapter 4), there is some precedence in the use of nonstandard analysis in this field, as Hoover [53, 54] studied the notions of exchangeability for multi-dimensional arrays using nonstandard methods in the guise of ultraproducts. In view of this work, Aldous [9, p. 179] had also expressed the hope of nonstandard analysis being useful in other topics in exchangeability. Another example is Dacunha-Castelle [27] who also used ultraproducts to study exchangeability in Banach spaces.

# Chapter 6. Some Results from Nonstandard and Topological Measure Theory

In this chapter, the main object of study is the space of probability measures $\mathfrak{P}(\mathrm{T})$ on a topological space $T$. We develop basic results on the so-called $A$-topology on $\mathfrak{P}(\mathrm{T})$. While some of this material can be viewed as a review of known results in topological measure theory (for which Topsøe [93] is our main reference), we provide a self-contained exposition that is aided by perspectives provided from nonstandard analysis. This leads to both new proofs of known results as well as some new results. A highlight of this chapter is a quick nonstandard proof of a generalization of Prokhorov's theorem (see Theorem 6.5.2; see also Section 6.5 for a historical discussion on Prokhorov's theorem).

An overarching goal of this discussion is to describe the method of pushing down Loeb measures, which is one of the main tools in our work as it allows us to precisely talk about when a nonstandard measure on the nonstandard extension of a topological space is, in a reasonable sense, infinitesimally close to a standard measure (this idea will be made more precise at the end of our discussion on Alexandroff topology in the next section; see, for example, Theorem 6.2.15 and Remark 6.2.16).

## 6.1. Pushing down Loeb measures

Recall the construction of the Loeb measure (see Section 1.3.3) corresponding to an internal probability space $(\mathfrak{T}, \mathcal{A}, \nu)$. In this section, we will work in the case when $\mathfrak{T}$ is the nonstandard extension of a topological space $T$ (that is, $\mathfrak{T} = {}^*T$, and $\mathcal{A}$ is the algebra ${}^*\mathcal{B}(T)$ of internally Borel subsets of ${}^*T$). Note that both here and in the sequel, we will use 'internally' as an adjective to describe nonstandard counterparts of certain standard

concepts. For instance, just as the Borel subsets of $T$ are the elements of $\mathcal{B}(T)$, the internally Borel subsets refer to elements of ${}^*\mathcal{B}(T)$. Similarly, an internally finite set will refer to a hyperfinite set, and an internally Radon probability measure on ${}^*T$ will refer to an element of ${}^*\mathfrak{P}_r(T)$, where $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is the space of Radon probability measures on $T$.

Given an internal probability space $({}^*T, {}^*\mathcal{B}(T), \nu)$, if we know that $\mathbf{st}^{-1}(B)$ is Loeb measurable with respect to the corresponding Loeb space $({}^*T, L({}^*\mathcal{B}(T)), L\nu)$ for all Borel sets $B \in \mathcal{B}(T)$, then one can define a Borel measure on $(T, \mathcal{B}(T))$ by defining the measure of a Borel set $B$ as $L\nu(\mathbf{st}^{-1}(B))$. The fact that this defines a Borel measure in this case is easily checked. This measure is not a probability measure, however, except in the case that the set of nearstandard points $\mathbf{Ns}({}^*T) := \mathbf{st}^{-1}(T)$ is Loeb measurable with Loeb measure equaling one.

Thus, in the setting of an internal probability space $({}^*T, {}^*\mathcal{B}(T), \nu)$, there are two things to ensure in order to obtain a natural standard probability measure on $(T, \mathcal{B}(T))$ corresponding to the internal measure $\nu$:

(i) The set $\mathbf{st}^{-1}(B)$ must be Loeb measurable for any Borel set $B \in \mathcal{B}(T)$.

(ii) It must be the case that $L\nu(\mathbf{Ns}({}^*T)) = 1$.

Verifying when $\mathbf{st}^{-1}(B)$ is Loeb measurable for all Borel sets $B \in \mathcal{B}(T)$ is a tricky endeavor in general, and has been studied extensively. It is interesting to note that if the underlying space $T$ is regular, then this condition is equivalent to the Loeb measurability of $\mathbf{Ns}({}^*T)$ (this was investigated by Landers and Rogge as part of a larger project on universal Loeb measurability—see [63, Corollary 3, p. 233]; see also Aldaz [7]). Prior to Landers and Rogge, the same result was proved for locally compact Hausdorff spaces by

Loeb [69]. Also, Henson [46] gave characterizations for measurability of $\mathbf{st}^{-1}(B)$ when the underlying space is either completely regular or compact. See also the discussion after Theorem 3.2 in Ross [84] for other relevant results in this context. We will, however, not assume any additional hypotheses on our spaces, and hence we must study sufficient conditions for (i) and (ii) that work for any Hausdorff space.

The results in Albeverio et al. [6, Section 3.4] are appropriate in the general setting of Hausdorff spaces. Their discussion is motivated by the works of Loeb [68, 69] and Anderson [12, 13]. We now outline the key ideas to motivate the main result in this theme (see Theorem 6.1.2, originally from [6, Theorem 3.4.6, p. 89]), which we will heavily use in the sequel.

If the underlying space $T$ is Hausdorff, then an application of Lemma 1.3.9 shows that the collection $\{B \in \mathcal{B}(T) : \mathbf{st}^{-1}(B) \in L({}^*\mathcal{B}(T))\}$ is a sigma algebra if and only if $\mathbf{Ns}({}^*T)$ is Loeb measurable. Thus in that case (that is, when $T$ is Hausdorff), one would need to show that $\mathbf{st}^{-1}(F)$ is Loeb measurable for all closed subsets $F \subseteq T$ (or the corresponding statement for all open subsets of $T$).

Thus, under the assumptions that $\mathbf{st}^{-1}(F)$ is Loeb measurable for all closed subsets $F \subseteq T$, and that $L\nu(\mathbf{Ns}({}^*T)) = 1$, the map $L\nu \circ \mathbf{st}^{-1} \colon \mathcal{B}(T) \to [0,1]$ does define a probability measure on $(T, \mathcal{B}(T))$ whenever $T$ is Hausdorff. This is the content of [6, Proposition 3.4.2, p. 87], which further uses the completeness of the Loeb measures and some nonstandard topology to show that $L\nu \circ \mathbf{st}^{-1}$ is actually a regular, complete measure on $(T, \mathcal{B}(T))$ in this case. Under what conditions can one guarantee that $\mathbf{st}^{-1}(F)$ is Loeb measurable for all closed subsets $F \subseteq T$? Note that if we replace $F$ by a compact set, then this is al-

ways true (for all sufficiently saturated nonstandard extensions):

**Lemma 6.1.1.** *Let $T$ be a topological space and let $\tau$ be the topology on $T$. Then we have, for any compact subset $K \subseteq T$:*

$$\mathbf{st}^{-1}(K) = \bigcap\{{}^*O : K \subseteq O \text{ and } O \in \tau\}.$$

*As a consequence, for any compact set $K \subseteq T$, the set $\mathbf{st}^{-1}(K)$ is universally Loeb measurable with respect to $({}^*T, {}^*\mathcal{B}(T))$. That is, for any internal probability measure $\nu$ on $({}^*T, {}^*\mathcal{B}(T))$ and any compact $K \subseteq T$, we have $\mathbf{st}^{-1}(K) \in L_\nu({}^*\mathcal{B}(T))$. Furthermore, we have:*

$$L\nu(\mathbf{st}^{-1}(K)) = \inf\{LP({}^*O) : K \subseteq O \text{ and } O \in \tau\} \text{ for all compact subsets } K \subseteq T.$$

See [6, Lemma 3.4.4 and Proposition 3.4.5, pp. 88-89] for a proof of Lemma 6.1.1 (note that $T$ is assumed to be Hausdorff in [6] but is not needed for this proof). Thus, if we require that there are arbitrarily large compact sets with respect to $({}^*T, {}^*\mathcal{B}(T), \nu)$ in the sense that

$$\sup\{L\nu(\mathbf{st}^{-1}(K)) : K \text{ is a compact subset of } T\} = 1, \tag{6.1}$$

then the completeness of the Loeb space $({}^*T, L({}^*\mathcal{B}(T)), L\nu)$ allows us to conclude that $L\nu(\mathbf{Ns}({}^*T)) = 1$ and that $\mathbf{st}^{-1}(F)$ is Loeb measurable for all closed sets $F \subseteq T$. In this case, if $T$ is also assumed to be Hausdorff, then $L\nu \circ \mathbf{st}^{-1}$ is thus shown to be a Radon measure on $(T, \mathcal{B}(T))$ (see [6, Corollary 3.4.3, p. 88] for a formal proof). In view of Lemma 6.1.1, we thus immediately obtain the following result; see also [6, Theorem 3.4.6, p. 89] for a detailed proof of a slightly more general form.

**Theorem 6.1.2.** *Let $T$ be a Hausdorff space with $\mathcal{B}(T)$ denoting the Borel sigma algebra on $T$. Let $({}^*T, {}^*\mathcal{B}(T), \nu)$ be an internal, finitely additive probability space and let $({}^*T, L({}^*\mathcal{B}(T)), L\nu)$ denote the corresponding Loeb space. Let $\tau$ denote the topology on $T$. Then $\mathbf{st}^{-1}(K) \in L({}^*\mathcal{B}(T))$ for all compact $K \subseteq T$.*

*Assume further that for each $\epsilon \in \mathbb{R}_{>0}$, there is a compact set $K_\epsilon$ with*

$$\inf\{L\nu({}^*O) : K_\epsilon \subseteq O \text{ and } O \in \tau\} \geq 1 - \epsilon. \tag{6.2}$$

*Then $L\nu \circ \mathbf{st}^{-1}$ is a Radon probability measure on $T$.*

Note that Theorem 6.1.2 is a special case of [6, Theorem 3.4.6, p. 89], which we have chosen to present here in this simplified form because we do not need the full power of the latter result in our current work. In the next section, we will study a natural topology on the space of all Borel probability measures on a topological space $T$. It will turn out that under the conditions of Theorem 6.1.2, the measure $\nu$ on $({}^*T, {}^*\mathcal{B}(T))$ is nearstandard to $L\nu \circ \mathbf{st}^{-1}$ in the nonstandard topological sense (see Theorem 6.2.15). Also, the subspace of Radon probability measures is always Hausdorff (see Theorem 6.3.4), so that Theorem 6.1.2 will allow us to push down, in a unique way, a natural nonstandard measure on the space of all (Radon) probability measures in our proof of de Finetti's theorem. We finish this section with a corollary that follows from the definition of tightness.

**Corollary 6.1.3.** *Let $T$ be a Hausdorff space and let $\mu$ be a tight probability measure on it. Then $L^*\mu \circ \mathbf{st}^{-1}$ is a Radon probability measure on $T$.*

### 6.2. The Alexandroff topology on the space of probability measures on a topological space

For a topological space $T$ and a function $f \colon T \to \mathbb{R}$, we say:

(i) $f$ is *upper semicontinuous* at $x_0 \in T$ if for every $\alpha \in \mathbb{R}$ with $\alpha > f(x_0)$, there is an open neighborhood $U$ of $x_0$ such that $\alpha > f(x)$ for all $x \in U$.

(ii) $f$ is *lower semicontinuous* at $x_0 \in T$ if for every $\alpha \in \mathbb{R}$ with $\alpha < f(x_0)$, there is an open neighborhood $U$ of $x_0$ such that $\alpha < f(x)$ for all $x \in U$.

A function $f\colon T \to \mathbb{R}$ is called *upper (respectively lower) semicontinuous* if $f$ is upper (respectively lower) semicontinuous at every point in $T$. The following characterization of upper/lower semicontinuity is immediate from the definition.

**Lemma 6.2.1.** *A function $f\colon T \to \mathbb{R}$ is upper semicontinuous if and only if the set $\{x \in T : f(x) < \alpha\}$ is open for every $\alpha \in \mathbb{R}$.*

*A function $f\colon T \to \mathbb{R}$ is lower semicontinuous if and only if the set $\{x \in T : f(x) > \alpha\}$ is open for every $\alpha \in \mathbb{R}$.*

*As a consequence, a function $f\colon T \to \mathbb{R}$ is upper semicontinuous if and only if $-f$ is lower semicontinuous.*

For a topological space $T$, we will denote the set of all bounded upper semicontinuous functions on $T$ by $USC_b(T)$. Similarly, $LSC_b(T)$ will denote the set of all bounded lower semicontinuous functions on $T$.

**Remark 6.2.2.** It is immediate from the definition that the indicator function of an open set is lower semicontinuous, and that the indicator function of a closed set is upper semicontinuous.

For a topological space $T$, let $\mathcal{B}(T)$ denote the Borel sigma algebra of $T$—that is, $\mathcal{B}(T)$ is the smallest sigma algebra containing all open sets. Consider the set $\mathfrak{P}(\mathrm{T})$ of all Borel probability measures on $T$. For each bounded measurable $f\colon T \to \mathbb{R}$, define the map

$E_f\colon \mathfrak{P}(\mathrm{T}) \to \mathbb{R}$ by

$$E_f(\mu) := \mathbb{E}_\mu(f) = \int_T f d\mu. \tag{6.3}$$

**Definition 6.2.3.** Let $T$ be a topological space. The *$A$-topology* on the space of Borel probability measures $\mathfrak{P}(\mathrm{T})$ is the weakest topology for which the maps $E_f$ are upper semi-continuous for all $f \in USC_b(T)$.

The "$A$" in $A$-topology refers to A.D. Alexandroff [10], who pioneered the study of weak convergence of measures and gave many of the results that we will use. In the literature, the term 'weak topology' is sometimes used in place of '$A$-topology'; see, for instance, Topsøe [93, p. 40]. However, following Kallianpur [58], Blau [18], and Bogachev [20], we will reserve the term weak topology for the smallest topology on $\mathfrak{P}(\mathrm{T})$ that makes the maps $E_f$ continuous for every bounded continuous function $f\colon T \to \mathbb{R}$. For a bounded Borel measurable function $f\colon T \to \mathbb{R}$ and $\alpha \in \mathbb{R}$, define the following sets:

$$\mathfrak{U}_{f,\alpha} := \{\mu \in \mathfrak{P}(\mathrm{T}) : \mathbb{E}_\mu(f) < \alpha\}, \tag{6.4}$$

$$\text{and } \mathfrak{L}_{f,\alpha} := \{\mu \in \mathfrak{P}(\mathrm{T}) : \mathbb{E}_\mu(f) > \alpha\}. \tag{6.5}$$

By Definition 6.2.3 and Lemma 6.2.1, the $A$-topology on $\mathfrak{P}(\mathrm{T})$ is the smallest topology under which $\mathfrak{U}_{f,\alpha}$ is open for all $f \in USC_b(T)$ and $\alpha \in \mathbb{R}$. More formally, the $A$-topology on $\mathfrak{P}(\mathrm{T})$ is induced by the subbasis $\{\mathfrak{U}_{f,\alpha} : f \in USC_b(T), \alpha \in \mathbb{R}\}$. Also, by the last part of Lemma 6.2.1, this collection is actually equal to the collection $\{\mathfrak{L}_{f,\alpha} : f \in LSC_b(T), \alpha \in \mathbb{R}\}$. These observations are summarized in the following useful description of the $A$-topology.

**Lemma 6.2.4.** *Let $T$ be a topological space, and $\mathfrak{P}(\mathrm{T})$ be the set of all Borel probability*

*measures on $T$. The $A$-topology on $\mathfrak{P}(\mathrm{T})$ is generated by the subbasis*

$$\{\mathfrak{U}_{f,\alpha} : f \in USC_b(T), \alpha \in \mathbb{R}\} = \{\mathfrak{L}_{f,\alpha} : f \in LSC_b(T), \alpha \in \mathbb{R}\}. \tag{6.6}$$

**Remark 6.2.5.** Note that, by Lemma 6.2.1, a function is continuous if and only if it is both upper and lower semicontinuous. Thus, by Lemma 6.2.4, the $A$-topology also makes the maps $E_f$ continuous for every bounded continuous function $f : T \to \mathbb{R}$, thus implying that the $A$-topology is, in general, finer than the weak topology on $\mathfrak{P}(\mathrm{T})$. The two topologies coincide if $T$ has a rich topological structure. For example, in Kallianpur [58, Theorem 2.1, p. 948], it is proved that the the $A$-topology and the weak topology on $\mathfrak{P}(\mathrm{T})$ are the same if $T$ is a completely regular Hausdorff space such that it can be embedded as a Borel subset of a compact Hausdorff space. This, in particular, means that the two topologies are the same if the underlying space $T$ is a Polish space (that is, a complete separable metric space) or is a locally compact Hausdorff space.

**Remark 6.2.6.** While we are focusing on Borel probability measures on topological spaces, we could have analogously defined the $A$-topology on the space of all finite Borel measures on a topological space as well. Although we will not work with non-probability measures, we are not losing too much generality in doing so. In fact, Blau [18, Theorem 1, p. 24] shows that the space of finite Borel measures on a topological space $T$ is naturally homeomorphic to the product of $\mathfrak{P}(\mathrm{T})$ and the space of positive reals. Thus, from a practical point of view, most results that we will obtain for $\mathfrak{P}(\mathrm{T})$ will also hold for the $A$-topology on the space of all finite measures (some results such as Prokhorov's theorem that talk about subsets of finite measures will hold in that setting with an added assumption of uniform boundedness that is inherently satisfied by all sets of probability

measures).

By Remark 6.2.2, we know that $\{\mu \in \mathfrak{P}(\mathrm{T}) : \mu(G) > \alpha\}$ is open for any open subset $G \subseteq T$ and $\alpha \in \mathbb{R}$; and similarly, $\{\mu \in \mathfrak{P}(\mathrm{T}) : \mu(F) < \alpha\}$ is open for any closed subset $F \subseteq T$ and $\alpha \in \mathbb{R}$. Lemma 6.2.9 will show that the $A$-topology is generated by either of these types of subbasic open sets as well. We first use the above facts to show that the evaluation maps are Borel measurable with respect to the $A$-topology.

**Theorem 6.2.7.** *Let $B$ be a Borel subset of a topological space $T$. Let $\mathfrak{P}(\mathrm{T})$ be the space of all Borel probability measures on $T$ equipped with the $A$-topology. Then the evaluation map $e_B \colon \mathfrak{P}(\mathrm{T}) \to [0,1]$ defined by $e_B(\mu) := \mu(B)$ is Borel measurable.*

*Proof.* Consider the collection

$$\mathcal{B} = \{B \in \mathcal{B}(T) : e_B \text{ is Borel measurable}\}.$$

This collection contains $T$, since $f_T$ is the constant function 1, which is continuous. It is also closed under taking relative complements. That is, if $A \subseteq B$ and $A, B \in \mathcal{B}$ then $B \backslash A \in \mathcal{B}$ as well, since $f_{B\backslash A} = f_B - f_A$ in that case. Finally, $\mathcal{B}$ is closed under countable increasing unions. That is, if $(B_n)_{n\in\mathbb{N}} \subseteq \mathcal{B}$ is a sequence of sets such that $B_n \subseteq B_{n+1}$ for all $n \in \mathbb{N}$, then $B := \cup_{n\in\mathbb{N}} B_n \in \mathcal{B}$ as well (this is because $f_B = \lim_{n\to\infty} f_{B_n}$ is a limit of Borel measurable functions in that case). Thus, $\mathcal{B}$ is a Dynkin system.

Furthermore, $\mathcal{B}$ contains all open sets since for any open set $G \subseteq T$, the set $\{\mu \in \mathfrak{P}(\mathrm{T}) : \mu(G) > \alpha\}$ is Borel measurable (in fact, open) for all $\alpha \in \mathbb{R}$. Thus, by Dynkin's $\pi$-$\lambda$ theorem, it contains, and hence is equal to, $\mathcal{B}(T)$, completing the proof. $\square$

Lemma 6.2.9 finds other useful subbases for the $A$-topology. We first need the fol-

lowing intuitive fact from probability theory as a tool in its proof.

**Lemma 6.2.8.** *Suppose $\mathbb{P}_1$ and $\mathbb{P}_2$ are probability measures on the same space and $X$ is a bounded random variable such that*

$$\mathbb{P}_1(X > x) \geq \mathbb{P}_2(X > x) \text{ for all } x \in \mathbb{R}. \tag{6.7}$$

*Then, we have $\mathbb{E}_{\mathbb{P}_1}(X) \geq \mathbb{E}_{\mathbb{P}_2}(X)$.*

*Proof.* With $\lambda$ denoting the Lebesgue measure on $\mathbb{R}$, we have the following representation of the expected value of any bounded random variable $X$ (see, for example, Lo [66, Proposition 2.1]):

$$\mathbb{E}_{\mathbb{P}}(X) = \int_{(0,\infty)} \mathbb{P}(X > x) d\lambda(x) - \int_{(-\infty,0)} \mathbb{P}(X < x) d\lambda(x). \tag{6.8}$$

Let $\mathbb{P}_1$, $\mathbb{P}_2$ and $X$ be as in the statement of the lemma. Then, using (6.7), we obtain the following for each $x \in \mathbb{R}$:

$$\begin{aligned}
\mathbb{P}_1(X < x) &= 1 - \mathbb{P}_1(X \geq x) \\
&= 1 - \mathbb{P}_1\left(\bigcap_{n\in\mathbb{N}} \left\{X > x - \frac{1}{n}\right\}\right) \\
&= 1 - \lim_{n\to\infty} \mathbb{P}_1\left(X > x - \frac{1}{n}\right) \\
&\leq 1 - \lim_{n\to\infty} \mathbb{P}_2\left(X > x - \frac{1}{n}\right) \\
&= \mathbb{P}_2(X < x). \qquad (6.9)
\end{aligned}$$

Using (6.8), (6.7) and (6.9), we thus obtain:

$$\begin{aligned}\mathbb{E}_{\mathbb{P}_1}(X) &= \int_{(0,\infty)} \mathbb{P}_1(X > x) d\lambda(x) - \int_{(-\infty,0)} \mathbb{P}_1(X < x) d\lambda(x)\\ &\geq \int_{(0,\infty)} \mathbb{P}_2(X > x) d\lambda(x) - \int_{(-\infty,0)} \mathbb{P}_2(X < x) d\lambda(x)\\ &= \mathbb{E}_{\mathbb{P}_2}(X),\end{aligned}$$

completing the proof. □

**Lemma 6.2.9.** *For each Borel set $B \in \mathcal{B}(T)$, let*

$$\mathfrak{U}_{B,\alpha} := \{\mu \in \mathfrak{P}(\mathrm{T}) : \mu(B) < \alpha\}, \tag{6.10}$$

$$\text{and } \mathfrak{L}_{B,\alpha} := \{\mu \in \mathfrak{P}(\mathrm{T}) : \mu(B) > \alpha\}. \tag{6.11}$$

*Then the topology on $\mathfrak{P}(\mathrm{T})$ generated by $\{\mathfrak{U}_{F,\alpha} : \alpha \in \mathbb{R}$ and $F$ is closed$\}$ as a subbasis is the same as the topology on $\mathfrak{P}(\mathrm{T})$ generated by $\{\mathfrak{L}_{G,\alpha} : \alpha \in \mathbb{R}$ and $G$ is open$\}$ as a subbasis. Both of these topologies equal the A-topology on $\mathfrak{P}(\mathrm{T})$.*

*Proof.* If $G$ is an open subset of $T$ and $\alpha \in \mathbb{R}$, then we have

$$\mathfrak{L}_{G,\alpha} = \bigcup_{\epsilon \in \mathbb{R}_{>0}} \mathfrak{U}_{T \setminus G, 1-\alpha+\epsilon}. \tag{6.12}$$

Since the complement of an open set is closed, this shows that a basic open set in the topology on $\mathfrak{P}(\mathrm{T})$ generated by $\{\mathfrak{L}_{G,\alpha} : \alpha \in \mathbb{R}$ and $G$ is open$\}$ as a subbasis, is a finite intersection of sets that are unions of elements in the collection $\{\mathfrak{U}_{F,\alpha} : \alpha \in \mathbb{R}$ and $F$ is closed$\}$. That is, a basic open set in the topology on $\mathfrak{P}(\mathrm{T})$ generated by $\{\mathfrak{L}_{G,\alpha} : \alpha \in \mathbb{R}$ and $G$ is open$\}$ as a subbasis, is also open in the topology on $\mathfrak{P}(\mathrm{T})$ generated by $\{\mathfrak{U}_{F,\alpha} : \alpha \in \mathbb{R}$ and $F$ is closed$\}$ as a subbasis. A similar argument shows that

a basic open set in the latter topology is also open in the former topology, thus proving that the two topologies are equal.

Let $\tau_1$ be the $A$-topology and $\tau_2$ be the topology induced by $\{\mathfrak{L}_{G,\alpha} : G \text{ open}, \alpha \in \mathbb{R}\}$ as a subbasis. From the discussion preceding this lemma, it is clear that $\tau_2 \subseteq \tau_1$. Conversely, let $U \in \tau_1$ and $\nu \in U$. By Lemma 6.2.4, there exist finitely many $f_1, \dots f_k \in LSC_b(T)$ and $\beta_1, \dots, \beta_k \in \mathbb{R}$ such that the following holds:

$$\nu \in \cap_{i=1}^{k} \mathfrak{L}_{f_i,\beta_i} \subseteq U. \tag{6.13}$$

Let $\mathbb{E}_\nu(f_i) = \delta_i > \beta_i$ for all $i \in \{1, \dots, k\}$. For each $i \in \{1, \dots, k\}$ and $\alpha \in \mathbb{R}$, let $G_{i,\alpha} = \{x \in T : f_i(x) > \alpha\}$, which is an open set by Lemma 6.2.1. Define

$$\mathfrak{L}_{\alpha,\epsilon} := \cap_{i=1}^{k} \mathfrak{L}_{G_{i,\alpha},\nu(G_{i,\alpha})-\epsilon} \text{ for all } \alpha \in \mathbb{R} \text{ and } \epsilon \in \mathbb{R}_{>0}. \tag{6.14}$$

Note that $\nu \in \mathfrak{L}_{\alpha,\epsilon}$ for all $\alpha \in \mathbb{R}$ and $\epsilon \in \mathbb{R}_{>0}$, where $\mathfrak{L}_{\alpha,\epsilon}$ is a subbasic set for the topology $\tau_2$. Thus it is sufficient to prove the following claim.

**Claim 6.2.10.** *There exists $n \in \mathbb{N}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$, $\epsilon_1, \dots, \epsilon_n \in \mathbb{R}_{>0}$ such that*

$$\cap_{j=1}^{n} \mathfrak{L}_{\alpha_j,\epsilon_j} \subseteq \cap_{i=1}^{k} \mathfrak{L}_{f_i,\beta_i} \subseteq U.$$

*Proof of Claim 6.2.10.* Suppose, if possible, that the claim is not true. Then for each $n \in \mathbb{N}$ and any $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ and $\epsilon_1, \dots, \epsilon_n \in \mathbb{R}_{>0}$, there must exist some $\mu \in \mathfrak{P}(\mathrm{T})$ such that $\mu \in \cap_{i=1}^{k} \mathfrak{L}_{G_{i,\alpha_j},\nu(G_{i,\alpha_j})-\epsilon_j}$ for all $j \in \{1, \dots, n\}$, but $\mu \notin \cap_{i=1}^{k} \mathfrak{L}_{f_i,\beta_i}$. By transfer, the following internal set is non-empty for each $n \in \mathbb{N}$, $\vec{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ and $\vec{\epsilon} :=$

$(\epsilon_1, \ldots, \epsilon_n) \in (\mathbb{R}_{>0})^n$.

$$B_{\vec{\alpha},\vec{\epsilon}} := \{\mu \in {}^*\mathfrak{P}(\mathrm{T}) : \mu({}^*G_{i,\alpha_j}) > \nu(G_{i,\alpha_j}) - \epsilon_j \text{ for all } i \in \{1, \ldots, k\}, j \in \{1, \ldots, n\}$$
$$\text{but } {}^*\mathbb{E}_\mu({}^*f_i) \leq \beta_i \text{ for some } i \in \{1, \ldots, k\}\}. \tag{6.15}$$

By the same argument (after concatenating different finite sequences of $\vec{\alpha}$'s and $\vec{\epsilon}$ 's, we note that the collection $\cup_{n\in\mathbb{N}}\{B_{\vec{\alpha},\vec{\epsilon}} : \vec{\alpha} \in \mathbb{R}^n, \vec{\epsilon} \in (\mathbb{R}_{>0})^n\}$ has the finite intersection property. By saturation, there exists $\mu \in {}^*\mathfrak{P}(\mathrm{T})$ such that the following holds:

$$\exists i_o \in \{1, \ldots, k\} \text{ such that } {}^*\mathbb{E}_\mu({}^*f_{i_0}) \leq \beta_{i_0} < \mathbb{E}_\nu(f_{i_0}) \text{ but}$$
$$\mu({}^*G_{i_0,\alpha}) > \nu(G_{i_0,\alpha}) - \epsilon \text{ for all } \alpha \in \mathbb{R}, \epsilon \in \mathbb{R}_{>0}. \tag{6.16}$$

But this implies that $L\mu({}^*G_{i_0,\alpha}) \geq L^*\nu({}^*G_{i_0,\alpha})$ for all $\alpha \in \mathbb{R}_{>0}$, which yields:

$$\begin{aligned} L\mu(\mathbf{st}({}^*f_{i_0}) > \alpha) &\geq \lim_{\epsilon\to 0} L\mu({}^*f_{i_0} > \alpha + \epsilon) \\ &\geq \lim_{\epsilon\to 0} L^*\nu({}^*f_{i_0} > \alpha + \epsilon) \\ &= L^*\nu(\mathbf{st}({}^*f_{i_0}) > \alpha). \end{aligned} \tag{6.17}$$

By Lemma 6.2.8 and (6.17), we thus obtain:

$$\mathbb{E}_{L\mu}(\mathbf{st}({}^*f_{i_0})) \geq \mathbb{E}_{L^*\nu}(\mathbf{st}({}^*f_{i_0})). \tag{6.18}$$

However, using the fact that finitely bounded internally measurable functions are $S$-integrable and that $\beta_{i_0}$ and $\mathbb{E}_\nu(f_{i_0})$ are real numbers, taking standard parts in the first inequality of (6.16) yields

$$\mathbb{E}_{L\mu}(\mathbf{st}({}^*f_{i_0})) < \mathbb{E}_{L^*\nu}(\mathbf{st}({}^*f_{i_0})),$$

which directly contradicts (6.18), completing the proof.

□

In the rest of the dissertation, we will interchangeably use either of the collections in Lemma 6.2.4 and Lemma 6.2.9 as a subbasis, depending on convenience.

If $T$ is a topological space, then for any subset $T' \subseteq T$, we can view $T'$ as a topological space under the subspace topology. By routine measure theoretic arguments, it is clear that the Borel sigma algebra on $T'$ with respect to the subspace topology contains precisely those sets that are intersections of $T'$ with Borel subsets of $T$. That is,

$$\mathcal{B}(T') = \{B \cap T' : B \in \mathcal{B}(T)\} \text{ for all } T' \subseteq T. \tag{6.19}$$

Indeed, the collection on the right side of (6.19) is a sigma algebra that contains all open subsets of $T'$ under the subspace topology (as any open subset of $T'$ is of the type $G \cap T'$ for some open, and hence Borel, subset of $T$). Using a similar argument, we can show the following functional version of (6.19):

**Lemma 6.2.11.** *Let $T'$ be a subspace of a topological space $T$. For any bounded $\mathcal{B}(T)$-measurable function $f\colon T \to \mathbb{R}$, its restriction $f\restriction_{T'}\colon T' \to \mathbb{R}$ is $\mathcal{B}(T')$-measurable.*

*Proof.* Consider the collection

$$\mathcal{C} := \{f\colon T \to \mathbb{R} : f\restriction_{T'} \text{ is } \mathcal{B}(T')\text{-measurable}\}. \tag{6.20}$$

By (6.19), the collection $\mathcal{C}$ contains the indicator function $\mathbb{1}_B$ of each $B \in \mathcal{B}(T)$. The collection $\mathcal{C}$ is clearly an $\mathbb{R}$-vector space closed under increasing limits. Thus $\mathcal{C}$ contains all bounded $\mathcal{B}(T)$-measurable functions by the monotone class theorem. □

Thus if $T'$ is a subspace of a topological space $T$ and $\mu \in \mathfrak{P}(T')$, then one can naturally define an "extension" $\mu' \in \mathfrak{P}(\mathrm{T})$ of $\mu$ as follows:

$$\mu'(B) := \mu(B \cap T') \text{ for all } B \in \mathcal{B}(T). \tag{6.21}$$

That $\mu'$ is well-defined follows from (6.19), and the fact that $\mu'$ is a Borel probability measure on $T$ follows from the fact that $\mu$ is a Borel probability measure on $T'$. We had put scare quotes around the word 'extension' to emphasize that $\mu$ is not necessarily a restriction of its extension $\mu'$ in this sense. Indeed, $T'$ could be a non-Borel subset of $T$ or it might not be known whether it is a Borel subset of $T$, in which cases $\mu'$ might not even be defined on a typical Borel subset of $T'$. This will be the situation in Section 7.3, when we will have to extend a probability measure defined on the space $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ of all Radon probability measures on a topological space $S$ to a Borel probability measure on $\mathfrak{P}(\mathrm{S})$, the space of all Borel probability measures on $S$ (thus $\mathfrak{P}(\mathrm{S})$ will play the role of $T$ and $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ will play the role of $T'$). We will study the subspace topology on the space of Radon probability measures in the next section. Let us now summarize our discussion on the extension of a Borel measure on a subspace so far and prove a natural correspondence of expected values in the following lemma.

**Lemma 6.2.12.** *Let $T$ be a topological space and let $T' \subseteq T$ be a subspace. Let $\mu \in \mathfrak{P}(T')$ be a Borel probability measure on $T'$ and let $\mu'$ be its extension, as defined in (6.21). Then $\mu' \in \mathfrak{P}(\mathrm{T})$. Furthermore, we have:*

$$\mathbb{E}_{\mu'}(f) = \mathbb{E}_{\mu}\left(f{\restriction}_{T'}\right) \text{ for all bounded } \mathcal{B}(T)\text{-measurable functions } f\colon T \to \mathbb{R}. \tag{6.22}$$

*Proof.* Only (6.22) remains to be proven. This follows from (6.21) and the monotone class theorem. $\square$

Before we proceed, let us recall the concept of nets which often play the same role in abstract topological spaces that sequences play in metric spaces. This discussion is mostly borrowed from a combination of Kelley [60, Chapter 2] and Bogachev [20, Chapter 2].

A *directed set* $D$ is a set with a partial order $\succcurlyeq$ on it such that for any pair of elements $i, j \in D$, there exists an element $k \in D$ having the property $k \succcurlyeq i$ and $k \succcurlyeq j$. For a topological space $T$, a *net* in $T$ is a function $f$ from a directed set $D$ into $T$, with $f(i)$ usually written as $x_i$ for each $i \in D$. Mimicking the notation for sequences, we denote a generic net by $(x_i)_{i \in D}$.

For a net $(c_i)_{i \in D}$ of real numbers, we define the *superior and inferior limits* as follows:

$$\limsup_{i \in D}(c_i) := \mathbf{lub}\{c \in \mathbb{R} : \ \forall k \in D \ \exists j \succcurlyeq k \text{ such that } c_j \geq c\}, \tag{6.23}$$

$$\text{and } \liminf_{i \in D}(c_i) = -\limsup_{i \in D}(-c_i), \tag{6.24}$$

where $\mathbf{lub}(A)$ (for a set $A \subseteq \mathbb{R}$) denotes the least upper bound of $A$.

A net $(x_i)_{i \in D}$ in a topological space $T$ is said to *converge* to a point $x \in T$ (written $(x_i)_{i \in D} \to x$) if for each open neighborhood $U$ of $x$, there exists $k \in D$ such that $x_i \in U$ for all $i \succcurlyeq k$. This definition clearly coincides with the usual definition of convergence of a sequence (thinking of $\mathbb{N}$ as a directed set with the usual order on it). The following generalizes the characterization of closure in metric spaces using sequences to abstract topological spaces using nets (see Kelley [60, Theorem 2.2] for a proof):

**Theorem 6.2.13.** *Let $T$ be a topological space and let $A \subseteq T$. A point $x$ belongs to the*

*closure of a $A$ if and only if there is a net in $A$ converging to $x$.*

With the language of nets, we can prove the following useful characterizations of convergence in the $A$-topology, originally due to Alexandroff (see Topsøe [93, Theorem 8.1, p. 40] for a similar result).

**Theorem 6.2.14.** *Let $T$ be a topological space and $\mathfrak{P}(\mathrm{T})$ be the space of Borel probability measures on $T$, equipped with the $A$-topology. For a net $(\mu_i)_{i\in D}$ in $\mathfrak{P}(\mathrm{T})$, the following are equivalent:*

*(i)* $(\mu_i)_{i\in D} \to \mu$.

*(ii)* $\limsup_{i\in D}(\mathbb{E}_{\mu_i}(f)) \le \mathbb{E}_\mu(f)$ *for all* $f \in USC_b(T)$.

*(iii)* $\liminf_{i\in D}(\mathbb{E}_{\mu_i}(f)) \ge \mathbb{E}_\mu(f)$ *for all* $f \in LSC_b(T)$.

*(iv)* $\limsup_{i\in D}(\mu_i(F)) \le \mu(F)$ *for all closed sets* $F \subseteq T$.

*(v)* $\liminf_{i\in D}(\mu_i(G)) \ge \mu(G)$ *for all open sets* $G \subseteq T$.

*Proof.* The equivalences *(ii)* $\iff$ *(iii)* and *(iv)* $\iff$ *(v)* are clear from (6.24) and the last part of Lemma 6.2.1 (along with the fact that a set is open if and only if its complement is closed). We will prove *(i)* $\iff$ *(ii)* and omit the very similar proof of *(i)* $\iff$ *(iv)*.

Throughout this proof, for any function $f \in USC_b(T)$, define

$$S_f := \{c \in \mathbb{R} :\ \forall k \in D\ \exists j \succcurlyeq k \text{ such that } \mathbb{E}_{\mu_j}(f) \ge c\}. \tag{6.25}$$

$\boxed{\text{Proof of } \textbf{\textit{(i)}} \implies \textbf{\textit{(ii)}}}$ Assume (i)—that is, $(\mu_i)_{i\in D} \to \mu$. Let $f \in USC_b(T)$ and $\beta := \mathbb{E}_\mu(f)$. We want to show that $\beta$ is at least as large as the least upper bound of $S_f$ (see (6.23)). In other words, we want to show that $\beta$ is an upper bound of $S_f$. To that

end, let $c \in S_f$. Suppose, if possible, that $c > \beta = \mathbb{E}_\mu(f)$. Then $\mu$ would be in the subbasic open set $\mathfrak{U}_{\mathfrak{f},\mathfrak{c}} = \{\gamma \in \mathfrak{P}(\mathrm{T}) : \mathbb{E}_\gamma(f) < c\}$. Since $(\mu_i)_{i\in D} \to \mu$, there would exist a $k \in D$ such that $\mu_i \in \mathfrak{U}_{\mathfrak{f},\mathfrak{c}}$ for all $i \succcurlyeq k$. That is,

$$\mathbb{E}_{\mu_i}(f) < c \text{ for all } i \succcurlyeq k. \tag{6.26}$$

Since $c \in S_f$, there would also exist $j \succcurlyeq k$ such that $\mathbb{E}_{\mu_j}(f) \geq c > \beta$. But this contradicts (6.26), so we know that it is not possible for $c > \beta$ to be true. Since $c$ was an arbitrary element of $S_f$, it is now clear that $\beta = \mathbb{E}_\mu(f)$ is an upper bound of $S_f$, completing the proof of *(i)* $\implies$ *(ii)*.

$\boxed{\text{Proof of } \boldsymbol{(ii)} \implies \boldsymbol{(i)}}$ Assume (ii)—that is, $\limsup_{i\in D}(\mathbb{E}_{\mu_i}(f)) \leq \mathbb{E}_\mu(f)$ for all $f \in USC_b(T)$. Suppose, if possible, that $(\mu_i)_{i\in D} \not\to \mu$. Then there would exist finitely many maps $f_1, \ldots, f_n \in USC_b(T)$ and real numbers $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, such that the set

$$U := \bigcap_{t=1}^{n} \{\gamma \in \mathfrak{P}(\mathrm{T}) : \mathbb{E}_\gamma(f_t) < \alpha_t\}$$

is a basic open neighborhood of $\mu$, and such that for any $k \in D$, one may find $j \succcurlyeq k$ such that $\mu_j \notin U$. Thus:

$$\text{For all } k \in D, \text{ there exists } j \succcurlyeq k \text{ such that } \mathbb{E}_{\mu_j}(f_t) \geq \alpha_t \text{ for some } t \in \{1, \ldots, n\}. \tag{6.27}$$

Since $\limsup_{i\in D}(\mathbb{E}_{\mu_i}(f_t)) \leq \mathbb{E}_\mu(f_t)$, we also know that $\mathbb{E}_\mu(f_t)$ is an upper bound of $S_{f_t}$ for all $t \in \{1, \ldots, n\}$. Since $\mu \in U$, we conclude that $\alpha_t$ is strictly larger than the least upper bound of $S_{f_t}$ for all $t \in \{1, \ldots, n\}$. In particular, $\alpha_t \notin S_{f_t}$ for any $t \in \{1, \ldots, n\}$. By the definition of $S_{f_t}$, this means that for each $t \in \{1, \ldots, n\}$, there exists a $k_t \in D$ such

that for all $j \succcurlyeq k_t$, we have $\mathbb{E}_{\mu_j}(f_t) < \alpha_t$. Since $D$ is a directed set, there exists $\tilde{k}$ such that $\tilde{k} \succcurlyeq k_t$ for all $t \in \{1, \ldots, n\}$. We thus conclude:

$$\mathbb{E}_{\mu_j}(f_t) < \alpha_t \text{ for all } j \succcurlyeq \tilde{k} \text{ and } t \in \{1, \ldots, n\}. \tag{6.28}$$

But (6.27) and (6.28) contradict each other, thus showing that the net $(\mu_i)_{i \in D}$ must in fact converge to $\mu$. This completes the proof of *(i)* $\implies$ *(ii)*. □

Returning to the theme of Loeb measures, we are now in a position to show that for any internal probability $\nu$ on $({}^*T, {}^*\mathcal{B}(T))$, if $L\nu \circ \mathbf{st}^{-1}$ is a legitimate Borel probability measure on $(T, \mathcal{B}(T))$, then $\nu$ is infinitesimally close to $L\nu \circ \mathbf{st}^{-1}$ in the sense that the former is nearstandard to the latter in ${}^*\mathfrak{P}(\mathrm{T})$. Combined with Theorem 6.1.2, we also have sufficient conditions for when this happens.

**Theorem 6.2.15.** *Let $T$ be a Hausdorff space. Suppose $({}^*T, {}^*\mathcal{B}(T), \nu)$ is an internal probability space, and let $({}^*T, L({}^*\mathcal{B}(T)), L\nu)$ be the associated Loeb space. If $L\nu \circ \mathbf{st}^{-1} \colon \mathcal{B}(T) \to [0, 1]$ is a Borel probability measure on $T$, then $\nu$ is nearstandard in ${}^*\mathfrak{P}(T)$ to $L\nu \circ \mathbf{st}^{-1}$. That is,*

$$\nu \in \mathbf{st}^{-1}(L\nu \circ \mathbf{st}^{-1}). \tag{6.29}$$

*Proof.* Let $\nu$ be as in the statement of the theorem. Thus, $L\nu \circ \mathbf{st}^{-1} \in \mathfrak{P}(\mathrm{T})$, which implicitly also requires that $\mathbf{st}^{-1}(B) \in L({}^*\mathcal{B}(T))$ for all $B \in \mathcal{B}(T)$. For brevity, denote $L\nu \circ \mathbf{st}^{-1}$ by $\mu$. Suppose $G_1, \ldots, G_n$ are finitely many open sets and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ are such that the set

$$\mathfrak{U} := \bigcap_{i=1}^{n} \{\gamma \in \mathfrak{P}(\mathrm{T}) : \gamma(G_i) > \alpha_i\} \tag{6.30}$$

is a basic open neighborhood of $\mu$ in $\mathfrak{P}(\mathrm{T})$.

Note that in a Hausdorff space, a subset $G$ is open if and only if $\mathbf{st}^{-1}(G) \subseteq {}^*G$ (see Theorem 1.3.10(i)). Since $\mu \in \mathfrak{U}$, we thus obtain:

$$L\nu({}^*G_i) \geq L\nu(\mathbf{st}^{-1} G_i) = \mu(G_i) > \alpha_i \text{ for all } i \in \{1, \ldots, n\}.$$

Since the $\alpha_i$ are real, it thus follows that

$$\nu({}^*G_i) > \alpha_i \text{ for all } i \in \{1, \ldots, n\}.$$

By the definition (6.30) of $\mathfrak{U}$, it is thus clear that $\nu \in {}^*\mathfrak{U}$. Since $\mathfrak{U}$ was an arbitrary neighborhood of $\mu$, it thus follows that $\nu \in \mathbf{st}^{-1}(\mu)$, completing the proof. □

**Remark 6.2.16.** For an internal probability measure $\nu$ on ${}^*T$, whenever $L\nu \circ \mathbf{st}^{-1}$ is a probability measure on the underlying topological space $T$, we typically call the measure $L\nu \circ \mathbf{st}^{-1}$ as being obtained by "pushing down" the Loeb measure $L\nu$. In fact, Albeverio et al. [6, Section 3.4] denotes $L\nu \circ \mathbf{st}^{-1}$ by $\mathbf{st}(L\nu)$, calling it the standard part of $\nu$. Theorem 6.2.15 makes this precise by showing that $L\nu \circ \mathbf{st}^{-1}$ is indeed nearstandard to $\nu \in {}^*\mathfrak{P}(\mathrm{T})$ when we equip the space of probability measures $\mathfrak{P}(\mathrm{T})$ with a natural topology. In Section 6.3, we show that the subset $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ of Radon probability measures on $T$ is Hausdorff, which will allow us to show that $L\nu \circ \mathbf{st}^{-1}$ is actually *the* standard part of $\nu$ as an element of ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ (see Theorem 6.3.5).

Theorem 6.2.15 applied together with Corollary 6.1.3 implies that the nonstandard extension of a tight measure is nearstandard to a Radon measure. Thus, while not all tight measures are Radon, each tight measure is close to a Radon measure from a topological point of view. More precisely, for each tight measure, there is a Radon measure

such that the former belongs to each open neighborhood of the latter. We record this as a corollary.

**Corollary 6.2.17.** *Let $T$ be a Hausdorff space and $\mu$ be a tight probability measure on it. Then there exists a Radon measure $\mu'$ on $T$ such that $\mu \in \mathfrak{U}$ for all open neighborhoods $\mathfrak{U}$ of $\mu'$ in $\mathfrak{P}(\mathrm{T})$.*

*Proof.* By Corollary 6.1.3 and Theorem 6.2.15, we have that $\mu' := L^*\mu \circ \mathbf{st}^{-1}$ is a Radon probability measure such that ${}^*\mu \in \mathbf{st}^{-1}(\mu')$. Also, by definition of $\mathbf{st}^{-1}$, we have that ${}^*\mu \in {}^*\mathfrak{U}$ for any open neighborhood $\mathfrak{U}$ of $\mu'$ in $\mathfrak{P}(\mathrm{T})$. By transfer, we have that $\mu \in \mathfrak{U}$ for any open neighborhood $\mathfrak{U}$ of $\mu'$ in $\mathfrak{P}(\mathrm{T})$. $\square$

This, in particular, shows that the $A$-topology is not always Hausdorff. We end this section with this corollary.

**Corollary 6.2.18.** *There exists a topological space $T$ such that the $A$-topology on its space of Borel probability measures $\mathfrak{P}(\mathrm{T})$ is not Hausdorff.*

*Proof.* There is a Hausdorff space $T$ and a Borel probability measure $\mu$ on it such that $\mu$ is tight but not Radon (in fact, $T$ may be taken to be a compact Hausdorff space; see Vakhania–Tarildaze–Chobanyan[96, Proposition 3.5, p.32] for an example/construction). By Corollary 6.2.17, there is a Radon probability measure $\mu'$ (thus $\mu \neq \mu'$ necessarily) such that $\mu$ and $\mu'$ cannot be separated by disjoint open sets in $\mathfrak{P}(\mathrm{T})$. As a consequence, $\mathfrak{P}(\mathrm{T})$ is not Hausdorff. $\square$

## 6.3. Space of Radon probability measures under the Alexandroff topology

In de Finetti's theorem, one wants to construct a second-order probability—a probability measure with certain properties on a space of probability measures. Our strategy will be to first create a nonstandard internal probability measure on the nonstandard extension of our space of probability measures and then "push it down" to get a standard Borel probability measure with the properties we desire of it. However, as is clear from the discussion in Section 6.2 (see, for example, Theorem 6.1.2), this general procedure usually requires the underlying space of probability measures that we are constructing our measure on to be Hausdorff. As Corollary 6.2.18 shows, the space $\mathfrak{P}(\mathrm{T})$ of *all* Borel probability measures that we have studied so far may be too wild! We want to identify a large collection of Borel measures that is Hausdorff under the subspace topology. The subspace of Radon probability measures on a Hausdorff space $T$ that we will focus on in this section serves our purposes adequately (see Theorem 6.3.4).

Recall the concept of Radon probability measures on an arbitrary Hausdorff space $T$ from Definition 1.2.3. The space of all Radon probability measures on $T$ is denoted by $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$, and we equip it with the subspace topology induced by the $A$-topology on $\mathfrak{P}(\mathrm{T})$. We require the Hausdorffness of $T$ to ensure that compact subsets are Borel measurable (as a compact subset of a Hausdorff space is closed).

Being a subspace of $\mathfrak{P}(\mathrm{T})$, a subbasis of $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ can be obtained by intersecting all sets of a given subbasis of $\mathfrak{P}(\mathrm{T})$ with $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$. Hence, by Lemma 6.2.4 and Lemma 6.2.9, we have the following result on various subbases of $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$.

**Lemma 6.3.1.** *Let $T$ be a Hausdorff space. Then the topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ as a subspace of*

$\mathfrak{P}(\mathrm{T})$ *under the $A$-topology is generated by either of the following collections as a subbasis:*

*(i)* $\{\{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \mu(G) > \alpha\} : G$ *an open subset of* $T$ *and* $\alpha \in \mathbb{R}\}$.

*(ii)* $\{\{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \mu(F) < \alpha\} : F$ *a closed subset of* $T$ *and* $\alpha \in \mathbb{R}\}$.

*(iii)* $\{\{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \mathbb{E}_{\mu}(f) > \alpha\} : f \in LSC_b(T)$ *and* $\alpha \in \mathbb{R}\}$.

*(iv)* $\{\{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \mathbb{E}_{\mu}(f) < \alpha\} : f \in USC_b(T)$ *and* $\alpha \in \mathbb{R}\}$.

Henceforth, we will call the subspace topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ as the $A$-topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$, and we will use either of the subbases from Lemma 6.3.1 for this topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$, depending on convenience. Using these subbases, the proofs of most of the results on $\mathfrak{P}(\mathrm{T})$ from Section 6.2 carry over to $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ almost immediately. We state below the analogs of Theorem 6.2.7 and Theorem 6.2.14 respectively (with the similar proofs omitted).

**Theorem 6.3.2.** *Let $B$ be a Borel subset of a Hausdorff space $T$. Let $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ be the space of all Radon probability measures on $T$. Then the evaluation map $e_B\colon \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) \to [0,1]$ defined by $e_B(\mu) := \mu(B)$ is $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{T}))$-measurable.*

**Theorem 6.3.3.** *Let $T$ be a Hausdorff space and $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ be the space of Radon probability measures on $T$, equipped with the $A$-topology. For a net $(\mu_i)_{i\in D}$ in $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$, the following are equivalent:*

*(i)* $(\mu_i)_{i\in D} \to \mu$.

*(ii)* $\limsup_{i\in D}(\mathbb{E}_{\mu_i}(f)) \leq \mathbb{E}_{\mu}(f)$ *for all* $f \in USC_b(T)$.

*(iii)* $\liminf_{i\in D}(\mathbb{E}_{\mu_i}(f)) \geq \mathbb{E}_{\mu}(f)$ *for all* $f \in LSC_b(T)$.

*(iv)* $\limsup_{i\in D}(\mu_i(F)) \leq \mu(F)$ *for all closed sets* $F \subseteq T$.

*(v)* $\liminf_{i\in D}(\mu_i(G)) \geq \mu(G)$ *for all open sets* $G \subseteq T$.

With these results motivated from the results in Section 6.2 out of the way, we now show why $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is inherently a better space to work with than $\mathfrak{P}(\mathrm{T})$—we show that $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is Hausdorff (see also Topsøe [93, Theorem 11.2, p. 49]).

**Theorem 6.3.4.** *If $T$ is a Hausdorff space, then $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is also Hausdorff.*

*Proof.* Let $T$ be a Hausdorff space. Suppose $\mu, \nu$ are two distinct elements of $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$. Since they are distinct Borel measures, there exists an open set $G \subseteq T$ such that $\alpha := \nu(G)$ and $\beta := \mu(G)$ are distinct. Without loss of generality, assume $\alpha < \beta$. Since $\mu$ and $\nu$ are Radon measures, we can find a compact set $K$ such that $K \subseteq G$ and the following holds:

$$\nu(K) \leq \nu(G) = \alpha < \alpha + \frac{3(\beta - \alpha)}{4} < \mu(K) \leq \beta = \mu(G). \tag{6.31}$$

Since $T$ is Hausdorff, all compact subsets of $T$ are closed. In particular, $K$ is closed. Consider the subbasic open set $\mathfrak{V}$ defined by:

$$\mathfrak{V} := \left\{ \gamma \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \gamma(K) < \alpha + \frac{\beta - \alpha}{4} \right\}.$$

By (6.31), it is clear that $\nu \in \mathfrak{V}$ and $\mu \notin \mathfrak{V}$. For each $\gamma \in \mathfrak{V}$, by Radonness, there exists an open set $G_\gamma$ such that $K \subseteq G_\gamma \subseteq G$ and we have:

$$\gamma(G_\gamma) < \alpha + \frac{\beta - \alpha}{2} \text{ for all } \gamma \in \mathfrak{V}. \tag{6.32}$$

Thus the following set, being the complement of a closed set (owing to the fact that an arbitrary intersection of closed sets is closed), is open:

$$\mathfrak{U} := \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) \setminus \left( \bigcap_{\gamma \in \mathfrak{V}} \left\{ \theta \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \theta(G_\gamma) \leq \alpha + \frac{\beta - \alpha}{2} \right\} \right).$$

By (6.31), it is clear that

$$\mu(G_\gamma) \geq \mu(K) > \alpha + \frac{3(\beta-\alpha)}{4} > \alpha + \frac{\beta-\alpha}{2} \text{ for all } \gamma \in \mathfrak{V}.$$

As a consequence, we have $\mu \in \mathfrak{U}$. Furthermore, by (6.32), it is clear that $\mathfrak{V} \cap \mathfrak{U} = \emptyset$, thus completing the proof. □

Since nonstandard extensions of Hausdorff spaces admit unique standard parts (of nearstandard elements), we have the following form of Theorem 6.2.15 for $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$:

**Theorem 6.3.5.** *Let $T$ be a Hausdorff space. Suppose $({}^*T, {}^*\mathcal{B}(T), \nu)$ is an internal probability space, and let $({}^*T, L({}^*\mathcal{B}(T)), L\nu)$ be the associated Loeb space. If $L\nu \circ \mathbf{st}^{-1} \colon \mathcal{B}(T) \to [0,1]$ is a Radon probability measure on $T$, then $\nu$ is nearstandard in ${}^*\mathfrak{P}_r(T)$ to $L\nu \circ \mathbf{st}^{-1}$. That is,*

$$\mathbf{st}(\nu) = L\nu \circ \mathbf{st}^{-1} \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T})\,. \tag{6.33}$$

*Proof.* We use $\mathbf{st}^{-1}_{\mathfrak{P}(\mathrm{T})}$ and $\mathbf{st}^{-1}_{\mathfrak{P}_{\mathrm{r}}(\mathrm{T})}$ to denote standard inverses on subsets of $\mathfrak{P}(\mathrm{T})$ and $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ respectively. By Theorem 6.2.15 and the given information, we have that

$$\nu \in \mathbf{st}^{-1}_{\mathfrak{P}(\mathrm{T})}(L\nu \circ \mathbf{st}^{-1}) \cap {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{T})\,.$$

By Lemma 1.3.13, we have

$$\nu \in \mathbf{st}^{-1}_{\mathfrak{P}_{\mathrm{r}}(\mathrm{T})}(L\nu \circ \mathbf{st}^{-1}).$$

Since $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is Hausdorff, this completes the proof. □

Knowing that $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is Hausdorff for any Hausdorff space $T$ thus allows us to apply results such as Theorem 6.1.2 to uniquely push down internal measures on

$({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{T}), {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{T}))$. In the next section, we will take $T = \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ for a Hausdorff topological space $S$, and construct a nonstandard measure living in ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ that we will be able to push down to a Radon measure on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$.

We begin this theme here with Theorem 6.3.7, which is a result about the uniqueness of the mixing measure in the context of Radon presentability (see Definition 5.3.3). This is different from the related uniqueness result of Hewitt–Savage [51, Theorem 9.4, p. 489] in two ways. Firstly, we are now focusing on the space of Radon probability measures (as opposed to the space of Baire probability measures), and secondly, we are working with the sigma algebra induced by the $A$-topology (as opposed to the cylinder sigma algebra induced by Baire sets). Our proof will use the following generalization of the monotone class theorem (see Dellacherie and Meyer [30, Theorem 21, p. 13-I] for a proof of this result).

**Theorem 6.3.6.** *Let $\mathbb{H}$ be an $\mathbb{R}$-vector space of bounded real-valued functions on some set $\mathcal{S}$ such that the following hold:*

*(i) $\mathbb{H}$ contains the constant functions.*

*(ii) $\mathbb{H}$ is closed under uniform convergence.*

*(iii) For every uniformly bounded increasing sequence of nonnegative functions $f_n \in \mathbb{H}$, the function $\lim_{n\to\infty} f_n$ belongs to $\mathbb{H}$.*

*If $\mathcal{C}$ is a subset of $\mathbb{H}$ which is closed under multiplication, then the space $\mathbb{H}$ contains all bounded functions measurable with respect to $\sigma(\mathcal{C})$ - the smallest sigma algebra with respect to which all functions in $\mathcal{C}$ are measurable.*

**Theorem 6.3.7.** *Let $S$ be a Hausdorff space and let $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ be the space of all Radon probability measures on $S$ under the $A$-topology. Suppose $\mathscr{P}, \mathscr{Q} \in \mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ are such that the*

*following holds:*

$$\int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_n) d\mathscr{P}(\mu) = \int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_n) d\mathscr{Q}(\mu)$$
$$\textit{for all } n \in \mathbb{N} \textit{ and } B_1, \ldots, B_n \in \mathcal{B}(S). \tag{6.34}$$

*Then it must be the case that* $\mathscr{P} = \mathscr{Q}$.

*Proof.* For $m \in \mathbb{N}$, let $\mathcal{M}([0,1]^m)$ denote the space of all bounded Borel measurable functions $f\colon [0,1]^m \to \mathbb{R}$. For each $m \in \mathbb{N}$, consider the following collection of functions:

$$\mathcal{G}_m := \{f \in \mathcal{M}([0,1]^m) : \mathbb{E}_{\mathscr{P}}\left[f\left(\mu(B_1), \ldots, \mu(B_m)\right)\right] = \mathbb{E}_{\mathscr{Q}}\left[f\left(\mu(B_1), \ldots, \mu(B_m)\right)\right]$$
$$\text{for all } B_1, \ldots, B_m \in \mathcal{B}(S)\}.$$

Note that the expected values in the definition of $\mathcal{G}_m$ are well-defined because of Theorem 6.3.2. It is clear that for each $m \in \mathbb{N}$, the collection $\mathcal{G}_m$ contains all polynomials over $m$ variables. Indeed, the collection $\mathcal{G}_m$ is an $\mathbb{R}$-vector space (that is, closed under finite linear combinations), and for a monomial $f\colon [0,1]^m \to \mathbb{R}$ of the type $f(x_1, \ldots, x_m) = {x_1}^{a_1} \cdot \ldots \cdot {x_m}^{a_m}$ (where $a_1, \ldots, a_m \in \mathbb{Z}_{\geq 0}$), the expectation $\mathbb{E}_{\mathscr{P}}\left[f\left(\mu(B_1), \ldots, \mu(B_m)\right)\right]$ is equal to $\mathbb{E}_{\mathscr{Q}}\left[f\left(\mu(B_1), \ldots, \mu(B_m)\right)\right]$ by (6.34). That $\mathcal{G}_m$ satisfies the conditions in Theorem 6.3.6 is also clear by dominated convergence theorem. It is straightforward to verify that the smallest sigma algebra on $[0,1]^m$ with respect to which all polynomials are measurable is the Borel sigma algebra on $[0,1]^m$. Since the set of polynomials over $m$ variables is closed under multiplication, it thus follows from Theorem 6.3.6 that for each $m \in \mathbb{N}$, the collection $\mathcal{G}_m$ contains all bounded Borel measurable functions $f\colon [0,1]^m \to \mathbb{R}$.

Let $\mathcal{G}$ be the collection of those Borel subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ that are assigned the same

measure by $\mathscr{P}$ and $\mathscr{Q}$. More formally, we define:

$$\mathcal{G} := \{\mathfrak{B} \in \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})) : \mathscr{P}(\mathfrak{B}) = \mathscr{Q}(\mathfrak{B})\}. \tag{6.35}$$

Taking $f$ to be the indicator function of a measurable rectangle in $[0,1]^m$, we have thus shown that $\mathcal{G}$ contains the following collection of cylinder sets:

$$\mathcal{C} := \{C_{(B_1,\ldots,B_m),(A_1,\ldots A_m)} : m \in \mathbb{N}; B_1, \ldots, B_m \in \mathcal{B}(S); A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R})\}, \tag{6.36}$$

where

$$\begin{aligned} C_{(B_1,\ldots,B_m),(A_1,\ldots A_m)} &:= \{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \mu(B_1) \in A_1, \ldots, \mu(B_m) \in A_m\} \\ &\qquad \text{for all } m \in \mathbb{N}; B_1, \ldots, B_m \in \mathcal{B}(S); A_1, \ldots, A_m \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

It is clear that the collection $\mathcal{C}$ contains the basic open subsets with respect to the subbasis (i) in Lemma 6.3.1. Thus all basic open subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ are elements of $\mathcal{G}$. Since $\mathcal{G}$ is a sigma algebra, all finite unions of basic open sets are in $\mathcal{G}$. (In fact, all countable unions are in $\mathcal{G}$, but we do not need this fact here.) Let $\mathfrak{C}$ be a compact subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ and let $\epsilon \in \mathbb{R}_{>0}$ be given. Since $\mathscr{P}$ and $\mathscr{Q}$ are Radon measures, we find an open subset $\mathfrak{U}$ of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ such that we have $\mathfrak{C} \subseteq \mathfrak{U}$ and

$$\mathscr{P}(\mathfrak{U}\backslash\mathfrak{C}) < \epsilon \text{ and } \mathscr{Q}(\mathfrak{U}\backslash\mathfrak{C}) < \epsilon. \tag{6.37}$$

Cover $\mathfrak{C}$ by finitely many basic open subsets contained in $\mathfrak{U}$ and let $\mathfrak{V}$ be the union of these basic open subsets. Then, we have (using (6.37)):

$$\mathscr{P}(\mathfrak{V}\backslash\mathfrak{C}) < \epsilon \text{ and } \mathscr{Q}(\mathfrak{V}\backslash\mathfrak{C}) < \epsilon. \tag{6.38}$$

Being, a finite union of basic open sets, we have $\mathfrak{V} \in \mathcal{G}$, or in other words:

$$\mathscr{P}(\mathfrak{V}) = \mathscr{Q}(\mathfrak{V}). \tag{6.39}$$

Using (6.38) and (6.39) (and the triangle inequality), we thus obtain:

$$|\mathscr{P}(\mathfrak{C}) - \mathscr{Q}(\mathfrak{C})| < 2\epsilon. \tag{6.40}$$

Since $\mathfrak{C}$ was an arbitrary compact subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ and $\epsilon \in \mathbb{R}_{>0}$ was arbitrary, this shows that the measures $\mathscr{P}$ and $\mathscr{Q}$ agree on all compact subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Since they are Radon measures, it is thus clear now that they agree on all Borel subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, completing the proof. □

**Remark 6.3.8.** Instead of using Theorem 6.3.6 (after showing that all polynomials in $m$ variables are in $\mathcal{G}_m$ for all $m \in \mathbb{N}$), we could have used the Stone–Weierstrass theorem to first show that all continuous functions on $[0,1]^m$ are in $\mathcal{G}_m$ for all $m \in \mathbb{N}$ and then approximate indicator functions of open subsets of $[0,1]^m$ by increasing sequences of continuous functions to complete the proof using the monotone class theorem. Theorem 6.3.6 achieved the same in a quicker manner.

In the above proof, the only place where Radonness was used was in extending the uniqueness result from the cylinder sigma algebra on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ to the Borel sigma algebra on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. In particular, the same argument shows that without working with Radon measures, one still has uniqueness if we focus on measures over the smallest sigma algebra generated by cylinder sets. We formally record this as a theorem in the next section that is devoted to other sigma algebras on $\mathfrak{P}(\mathrm{S})$.

## 6.4. Useful sigma algebras on spaces of probability measures

Let $S$ be a topological space and $\mathfrak{P}(\mathrm{S})$ be the space of all Borel probability measures on $S$. So far, we have studied the $A$-topology and the Borel sigma algebra $\mathcal{B}(\mathfrak{P}(\mathrm{S}))$ on $\mathfrak{P}(\mathrm{S})$ arising out of it. As Remark 6.2.5 shows, the $A$-topology coincides with the more commonly studied weak topology (which is the smallest topology that makes the map $\mu \mapsto \mathbb{E}_\mu(f)$ continuous for each bounded continuous $f\colon S \to \mathbb{R}$) in the cases when $S$ is a Polish space or when $S$ is a locally compact Hausdorff space. Let $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$ denote the Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ with respect to the weak topology.

For general spaces, the $A$-topology is typically richer than the weak topology, and the corresponding Borel sigma algebra on the space of all probability measures is a very natural sigma algebra to work with from a topological measure theoretic standpoint. However, the Borel sigma algebra arising from the $A$-topology might be too large in some cases—it might contain more events than we might hope to have a grip on in some applications. There are other sigma algebras on spaces of probability measures on $S$ that are also used in practice, some that make sense even if $S$ is *not* a topological space. In fact, constructing a measurable space out of the space of all probability measures (on some space) is the first foundational step needed to talk about prior distributions in a Bayesian nonparametric setting. In Bayesian nonparametrics, it is generally agreed that any reasonable sigma algebra on the space of all probability measures on some measurable space $(S, \mathfrak{S})$ must make the evaluation functions (that is, the functions $\mu \mapsto \mu(B)$ for each $B \in \mathfrak{S}$) measurable. Let us give a name for the smallest sigma algebra with this property.

**Definition 6.4.1.** Let $(S, \mathfrak{S})$ be a measurable space and let $\mathcal{C}(S)$ be the smallest sigma

algebra on $\mathfrak{P}(\mathrm{S})$, the space of all probability measures on $S$, such that for each $B \in \mathfrak{A}$, the evaluation function $\mu \mapsto \mu(B)$ is measurable.

As explained above, the sigma algebra $\mathcal{C}(S)$ is ubiquitous in the nonparametric Bayesian analysis literature. To mention just one classic example, this was the sigma algebra used by Ferguson [38] in his pioneering work on the Dirichlet processes.

When the underlying space $S$ has a topological structure, then it is useful to see how this sigma algebra relates to the Borel sigma algebras arising out of the natural topologies on $\mathfrak{P}(\mathrm{S})$ (namely the $A$-topology and the weak topology). Theorem 6.2.7 and Remark 6.2.5 show that $\mathcal{B}(\mathfrak{P}(\mathrm{S}))$ contains both $\mathcal{C}(S)$ and $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$. In a metric space, the indicator function of an open set is a pointwise limit of uniformly bounded continuous functions, so that by routine measure theory we obtain the following whenever $S$ is a metric space:

$$\{\{\mu \in \mathfrak{P}(\mathrm{S}) : \mu(G) > \alpha\} : G \text{ open in } S \text{ and } \alpha \in \mathbb{R}\} \subseteq \mathcal{B}_w(\mathfrak{P}(\mathrm{S})).$$

In particular, the proof of Theorem 6.2.7 also shows that if $S$ is a metric space, then $\mathcal{C}(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$. Finally, it is not very difficult to observe (for example, see Gaudard and Hadwin [45, Theorem 2.3, p. 171]) that these two sigma algebras actually coincide if $S$ is a separable metric space. We summarize this discussion in the next theorem.

**Theorem 6.4.2.** *Let $S$ be a topological space and let $\mathfrak{P}(\mathrm{S})$ denote the space of all Borel probability measures on $S$. Let $\mathcal{B}(\mathfrak{P}(\mathrm{S}))$ and $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$ be the Borel sigma algebras on $\mathfrak{P}(\mathrm{S})$ with respect to the $A$-topology and the weak topology respectively. Let $\mathcal{C}(S)$ be the smallest sigma algebra on $\mathfrak{P}(\mathrm{S})$ that makes the evaluation functions measurable. Then we*

*have:*

*(i)* $\mathcal{C}(S) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$ *and* $\mathcal{B}_w(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$.

*(ii) If* $S$ *is metrizable, then* $\mathcal{C}(S) \subseteq \mathcal{B}_w(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$.

*(iii) If* $S$ *is a separable metric space, then* $\mathcal{C}(S) = \mathcal{B}_w(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$.

*(iv) If* $S$ *is a complete separable metric space, then* $\mathcal{C}(S) = \mathcal{B}_w(\mathfrak{P}(\mathrm{S})) = \mathcal{B}(\mathfrak{P}(\mathrm{S}))$.

With the requisite terminology now established, we finish this section by formally writing our observations at the end of Section 6.3 as a version of Theorem 6.3.7 for the space of all probability measures (not necessarily Radon). Theorem 6.4.2(iii) allows us to say something more in the case when $S$ is a separable metric space.

**Theorem 6.4.3.** *Let* $S$ *be a topological space and let* $\mathfrak{P}(\mathrm{S})$ *be the space of all Borel probability measures on* $S$ *under the A-topology. Let* $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ *be the smallest sigma algebra such that for any* $B \in \mathcal{B}(S)$*, the evaluation function* $e_B \colon \mathfrak{P}(\mathrm{S}) \to \mathbb{R}$*, defined by* $e_B(\nu) = \nu(B)$*, is measurable. Then* $\mathcal{C}(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$.

*Suppose* $\mathscr{P}, \mathscr{Q}$ *are two probability measures on* $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ *such that the following holds:*

$$\int_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_n) d\mathscr{P}(\mu) = \int_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_n) d\mathscr{Q}(\mu)$$

$$\text{for all } n \in \mathbb{N} \text{ and } B_1, \ldots, B_n \in \mathcal{B}(S).$$

*Then it must be the case that* $\mathscr{P} = \mathscr{Q}$.

*Furthermore, if* $S$ *is a separable metric space, then* $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ *in the above result may be replaced by the Borel sigma algebra* $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$ *induced by the weak topology on* $\mathfrak{P}(\mathrm{S})$.

### 6.5. Generalizing Prokhorov's theorem—tightness implies relative compactness for probability measures on any Hausdorff space

Prokhorov [82, Theorem 1.12] famously proved that a collection $\mathfrak{A}$ of Borel probability measures on a Polish space $T$ (that is, a complete and separable metric space) is relatively compact (that is, the closure $\bar{\mathfrak{A}}$ of $\mathfrak{A}$ is compact) if and only if $\mathfrak{A}$ satisfies the following property that is now known as *tightness* (being a property that is uniformly satisfied by all measures in $\mathfrak{A}$, it is sometimes called "uniform tightness" to avoid confusion with tightness of a particular measure as defined in Definition 1.2.2).

$$\text{(Tightness of } \mathfrak{A}) : \text{For each } \epsilon \in \mathbb{R}_{>0}, \text{ there exists a compact set } K_\epsilon \subseteq T \text{ such that}$$
$$\mu(K_\epsilon) \geq 1 - \epsilon \text{ for all } \mu \in \mathfrak{A}.$$

Those topological spaces $T$ for which a collection $\mathfrak{A} \subseteq \mathfrak{P}(\mathrm{T})$ is relatively compact if and only if $\mathfrak{A}$ is tight are called *Prokhorov spaces*. Thus, Prokhorov [82] proved that all Polish spaces are Prokhorov spaces. Anachronistically, Alexandroff [10, Theorem V.4] had earlier shown that all locally compact Hausdorff spaces are also Prokhorov spaces. What is the topology on $\mathfrak{P}(\mathrm{T})$ that is under consideration in the above results? As is clear from Remark 6.2.5, there is not a lot of choice in the results described so far, as the $A$-topology and the weak topology on $\mathfrak{P}(\mathrm{T})$ are the same when $T$ is a Polish space or a locally compact Hausdorff space.

With respect to the $A$-topology, tightness of a set $\mathfrak{A} \subseteq \mathfrak{P}(\mathrm{T})$ is known to *not* be a necessary condition for the relative compactness of $\mathfrak{A}$. Nice counterexamples were independently constructed by Varadarajan [97], Fernique [39], and Preiss [81]. See Topsøe [94, p. 191] for a description of these counterexamples, and also for further history of Prokhorov's

theorem. The situation is slightly better when we restrict to the space of Radon probability measures (and look for relative compactness in that space). For example, Topsøe (see the comments following Theorem 3.1 in [94]) proves Prokhorov's theorem for the space $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ of all Radon probability measures on a regular topological space $T$. (Thus for a regular space $T$, the set of probability measures $\mathfrak{A}$ is relatively compact in $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ equipped with the $A$-topology if and only if it is tight.)

With the knowledge that tightness is not a necessary condition for relative compactness in $\mathfrak{P}(\mathrm{T})$ in general, our focus here is on a result in the other direction—to see if tightness is still sufficient for relative compactness without too many additional assumptions. It is in this sense that we are looking for a generalization of Prokhorov's theorem. The sufficiency of tightness seems to be known, in many cases, for the relative compactness on spaces of Radon measures equipped with either the weak topology or the $A$-topology. For example, Bogachev [19, Theorem 8.6.7, p. 206, vol. 2] shows that tightness is sufficient for relative compactness in the space of Radon probability measures, equipped with the weak topology, on any completely regular Hausdorff space. Under the $A$-topology, Topsøe [93, Theorem 9.1(iii), p. 43] (see also [92]) has proved that tightness is sufficient for relative compactness in the space of Radon probability measures over any Hausdorff space.

**Remark 6.5.1.** The above discussion seems to allude to the fact that relative compactness under the weak topology is a more restrictive notion than under the $A$-topology. This is technically correct, even though compactness in the weak topology is less restrictive than in the $A$-topology. Indeed, by Remark 6.2.5, it is clear that the weak topology on

$\mathfrak{P}(\mathrm{T})$ (and hence on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$) is coarser than the $A$-topology. Hence any set that is compact in $\mathfrak{P}(\mathrm{T})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$) with the $A$-topology is also compact in $\mathfrak{P}(\mathrm{T})$ (respectively to $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$) with the weak topology. On the other hand, the closure of a set with respect to the $A$-topology on $\mathfrak{P}(\mathrm{T})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$) is contained in the closure of that set with respect to the weak topology on $\mathfrak{P}(\mathrm{T})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$). This last fact, which can be seen by Theorem 6.2.13 and Remark 6.2.5, shows that a set that is relatively compact under the $A$-topology might fail to be so under the weak topology.

Our next result (Theorem 6.5.2) proves the sufficiency of tightness for relative compactness in the $A$-topology on the space of all probability measures on a Hausdorff space $T$. It is a slight variation of the same result that is known for the space of all Radon probability measures, and its proof can be readily adapted to show the latter result as well (see Theorem 6.5.4). The proof of Theorem 6.5.2 is short as most of the work has already been done in setting up the convenient framework of topological and nonstandard measure theory in the previous sections. To the best of the author's knowledge, this generalization of Prokhorov's theorem is new.

**Theorem 6.5.2** (Prokhorov's theorem for the space of probability measures on any Hausdorff space)**.** *Let $T$ be a Hausdorff space, and let $\mathfrak{P}(\mathrm{T})$ be the space of all Borel probability measures on $T$, equipped with the $A$-topology. Let $\mathfrak{A} \subseteq \mathfrak{P}(\mathrm{T})$ be such that for any $\epsilon \in \mathbb{R}_{>0}$, there exists a compact set $K_\epsilon \subseteq T$ for which*

$$\mu(K_\epsilon) \geq 1 - \epsilon \text{ for all } \mu \in \mathfrak{A}. \tag{6.41}$$

*Then the closure of $\mathfrak{A}$ in $\mathfrak{P}(\mathrm{T})$ is compact.*

*Proof.* Let $\mathfrak{A}$ be as in the statement of the theorem. Let $\bar{\mathfrak{A}}$ be its closure in $\mathfrak{P}(\mathrm{T})$ with

respect to the $A$-topology. By the nonstandard characterization of compactness (see Theorem 1.3.10(iii)), it suffices to show that ${}^*\bar{\mathfrak{A}} \subseteq \mathbf{st}^{-1}(\bar{\mathfrak{A}})$. Since $\bar{\mathfrak{A}}$ is closed, any nearstandard element in ${}^*\bar{\mathfrak{A}}$ must be nearstandard to an element of $\bar{\mathfrak{A}}$ (this follows from the nonstandard characterization of closed sets; see Theorem 1.3.10(ii)). Thus, it suffices to show that all elements in ${}^*\bar{\mathfrak{A}}$ are nearstandard. Toward that end, let $\nu \in {}^*\bar{\mathfrak{A}}$. For each $\epsilon \in \mathbb{R}_{>0}$, let $K_\epsilon$ be as in the statement of the theorem. We now prove the following claim.

**Claim 6.5.3.** *$L\nu({}^*K_\epsilon) \geq 1 - \epsilon$ for all $\epsilon \in \mathbb{R}_{>0}$.*

*Proof of Claim 6.5.3.* Suppose, if possible, that there is some $\epsilon \in \mathbb{R}_{>0}$ such that $L\nu({}^*K_\epsilon) < 1 - \epsilon$. Since $\epsilon \in \mathbb{R}_{>0}$, this implies that $\nu({}^*K_\epsilon) < 1 - \epsilon$ as well. By transfer, we conclude that $\nu$ belongs to ${}^*\mathfrak{U}$, where $\mathfrak{U}$ is the following subbasic open subset of $\mathfrak{P}(\mathrm{T})$.

$$\mathfrak{U} := \{\gamma \in \mathfrak{P}(\mathrm{T}) : \gamma(K_\epsilon) < 1 - \epsilon\}. \tag{6.42}$$

Note that $\mathfrak{U}$ is indeed a subbasic open subset of $\mathfrak{P}(\mathrm{T})$, since $K_\epsilon$, being a compact subset of the Hausdorff space $T$, is closed in $T$. By the definition of closure, we know that any open neighborhood of an element in the closure of $\mathfrak{A}$ must have a nonempty intersection with $\mathfrak{A}$. By transfer, we thus find an element $\mu \in \mathfrak{U} \cap \mathfrak{A}$. But this is a contradiction (in view of (6.41) and (6.42)), thus completing the proof of the claim.

Claim 6.5.3 now completes the proof using Theorems 6.1.2 and 6.2.15 (in view of the fact that ${}^*K \subseteq \mathbf{st}^{-1}(K)$ for all compact $K \subseteq T$). $\square$

Using Lemma 6.3.1, the proof of Theorem 6.5.2 carries over immediately to give Prokhorov's theorem for the space of Radon probability measures.

**Theorem 6.5.4** (Prokhorov's theorem for the space of Radon probability measures on any Hausdorff space)**.** *Let $T$ be a Hausdorff space and let $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ be the space of all Radon probability measures on $T$, equipped with the $A$-topology. Let $\mathfrak{A} \subseteq \mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ be such that for any $\epsilon \in \mathbb{R}_{>0}$, there exists a compact set $K_\epsilon \subseteq T$ for which*

$$\mu(K_\epsilon) \geq 1 - \epsilon \text{ for all } \mu \in \mathfrak{A}. \tag{6.43}$$

*Then the closure of $\mathfrak{A}$ in $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$ is compact.*

*Proof.* As in the proof of Theorem 6.5.2, it suffices to show that all elements in ${}^*\bar{\mathfrak{A}}$ are nearstandard (in the current setting, $\bar{\mathfrak{A}}$ is the closure of $\mathfrak{A}$ in the space $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$, and the nearstandardness in question is with respect to the $A$-topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{T})$).

Toward that end, let $\nu \in {}^*\bar{\mathfrak{A}}$. Then we see that $\nu$ is nearstandard by Theorem 6.1.2 and Theorem 6.3.5, in view of the following analog of Claim 6.5.3 (which has the same proof as that of Claim 6.5.3, with the subbasic open set $\{\gamma \in \mathfrak{P}_{\mathrm{r}}(\mathrm{T}) : \gamma(K_\epsilon) < 1 - \epsilon\}$ used as the analog of (6.42) from the earlier proof):

$$L\nu({}^*K_\epsilon) \geq 1 - \epsilon \text{ for all } \epsilon \in \mathbb{R}_{>0}.$$

$\square$

# Chapter 7. Proving Our Generalization of De Finetti–Hewitt–Savage Theorem

## 7.1. Introduction

The goal of this chapter is to establish the generalization of de Finetti–Hewitt–Savage theorem that was promised in Chapter 5. Recall that the original formulation of this theorem states that a sequence of exchangeable random variables taking values in $\{0, 1\}$ is uniquely representable as a mixture of independent and identically distributed (iid) random variables. We show that the same conclusion holds for any sequence of Radon distributed exchangeable random variables taking values in any Hausdorff space equipped with its Borel sigma algebra (see Theorem 7.3.7). This includes and extends the previously known generalizations of de Finetti's theorem following the works of Hewitt and Savage [51] (who proved de Finetti's theorem in the case when the state space is a compact Hausdorff space equipped with its Baire sigma algebra). An analysis of our proof reveals that a slightly weaker condition than Radonness of the underlying common distribution is sufficient—we only need the common distribution of the random variables to be tight and outer regular on compact sets (see the discussion following Theorem 7.3.7).

Dubins and Freedman [35] had constructed a counterexample that showed that de Finetti's theorem does not hold for a particular exchangeable sequence of Borel measurable random variables taking values in some separable metric space. Thus, one consequence of the current work is to show that the random variables in their counterexample did not have a tight distribution (as any tight probability measure on a metric space is also Radon). In general, there is a large class of Hausdorff spaces such that de Finetti's

theorem holds for any sequence of tightly distributed exchangeable random variables taking values in any such Hausdorff space equipped with its Borel sigma algebra (see the discussion following Theorem 7.3.7). Another consequence is that de Finetti's theorem holds whenever the state space is a Radon space equipped with its Borel sigma algebra (see Corollary 7.3.10).

In Section 7.2, we only assume that the sequence $(X_n)_{n\in\mathbb{N}}$ is identically distributed and derive several useful foundational results as applications of the theory built in Section 2.2. In particular, we study the structure of the hyperfinite empirical distributions derived from (the nonstandard extension of) an identically distributed sequence of random variables. We also study the properties of the measures that these hyperfinite empirical distributions induce on the space of all Radon probability measures on the state space.

In Section 7.3, we exploit the added structure provided by exchangeability that allows us to use the results from Section 7.2 to prove our generalizations of de Finetti's theorem. Section 7.4 briefly mentions some other possible versions and generalizations of de Finetti's theorem that we did not consider in this dissertation, along with a discussion on potential future work.

## 7.2. Hyperfinite empirical measures induced by identically Radon distributed random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $S$ be a Hausdorff space equipped with its Borel sigma algebra $\mathcal{B}(S)$. Suppose $X_1, X_2, \ldots$ is a sequence of identically distributed $S$-valued random variables on $\Omega$—that is, the pushforward measure $\mathbb{P} \circ {X_i}^{-1}$ on $(S, \mathcal{B}(S))$ is the same for all $i \in \mathbb{N}$. Note that de Finetti–Hewitt–Savage theorem requires the stronger

condition of exchangeability, which we will assume in the next section when we prove our generalization of that theorem. However, the results in this section are more abstract and preparatory in nature, and they are applicable to all identically distributed sequences of random variables.

Throughout this section, we will further assume that the common distribution of the $X_i$ is Radon. This is for ease of presentation as we will, however, not use the full strength of this hypothesis—we will only have occasion to use the fact that this distribution is tight and outer regular on compact subsets of $S$. By tightness, there exists an increasing sequence of compact subsets $(K_n)_{n\in\mathbb{N}}$ of $S$ such that:

$$\mathbb{P}(X_1 \in K_n) > 1 - \frac{1}{n} \text{ for all } n \in \mathbb{N}. \tag{7.1}$$

The results up to Lemma 7.2.14 only require tightness of the underlying distribution. We will also need outer regularity on compact subsets from Lemma 7.2.15 onwards.

For each $\omega \in \Omega$ and $n \in \mathbb{N}$, define the *empirical measure* $\mu_{\omega,n}$ on $\mathcal{B}(S)$ as follows:

$$\mu_{\omega,n}(B) := \frac{\#\{i \in [n] : X_i(\omega) \in B\}}{n} \text{ for all } B \in \mathcal{B}(S). \tag{7.2}$$

Nonstandardly, we also have for each $\omega \in {}^*\Omega$ and each $N \in {}^*\mathbb{N}$, the hyperfinite empirical measure $\mu_{\omega,N}$ defined by the following:

$$\mu_{\omega,N}(B) := \frac{\#\{i \in [N] : X_i(\omega) \in B\}}{N} \text{ for all } B \in {}^*\mathcal{B}(S). \tag{7.3}$$

Although we are calling $\mu_{\omega,N}$ a *hyperfinite* empirical measure because $N \in {}^*\mathbb{N}$, we do not need to assume $N > \mathbb{N}$ (that is, $N \in {}^*\mathbb{N}\backslash\mathbb{N}$) in this section. Also, like in Chapter 4, we are abusing notation by using $(X_i)$ to denote both the standard sequence $(X_i)_{i\in\mathbb{N}}$ of random variables and the nonstandard extension of this sequence. More precisely, if $\mathfrak{X}\colon \Omega \times \mathbb{N} \to S$ is defined by $\mathfrak{X}(\omega, i) := X_i(\omega)$ for all $\omega \in \Omega$ and $i \in \mathbb{N}$, then for any $i \in {}^*\mathbb{N}$, the internal random variable $X_i\colon {}^*\Omega \to {}^*S$ is defined as follows:

$$X_i(\omega) = {}^*\mathfrak{X}(\omega, i) \text{ for all } \omega \in {}^*\Omega \text{ and } i \in {}^*\mathbb{N}.$$

The notation fixed above will be valid for the rest of this section which studies the structure of these empirical measures within the space of all Radon probability measures on $S$. We divide the exposition into four subsections. Section 7.2.1 deals with some basic properties that are satisfied by almost all hyperfinite empirical measures. Section 7.2.2 deals with the study of the pushforward measure induced on the space ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ of internal Radon measures on ${}^*S$ by the map $\omega \mapsto \mu_{\omega,N}$. The goal of Section 7.2.3 is to show in a precise sense that the standard part of a hyperfinite empirical measure evaluated at a Borel set is almost surely given by the standard part of the measure of the nonstandard extension of that Borel set (see Theorem 7.2.19). Section 7.2.4 synthesizes the theory built so far in order to express some Loeb integrals on the space of all internal Radon probability measures in terms of the corresponding integrals on the standard space of Radon probability measures on $S$.

### 7.2.1. Hyperfinite empirical measures as random elements in the space of all internal Radon measures

Being supported on a finite set, it is clear that $\mu_{\omega,n}$ is, in fact, a Radon probability measure on $S$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$. Furthermore, for each $n \in \mathbb{N}$, the map $\omega \mapsto \mu_{\omega,n}$ is a measurable function from $(\Omega, \mathcal{F})$ to $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$. We record this as a lemma.

**Lemma 7.2.1.** *For each $n \in \mathbb{N}$, the map $\mu_{\cdot,n} \colon \Omega \to \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ defined by (7.2) is Borel measurable. Furthermore, for any $B \in \mathcal{B}(S)$, the map $\mu_{\cdot,n}(B) \colon \Omega \to [0,1]$ (that is, $\omega \mapsto \mu_{\omega,n}(B)$) is Borel measurable for each $n \in \mathbb{N}$.*

*Proof.* The proof is immediate from the measurability of the $X_i$, in view of the observation that for each $n \in \mathbb{N}, \omega \in \Omega$, and $B \in \mathcal{B}(S)$, we have:

$$\mu_{\omega,n}(B) = \frac{1}{n}\left(\sum_{i \in [n]} \mathbb{1}_B(X_i(\omega))\right). \tag{7.4}$$

$\square$

By transfer, we obtain the following immediate consequence.

**Corollary 7.2.2.** *For each $N \in {}^*\mathbb{N}$, the map $\mu_{\cdot,N} \colon {}^*\Omega \to {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ is an internally Borel measurable function from ${}^*\Omega$ to ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. That is, $\mu_{\cdot,N} \colon {}^*\Omega \to {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ is internal and the set $\{\omega \in {}^*\Omega : \mu_{\omega,N} \in \mathfrak{B}\}$ belongs to ${}^*\mathcal{F}$ whenever $\mathfrak{B} \in {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$. Furthermore, for each $B \in {}^*\mathcal{B}(S)$, the map $\mu_{\cdot,N}(B) \colon {}^*\Omega \to {}^*[0,1]$ is internally Borel measurable.*

By the usual Loeb measure construction, we have a collection of complete probability spaces indexed by ${}^*\Omega$, namely $({}^*S, L_{\omega,N}({}^*\mathcal{B}(S)), L\mu_{\omega,N})_{\omega \in {}^*\Omega}$.

We now prove that with respect to the Loeb measure $L^*\mathbb{P}$, almost all $L\mu_{\omega,N}$ assign full mass to the set $\mathbf{Ns}({}^*S)$ of nearstandard elements of ${}^*S$. This implicitly requires

us to first show that for all $\omega$ in an $L^*\mathbb{P}$ almost sure subset of ${}^*\Omega$, the set $\mathbf{Ns}({}^*S)$ is in the Loeb sigma algebra $L_{\omega,N}({}^*\mathcal{B}(S))$ corresponding to the internal probability space $({}^*S, {}^*\mathcal{B}(S), \mu_{\omega,N})$.

**Lemma 7.2.3.** *Let $S$ be a Hausdorff space and $N \in {}^*\mathbb{N}$. There is a set $E_N \in L({}^*\mathcal{F})$ with $L^*\mathbb{P}(E_N) = 1$ such that for any $\omega \in E_N$, we have $L\mu_{\omega,N}(\mathbf{Ns}({}^*S)) = 1$.*

*Proof.* Let $(K_n)_{n\in\mathbb{N}}$ be as in (7.1). By the transfer of the second part of Lemma 7.2.1, the function $\omega \mapsto \mu_{\omega,N}({}^*K_n)$ is an internal random variable for each $n \in \mathbb{N}$. Since it is finitely bounded, it is $S$-integrable with respect to the Loeb measure $L^*\mathbb{P}$. Thus, for each $n \in \mathbb{N}$, the $[0,1]$-valued function $L\mu_{\cdot,N}({}^*K_n)$ defined by $\omega \mapsto L\mu_{\omega,N}({}^*K_n)$, is Loeb measurable, and furthermore we have:

$$\begin{aligned}
\mathbb{E}_{L^*\mathbb{P}}(L\mu_{\cdot,N}({}^*K_n)) &\approx {}^*\mathbb{E}_{^*\mathbb{P}}(\mu_{\cdot,N}({}^*K_n)) \\
&= {}^*\mathbb{E}_{^*\mathbb{P}}\left[\sum_{i=1}^{N} \frac{1}{N} \mathbb{1}_{^*K_n}(X_i)\right] \\
&= \frac{1}{N}\left[\sum_{i=1}^{N} {}^*\mathbb{P}(X_i \in {}^*K_n)\right] \\
&> \frac{1}{N}\left[N\left(1 - \frac{1}{n}\right)\right] \\
&= 1 - \frac{1}{n},
\end{aligned}$$

where the last line follows from (7.1) and the fact that each $X_i$ has the same distribution.

For each $\omega \in {}^*\Omega$, the upper monotonicity of the measure $L_{\omega,N}$ implies that $\lim\limits_{n\to\infty} L\mu_{\omega,N}({}^*K_n) = L\mu_{\omega,N}\left(\cup_{n\in\mathbb{N}}{}^*K_n\right)$. Thus, being a limit of Loeb measurable functions, $\lim\limits_{n\to\infty} L\mu_{\cdot,N}({}^*K_n) = L\mu_{\cdot,N}\left(\cup_{n\in\mathbb{N}}{}^*K_n\right)$, is also Loeb measurable. Therefore, by the monotone convergence theorem, we obtain:

$$
\begin{aligned}
\mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left(\cup_{n\in\mathbb{N}}{}^*K_n\right)\right] &= \mathbb{E}_{L^*\mathbb{P}}\left[\lim_{n\to\infty} L\mu_{\cdot,N}({}^*K_n)\right] \\
&= \lim_{n\to\infty} \mathbb{E}_{L^*\mathbb{P}}(L\mu_{\cdot,N}({}^*K_n)) \\
&\geq \lim_{n\to\infty}\left(1-\frac{1}{n}\right) = 1.
\end{aligned} \tag{7.5}
$$

But $L\mu_{\omega,N}\left[\cup_{n\in\mathbb{N}}{}^*K_n\right] \leq 1$ for all $\omega \in {}^*\Omega$. Therefore, by (7.5), we get:

$$
L^*\mathbb{P}(E_N) = 1, \tag{7.6}
$$

where

$$
E_N = \{\omega : L\mu_{\omega,N}\left[\cup_{n\in\mathbb{N}}{}^*K_n\right] = 1\} \in L({}^*\mathcal{F}). \tag{7.7}
$$

Since each $K_n$ is compact, we have ${}^*K_n \subseteq \mathbf{Ns}({}^*S)$ for all $n \in \mathbb{N}$. Thus for each $\omega \in E_N$, we have the following inequality for the inner measure with respect to $\mu_{\omega,N}$ (see (1.10)):

$$
\underline{\mu_{\omega,N}}\left[\mathbf{Ns}({}^*S)\right] \geq L\mu_{\omega,N}({}^*K_n) \text{ for all } n \in \mathbb{N},
$$

By taking the limit as $n \to \infty$ on the right side and using the definition (7.7) of $E_N$, we obtain:

$$
\underline{\mu_{\omega,N}}\left[\mathbf{Ns}({}^*S)\right] \geq \lim_{n\to\infty} L\mu_{\omega,N}({}^*K_n) = L\mu_{\omega,N}\left[\cup_{n\in\mathbb{N}}{}^*K_n\right] = 1 \text{ for all } \omega \in E_N.
$$

Since

$$
1 = \underline{\mu_{\omega,N}}\left[\mathbf{Ns}({}^*S)\right] \leq \overline{\mu_{\omega,N}}\left[\mathbf{Ns}({}^*S)\right] \leq 1,
$$

it follows that $\mathbf{Ns}(^*S)$ is Loeb measurable, and that $L\mu_{\omega,N}\left[\mathbf{Ns}(^*S)\right] = 1$ for all $\omega \in E_N$.

□

The idea, used in the above proof, of showing that the expected value of a probability is one in order to conclude that the concerned probability is equal to one almost surely, can be turned around and used to show that a certain probability is zero almost surely, by showing that the expected value of that probability is zero. We use this idea to prove next that almost surely, $L_{\omega,N}$ treats the nonstandard extension of a countable disjoint union as if it were the disjoint union of the nonstandard extensions, the leftover portion being assigned zero mass.

**Lemma 7.2.4.** *Let $S$ be a Hausdorff space and $N \in {}^*\mathbb{N}$. Let $(B_n)_{n\in\mathbb{N}}$ be a sequence of disjoint Borel sets. There is a set $E_{(B_n)_{n\in\mathbb{N}}} \in L(^*\mathcal{F})$ with $L^*\mathbb{P}(E_{(B_n)_{n\in\mathbb{N}}}) = 1$ such that*

$$L\mu_{\omega,N}\left[^*\left(\sqcup_{n\in\mathbb{N}}B_n\right)\right] = \sum_{n\in\mathbb{N}} L\mu_{\omega,N}\left(^*B_n\right) \text{ for all } \omega \in E_{(B_n)_{n\in\mathbb{N}}}, \tag{7.8}$$

*where $\sqcup$ denotes a disjoint union.*

**Remark 7.2.5.** Note that the above lemma does not follow from the disjoint additivity of the measure $L\mu_{\omega,N}$, because $\sqcup_{n\in\mathbb{N}}{}^*B_n \subseteq {}^*\left(\sqcup_{n\in\mathbb{N}}B_n\right)$ with equality if and only if the $B_n$ are empty for all but finitely many $n$. Also, the almost sure set $E_{(B_n)_{n\in\mathbb{N}}}$ depends on the sequence $(B_n)_{n\in\mathbb{N}}$. Since there are potentially uncountably many such sequences, therefore we cannot expect to find a single $L^*\mathbb{P}$-almost sure set on which equation (7.8) is always valid for all disjoint sequences $(B_n)_{n\in\mathbb{N}}$ of Borel sets.

*Proof of Lemma 7.2.4.* Let $(B_n)_{n\in\mathbb{N}}$ be a disjoint sequence of Borel sets and let

$$B := \sqcup_{n\in\mathbb{N}}B_n.$$

For each $m \in \mathbb{N}$, let $B_{(m)} := \sqcup_{n \in [m]} B_n$. Consider the map $\omega \mapsto \mu_{\omega,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]$, which is internally Borel measurable by Corollary 7.2.2. Since this map is finitely bounded, it is $S$-integrable with respect to the Loeb measure $L^*\mathbb{P}$. In particular, for each $m \in \mathbb{N}$, the $[0,1]$-valued function $L\mu_{\cdot,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]$, defined by $\omega \mapsto L\mu_{\omega,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]$, is Loeb measurable. Taking expected values and using $S$-integrability, we obtain:

$$\begin{aligned}
\mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]\right] &\approx {}^*\mathbb{E}_{{}^*\mathbb{P}}\left[\mu_{\cdot,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]\right] \\
&= {}^*\mathbb{E}_{{}^*\mathbb{P}}\left[\sum_{i=1}^{N} \frac{1}{N} \mathbb{1}_{{}^*\left(B \backslash B_{(m)}\right)}(X_i)\right] \\
&= \frac{1}{N}\left[\sum_{i=1}^{N} {}^*\mathbb{P}(X_i \in {}^*\left(B \backslash B_{(m)}\right))\right] \\
&= \frac{1}{N}\left[N {}^*\mathbb{P}(X_1 \in {}^*\left(B \backslash B_{(m)}\right))\right] \\
&= {}^*\mathbb{P}(X_1 \in {}^*\left(B \backslash B_{(m)}\right)) \\
&= \mathbb{P}(X_1 \in B \backslash B_{(m)}) \\
&= \mathbb{P}(X_1 \in B) - \mathbb{P}(X_1 \in B_{(m)}). \qquad (7.9)
\end{aligned}$$

Since the expression in (7.9) is a real number, we have the following equality:

$$\mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]\right] = \mathbb{P}(X_1 \in B) - \mathbb{P}(X_1 \in B_{(m)}) \text{ for all } m \in \mathbb{N}. \qquad (7.10)$$

Note that for each $\omega \in {}^*\Omega$, the limit

$$\lim_{m \to \infty} L\mu_{\omega,N}\left[{}^*\left(B \backslash B_{(m)}\right)\right]$$

exists and is equal to $L\mu_{\omega,N}\left[\cap_{m \in \mathbb{N}} {}^*\left(B \backslash B_{(m)}\right)\right]$, because $({}^*(B \backslash B_{(m)}))_{m \in \mathbb{N}}$ is a decreasing sequence of measurable sets. Also, by the upper monotonicity of the measure induced by $X_1$ on $S$, we know that

$$\lim_{m \to \infty} \mathbb{P}(X_1 \in B_{(m)}) = \mathbb{P}\left(X_1 \in \cup_{m \in \mathbb{N}} B_{(m)}\right) = \mathbb{P}(X_1 \in B).$$

Using this in (7.10), followed by an application of the dominated convergence theorem, we thus obtain the following:

$$
\begin{aligned}
0 &= \lim_{m\to\infty} \mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left[{}^*\left(B\backslash B_{(m)}\right)\right]\right] \\
&= \mathbb{E}_{L^*\mathbb{P}}\left[\lim_{m\to\infty} L\mu_{\cdot,N}\left[{}^*\left(B\backslash B_{(m)}\right)\right]\right]. \qquad (7.11)
\end{aligned}
$$

Also, since $\lim_{m\to\infty} L\mu_{\omega,N}\left[{}^*\left(B\backslash B_{(m)}\right)\right] \geq 0$, it follows from (7.11) that there is an $L^*\mathbb{P}$-almost sure set $E_{(B_n)_{n\in\mathbb{N}}}$ such that

$$
\lim_{m\to\infty} L\mu_{\omega,N}\left[{}^*\left(B\backslash B_{(m)}\right)\right] = 0 \text{ for all } \omega \in E_{(B_n)_{n\in\mathbb{N}}}. \qquad (7.12)
$$

But for each $\omega \in E_{(B_n)_{n\in\mathbb{N}}}$, we have the following:

$$
\begin{aligned}
L\mu_{\omega,N}\left[{}^*\left(B\backslash B_{(m)}\right)\right] &= L\mu_{\omega,N}({}^*B) - L\mu_{\omega,N}\left(B_{(m)}\right) \\
&= L\mu_{\omega,N}({}^*B) - L\mu_{\omega,N}\left(\sqcup_{n\in[m]B_n}\right) \\
&= L\mu_{\omega,N}({}^*B) - \sum_{n\in[m]} L\mu_{\omega,N}({}^*B_m) \text{ for all } m \in \mathbb{N}. \qquad (7.13)
\end{aligned}
$$

The proof is completed by letting $m \to \infty$ in (7.13), followed by an application of (7.12).

□

The specific form of the set $E_N$ allows us to use Theorem 6.1.2 to show that for each $N \in {}^*\mathbb{N}$, the measure $L\mu_{\omega,N} \circ \mathbf{st}^{-1}$ is Radon for all $\omega \in E_N$, and that $\mu_{\omega,N}$ is nearstandard in ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ to this measure. This is proved in the next lemma.

**Lemma 7.2.6.** *Let $S$ be a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and $E_N$ be as in (7.7). For all $\omega \in E_N$, we have:*

*(i)* $L\mu_{\omega,N} \circ \mathbf{st}^{-1} \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$.

*(ii)* $\mu_{\omega,N} \in \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, *with* $\mathbf{st}(\mu_{\omega,N}) = L\mu_{\omega,N} \circ \mathbf{st}^{-1}$.

*Proof.* By the definition (7.7), we know that

$$L\mu_{\omega,N}\left(\cup_{n\in\mathbb{N}} {}^*K_n\right) = 1 \text{ for all } \omega \in E_N,$$

where the $K_n$ are compact subsets of $S$.

By the upper monotonicity of the probability measure $L\mu_{\omega,N}$ and the fact that $({}^*K_n)_{n\in\mathbb{N}}$ is an increasing sequence, we obtain:

$$\lim_{n\to\infty} L\mu_{\omega,N}\left({}^*K_n\right) = 1 \text{ for all } \omega \in E_N. \tag{7.14}$$

Therefore, given $\epsilon \in \mathbb{R}_{>0}$, there exists an $n_\epsilon$ such that $L\mu_{\omega,N}\left({}^*K_n\right) > 1-\epsilon$ for all $\omega \in E_N$ and $n \in \mathbb{N}_{>n_\epsilon}$. Thus the tightness condition (6.2) holds for $\mu_{\omega,N}$ whenever $\omega \in E_N$. Theorem 6.1.2 now completes the proof. $\square$

Let $\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$ denote the $A$-topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. For $\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, let $\tau_\mu$ denote the set of all open neighborhoods of $\mu$ in $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. That is,

$$\tau_\mu := \{\mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} : \mu \in \mathfrak{U}\}.$$

Also, for any open set $\mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$, let $\tau_{\mathfrak{U}}$ be the subspace topology on $\mathfrak{U}$. In other words, we define

$$\tau_{\mathfrak{U}} := \{\mathfrak{V} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} : \mathfrak{V} = \mathfrak{W} \cap \mathfrak{U} \text{ for some } \mathfrak{W} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}\} = \{\mathfrak{V} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} : \mathfrak{V} \subseteq \mathfrak{U}\}.$$

For internal sets $A, B$, we use $\mathfrak{F}(A, B)$ to denote the internal set of all internal functions from $A$ to $B$.

**Lemma 7.2.7.** *Let $S$ be Hausdorff and $N \in {}^*\mathbb{N}$. Let $E_N$ be as defined in (7.7). For each internal subset $E \subseteq E_N$, there exists an internal function $U_\cdot \colon E \to {}^*\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$ such that*

$$\mu_{\omega,N} \in U_\omega \text{ and } U_\omega \subseteq \mathbf{st}^{-1}(L\mu_{\omega,N} \circ \mathbf{st}^{-1}) \text{ for all } \omega \in E.$$

*Proof.* Fix an internal set $E \subseteq E_N$. For each open set $\mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$, define the following set of internal functions:

$$\mathcal{G}_{\mathfrak{U}} := \left\{ f \in \mathfrak{F}(E, {}^*\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}) : f(\omega) \in {}^*\tau_{\mathfrak{U}} \text{ and } \mu_{\omega,N} \in f(\omega) \text{ for all } \omega \in E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}) \right\}.$$

Since $E$ is internal and ${\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U})$ is internal by Lemma 7.2.1, therefore the set $\mathcal{G}_{\mathfrak{U}}$ is internal for all $\mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$ by the internal definition principle (see, for example, Loeb [70, Theorem 2.8.4, p. 54]). Also, $\mathcal{G}_{\mathfrak{U}}$ is nonempty for each $\mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$. Indeed, if $E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}) = \emptyset$, then $\mathcal{G}_{\mathfrak{U}} = \mathfrak{F}(E, {}^*\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})})$. Otherwise, if $\omega \in E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U})$, then define $f(\omega) := {}^*\mathfrak{U}$, and define $f$ (internally) arbitrarily on the remainder of $E$. It is clear that this function $f$ is an element of $\mathcal{G}_{\mathfrak{U}}$.

Now let $\mathfrak{U}_1, \mathfrak{U}_2$ be two distinct open subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Define a function $f$ on $E$ as follows:

$$f(\omega) := \begin{cases} {}^*\mathfrak{U}_1 \cap {}^*\mathfrak{U}_2 & \text{if} \quad \omega \in E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_1) \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_2) \\ {}^*\mathfrak{U}_1 & \text{if} \quad \omega \in [E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_1)] \backslash {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_2) \\ {}^*\mathfrak{U}_2 & \text{if} \quad \omega \in [E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_2)] \backslash {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_1) \\ {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}) & \text{if} \quad \omega \in E \backslash \left[ {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_1) \cup {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U}_2) \right]. \end{cases}$$

The above function is clearly in $\mathcal{G}_{\mathfrak{U}_1} \cap \mathcal{G}_{\mathfrak{U}_2}$. In general, to show the finite intersection property of the collection $\{\mathcal{G}_{\mathfrak{U}} : \mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}\}$, the same recipe of “disjointifying” the union of finitely many open sets $\mathfrak{U}_1, \ldots, \mathfrak{U}_k$ works. More precisely, for a subset $\mathfrak{A} \subseteq \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, let $\mathfrak{A}^{(0)}$ denote $\mathfrak{A}$ and $\mathfrak{A}^{(1)}$ denote the complement $\mathfrak{P}_{\mathrm{r}}(\mathrm{S}) \backslash \mathfrak{A}$. If $\mathfrak{U}_1, \ldots, \mathfrak{U}_k$ are finitely many open subsets of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, then for each $\omega \in E$, define $(i_1(\omega), \ldots, i_k(\omega)) \in \{0,1\}^k$ to be the unique tuple such that $\omega \in E \cap \left( \cap_{j \in [k]} {\mu_{\cdot,N}}^{-1}({}^*{\mathfrak{U}_j}^{(i_j(\omega))}) \right)$. Then the

function $f$ on $E$ defined as follows is immediately seen to be a member of $\cap_{j\in[k]}\mathcal{G}_{\mathfrak{U}_j}$:

$$f(\omega) := \bigcap_{\{j\in[k]:i_j(\omega)=1\}} {}^*\mathfrak{U}_j \text{ for all } \omega \in E.$$

Thus the collection $\{\mathcal{G}_{\mathfrak{U}} : \mathfrak{U} \in \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}\}$ has the finite intersection property. Let $U.$ be in the intersection of the $\mathcal{G}_{\mathfrak{U}}$ (which is nonempty by saturation). It is clear from the definition of the sets $\mathcal{G}_{\mathfrak{U}}$ that $\mu_{\omega,N} \in U_\omega$ for all $\omega \in E$. We now show that $U_\omega \subseteq \mathbf{st}^{-1}(L\mu_{\omega,N} \circ \mathbf{st}^{-1})$ for all $\omega \in E$

By Lemma 7.2.6, we know that $\mu_{\omega,N} \in \mathbf{st}^{-1}(L\mu_{\omega,N} \circ \mathbf{st}^{-1})$ for all $\omega \in E$. Thus for each $\omega \in E$, we have $\mu_{\omega,N} \in {}^*\mathfrak{U}$ for all $\mathfrak{U} \in \tau_{L\mu_{\omega,N}\circ\mathbf{st}^{-1}}$. Hence, for each $\omega \in E$, we have $\omega \in E \cap {\mu_{\cdot,N}}^{-1}({}^*\mathfrak{U})$ for all $\mathfrak{U} \in \tau_{L\mu_\omega\circ\mathbf{st}^{-1}}$. Therefore, by the definition of the collections $\mathcal{G}_{\mathfrak{U}}$, we deduce that $U_\omega \in {}^*\tau_{\mathfrak{U}}$ for all $\mathfrak{U} \in \tau_{L\mu_{\omega,N}\circ\mathbf{st}^{-1}}$. As a consequence, $U_\omega \subseteq {}^*\mathfrak{U}$ for all $\mathfrak{U} \in \tau_{L\mu_{\omega,N}\circ\mathbf{st}^{-1}}$ and $\omega \in E$. Hence,

$$U_\omega \subseteq \cap_{\mathfrak{U}\in\tau_{L\mu_{\omega,N}\circ\mathbf{st}^{-1}}} {}^*\mathfrak{U} = \mathbf{st}^{-1}(L\mu_{\omega,N} \circ \mathbf{st}^{-1}) \text{ for all } \omega \in E,$$

as desired. □

For each $N \in {}^*\mathbb{N}$, since $E_N$ is a Loeb measurable set of (inner) measure equaling one, there exists an increasing sequence $(E_{N,n})_{n\in\mathbb{N}}$ of internal subsets of $E_N$ such that the following holds:

$$ {}^*\mathbb{P}(E_{N,n}) > 1 - \frac{1}{n} \text{ for all } n \in \mathbb{N}. \tag{7.15}$$

Lemma 7.2.7 applied to the internal sets $E_{N,n}$ will imply that the pushforward (internal) measure on ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ induced by the random variable $\mu_{\cdot,N}$ is such that its Loeb mea-

sure assigns full measure to $\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$. This is the content of our next result.

More precisely, for each $N \in {}^*\mathbb{N}$, define an internal finitely additive probability $P_N$ on $({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ as follows:

$$P_N(\mathfrak{B}) := {}^*\mathbb{P}\left(\{\omega \in {}^*\Omega : \mu_{\omega,N} \in \mathfrak{B}\}\right) = {}^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{B})\right) \text{ for all } \mathfrak{B} \in {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})). \tag{7.16}$$

That this is indeed an internal probability follows from Corollary 7.2.2. As promised, we now show that the corresponding Loeb measure $LP_N$ is concentrated on nearstandard elements of ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$.

**Theorem 7.2.8.** *Let $S$ be a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $P_N$ be as in (7.16). Let*

$$({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))), LP_N)$$

*be the associated Loeb space. Then the set $\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ is Loeb measurable, with*

$$LP_N(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) = 1.$$

*Proof.* Let $E_N$ be as in (7.7) and let $(E_{N,n})_{n\in\mathbb{N}} \subseteq E_N$ be as in (7.15). Fix $n \in \mathbb{N}$. With $E := E_{N,n}$, apply Lemma 7.2.7 to obtain an internal function $U_\cdot \colon E_{N,n} \to {}^*\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$ such that

$$\mu_{\omega,N} \in U_\omega \text{ and } U_\omega \subseteq \mathbf{st}^{-1}(L\mu_{\omega,N} \circ \mathbf{st}^{-1}) \text{ for all } \omega \in E_{N,n}.$$

In particular, $U_\omega \subseteq \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ for all $\omega \in F_{N,n}$, so that $\cup_{\omega\in F_{N,n}} U_\omega \subseteq \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$. By transfer (of the fact that if $f \colon I \to \tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})}$ is a function, then the set $U := \cup_{i\in I} f(i)$, with the membership relation given by $x \in U$ if and only if there exists $i \in I$ with $x \in f(i)$, is open), we have the following conclusions:

$$U := \cup_{\omega\in E_{N,n}} U_\omega \subseteq \mathbf{Ns}({}^*S) \text{ and } U \in {}^*\tau_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \subseteq {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})).$$

Since $\mu_{\omega,N} \in U_\omega$ for all $\omega \in E_{N,n}$, we have $E_{N,n} \subseteq {\mu_{\cdot,N}}^{-1}(U)$. Hence it follows from (7.16) that

$$\underline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) \geq LP_N(U) = L^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(U)\right) \geq L^*\mathbb{P}(E_{N,n}).$$

Using (7.15) and observing that $n \in \mathbb{N}$ was arbitrary, we thus obtain the following:

$$\underline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) \geq 1 - \frac{1}{n} \text{ for all } n \in \mathbb{N}.$$

This clearly implies that

$$1 = \underline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) \leq \overline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) \leq 1,$$

so that $\underline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) = \overline{P_N}(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) = 1$. As a consequence, $\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ is Loeb measurable with $LP_N(\mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) = 1$, completing the proof. □

The next lemma provides a useful dictionary between Loeb integrals with respect to $LP_N$ and those with respect to $L^*\mathbb{P}$:

**Lemma 7.2.9.** *Let $S$ be a Hausdorff space and $N \in {}^*\mathbb{N}$. Let $P_N$ be as in (7.16). For any bounded $LP_N$-measurable function $f\colon {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}) \to \mathbb{R}$, we have:*

$$\int_{{}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} f(\mu) dLP_N(\mu) = \int_{{}^*\Omega} f(\mu_{\omega,N}) dL^*\mathbb{P}(\omega). \tag{7.17}$$

*Proof.* First fix an internally Borel set $\mathfrak{B} \in {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ and let $f = \mathbb{1}_{\mathfrak{B}}$. Then the left side of (7.17) is equal to $LP_N(\mathfrak{B}) = \mathbf{st}(P_N(\mathfrak{B}))$, which also equals the following by (7.16):

$$\mathbf{st}\left[{}^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{B})\right)\right] = L^*\mathbb{P}\left[\{\omega \in {}^*\Omega : \mu_{\omega,N} \in \mathfrak{B}\}\right] = \int_{{}^*\Omega} \mathbb{1}_{\mathfrak{B}}(\mu_{\omega,N}) dL^*\mathbb{P}(\omega).$$

Thus (7.17) is true when $f$ is the indicator function of an internally Borel subset of ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. That is:

$$LP_N(\mathfrak{B}) = L^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{B})\right) \text{ for all } \mathfrak{B} \in {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})). \tag{7.18}$$

Now, let $\mathfrak{A}$ be a Loeb measurable set—that is, $\mathfrak{A} \in L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ and $f = \mathbb{1}_{\mathfrak{A}}$. By the fact that the Loeb measure of a Loeb measurable set equals its inner and outer measure with respect to the internal algebra ${}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, we obtain sets $\mathfrak{A}_\epsilon, \mathfrak{A}^\epsilon \in {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ for each $\epsilon \in \mathbb{R}_{>0}$, such that $\mathfrak{A}_\epsilon \subseteq \mathfrak{A} \subseteq \mathfrak{A}^\epsilon$ and such that the following holds:

$$LP_N(\mathfrak{A}) - \epsilon < LP_N(\mathfrak{A}_\epsilon) \leq LP_N(\mathfrak{A}) \leq LP_N(\mathfrak{A}^\epsilon) < LP_N(\mathfrak{A}) + \epsilon. \tag{7.19}$$

Using (7.18) in (7.19) yields the following for each $\epsilon \in \mathbb{R}_{>0}$:

$$LP_N(\mathfrak{A}) - \epsilon < L^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A}_\epsilon)\right) \leq LP_N(\mathfrak{A}) \leq L^*\mathbb{P}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A}^\epsilon)\right) < LP_N(\mathfrak{A}) + \epsilon. \tag{7.20}$$

Since ${\mu_{\cdot,N}}^{-1}(\mathfrak{A}_\epsilon), {\mu_{\cdot,N}}^{-1}(\mathfrak{A}^\epsilon)$ are members of ${}^*\mathcal{F}$ by Lemma 7.2.1, it follows from (7.20) that for any $\epsilon \in \mathbb{R}_{>0}$ we have:

$$\begin{aligned} LP_N(\mathfrak{A}) - \epsilon &\leq \sup\{L^*\mathbb{P}(E) : E \in {}^*\mathcal{F} \text{ and } E \subseteq {\mu_{\cdot,N}}^{-1}(\mathfrak{A}_\epsilon)\} \\ &\leq \sup\{L^*\mathbb{P}(E) : E \in {}^*\mathcal{F} \text{ and } E \subseteq {\mu_{\cdot,N}}^{-1}(\mathfrak{A})\} \\ &= \underline{{}^*\mathbb{P}}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right), \end{aligned}$$

and

$$\begin{aligned} LP_N(\mathfrak{A}) + \epsilon &\geq \inf\{L^*\mathbb{P}(E) : E \in {}^*\mathcal{F} \text{ and } {\mu_{\cdot,N}}^{-1}(\mathfrak{A}^\epsilon) \subseteq E\} \\ &\geq \inf\{L^*\mathbb{P}(E) : E \in {}^*\mathcal{F} \text{ and } {\mu_{\cdot,N}}^{-1}(\mathfrak{A}) \subseteq E\} \\ &= \overline{{}^*\mathbb{P}}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right). \end{aligned}$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary, it thus follows that $\underline{{}^*\mathbb{P}}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right) = \overline{{}^*\mathbb{P}}\left({\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right)$, both being equal to $LP_N(\mathfrak{A})$. This shows that ${\mu_{\cdot,N}}^{-1}(\mathfrak{A})$ is Loeb measurable and that the following holds:

$$LP_N(\mathfrak{A}) = L^*\mathbb{P}\left[{\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right] \text{ for all } \mathfrak{A} \in L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))). \tag{7.21}$$

This proves (7.17) for indicator functions of Loeb measurable sets. Since the functions $f$ satisfying (7.17) are clearly closed under taking $\mathbb{R}$-linear combinations, the result is true for simple functions (that is, those Loeb measurable functions that take finitely many values). The result for general bounded Loeb measurable functions follows from this (and the dominated convergence theorem) since any bounded measurable function can be uniformly approximated by a sequence of simple functions. $\square$

The result in (7.21) is interesting and useful in its own right. We record this observation as a corollary of the above proof.

**Corollary 7.2.10.** *Let $S$ be a Hausdorff space and let $N \in {}^*\mathbb{N}$. Let $P_N$ be as in (7.16). For any $\mathfrak{A} \in L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$, the set ${\mu_{\cdot,N}}^{-1}(\mathfrak{A})$ is $L^*\mathbb{P}$-measurable. Furthermore, we have:*

$$LP_N(\mathfrak{A}) = L^*\mathbb{P}\left[{\mu_{\cdot,N}}^{-1}(\mathfrak{A})\right] \text{ for all } \mathfrak{A} \in L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))).$$

#### 7.2.2. An internal measure induced on the space of all internal Radon probability measures

Armed with a way to compute the $LP_N$ measure of a large collection of sets, we are in a position to use Prokhorov's theorem (Theorem 6.5.4) to verify that $P_N$ satisfies the tightness condition (6.41) from Theorem 6.1.2.

**Theorem 7.2.11.** *Let $S$ be a Hausdorff space and let $N \in {}^*\mathbb{N}$. Let $P_N$ be as in (7.16). Given $\epsilon \in \mathbb{R}_{>0}$, there exists a compact set $\mathfrak{K}_{(\epsilon)} \subseteq \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ such that*

$$LP_N({}^*\mathfrak{U}) \geq 1 - \epsilon \text{ for all open sets } \mathfrak{U} \text{ such that } \mathfrak{K}_{(\epsilon)} \subseteq \mathfrak{U}.$$

*Proof.* Let $(K_n)_{n\in\mathbb{N}}$ be the increasing sequence of compact subsets of $S$ fixed in (7.1). Re-

call the $L^*\mathbb{P}$ almost sure set $E_N$ from (7.7):

$$\begin{aligned}
E_N &= \{\omega \in {}^*\Omega : L\mu_{\omega,N}\left[\cup_{n\in\mathbb{N}} {}^*K_n\right] = 1\} \\
&= \left\{\omega \in {}^*\Omega : \lim_{n\to\infty} L\mu_{\omega,N}\left({}^*K_n\right) = 1\right\} \\
&= \bigcap_{\ell\in\mathbb{N}} \left( \bigcup_{m\in\mathbb{N}} \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right).
\end{aligned}$$

Note that $\left( \bigcup_{m\in\mathbb{N}} \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right)_{\ell\in\mathbb{N}}$ is a decreasing sequence of Loeb measurable sets. Hence the fact that $L^*\mathbb{P}(E_N) = 1$ implies the following:

$$1 = \lim_{\ell\to\infty} L^*\mathbb{P}\left( \bigcup_{m\in\mathbb{N}} \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right). \tag{7.22}$$

Let $\epsilon \in \mathbb{R}_{>0}$ be given. By (7.22), there exists an $\ell_\epsilon \in \mathbb{N}$ such that we have

$$L^*\mathbb{P}\left( \bigcup_{m\in\mathbb{N}} \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right) > 1 - \frac{\epsilon}{4} \text{ for all } \ell \in \mathbb{N}_{\geq \ell_\epsilon}. \tag{7.23}$$

Now $\left( \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right)_{m\in\mathbb{N}}$ is an increasing sequence of Loeb measurable sets. By (7.23), we thus find an $m_\epsilon \in \mathbb{N}$ for which the following holds:

$$L^*\mathbb{P}\left( \bigcap_{n\in\mathbb{N}_{\geq m}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{\ell}\right\}\right) > 1 - \frac{\epsilon}{2}$$

$$\text{for all } \ell \in \mathbb{N}_{\geq \ell_\epsilon} \text{ and } m \in \mathbb{N}_{\geq m_\epsilon}. \tag{7.24}$$

Let $n_\epsilon = \max\{\ell_\epsilon, m_\epsilon\} \in \mathbb{N}$. By (7.24), the following internal set contains $\mathbb{N}_{\geq n_\epsilon}$:

$$\mathcal{G}_\epsilon := \left\{ n_0 \in {}^*\mathbb{N}_{\geq n_\epsilon} : {}^*\mathbb{P}\left[ \bigcap_{\substack{n\in{}^*\mathbb{N} \\ n_\epsilon \leq n \leq n_0}} \left\{\omega \in {}^*\Omega : \mu_{\omega,N}\left({}^*K_n\right) \geq 1 - \frac{1}{n_0}\right\}\right] > 1 - \epsilon \right\}. \tag{7.25}$$

By overflow, we obtain an $N_\epsilon > \mathbb{N}$ in $\mathcal{G}_\epsilon$. As a consequence, we conclude that for

any $n_0 \in \mathbb{N}_{\geq n_\epsilon}$ we have the following:

$$
\begin{aligned}
&L^*\mathbb{P}\left[\bigcap_{\substack{n\in{}^*\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}\left\{\omega\in{}^*\Omega:\mu_{\omega,N}\left({}^*K_n\right)\geq 1-\frac{1}{n}\right\}\right]\\
\geq &L^*\mathbb{P}\left[\bigcap_{\substack{n\in{}^*\mathbb{N}\\ n_\epsilon\leq n\leq N_\epsilon}}\left\{\omega\in{}^*\Omega:\mu_{\omega,N}\left({}^*K_n\right)\geq 1-\frac{1}{N_0}\right\}\right]\\
\geq &1-\epsilon.
\end{aligned}
\tag{7.26}
$$

. For each $n \in \mathbb{N}$, consider the set $\mathfrak{F}_n$ defined as follows:

$$
\mathfrak{F}_n := \left\{\gamma \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \gamma(K_n) \geq 1 - \frac{1}{n}\right\}.
$$

Since compact subsets of a Hausdorff space are closed, the set $\mathfrak{F}_n$ is the complement of a subbasic open subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, and is hence closed for each $n \in \mathbb{N}$. Since the nonstandard extension of a finite intersection is the intersection of the nonstandard extensions, Corollary 7.2.10 implies that for each $n_0 \in \mathbb{N}_{\geq n_\epsilon}$, we have:

$$
\begin{aligned}
LP_N\left(\bigcap_{\substack{n\in\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}{}^*\mathfrak{F}_n\right) &= LP_N\left({}^*\bigcap_{\substack{n\in\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}\mathfrak{F}_n\right)\\
&= L^*\mathbb{P}\left(\left\{\omega\in{}^*\Omega:\mu_{\omega,N}\in{}^*\bigcap_{\substack{n\in\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}\mathfrak{F}_n\right\}\right)\\
&= L^*\mathbb{P}\left(\left\{\omega\in{}^*\Omega:\mu_{\omega,N}\in\bigcap_{\substack{n\in\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}{}^*\mathfrak{F}_n\right\}\right).
\end{aligned}
\tag{7.27}
$$

Using (7.27) and (7.26), we thus conclude the following:

$$
LP_N\left(\bigcap_{\substack{n\in\mathbb{N}\\ n_\epsilon\leq n\leq n_0}}{}^*\mathfrak{F}_n\right) \geq 1-\epsilon \text{ for all } n_0 \in \mathbb{N}_{\geq n_\epsilon}. \tag{7.28}
$$

Since $LP_N$ is a finite measure and $\left( {}^* \bigcap_{\substack{n \in \mathbb{N} \\ n_\epsilon \le n \le n_0}} \mathfrak{F}_n \right)_{n_0 \in \mathbb{N}_{\ge n_\epsilon}}$ is a decreasing sequence of $LP_N$-measurable sets, we may take the limit as $n_0 \to \infty$ in (7.28) to obtain the following:

$$LP_N \left( \bigcap_{n \in \mathbb{N}_{\ge n_\epsilon}} {}^*\mathfrak{F}_n \right) \ge 1 - \epsilon. \tag{7.29}$$

Define $\mathfrak{K}_{(\epsilon)}$ as follows:

$$\mathfrak{K}_{(\epsilon)} := \bigcap_{n \in \mathbb{N}_{\ge n_\epsilon}} \mathfrak{F}_n. \tag{7.30}$$

Since arbitrary intersections of closed sets are closed, it follows that $\mathfrak{K}_{(\epsilon)}$ is a closed subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. It is also relatively compact by Theorem 6.5.4. Being a closed set that is relatively compact, it follows that $\mathfrak{K}_{(\epsilon)}$ is a compact subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Let $\mathfrak{U}$ be any open subset of $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ containing $\mathfrak{K}_{(\epsilon)}$. We make the following immediate observation using Lemma 1.3.12:

$${}^*\mathfrak{K}_{(\epsilon)} \subseteq \left[ \left( \bigcap_{n \in \mathbb{N}_{\ge n_\epsilon}} {}^*\mathfrak{F}_n \right) \cap \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})) \right] \subseteq {}^*\mathfrak{U}. \tag{7.31}$$

By (7.31) and Theorem 7.2.8, we thus obtain:

$$LP_N({}^*\mathfrak{U}) \ge LP_N \left[ \left( \bigcap_{n \in \mathbb{N}_{\ge n_\epsilon}} {}^*\mathfrak{F}_n \right) \cap \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})) \right] = LP_N \left( \bigcap_{n \in \mathbb{N}_{\ge n_\epsilon}} {}^*\mathfrak{F}_n \right).$$

Using (7.29) now shows that $LP_N({}^*\mathfrak{U}) \ge 1 - \epsilon$, thus completing the proof. $\square$

Theorem 7.2.11, Theorem 6.1.1, and Theorem 6.2.15 now immediately lead to the following result.

**Theorem 7.2.12.** *Suppose that $S$ is a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $P_N$ be as in (7.16). Let*

$$({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))), LP_N)$$

*be the associated Loeb space. Then $LP_N \circ \mathbf{st}^{-1}$ is a Radon measure on the Hausdorff space $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Furthermore, $P_N$ is nearstandard to $LP_N \circ \mathbf{st}^{-1}$ in ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$—that is, we have:*

$$P_N \in \mathbf{st}^{-1}(LP_N \circ \mathbf{st}^{-1}) \subseteq {}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})).$$

It is worthwhile to point out two useful observations arising from the statement of Theorem 7.2.12. Firstly, we were able to say that $P_N$ is nearstandard to $LP_N \circ \mathbf{st}^{-1}$ in ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, but we can still not say that the standard part of $P_N$ is $LP_N \circ \mathbf{st}^{-1}$. This is because ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ is not necessarily Hausdorff and even though $LP_N \circ \mathbf{st}^{-1} \in \mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, we do not know whether $P_N$ belongs to ${}^*\mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ or not (so we are not able to use the standard part map $\mathbf{st}\colon \mathbf{Ns}({}^*\mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))) \to \mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ in this context).

Secondly, since $LP_N \circ \mathbf{st}^{-1}$ is a measure on $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, it is (in particular) the case that $\mathbf{st}^{-1}(\mathfrak{B})$ is $LP_N$-measurable for all $\mathfrak{B} \in \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$. This observation is useful enough that we record it as a corollary.

**Corollary 7.2.13.** *Let $S$ be a Hausdorff space and let $P_N$ be as in (7.16). For each $\mathfrak{B} \in \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, the set $\mathbf{st}^{-1}(\mathfrak{B}) \subseteq {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ is $LP_N$-measurable.*

### 7.2.3. Almost sure standard parts of hyperfinite empirical measures

We now return to studying properties of the measures $L\mu_{\omega,N}$ for $N \in {}^*\mathbb{N}$. Corollary 7.2.13 immediately leads us to the following.

**Lemma 7.2.14.** *Let $S$ be a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $E_N$ be the $L^*\mathbb{P}$-almost sure set fixed in (7.7). Then for each $B \in \mathcal{B}(S)$, the set $\mathbf{st}^{-1}(B)$ is $L\mu_{\omega,N}$-measurable for all $\omega \in E_N$. Furthermore, for each $B \in \mathcal{B}(S)$, the function $\omega \mapsto L\mu_{\omega,N}(\mathbf{st}^{-1}(B))$ thus defines a $[0,1]$-valued random variable almost everywhere on $({}^*\Omega, L({}^*\mathcal{F}), L^*\mathbb{P})$.*

*Proof.* It was proved as part of Lemma 7.2.6 that for each $B \in \mathcal{B}(S)$, the set $\mathbf{st}^{-1}(B)$ is $L\mu_{\omega,N}$-measurable for all $\omega \in E_N$. Thus, the function $\omega \mapsto L\mu_{\omega,N}(\mathbf{st}^{-1}(B))$ is defined $L^*\mathbb{P}$-almost surely on ${}^*\Omega$ for all $B \in \mathcal{B}(S)$.

Now fix $B \in \mathcal{B}(S)$. Since $L^*\mathbb{P}(E_N) = 1$ and $({}^*\Omega, L({}^*\mathcal{F}), L^*\mathbb{P})$ is a complete probability space, showing that the map $\omega \mapsto L\mu_{\omega,N}(\mathbf{st}^{-1}(B))$ is Loeb measurable is equivalent to showing that for any $\alpha \in \mathbb{R}$, the set $\{\omega \in E_N : L\mu_{\omega,N}\left[\mathbf{st}^{-1}(B)\right] > \alpha\}$ is Loeb measurable. Toward that end, fix $\alpha \in \mathbb{R}$. Note that by Lemma 7.2.6, we obtain the following:

$$\begin{aligned}\{\omega \in E_N : L\mu_{\omega,N}\left[\mathbf{st}^{-1}(B)\right] > \alpha\} &= \{\omega \in E_N : \left[\mathbf{st}(\mu_{\omega,N})\right](B) > \alpha\} \\ &= E_N \cap \left[\mu_{\cdot,N}{}^{-1}\left(\mathbf{st}^{-1}\left(\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > \alpha\}\right)\right)\right].\end{aligned}$$

By Theorem 6.3.2 and Corollary 7.2.13, we also have the following:

$$\mathbf{st}^{-1}\left(\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > \alpha\}\right) \in L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))).$$

The proof is now completed by Corollary 7.2.10. $\square$

The next two lemmas are preparatory for Theorem 7.2.18 that shows that for each Borel set $B \in \mathcal{B}(S)$, the $L\mu_{\omega,N}$ measures of $\mathbf{st}^{-1}(B)$ and ${}^*B$ are almost surely equal to each other.

**Lemma 7.2.15.** *Let $S$ be a Hausdorff space and let $N \in {}^*\mathbb{N}$. Let $K$ be a compact subset of $S$. Then,*

$$L\mu_{\omega,N}(\mathbf{st}^{-1}(K)) = L\mu_{\omega,N}({}^*K) \text{ for } L^*\mathbb{P}\text{-almost all } \omega \in {}^*\Omega.$$

*Proof.* Let $K \subseteq S$ be a compact set. Let $E_N \subseteq {}^*\Omega$ be as in (7.7). By Lemma 7.2.6, we know that $\mathbf{st}^{-1}(K)$ is $L\mu_{\omega,N}$-measurable for all $\omega \in E_N$. Since $K$ is compact, we also have

${}^*K \subseteq \mathbf{st}^{-1}(K)$. It is thus clear from the definition of standard parts that the following holds:

$$\mathbf{st}^{-1}(K) \backslash {}^*K \subseteq {}^*O \backslash {}^*K = {}^*(O \backslash K) \text{ for all open sets } O \text{ such that } K \subseteq O. \tag{7.32}$$

Using Lemma 7.2.14 and Corollary 7.2.2 respectively, we know that the maps $\omega \mapsto L\mu_{\omega,N}\left[\mathbf{st}^{-1}(K) \backslash {}^*K\right]$ and $\omega \mapsto L\mu_{\omega,N}({}^*O \backslash {}^*K)$ are $L^*\mathbb{P}$ measurable for all open sets $O$ containing $K$. Taking expected values and using (7.32), we obtain the following for any open set $O$ containing $K$:

$$\mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left(\mathbf{st}^{-1}(K) \backslash {}^*K\right)\right] \leq \mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left({}^*O \backslash {}^*K\right)\right]. \tag{7.33}$$

But, by $S$-integrability of the map $\omega \to \mu_{\omega,N}({}^*O \backslash {}^*K)$, we also obtain the following:

$$\begin{aligned} \mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left({}^*O \backslash {}^*K\right)\right] &\approx {}^*\mathbb{E}(\mu_{\cdot}({}^*O \backslash {}^*K)) \\ &= \frac{1}{N} \sum_{i \in [N]} \mathbb{P}[X_i \in O \backslash K] \\ &= \mathbb{P}[X_1 \in O \backslash K]. \end{aligned}$$

Using this in (7.33), taking infimum as $O$ varies over open sets containing $K$, and using the fact that the distribution of $X_1$ is outer regular on compact subsets of $S$, we obtain the following:

$$\mathbb{E}_{L^*\mathbb{P}}\left[L\mu_{\cdot,N}\left(\mathbf{st}^{-1}(K) \backslash {}^*K\right)\right] = 0. \tag{7.34}$$

As a result, there exists a Loeb measurable set $E_{K,N} \in L({}^*\mathcal{F})$ such that

$$\left[L\mu_{\omega,N}\left(\mathbf{st}^{-1}(K) \backslash {}^*K\right)\right] = 0 \text{ for all } \omega \in E_{K,N},$$

completing the proof. □

**Remark 7.2.16.** So far, we have only used the facts that the common distribution of the random variables $X_1, X_2, \ldots$ is tight and that it is outer regular on compact subsets of $S$. Tightness was used in (7.1) and all subsequent results that depended on it, while outer regularity on compact subsets was used to obtain (7.34). The results that follow are consequences of the results obtained so far, and, as such, they also only require the common distribution to be tight and outer regular on compact subsets. For simplicity, however, we will continue working under the assumption that the common distribution of the random variables $X_1, X_2, \ldots$ is Radon.

We can strengthen Lemma 7.2.15 to work for all closed sets, as we show next.

**Lemma 7.2.17.** *Let $S$ be a Hausdorff space and let $N \in {}^*\mathbb{N}$. Let $F$ be a closed subset of $S$. Then we have the following:*

$$L\mu_{\omega,N}(\mathbf{st}^{-1}(F)) = L\mu_{\omega,N}({}^*F) \text{ for } L^*\mathbb{P}\text{-almost all } \omega \in {}^*\Omega. \tag{7.35}$$

*Proof.* Let $(K_n)_{n\in\mathbb{N}}$ be the increasing sequence of compact subsets of $S$ fixed in (7.1), and let $E_N$ be as in (7.7). Thus, we have:

$$L\mu_{\omega,N}\left(\cup_{n\in\mathbb{N}}{}^*K_n\right) = 1 \text{ for all } \omega \in E_N.$$

Using the upper monotonicity of $L\mu_{\omega,N}$, we rewrite the above as follows:

$$\lim_{n\to\infty} L\mu_{\omega,N}\left({}^*K_n\right) = 1 \text{ for all } \omega \in E_N. \tag{7.36}$$

Let $F \subseteq S$ be closed. Since $F \cap K_n$ is compact for all $n \in \mathbb{N}$, by Lemma 7.2.15, there exist $L^*\mathbb{P}$-almost sure sets $(E^{(n)})_{n\in\mathbb{N}}$ such that the following holds:

$$L\mu_{\omega,N}(\mathbf{st}^{-1}(F \cap K_n)) = L\mu_{\omega,N}({}^*F \cap {}^*K_n) \text{ for all } \omega \in E^{(n)}, \text{ where } n \in \mathbb{N}. \tag{7.37}$$

Let $E_F := E_N \cap \left(\cap_{n\in\mathbb{N}} E^{(n)}\right)$. Being a countable intersection of almost sure sets, $E_F$ is also $L^*\mathbb{P}$-almost sure. Letting $\omega \in E_F$ and taking limits as $n \to \infty$ on both sides of (7.37), we obtain the following in view of (7.36):

$$\lim_{n\to\infty} L\mu_{\omega,N}(\mathbf{st}^{-1}(F \cap K_n)) = L\mu_{\omega,N}({}^*F) \text{ for all } \omega \in E_F. \tag{7.38}$$

Using the upper monotonicity of the measure $L\mu_{\omega,N}$ on the left side of (7.38), we obtain the following:

$$L\mu_{\omega,N}\left(\cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n)\right) = L\mu_{\omega,N}({}^*F) \text{ for all } \omega \in E_F. \tag{7.39}$$

But, we also have the following:

$$\begin{aligned}\cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n) &= \mathbf{st}^{-1}\left(\cup_{n\in\mathbb{N}}(F \cap K_n)\right)\\ &= \mathbf{st}^{-1}\left(F \cap \left(\cup_{n\in\mathbb{N}} K_n\right)\right),\end{aligned}$$

so that

$$\begin{aligned}\mathbf{st}^{-1}(F)\backslash \cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n) &= \mathbf{st}^{-1}(F)\backslash\, \mathbf{st}^{-1}\left(F \cap \left(\cup_{n\in\mathbb{N}} K_n\right)\right)\\ &= \mathbf{st}^{-1}\left(F \cap \left(\cap_{n\in\mathbb{N}} S\backslash K_n\right)\right)\\ &\subseteq \cap_{n\in\mathbb{N}} \mathbf{st}^{-1}(S\backslash K_n)\\ &= \cap_{n\in\mathbb{N}}\left[\mathbf{st}^{-1}(S)\backslash\, \mathbf{st}^{-1}(K_n)\right].\end{aligned}$$

Thus, for any $\omega \in E_F$, the following holds:

$$\begin{aligned}L\mu_{\omega,N}\left[\mathbf{st}^{-1}(F)\backslash \cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n)\right] &\leq \lim_{n\to\infty} L\mu_{\omega,N}\left[\mathbf{st}^{-1}(S)\backslash\, \mathbf{st}^{-1}(K_n))\right]\\ &= \lim_{n\to\infty}\left[L\mu_{\omega,N}(\mathbf{Ns}({}^*S)) - L\mu_{\omega,N}(\mathbf{st}^{-1}(K_n))\right]\\ &= \lim_{n\to\infty}\left[1 - L\mu_{\omega,N}({}^*K_n)\right],\end{aligned} \tag{7.40}$$

where the last line follows from Lemma 7.2.15 and the fact that $L\mu_{\omega,N}(\mathbf{Ns}(^*S)) = 1$ for all $\omega \in E_F \subseteq E_N$. Using (7.36) and (7.40), we thus obtain the following:

$$L\mu_{\omega,N}\left[\mathbf{st}^{-1}(F) \setminus \cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n)\right] \leq 1 - \lim_{n\to\infty} L\mu_{\omega,N}(^*K_n) = 1 - 1 = 0.$$

Since $\cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n) \subseteq \mathbf{st}^{-1}(F)$, we thus conclude that

$$L\mu_{\omega,N}\left[\cup_{n\in\mathbb{N}} \mathbf{st}^{-1}(F \cap K_n)\right] = L\mu_{\omega,N}(\mathbf{st}^{-1}(F)). \tag{7.41}$$

Using (7.41) in (7.39) completes the proof. □

Having proved (7.35) for closed sets, it is easy to generalize it for all Borel sets using the standard measure theory trick of showing that the collection of sets satisfying (7.35) forms a sigma algebra. This is the next result.

**Theorem 7.2.18.** *Let $S$ be a Hausdorff space and let $N \in {}^*\mathbb{N}$. Let $B$ be a Borel subset of $S$. Then we have the following:*

$$L\mu_{\omega,N}(\mathbf{st}^{-1}(B)) = L\mu_{\omega,N}(^*B) \text{ for } L^*\mathbb{P}\text{-almost all } \omega \in {}^*\Omega. \tag{7.42}$$

*Proof.* Let $E_N$ be as in (7.7). By Lemma 7.2.6, we know that $\mathbf{st}^{-1}(B)$ is $L\mu_{\omega,N}$-measurable for all $\omega \in E_N$ and $B \in \mathcal{B}(S)$. Consider the following collection:

$$\mathcal{G} := \{B \in \mathcal{B}(S) : \ \exists E_B \in L(^*\mathcal{F}) \\ [(L^*\mathbb{P}(E_B) = 1) \wedge \left(\forall \omega \in E_B \cap E_N \left(L\mu_{\omega,N}(\mathbf{st}^{-1}(B)) = L\mu_{\omega,N}(^*B)\right)\right)]\}. \tag{7.43}$$

By Lemma 7.2.17, we know that $\mathcal{G}$ contains all closed sets. In order to show that $\mathcal{G}$ contains all Borel sets, by Dynkin's $\pi$-$\lambda$ theorem, it thus suffices to show that $\mathcal{G}$ is a Dynkin system. In other words, it suffices to show the following:

(i) $S \in \mathcal{G}$.

(ii) If $B \in \mathcal{G}$, then $S \backslash B \in \mathcal{G}$ as well.

(iii) If $(B_n)_{n\in\mathbb{N}}$ is a sequence of mutually disjoint elements of $\mathcal{G}$, then $\cup_{n\in\mathbb{N}} B_n \in \mathcal{G}$.

(i) is immediate from Lemma 7.2.17, with $E_S := E_N$. To see (ii), take $B \in \mathcal{G}$ and let $E_B$ be as (7.43). Note that for any $\omega \in E_B \cap E_N$, we have:

$$\begin{aligned}
L\mu_{\omega,N}\left({}^*(S\backslash B)\right) &= L\mu_{\omega,N}\left({}^*S\backslash {}^*B\right) \\
&= L\mu_{\omega,N}({}^*S) - L\mu_{\omega,N}({}^*B) \\
&= L\mu_{\omega,N}(\mathbf{st}^{-1}(S)) - L\mu_{\omega,N}(\mathbf{st}^{-1}(B)) \\
&= L\mu_{\omega,N}\left(\mathbf{st}^{-1}(S)\backslash\, \mathbf{st}^{-1}(B)\right) \\
&= L\mu_{\omega,N}\left(\mathbf{st}^{-1}(S\backslash B)\right).
\end{aligned}$$

In the above argument, the third line used the fact that $S$ and $B$ are in $\mathcal{G}$, the fourth line used the fact that $\mathbf{st}^{-1}(B) \subseteq \mathbf{st}^{-1}(S)$, and the fifth line used the fact that $\mathbf{st}^{-1}(S)\backslash\, \mathbf{st}^{-1}(B) = \mathbf{st}^{-1}(S\backslash B)$ (which can be seen to follow from Lemma 1.3.9 since $S$ is Hausdorff).

We now prove (iii). Let $(B_n)_{n\in\mathbb{N}}$ be a sequence of mutually disjoint elements of $\mathcal{G}$ and let $B := \sqcup_{n\in\mathbb{N}} B_n$. By Lemma 1.3.9 and the fact that $B_n \in \mathcal{G}$ for all $n \in \mathbb{N}$, we have

the following for all $\omega \in {}^*\Omega$:

$$\begin{aligned}
L\mu_{\omega,N}\left(\mathbf{st}^{-1}(B)\right) &= L\mu_{\omega,N}\left(\mathbf{st}^{-1}\left(\sqcup_{n\in\mathbb{N}} B_n\right)\right) \\
&= L\mu_{\omega,N}\left(\sqcup_{n\in\mathbb{N}}\,\mathbf{st}^{-1}(B_n)\right) \\
&= \sum_{n\in\mathbb{N}} L\mu_{\omega,N}\left(\mathbf{st}^{-1}(B_n)\right) \\
&= \sum_{n\in\mathbb{N}} L\mu_{\omega,N}\left({}^*B_n\right). \qquad (7.44)
\end{aligned}$$

Let $E_{(B_n)_{n\in\mathbb{N}}}$ be as in Lemma 7.2.4 and define $E_B := E_{(B_n)_{n\in\mathbb{N}}}$. Using (7.44) and (7.8), we thus obtain the following:

$$L\mu_{\omega,N}\left(\mathbf{st}^{-1}(B)\right) = L\mu_{\omega,N}\left[{}^*\left(\sqcup_{n\in\mathbb{N}} B_n\right)\right] = L\mu_{\omega,N}({}^*B) \text{ for any } \omega \in E_B \cap E_N,$$

completing the proof. $\square$

Recall that by Lemma 7.2.6, if $S$ is Hausdorff then $\mu_{\omega,N} \in \mathbf{Ns}({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, with $\mathbf{st}(\mu_{\omega,N}) = L\mu_{\omega,N} \circ \mathbf{st}^{-1}$ for all $\omega \in E_N$. Thus Theorem 7.2.18 shows the following:

**Theorem 7.2.19.** *Let $S$ be a Hausdorff space. For any Borel set $B \in \mathcal{B}(S)$, we have*

$$\mathbf{st}(\mu_{\omega,N}({}^*B)) = (\mathbf{st}(\mu_{\omega,N}))(B) \text{ for almost all } \omega \in {}^*\Omega. \qquad (7.45)$$

We point out an interesting interpretation of Theorem 7.2.19. For each Borel set $B \in \mathcal{B}(S)$, the Loeb measure $L\mu_{\omega,N}({}^*B)$ can almost surely be computed by either of the following two-step procedures:

(i) First find $\mu_{\omega,N}({}^*B) \in {}^*[0,1]$ and then take the standard part of this finite nonstandard real number, which is the direct way.

(ii) First take the standard part of the internal measure $\mu_{\omega,N} \in {}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, and then compute the measure $\mathbf{st}(\mu_{\omega,N})(B)$ of $B$ with respect to this standard part.

Since the intersection of countably many almost sure sets is almost sure, we have thus shown the almost sure commutativity of the following diagram for any countable subset $\mathcal{C} \subseteq \mathcal{B}(S)$:

$$\begin{tikzcd} & {}^*[0,1] \arrow[dr, "\mathbf{st}"] & \\ \mathcal{C} \arrow[ur, "B \mapsto \mu_{\omega,N}({}^*B)"] \arrow[rr, "\mathbf{st}(\mu_{\omega,N})"] & & {[0,1]} \end{tikzcd}$$

It is also interesting to remark that equation (7.42) in the conclusion of Theorem 7.2.18 is related to the notion of the so-called *standardly distributed* internal measures, first defined in Anderson [13, Definition 8.1, p. 683] as a concept motivated by an application to mathematical economics á la Anderson [14].

**Definition 7.2.20.** An internal probability measure $\nu$ on $({}^*S, {}^*\mathcal{B}(S))$ is said to be *standardly distributed* if the following holds:

$$L\nu({}^*B) = L\nu(\mathbf{st}^{-1}(B)) \text{ for all } B \in \mathcal{B}(S). \tag{7.46}$$

Theorem 7.2.18 shows that given a particular $B \in \mathcal{B}(S)$ and $N \in {}^*N$, equation (7.46) holds for $\nu$ of the type $\mu_{\omega,N}$ for $L^*\mathbb{P}$-almost all $\omega$. Using a more quantitative approach, Anderson [13, Theorem 8.7(i), p. 685] shows a stronger version of this result with the added hypothesis that the $(X_n)_{n\in\mathbb{N}}$ are independent.

### 7.2.4. Pushing down certain Loeb integrals on the space of all Radon probability measures

We finish this section by relating certain nonstandard integrals over the space $({}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), {}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})), P_N)$ to those over $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})), LP_N \circ \mathbf{st}^{-1})$.

**Theorem 7.2.21.** *Suppose $S$ is a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $P_N$ be as in*

(7.16). *Let* $({}^*\mathfrak{P}(\mathrm{S}), L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))), LP_N)$ *be the associated Loeb space. Then for any Borel subset* $B$ *of* $S$*, we have:*

$$^*\int_{^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(^*B)dP_N(\mu) \approx \int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B)d\mathscr{P}_N(\mu), \tag{7.47}$$

*where* $\mathscr{P}_N = LP_N \circ \mathbf{st}^{-1} \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$.

*Proof.* Fix $B \in \mathcal{B}(S)$. By Corollary 7.2.2 and (7.16), the function $\mu \mapsto \mu(^*B)$ is internally Borel measurable on ${}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Since it is finitely bounded (by one), it is $S$-integrable. Using this and Lemma 7.2.9, we thus obtain the following:

$$\begin{aligned} {}^*\mathbb{E}_{P_N}(\mu(^*B)) &\approx \int_{^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mathbf{st}(\mu(^*B))dLP_N(\mu) \\ &= \int_{^*\Omega} \mathbf{st}(\mu_{\omega,N}(^*B))dL^*\mathbb{P}(\omega) \\ &= \int_{^*\Omega} (\mathbf{st}(\mu_{\omega,N}))(B)dL^*\mathbb{P}(\omega), \end{aligned}$$

where we used Theorem 7.2.19 in the last line. Writing the last integral as a Lebesgue integral of tail probabilities, we make the following conclusion:

$$\begin{aligned} {}^*\mathbb{E}_{P_N}(\mu(^*B)) &\approx \int_{[0,1]} L^*\mathbb{P}((\mathbf{st}(\mu_{\omega,N}))(B) > y)d\lambda(y) \\ &= \int_{[0,1]} L^*\mathbb{P}\left[{\mu_{\cdot,N}}^{-1}\left(\mathbf{st}^{-1}\left(\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > y\}\right)\right)\right] d\lambda(y) \\ &= \int_{[0,1]} LP_N\left(\mathbf{st}^{-1}\left(\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > y\}\right)\right) d\lambda(y), \end{aligned}$$

where the last line follows from Corollary 7.2.10. (This also uses the fact that the set $\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > y\}$ is Borel measurable, in view of Theorem 6.3.2.)

Defining $\mathscr{P}_N := LP_N \circ \mathbf{st}^{-1}$ and noting that $\mathscr{P}_N$ is a Radon probability measure on

$\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ (by Theorem 7.2.12), we obtain the following:

$$
\begin{aligned}
{}^*\mathbb{E}_{P_N}(\mu({}^*B)) &\approx \int_{[0,1]} \mathscr{P}_N\left(\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B) > y\}\right) d\lambda(y) \\
&= \int_{\mathfrak{P}(\mathrm{S})} \mu(B) d\mathscr{P}_N(\mu),
\end{aligned}
$$

thus completing the proof. □

Note that the same proof idea can be used to prove the version of (7.47) for multiple closed sets. Indeed, we have the following theorem.

**Theorem 7.2.22.** *Suppose $S$ is a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $P_N$ be as in (7.16). Let $({}^*\mathfrak{P}(\mathrm{S}), L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))), LP_N)$ be the associated Loeb space. Then for finitely many Borel subsets $B_1, \ldots, B_k$ of $S$, we have:*

$$
{}^*\!\!\int_{{}^*\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu({}^*B_1)\cdots\mu({}^*B_k) dP_N(\mu) \approx \int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1)\cdots\mu(B_k) d\mathscr{P}_N(\mu), \tag{7.48}
$$

*where $\mathscr{P}_N = LP_N \circ \mathbf{st}^{-1}$.*

The proof goes exactly the same way as that of Theorem 7.2.21, once we know that the set $\{\nu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \nu(B_1)\cdots\nu(B_k) > y\}$ is Borel measurable in $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ for all $y \in [0,1]$. But this follows from the fact that a product of measurable functions is measurable (and that for each $i \in [k]$, the function $\nu \mapsto \nu(B_i)$ is measurable by Theorem 6.2.7).

Combining with Lemma 7.2.9, we can interject a ${}^*\mathbb{P}$-integral in the approximate equation (7.48), which will be useful in our proof of de Finetti's theorem in the next section. We state that as a corollary,

**Corollary 7.2.23.** *Suppose $S$ is a Hausdorff space. Let $N \in {}^*\mathbb{N}$ and let $P_N$ be as in (7.16). Let $({}^*\mathfrak{P}(\mathrm{S}), L_{P_N}({}^*\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))), LP_N)$ be the associated Loeb space. Let $\mathscr{P}_N = LP_N \circ$*

$\mathbf{st}^{-1}$, *which is a Radon measure on* $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. *Then for finitely many Borel subsets* $B_1, \ldots, B_k$ *of* $S$, *we have:*

$$^{*}\int_{^*\Omega} \mu_{\omega,N}(^*B_1) \cdots \mu_{\omega,N}(^*B_k) d^*\mathbb{P}(\omega) \approx \int_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdots \mu(B_k) d\mathscr{P}_N(\mu). \tag{7.49}$$

### 7.3. de Finetti–Hewitt–Savage theorem

#### 7.3.1. Uses of exchangeability and a generalization of Ressel's Radon presentability

The previous section built a theory of hyperfinite empirical measures arising out of any sequence of identically Radon distributed random variables taking values in a Hausdorff space. If we further require the random variables to be exchangeable, then the theory from Section 7.2 gives new tools to attack de Finetti style theorems in great generality. Let us first consider an exchangeable sequence of random variables taking values in *any* measurable space $S$. We define hyperfinite empirical measures $\mu_{\omega,N}$ in the same manner as in the previous section. If $N > \mathbb{N}$, then the joint distribution of any finite subcollection of the random variables is given by the expected values of products of hyperfinite empirical measures. This is proved in the next theorem, which is the main technical result that yields general forms of de Finetti's theorem in view of Corollary 7.2.23.

**Theorem 7.3.1.** *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Let* $(X_n)_{n\in\mathbb{N}}$ *be a sequence of* $S$-*valued exchangeable random variables, where* $(S, \mathfrak{S})$ *is some measurable space. For each* $N > \mathbb{N}$ *and* $\omega \in {^*\Omega}$, *define the internal probability measure* $\mu_{\omega,N}$ *as follows:*

$$\mu_{\omega,N}(B) := \frac{\#\{i \in [N] : X_i(\omega) \in B\}}{n} \text{ for all } B \in {^*\mathfrak{S}}. \tag{7.50}$$

*Then we have:*

$$^*\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) \approx {}^*\!\!\int_{^*\Omega} \mu_{\omega,N}(B_1) \cdots \mu_{\omega,N}(B_k) d^*\mathbb{P}(\omega)$$
$$\textit{for all } k \in \mathbb{N} \textit{ and } B_1, \ldots, B_k \in {}^*\mathfrak{S}. \tag{7.51}$$

It should be pointed out that Theorem 7.3.1 may be viewed as a consequence of transferring Diaconis–Freedman's finite, approximate version of de Finetti's theorem [31, Theorem (13)] into the hyperfinite setting. We will provide two alternate proofs that underscore other ways of thinking about this result. The proof of Theorem 7.3.1 in the main body of the dissertation uses a similar combinatorial construction as Diaconis–Freedman's proof, with a key difference being that we can use inclusion-exclusion to give softer combinatorial arguments while still obtaining the same bounds. This proof does not use the hyperfiniteness of $N$ in an essential way, and, as such, it can actually be thought of as a proof of the aforementioned result in Diaconis–Freedman (see (7.56), (7.59), (7.60), (7.61), and compare with [32, Theorem (13), p. 749]).

Our second proof of Theorem 7.3.1 is carried out in Appendix E. This proof illustrates an important explanatory advantage of stating Theorem 7.3.1 as a less quantitative version of Diaconis–Freedman's result in the hyperfinite setting—such a statement is still strong enough to be sufficient in the proof of the infinitary de Finetti's theorem, while the particular form of the statement ensures that it can be both predicted and understood by a reasoning based on Bayes' theorem. This nicely ties in with the fact that de Finetti's theorem is often interpreted as a foundational result for Bayesian statistics (see, for example, Savage [88, Section 3.7]; see also Orbanz and Roy [76] for a recent discussion in

connection with the foundations of statistical modeling).

To better understand this idea, let us analyze (7.51) from the perspective of Bayes' theorem. Instead of the sets $B_1, \ldots, B_k \in {}^*\mathfrak{S}$ that appear there, suppose we consider $A_1, \ldots, A_k \in {}^*\mathfrak{S}$ such that any two of them are either disjoint or equal. Let $C_1, \ldots, C_n$ be the distinct sets appearing in the finite sequence $A_1, \ldots, A_k$. In that case, writing the Cartesian product $A_1 \times \ldots \times A_k$ as $\vec{A}$ and the random vector $(X_1, \ldots, X_k)$ as $\vec{X}$, the internal Bayes' theorem expansion (conditioning on the various possible values of the empirical sample means of the distinct sets $C_1, \ldots, C_n$) of the left side of (7.51) is the following:

$$\begin{aligned} &{}^*\mathbb{P}((X_1, \ldots, X_k) \in \vec{A}) \\ &= \sum_{(t_1,\ldots,t_n)\in[N]^n} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right). \end{aligned} \tag{7.52}$$

In this case, assuming that the set $C_i$ appears in the finite sequence $A_1, \ldots, A_k$ with a frequency $k_i$ (where $i \in [n]$), the right side of (7.51) can be written as the following hyperfinite sum by the (transfer of the) definition of expected values:

$$\begin{aligned} &{}^*\!\!\int_{{}^*\Omega} \mu_{\omega,N}(A_1) \cdots \mu_{\omega,N}(A_k) d^*\mathbb{P}(\omega) \\ &= \sum_{(t_1,\ldots,t_n)\in[N]^n} \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right). \end{aligned} \tag{7.53}$$

If $t_1, \ldots, t_n > \mathbb{N}$ are such that the corresponding term in the internal sum (7.52) is nonzero, then the ratio of that term with the corresponding term on the right side of (7.53) can be shown to be infinitesimally close to one. By an application of underflow and the fact that the partial sums in (7.52) and (7.53) are both infinitesimals when $t_1, \ldots, t_n$ are all bounded by a standard natural number, it can be shown that the two expansions

(7.52) and (7.53) are infinitesimally close, proving (7.51) in the case when any two of the measurable sets being considered are either disjoint or equal. This was the idea in the nonstandard proof of de Finetti's theorem for exchangeable Bernoulli random variables in Alam [5]. Such an argument can then be modified to a proof of Theorem 7.3.1 by writing the event $\{X_1 \in B_k, \ldots, X_k \in B_k\}$ represented by arbitrary sets $B_1, \ldots, B_k \in {}^*\mathfrak{S}$ as a finite disjoint union of events represented by sets of the above type.

A conceptual benefit of this approach is that the idea of the proof is in some sense immediate after expressing the expansions (7.52) and (7.53). Indeed, the two expansions should be expected to be close to each other since the "majority" of the terms are very close to each other, while the rest add up to infinitesimals! While this is a quick way of understanding why Theorem 7.3.1 holds, the details of the term-by-term comparison between (7.52) and (7.53) may get computationally involved. We therefore present a shorter proof below that replaces the exact combinatorial formulas by simpler estimates using inclusion-exclusion. A complete proof based on the above Bayes' theorem idea is included in Appendix E as an alternative.

*Proof of Theorem 7.3.1.* Let $N > \mathbb{N}$ and $(B_1, \ldots, B_k) \in {}^*\mathfrak{S}^k$ be a finite sequence of internal events. Consider the following equation obtained by rewriting the internal product of internal sums on the left as an internal sum of internal products by (the transfer of) distributivity:

$$\prod_{i \in [k]} \left( \sum_{j \in [N]} \mathbb{1}_{B_i}(X_j) \right) = \sum_{(\ell_1, \ldots, \ell_k) \in [N]^k} \left( \prod_{i \in [k]} \mathbb{1}_{B_i}(X_{\ell_i}) \right). \tag{7.54}$$

We separate the terms in the sum on the right of (7.54) according to whether there

is any repetition in $(\ell_1, \ldots, \ell_k)$ or not. Let

$$\mathcal{R} := \{(\ell_1, \ldots, \ell_k) \in [N]^k : \ell_\alpha = \ell_\beta \text{ for some } \alpha \neq \beta\}.$$

An exact value of $\#(\mathcal{R})$ can be found using the (internal) inclusion-exclusion principle. However, the following immediate combinatorial estimate will be sufficient for our needs (for each of the $N$ numbers in $[N]$, there are at most $\binom{k}{2} N^{k-2}$ elements of $[N]^k$ in which that number is repeated at least twice):

$$\#(\mathcal{R}) \leq N \binom{k}{2} N^{k-2} = \binom{k}{2} N^{k-1}. \tag{7.55}$$

Dividing both sides of (7.54) by $N^k$ and noting that $\frac{1}{N} \sum_{j \in [N]} \mathbb{1}_{B_i}(X_j)$ is the same as $\mu_{\cdot,N}(B_i)$ for each $i \in [k]$, we obtain the following:

$$\prod_{i \in [k]} \mu_{\cdot,N}(B_i) = \frac{1}{N^k} \sum_{\ell_1, \ldots, \ell_k \in \mathcal{R}} \left( \prod_{i \in [k]} \mathbb{1}_{B_i}(X_{\ell_i}) \right) + \frac{1}{N^k} \sum_{\ell_1, \ldots, \ell_k \in [N]^k \setminus \mathcal{R}} \left( \prod_{i \in [k]} \mathbb{1}_{B_i}(X_{\ell_i}) \right).$$

Taking expected values and using (7.55) thus yields:

$$\begin{aligned}
0 \leq & {}^*\mathbb{E} \left( \prod_{i \in [k]} \mu_{\cdot,N}(B_i) \right) - {}^*\mathbb{E} \left[ \frac{1}{N^k} \sum_{(\ell_1, \ldots, \ell_k) \in [N]^k \setminus \mathcal{R}} \left( \prod_{i \in [k]} \mathbb{1}_{B_i}(X_{\ell_i}) \right) \right] \\
= & {}^*\mathbb{E} \left[ \frac{1}{N^k} \sum_{\ell_1, \ldots, \ell_k \in \mathcal{R}} \left( \prod_{i \in [k]} \mathbb{1}_{B_i}(X_{\ell_i}) \right) \right] \\
\leq & \frac{\#(\mathcal{R})}{N^k} \\
\leq & \frac{\binom{k}{2} N^{k-1}}{N^k} \\
= & \frac{\binom{k}{2}}{N} \qquad (7.56) \\
\approx & 0. \qquad (7.57)
\end{aligned}$$

As a consequence of (7.57), and using the linearity of expectation, we thus obtain

the following:

$$ {}^*\mathbb{E}\left(\prod_{i\in[k]}\mu_{\cdot,N}(A_i)\right)\approx\frac{1}{N^k}\sum_{(\ell_1,\ldots,\ell_k)\in[N]^k\setminus\mathcal{R}}{}^*\mathbb{E}\left(\prod_{i\in[k]}\mathbb{1}_{A_i}(X_{\ell_i})\right). \tag{7.58} $$

By exchangeability, we also have the following:

$$ \begin{aligned} {}^*\mathbb{E}\left(\prod_{i\in[k]}\mathbb{1}_{B_i}(X_{\ell_i})\right) &= {}^*\mathbb{P}(X_{\ell_1}\in B_1,\ldots,X_{\ell_k}\in B_k) \\ &= {}^*\mathbb{P}(X_1\in B_1,\ldots,X_k\in B_k) \text{ for all } (\ell_1,\ldots,\ell_k)\in[N]^k\setminus\mathcal{R}, \end{aligned} \tag{7.59} $$

which allows us to conclude the following from (7.58):

$$ {}^*\mathbb{E}\left(\prod_{i\in[k]}\mu_{\cdot,N}(B_i)\right)\approx\frac{\#([N]^k\setminus\mathcal{R})}{N^k}{}^*\mathbb{P}(X_1\in B_1,\ldots,X_k\in B_k). \tag{7.60} $$

From (7.55), it is clear that

$$ 1>\frac{\#([N]^k\setminus\mathcal{R})}{N^k}\geq\frac{N^k-\binom{k}{2}N^{k-1}}{N^k}=1-\frac{\binom{k}{2}}{N}\approx 1, \tag{7.61} $$

so that

$$ \frac{\#([N]^k\setminus\mathcal{R})}{N^k}\approx 1. \tag{7.62} $$

Using (7.62) in (7.60) yields the following:

$$ {}^*\mathbb{E}\left(\prod_{i\in[k]}\mu_{\cdot,N}(B_i)\right)\approx{}^*\mathbb{P}(X_1\in B_1,\ldots,X_k\in B_k), $$

thus completing the proof. □

We are in a position to prove the following generalization of Ressel [83, Theorem 3, p. 906].

**Theorem 7.3.2.** *Let $S$ be a Hausdorff topological space, with $\mathcal{B}(S)$ denoting its Borel sigma algebra. Let $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ be the space of all Radon probability measures on $S$ and $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ be the Borel sigma algebra on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ with respect to the A-topology on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$.*

*Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X_1, X_2, \ldots$ be a sequence of exchangeable $S$-valued random variables such that the common distribution of the $X_i$ is Radon on $S$. Then there exists a unique probability measure $\mathscr{P}$ on $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ such that the following holds for all $k \in \mathbb{N}$:*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) = \int\limits_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{P}(\mu)$$

$$\text{for all } B_1, \ldots, B_k \in \mathcal{B}(S). \tag{7.63}$$

*Proof.* Let $N > \mathbb{N}$ and let $P_N$ be as in (7.16). Let $\mathscr{P}$ be $LP_N \circ \mathbf{st}^{-1}$, which is a Radon probability measure on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ by Theorem 7.2.12. The right side of (7.63) is the same as the right side of (7.49), while the left sides of the two equations are infinitesimally close in view of Theorem 7.3.1. This shows the existence of a measure $\mathscr{P} \in \mathfrak{P}_r(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ satisfying (7.63). The uniqueness follows from Theorem 6.3.7. $\square$

We end this subsection with some immediate remarks on the proof of Theorem 7.3.2.

**Remark 7.3.3.** Note that the proof of Theorem 7.3.2 showed that $\mathscr{P}$ could be taken as $LP_N \circ \mathbf{st}^{-1}$ for any $N > \mathbb{N}$, and all of these would have given the same (Radon) measure on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$. Following Theorem 7.2.12, this shows that in the nonstandard extension ${}^*\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ of $\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, the internal measures $P_N$ are nearstandard to $\mathscr{P}$ for *all* $N > \mathbb{N}$.

From the nonstandard characterization of limits in topological spaces, it thus follows that $\mathscr{P}$ is a limit of the sequence $(P_n)_{n\in\mathbb{N}}$ in the $A$-topology on $\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ (and hence in the weak topology as well, since the $A$-topology is finer than the weak topology), where for each $n \in \mathbb{N}$, the probability measure $P_n$ on $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ is defined as follows (this definition of $(P_n)_{n\in\mathbb{N}}$ ensures, by (7.16) and transfer, that $P_N$ is the $N^{\text{th}}$ term in the nonstandard extension of the sequence $(P_n)_{n\in\mathbb{N}}$ for each $N > \mathbb{N}$):

$$P_n(\mathfrak{B}) := \mathbb{P}\left(\{\omega \in \Omega : \mu_{\omega,n} \in \mathfrak{B}\}\right) = \mathbb{P}\left({\mu_{\cdot,n}}^{-1}(\mathfrak{B})\right) \text{ for all } \mathfrak{B} \in \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})). \tag{7.64}$$

Thus our proof shows that the canonical (pushforward) measure on $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ induced by the empirical distribution of the first $n$ random variables does converge (as $n \to \infty$) to a (Radon) measure on $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ which witnesses the truth of Radon presentability. This gives a different (standard) way to understand the measure $\mathscr{P}$ in Theorem 7.3.2, and also connects the proof to the heuristics from statistics described in Section 5.2.

**Remark 7.3.4.** While Remark 7.3.3 shows that the measure $\mathscr{P}$ in Theorem 7.3.2 can be thought of as *a* limit of the sequence $(P_n)_{n\in\mathbb{N}}$, we cannot say that it is *the* limit of this sequence (as the space $\mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, where this sequence lives, may not be Hausdorff). While this was not intended, the use of nonstandard analysis allowed us to canonically find a useful limit point of this sequence using the machinery built in Theorem 6.1.2 and Theorem 6.5.4. The usefulness of nonstandard analysis in this context is thus highlighted by the observation that without invoking this machinery, it is not clear why there should be a Radon limit of this sequence at all.

**Remark 7.3.5.** Following Lemma 6.2.12 (thinking of $T'$ as $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ and $T$ as $\mathfrak{P}(\mathrm{S})$), we can canonically get a sequence $({P_n}')_{n\in\mathbb{N}}$ in $\mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ that can be seen to have $\mathscr{P}' \in \mathfrak{P}(\mathfrak{P}(\mathrm{S}))$ as

a limit point. We make this way of thinking precise when we next prove a generalization of the classical version of de Finetti's theorem (as opposed to Ressel's "Radon presentable" version).

### 7.3.2. Generalizing classical de Finetti's theorem

While Theorem 7.3.2 is already a generalization of de Finetti's theorem, its conclusion is slightly different from classical statements of de Finetti's theorem that postulate the existence of a probability measure on the space of all probability measures (as opposed to a Radon measure on the space of all Radon measures). This can be easily remedied using ideas from Lemma 6.2.11 and Lemma 6.2.12, but at the cost of uniqueness. By Theorem 6.4.3, we still have uniqueness if we focus on probability measures on the smallest sigma algebra on $\mathfrak{P}(\mathrm{S})$ that makes all evaluation functions measurable. As pointed out in Theorem 6.4.3, this is the same as uniqueness for Borel measures on $\mathfrak{P}(\mathrm{S})$ if $S$ is a separable metric space. We prove this generalization next. In fact, we prove a slightly stronger result that has the above conclusion for any sequence $(X_n)_{n\in\mathbb{N}}$ of random variables satisfying (7.63).

**Theorem 7.3.6.** *Let $S$ be a Hausdorff topological space, with $\mathcal{B}(S)$ denoting its Borel sigma algebra. Let $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$) be the space of all Borel probability measures (respectively Radon probability measures) on $S$, and let $\mathcal{B}(\mathfrak{P}(\mathrm{S}))$ (respectively $\mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$) be the Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$) with respect to the A-topology on $\mathfrak{P}(\mathrm{S})$ (respectively $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$). Let $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ be the smallest sigma algebra on $\mathfrak{P}(\mathrm{S})$ such that for any $B \in \mathcal{B}(S)$, the evaluation function $e_B\colon \mathfrak{P}(\mathrm{S}) \to \mathbb{R}$, defined by $e_B(\nu) = \nu(B)$, is measurable. Also let $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$ be the Borel sigma algebra induced by the weak topology on*

$\mathfrak{P}(\mathrm{S})$.

*Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Let* $X_1, X_2, \dots$ *be a sequence of* $S$*-valued random variables. Suppose that there exists a unique probability measure* $\mathscr{P}$ *on* $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{B}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ *such that the following holds for all* $k \in \mathbb{N}$*:*

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \int\limits_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \dots \cdot \mu(B_k) d\mathscr{P}(\mu)$$

$$\textit{for all } B_1, \dots, B_k \in \mathcal{B}(S). \tag{7.65}$$

*Then there exists a probability measure* $\mathscr{Q}$ *on* $(\mathfrak{P}(\mathrm{S}), \mathcal{B}(\mathfrak{P}(\mathrm{S})))$ *such that the following holds for all* $k \in \mathbb{N}$*:*

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \dots \cdot \mu(B_k) d\mathscr{Q}(\mu)$$

$$\textit{for all } B_1, \dots, B_k \in \mathcal{B}(S). \tag{7.66}$$

*Also, there is a unique probability measure* $\mathscr{Q}_c$ *on* $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ *satisfying the following for all* $k \in \mathbb{N}$*:*

$$\mathbb{P}(X_1 \in B_1, \dots, X_k \in B_k) = \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \dots \cdot \mu(B_k) d\mathscr{Q}_c(\mu)$$

$$\textit{for all } B_1, \dots, B_k \in \mathcal{B}(S). \tag{7.67}$$

*Furthermore, if* $S$ *is a separable metric space, then* $\mathcal{C}(\mathfrak{P}(\mathrm{S})) = \mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$*, so that there is a unique probability measure* $\mathscr{Q}_c$ *on* $(\mathfrak{P}(\mathrm{S}), \mathcal{B}_w(\mathfrak{P}(\mathrm{S})))$ *satisfying* (7.67)*.*

*Proof.* Let $\mathscr{P} \in \mathfrak{P}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ be the (Radon) measure obtained in (7.65). Define $\mathscr{Q} \colon \mathcal{B}(\mathfrak{P}(\mathrm{S})) \to$

$[0,1]$ as follows:

$$\mathscr{Q}(\mathfrak{B}) := \mathscr{P}(\mathfrak{B} \cap \mathfrak{P}_{\mathrm{r}}(\mathrm{S})) \text{ for all } \mathfrak{B} \in \mathcal{B}(\mathfrak{P}(\mathrm{S})). \tag{7.68}$$

By Lemma 6.2.12, this defines a probability measure on $(\mathfrak{P}(\mathrm{S}), \mathcal{B}(\mathfrak{P}(\mathrm{S})))$ (in fact, $\mathscr{Q}$ is the same as $\mathscr{P}'$ in the terminology of Lemma 6.2.12). Equation (7.66) now follows from (7.65) and (6.22) (within Lemma 6.2.12).

Call $\mathscr{Q}_c$ the restriction of $\mathscr{Q}$ to $\mathcal{C}(\mathfrak{P}(\mathrm{S})) \subseteq \mathcal{B}(\mathfrak{P}(\mathrm{S}))$. Note that for each $k \in \mathbb{N}$ and $B_1, \ldots, B_k \in \mathcal{B}(S)$, the map $\mu \mapsto \mu(B_1) \cdot \ldots \cdot \mu(B_k)$ is $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ measurable as well, so that we have the following:

$$\begin{aligned}
\int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{Q}_c(\mu) &= \int_{[0,1]} \mathscr{Q}_c \left[\mu(B_1) \cdot \ldots \cdot \mu(B_k) > y\right] d\lambda(y) \\
&= \int_{[0,1]} \mathscr{Q} \left[\mu(B_1) \cdot \ldots \cdot \mu(B_k) > y\right] d\lambda(y) \\
&= \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{Q}_c(\mu).
\end{aligned}$$

Together with Theorem 6.4.3, this shows that there is a unique probability measure $\mathscr{Q}_c$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ satisfying (7.70). Theorem 6.4.2(iii) now completes the proof. □

In view of Theorem 7.3.2, the above result immediately yields our main theorem.

**Theorem 7.3.7.** *Let $S$ be a Hausdorff topological space, with $\mathcal{B}(S)$ denoting its Borel sigma algebra. Let $\mathfrak{P}(\mathrm{S})$ be the space of all Borel probability measures on $S$ and $\mathcal{B}(\mathfrak{P}(\mathrm{S}))$ be the Borel sigma algebra on $\mathfrak{P}(\mathrm{S})$ with respect to the A-topology on $\mathfrak{P}(\mathrm{S})$. Let $\mathcal{C}(\mathfrak{P}(\mathrm{S}))$ be the smallest sigma algebra on $\mathfrak{P}(\mathrm{S})$ such that for any $B \in \mathcal{B}(S)$, the evaluation function $e_B \colon \mathfrak{P}(\mathrm{S}) \to \mathbb{R}$, defined by $e_B(\nu) = \nu(B)$, is measurable. Also let $\mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$ be the Borel sigma algebra induced by the weak topology on $\mathfrak{P}(\mathrm{S})$.*

*Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X_1, X_2, \ldots$ be a sequence of exchangeable $S$-valued random variables such that the common distribution of the $X_i$ is Radon on $S$. Then there exists a probability measure $\mathscr{Q}$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{B}(\mathfrak{P}(\mathrm{S})))$ such that the following holds for all $k \in \mathbb{N}$:*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) = \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{Q}(\mu)$$

$$\textit{for all } B_1, \ldots, B_k \in \mathcal{B}(S). \tag{7.69}$$

*There is a unique probability measure $\mathscr{Q}_c$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ satisfying the following for all $k \in \mathbb{N}$:*

$$\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) = \int\limits_{\mathfrak{P}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{Q}_c(\mu)$$

$$\textit{for all } B_1, \ldots, B_k \in \mathcal{B}(S). \tag{7.70}$$

*Furthermore, if $S$ is a separable metric space, then $\mathcal{C}(\mathfrak{P}(\mathrm{S})) = \mathcal{B}_w(\mathfrak{P}(\mathrm{S}))$, so that there is a unique probability measure $\mathscr{Q}_c$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{B}_w(\mathfrak{P}(\mathrm{S})))$ satisfying (7.70).*

As explained in Remark 7.2.16, our proof of Theorem 7.3.7 did not use the full strength of the assumption that the common distribution of the exchangeable random variables $X_1, X_2, \ldots$ is Radon. The same proof would work if we assumed this common distribution to be tight and outer regular on compact subsets of $S$ (indeed, the proof of Theorem 7.3.2 would go through under these assumptions, while the rest of the steps in our proof of Theorem 7.3.7 are consequences of the conclusion of Theorem 7.3.2).

In practice, a natural situation in which the latter condition always holds is when $S$ is a Hausdorff $G_\delta$ space—that is, when all closed subsets of $S$ are $G_\delta$ sets (as any finite

Borel measure on such a space is actually outer regular on all closed subsets, and in particular on all compact subsets).

In the point-set topology literature, $G_\delta$ spaces typically arise in discussions on *perfectly normal spaces*. Following are some commonly studied examples of spaces that are perfectly normal (as described in Gartside [44, p. 274], these are actually examples of *stratifiable spaces*, which are automatically perfectly normal):

(i) All CW complexes are perfectly normal. See Lundell and Weingram [71, Proposition 4.3, p. 55].

(ii) All Lašnev spaces (that is, all continuous closed images of metric spaces, where a continuous map $g\colon T \to T'$ is called *closed* if $g(F)$ is closed in $T'$ whenever $F$ is closed in $T$) are perfectly normal. This, in particular, includes all metric spaces. See Slaughter [91] for more details.

(iii) If $T$ is a compact-covering image of a Polish space (here, a continuous map $f\colon T \to T'$ is called a *compact-covering* if every compact subset of $T'$ is the image of a compact subset of $T$; see Michael–Nagami 1973 and the references therein for more details on compact-covering images of metric spaces), then the space $C_k(T)$ of continuous real-valued functions on $T$ (equipped with the compact-open topology) is perfectly normal. In particular, this implies that $C_k(T)$ is perfectly normal whenever $T$ is a Polish space. See Gartside and Reznichenko [Theorem 34, p. 111][43].

The above discussion shows that we could have stated Theorem 7.3.7 for any exchangeable sequence of tightly distributed random variables taking values in a Hausdorff state space that is either a CW complex, a Lašnev space, or a space of continuous real-valued functions on a Polish space (with the compact-open topology). This, however, would not be a more general statement than that of Theorem 7.3.7, as it is easy to see that any tight finite measure on a Hausdorff $G_\delta$ space is automatically Radon. It is still instructive to keep in mind these settings where one only needs to verify tightness of the common distribution in order for de Finetti–Hewitt–Savage theorem to hold.

**Remark 7.3.8.** Dubins and Freedman [35] had constructed an exchangeable sequence of random variables taking values in a separable metric space for which the conclusion of de Finetti's theorem does not hold. An indirect consequence of the above discussion is that any random variable in such an example must not have a tight distribution.

**Remark 7.3.9.** We emphasize again that besides tightness of the underlying common distribution, one only needs outer regularity on compact subsets in order for de Finetti-Hewitt–Savage theorem to hold. Though we have not been able to find any natural examples of Hausdorff spaces in which all compact subsets (but not all closed subsets) are $G_\delta$ sets, such spaces (if they exist) might yield more classes of examples where de Finetti–Hewitt–Savage theorem holds for any exchangeable sequence of tightly distributed random variables.

Note that all finite Borel measures on any $\sigma$-compact space are tight. Combined with the above examples of perfectly normal spaces, this gives us classes of state spaces for which de Finetti–Hewitt–Savage theorem holds unconditionally (namely, any $\sigma$-compact perfectly normal space would be an example). While instructive from the point of view of examples, this is not surprising as such spaces are also examples of *Radon spaces* (that is, spaces on which every finite Borel measure is Radon), so that Theorem 7.3.7 automatically holds for any exchangeable sequence of random variables on such state spaces. Other examples of Radon spaces are Polish spaces, which is the setting for modern treatments of de Finetti's theorem. In this sense, Theorem 7.3.7 includes and generalizes the currently known versions of de Finetti's theorem for sequences of Borel measurable exchangeable random variables taking values in a Hausdorff state space. We finish this subsection

by recording the observation that Theorem 7.3.7 theorem holds unconditionally for any Radon state space.

**Corollary 7.3.10.** *Let $S$ be a Radon space. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $X_1, X_2, \ldots$ be a sequence of exchangeable $S$-valued random variables. Then there exists a probability measure $\mathscr{Q}$ on the space $(\mathfrak{P}(\mathrm{S}), \mathcal{B}(\mathfrak{P}(\mathrm{S})))$ such that (7.69) holds. Also, there is a unique probability measure $\mathscr{Q}_c$ on $(\mathfrak{P}(\mathrm{S}), \mathcal{C}(\mathfrak{P}(\mathrm{S})))$ such that (7.70) holds.*

### 7.4. Comments and possible future work

Starting from a result on an exchangeable sequence of $\{0, 1\}$-valued random variables, de Finetti's theorem has had generalizations in several directions. While the classical form of de Finetti's theorem was known to be true for Polish spaces, Dubins and Freedman [35] had shown that some form of topological condition on the state space is necessary. Theorem 7.3.7 shows that we actually do not need any topological conditions on the state space besides Hausdorffness as long as we focus on exchangeable sequences of Radon distributed random variables (by the discussion following Theorem 7.3.7, we actually only need to assume that the common distribution of the random variables is tight and outer regular on compact subsets).

Since properties of the common distribution were crucially used in our proof, the question of the most general state space under which de Finetti's theorem holds (without any assumptions on the common distribution) is quite natural. Corollary 7.3.10 provides some answers (in the form of Radon spaces), but the leap from Theorem 7.3.7 to Corollary 7.3.10 is rather trivial. It would be instructive to investigate if there are other classes of state spaces for which de Finetti's theorem holds unconditionally. Along these lines, it

would also be instructive to find examples of state spaces for which tightness of the underlying common distribution is sufficient for an exchangeable sequence of random variables to be presentable. Radon spaces are again trivial examples, while Hausdorff $G_\delta$ spaces (see examples in (i), (ii), and (iii)) provide some non-trivial examples. Remark 7.3.9 provides a potential strategy for finding more examples, though carrying out this project seems to be beyond the scope of the current dissertation.

There are other formulations of de Finetti's theorem that we have completely ignored in the present treatment. For example, a useful formulation says that an infinite sequence of exchangeable random variables is conditionally independent with respect to certain sigma algebras. See Kingman [61] for a description of such a version of de Finetti's theorem along with some applications.

Another setting in which de Finetti's theorem is traditionally generalized is the setting of exchangeable arrays, with the main result in that setting sometimes called the Aldous–Hoover–Kallenberg representation theorem (See Aldous [8, 9], Hoover [54, 53], and Kallenberg [55, 57]). This is a highly fruitful setting from the point of view of both theoretical and practical applications. Indeed, it has been recently used in graph limits, random graphs, and ergodic theory (see Diaconis and Janson [33], and also Austin [16]) on one hand, and statistical network modeling (see Caron and Fox [23], as well as Veitch and Roy [99]) on the other. While we did not cover exchangeable arrays, an obvious future direction is to try to see if similar techniques allow us to treat that setting as well. In view of Hoover's existing work based on ultraproducts in this setting, it seems likely that there are areas that would benefit from a more concerted nonstandard analytic treatment.

Finally, there are existing generalizations of de Finetti's theorem for random variables indexed by continuous time as well (see Bühlmann [22], Freedman [41], as well as Accardi and Lu [1]), which is yet another area where a nonstandard analytic treatment using hyperfinite time intervals could be useful.

# Appendix A. The Kinetic Theory of Gases and Spherical Surface Measures

This appendix is devoted to the physical motivation behind viewing a high-dimensional spherical integral as a Gaussian mean. We will give an outline of the usual derivation of the Maxwell–Boltzmann distribution (originally discovered by Maxwell in [74] and improved by Boltzmann in [21]), and explain its connection with the problem on limiting spherical integrals studied in this dissertation. We recommend Chapter 5 of Pauli and Enz [77] (which we also roughly follow for our outline) for more details on the underlying physics.

We work under the assumption that a statistically large number (which we shall denote by $N$) of particles of a monatomic gas are moving randomly in a container of a given volume. Each particle has a mass $m$. We further assume that the velocity of a given particle behaves like a random vector following an isotropic continuous probability density function $f\colon \mathbb{R}^3 \to \mathbb{R}$, where the isotropicity of $f$ just means the following:

$$\exists\, g\colon \mathbb{R} \to \mathbb{R} \text{ such that } f(v_1, v_2, v_3) = g({v_1}^2 + {v_2}^2 + {v_3}^2) \text{ for all } v_1, v_2, v_3 \in \mathbb{R}. \tag{A.1}$$

Newtonian mechanics can be used to postulate that the pressure on any wall of the container is directly proportional to the mean squared speed of the gas particles. Combining this with the ideal gas law, it then follows that the average kinetic energy of the particles should be directly proportional to the temperature $T$ of the system. This is typically described by the following equation, where $\vec{v}_i$ is the velocity of the $i^{\text{th}}$ particle, and $k$ is a constant called the *Boltzmann constant*. Note that the factor of $\dfrac{3}{2}$ appears in the fol-

lowing in order to make sure that our $k$ agrees with the traditional value of the Boltzmann constant.

$$\sum_{i=1}^{N} \frac{1}{2} m \left|\left|\vec{v}_i\right|\right|^2 = \frac{3}{2} kTN, \text{ that is, } \frac{\sum_{i=1}^{N} \left|\left|\vec{v}_i\right|\right|^2}{N} = \frac{3kT}{m}. \tag{A.2}$$

We also assume that the three components of the velocity vector of a given particle are independent and identically distributed, with a continuous density function $h \colon \mathbb{R} \to \mathbb{R}$ satisfying the following condition.

$$f(v_1, v_2, v_3) = h(v_1)h(v_2)h(v_3) \text{ for all } v_1, v_2, v_3 \in \mathbb{R}. \tag{A.3}$$

We define new functions $\psi \colon \mathbb{R} \to \mathbb{R}$ and $\phi \colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ by the following formulae:

$$\psi(v_i) := \log(h(v_i)) \text{ for all } v_i \in \mathbb{R}, \text{ and}$$

$$\phi(v^2) := \log(g(v^2)) \text{ for all } v \in \mathbb{R}.$$

Then $\phi$ and $\psi$ satisfy the following functional equation:

$$\phi({v_1}^2 + {v_2}^2 + {v_3}^2) = \psi(v_1) + \psi(v_2) + \psi(v_3). \tag{A.4}$$

Assuming that $\phi$ and $\psi$ are sufficiently differentiable, it can be shown that (A.4) can be satisfied only if $\phi$ is linear. After some simplifications, we obtain the following.

$$f(v_1, v_2, v_3) = g({v_1}^2 + {v_2}^2 + {v_3}^2) = Ce^{-\alpha({v_1}^2 + {v_2}^2 + {v_3}^2)}, \tag{A.5}$$

for some constants $C, \alpha \in \mathbb{R}_{>0}$.

The constant $C$ is obtained to be $\left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}}$ by integrating both sides of (A.5) and noting that the integral of $f$ is equal to 1 as $f$ is a probability density function. We then compute the expected value of the square of the speed ${v_1}^2 + {v_2}^2 + {v_3}^2$, and equate it with $\frac{3kT}{m}$ (which comes (A.2), using our underlying hypothesis of $N$ being statistically large so that the mean of the individual particles' squared speed should be very close to the theoretical expected value – more precisely, one can let $N \to \infty$ and use the strong law of large numbers). From that, we find $\frac{3}{2\alpha} = \frac{3kT}{m}$, so that $\alpha = \frac{m}{2kT}$. We thus obtain the famous Maxwell–Boltzmann distribution for velocity:

$$f(v_1, v_2, v_3) = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{m}{2kT}({v_1}^2+{v_2}^2+{v_3}^2)} \text{ for all } v_1, v_2, v_3 \in \mathbb{R}. \tag{A.6}$$

From the above formula, Maxwell and Boltzmann proceeded to derive probability distributions of other important functions (such as speed) of velocity. These distributions are heavily used in statistical mechanics and thermodynamics.

The problem of statistically estimating the behavior of a function of the velocity of a random gas particle can be reinterpreted in a useful way with the notion of surface area measures on Euclidean spheres. For simplicity of terms we let $N_0 := 3N$, and renormalize the constants in equation (A.2) (by assuming that $kT = m$). Writing $\vec{v}_i = (v_{i,x}, v_{i,y}, v_{i,z}) \in \mathbb{R}^3$, we then get:

$$\sum_{i=1}^{N} \left({v_{i,x}}^2 + {v_{i,y}}^2 + {v_{i,z}}^2\right) = N_0. \tag{A.7}$$

Hence $(\vec{v}_1, \ldots, \vec{v}_N)$ is a vector in $\mathbb{R}^{N_0}$ of norm $\sqrt{N_0}$. In other words, $(\vec{v}_1, \ldots, \vec{v}_N)$

is an element of $S^{N_0-1}(\sqrt{N_0})$. Since we do not have any information about the motion of these particles other than what is contained in equation (A.2), it is reasonable to assume that the value of $(\vec{v}_1, \ldots, \vec{v}_N)$ at a given time is a "random point" of $S^{N_0-1}(\sqrt{N_0})$. The surface area measure $\bar{\sigma}_S$ for a sphere $S$ serves as a notion of a uniform probability measure on $S$. Thus we can make the observation regarding $(\vec{v}_1, \ldots, \vec{v}_N)$ being a random point of $S^{N_0-1}(\sqrt{N_0})$ more precise by postulating that the probability that $(\vec{v}_1, \ldots, \vec{v}_N)$ lies in a Borel set $B \subseteq S^{N_0-1}(\sqrt{N_0})$ is given by $\bar{\sigma}_{S^{N_0-1}(\sqrt{N_0})}(B)$.

Since we are working under the assumption that the number of particles is very large, the probability that the first component of the velocity of the first particle, and hence of a random particle (by symmetry), is in a Borel set $B_1 \subseteq \mathbb{R}^1$, should be given by $\lim_{N_0 \to \infty} \bar{\sigma}((B_1 \times \mathbb{R}^{N_0-1}) \cap S^{N_0-1}(\sqrt{N_0}))$. Also, the expected or mean value of the first component of its velocity should be given by the following integral:

$$\lim_{N_0 \to \infty} \int_{S^{N_0-1}(\sqrt{N_0})} v_{1,x} d\bar{\sigma}(v_{1,x}, v_{1,y}, v_{1,z}, \ldots, v_{N,x}, v_{N,y}, v_{N,z}).$$

Similarly, the expected value of speed would be given by the limit of the integrals of $\sqrt{v_{1,x}^2 + v_{1,y}^2 + v_{1,z}^2}$. In fact, the limit of integrals of any finite-dimensional function on these spheres can be interpreted as the expected value of some function of velocities of randomly chosen particles in our gaseous system.

If there were a way to directly compute these limits, then we would be able to evaluate various probabilities associated with values taken by the velocity components, as well as recover the expected values of many functions of velocities of the particles. Furthermore, such a derivation would have the benefit of being less circular as we would not be making any assumptions on the nature (or even existence) of the density $f$ that was de-

rived in (A.6).

Thus the problem of generalizing Theorem 2.1.1 to the largest class of functions possible is intimately connected to, and has implications on our understanding of the kinetic theory of gases. Furthermore, the fact that there already exist distributions for functions of velocity such as speed (which, being equal to $\sqrt{{v_1}^2 + {v_2}^2 + {v_3}^2}$, is clearly not a bounded function) suggests that (2.1.1) should, in principle, be generalizable to at least some unbounded functions, which in turn makes the problem of finding all such functions naturally appealing.

Mathematically, (2.2) tells us that the Gaussian measure $\mu$ is well-equipped to measure the limiting expected value of any bounded measurable function of a given collection of coordinates. In some sense, it retains all probabilistic information of the manner in which such functions behave over these spheres in the large-$N$ limit. From this point of view as well, it becomes a natural question to find out for which class of functions does it retain all such information.

Nonstandard analysis gives access to hyperfinite natural numbers which provide a natural model for statistically large number of particles. The probability that the velocity of a random particle lies in some set could actually be thought of as the uniform surface area of the portion of a hyperfinite-dimensional sphere corresponding to this set.

## Appendix B. Some Results on Linear Independence

**Lemma B.1.** *Let $u^{(1)}, \ldots, u^{(\gamma)}$ be $\mathbb{R}$-linearly independent vectors in $\mathbb{R}^{\mathbb{N}}$. For all $K > \mathbb{N}$, $(u^{(1)})_{(K)}, \ldots, (u^{(\gamma)})_{(K)}$ are ${}^*\mathbb{R}$-linearly independent. As a consequence, $(u^{(1)})_{(m)}, \ldots, (u^{(\gamma)})_{(m)}$ are $\mathbb{R}$-linearly independent for all large $m \in \mathbb{N}$.*

*Proof.* Fix $M > \mathbb{N}$. Suppose, if possible, that $(u^{(1)})_{(M)}, \ldots, (u^{(\gamma)})_{(M)}$ are not ${}^*\mathbb{R}$-linearly independent. Then there exist $a_1, \ldots, a_\gamma \in {}^*\mathbb{R}$ such that not all the $a_i$ are zero and

$$\sum_{i=1}^{\gamma} a_i (u^{(i)})_{(M)} = \mathbf{0} \in {}^*\mathbb{R}^M. \tag{B.1}$$

Let $a = \max_{i \in \{1,\ldots,\gamma\}} |a_i|$. Since the $a_i$ are not all equal to zero, we have $a > 0$. For each $i \in \{1, \ldots, \gamma\}$, let $b_i = \frac{a_i}{a}$. Divide both sides of (B.1) by $a$ to get

$$\sum_{i=1}^{\gamma} b_i (u^{(i)})_{(M)} = \mathbf{0} \in {}^*\mathbb{R}^M. \tag{B.2}$$

All the $b_i$ are bounded above in absolute value by 1. Thus taking standard parts along the coordinates in $\mathbb{N}$ on both sides of (B.2) gives

$$\sum_{i=1}^{\gamma} \mathbf{st}(b_i) u^{(i)} = \mathbf{0} \in \mathbb{R}^{\mathbb{N}}.$$

Since $u^{(1)}, \ldots, u^{(\gamma)}$ are $\mathbb{R}$-linearly independent, it follows that $\mathbf{st}(b_i) = 0$ for all $i \in \{1, \ldots, \gamma\}$. But this contradicts the fact that $|b_i| = 1$ for at least one $i$ (namely for that index $i$ which makes $|a_i|$ maximum).

Hence $(u^{(1)})_{(M)}, \ldots, (u^{(\gamma)})_{(M)}$ are ${}^*\mathbb{R}$-linearly independent for all $M > \mathbb{N}$. By underflow, $(u^{(1)})_{(m)}, \ldots, (u^{(\gamma)})_{(m)}$ are ${}^*\mathbb{R}$-linearly independent for all large $m \in \mathbb{N}$. But if $m \in \mathbb{N}$, then $(u^{(1)})_{(m)}, \ldots, (u^{(\gamma)})_{(m)}$ are all vectors in $\mathbb{R}^m$, so that the ${}^*\mathbb{R}$-linear independence of these vectors implies that they are also $\mathbb{R}$-linearly independent by transfer. ∎

It is interesting to note that Lemma B.1 is a special case of a more general result on linear independence in infinite-dimensional functional spaces, which we include below. Techniques of this nature have been used in the past in the works of Ross. See [85, Theorem 3] for a related idea used to give a nonstandard proof of the Riesz Representation Theorem.

**Theorem B.2.** *Let $X$ be an infinite set and let $F(X,\mathbb{R})$ denote the vector space of functions from $X$ to $\mathbb{R}$. Suppose $f_1,\dots,f_\gamma$ are linearly independent in $F(X,\mathbb{R})$. Then there is a number $m_0 \in \mathbb{N}$ such that for all $m \in \mathbb{N}_{>m_0}$, there are points $x_1,\dots,x_m \in X$ for which the vectors $(f_i(x_1),\dots,f_i(x_m))_{i\in\{1,\dots,\gamma\}}$ are linearly independent in $\mathbb{R}^m$.*

**Remark B.3.** If $X$ is countable, say, $X = \mathbb{N}$ (in which case $F(X,\mathbb{R})$ is just the vector space of sequences of real numbers), then for any linearly independent vectors $u^{(1)},\dots,u^{(\gamma)}$, there is an $m_0 \in \mathbb{N}$ such that for all $m \in \mathbb{N}_{>m_0}$, the vectors $\left(u^{(i)}{}_1,\dots,u^{(i)}{}_m\right)_{i\in\{1,\dots,\gamma\}}$ are linearly independent in $\mathbb{R}^m$. In this sense, Lemma B.1 is a corollary of this theorem.

*Proof.* For brevity, we will write $f_i(x_1,\dots,x_m)$ to denote $f_i((x_1),\dots,f_i(x_m))$. Let $F_0$ be a hyperfinite set such that $X \subseteq F_0 \subseteq {}^*X$. One obtains $F_0$ via saturation as an element of the following set:

$$\cap_{x\in X}\{F \in {}^*\mathcal{P}_{\text{fin}}(X) : x \in F\}.$$

Suppose that the internal cardinality of $F_0$ is $|F_0| = N$. Since $X$ is infinite, there is an injective map $s\colon \mathbb{N} \to X$. Extend this map to get a hyperfinite sequence $(x_i)_{i\in{}^*\mathbb{N}}$ of distinct elements in ${}^*X$ (by taking $x_i := {}^*s(i)$). For each natural number $m$, we let $[m]$ denote the set $\{1,\dots,m\}$. Also, for two subsets $A,B$ in the standard universe, we let

Bij$(A, B)$ denote the set of bijections between $A$ and $B$ (so it is empty if $A$ and $B$ have different cardinalities). Consider the following internal set:

$$\mathcal{G} := \left\{m \in {}^*\mathbb{N} : \exists \phi \in {}^*\mathrm{Bij}([N], F_0) \text{ such that } {}^*s|_{[m]} = \phi|_{[m]}\right\}.$$

Clearly, $\mathcal{G}$ contains $\mathbb{N}$ and hence contains an $M > \mathbb{N}$ by overflow. Let $\phi$ be the bijection that witnesses the inclusion of $M$ in $\mathcal{G}$. Let $y_i = \phi(i)(= x_i)$ for all $i \in [M]$. Extend $\phi$ (by transfer of the fact that any injective map from an initial set of $\mathbb{N}$ to $X$ can be extended to an injective map from $\mathbb{N}$ to $X$) to an internal injective map $\Phi\colon {}^*\mathbb{N} \to {}^*X$, and still call $y_i = \Phi(i)$ for all $i \in {}^*\mathbb{N}$. Recall that $y_i = x_i$ for all $i \in [M]$ (in particular for all $i \in \mathbb{N}$).

We claim that $({}^*f_i(y_1), \ldots, {}^*f_i(y_N))_{i\in[\gamma]}$ are ${}^*\mathbb{R}$-linearly independent in ${}^*\mathbb{R}^N$. For if not, then there exist $a_1, \ldots, a_\gamma \in {}^*\mathbb{R}$, not all zero, such that $\sum_{i\in[\gamma]} a_i {}^*f_i|_{F_0} = 0$. As in the proof of Lemma B.1, we divide both sides by $\max\{|a_1|, \ldots, |a_\gamma|\}$ and restrict the functions to $X$ to get a contradiction to the linear independence of $f_1, \ldots, f_\gamma$.

Since $N > \mathbb{N}$ was arbitrary, the following internal set contains ${}^*\mathbb{N}\backslash\mathbb{N}$:

$$\{n \in {}^*\mathbb{N} :\ \exists A \in {}^*\mathcal{P}_{\text{fin}}(X)[(|A| = n) \wedge (f_1|_A, \ldots, f_\gamma|_A \ {}^*\mathbb{R}\text{-linearly independent})]\}.$$

By underflow, there is an $m_0 \in \mathbb{N}$ in this set. For any $m \in \mathbb{N}_0$, the transfer of the following sentence completes the proof of the first part of this theorem:

$$\exists A \in {}^*\mathcal{P}_{\text{fin}}(X)[(|A| = n) \wedge (f_1|_A, \ldots, f_\gamma|_A \text{ are } {}^*\mathbb{R}\text{-linearly independent})].$$

For the second part of the theorem, replace $A$ by $[n]$ in the above underflow argument and then proceed as before. $\square$

# Appendix C. Working with Infinitesimally Separated Linear Spaces

In an internal inner product space $V$ (over ${}^*\mathbb{R}$ or ${}^*\mathbb{C}$), a collection of vectors $\mathcal{V}$ is said to satisfy the *separation property (SP)* if the following holds:

$$\text{For any } v \in \mathcal{V}, \;\; \left|\left|v - P_{\mathrm{span}(\mathcal{V}\setminus\{v\})}(v)\right|\right| \not\approx 0. \tag{C.1}$$

Here, for a subspace $H$, the vector $P_H(v)$ denotes the orthogonal projection of the vector $v$ onto $H$. The following equivalent version of SP is more convenient for our applications (the equivalence follows from the linear algebraic fact that distance of a vector from its projection onto a larger subspace cannot be bigger than the distance from its projection onto a smaller subspace).

$$\text{For any } v \in \mathcal{V} \text{ and any subcollection } \mathcal{V}' \subseteq \mathcal{V}\setminus\{v\} : \;\; \left|\left|v - P_{\mathrm{span}(\mathcal{V}')}(v)\right|\right| \not\approx 0. \tag{C.2}$$

When working with spheres intersected by hyperplanes, we often need to orthonormalize different sets of linearly independent vectors (corresponding to two different hyperplanes). If two such sets of vectors can be matched with each other in the sense that any pair is only infinitesimally apart, then we can make such a matching with their orthonormalizations as well, provided the original set of vectors satisfies the Separation Property (this is proved in Theorem C.2).

We first prove a preliminary result that shows that any collection of vectors satisfying SP must be linearly independent. Note that the converse is not true—one could take vectors $\{e_1, \epsilon e_2\}$, or $\{e_1, e_1 + \epsilon e_2\}$ in ${}^*\mathbb{R}^2$, where $\epsilon$ is an infinitesimal. In what follows, we call a vector $v$ *infinitesimal* if $||v|| \approx 0$.

**Proposition C.1.** *Suppose a collection of vectors $\mathcal{V}$ satisfies SP. Then $\mathcal{V}$ does not contain any infinitesimal. Furthermore, $\mathcal{V}$ is ${}^*\mathbb{R}$-linearly independent.*

*Proof.* The first part follows from the fact that any orthogonal projection operator has norm at most 1. Indeed, for any $v \in \mathcal{V}$,

$$\left|\left|v - P_{\mathrm{span}(\mathcal{V}\setminus\{v\})}(v)\right|\right| \leq ||v|| + \left|\left|P_{\mathrm{span}(\mathcal{V}\setminus\{v\})}(v)\right|\right| \leq 2\,||v||\,,$$

which would be infinitesimal if $v$ is an infinitesimal vector.

Now if $\mathcal{V}$ were not linearly independent, then there would exist a vector $v \in \mathcal{V}$ which could be written as a linear combination of vectors from some subcollection $\mathcal{V}' \subseteq V\setminus\{v\}$. But then we would have $P_{\mathrm{span}(\mathcal{V}')}(v) = v$, violating the Separation Property, since we have already shown $v$ to be non-infinitesimal. □

**Theorem C.2.** *Let $V$ be an internal inner product space. Let $\gamma \in \mathbb{N}$. For each $i \in \{1,\ldots,\gamma\}$, let $v^{(i)}, v'^{(i)} \in V$ be such that the following conditions hold:*

*(i) The collections $\{v^{(1)},\ldots,v^{(\gamma)}\}$ and $\{v'^{(1)},\ldots,v'^{(\gamma)}\}$ both satisfy the Separation Property. .*

*(ii) $\left|\left|v^{(i)}\right|\right|, \left|\left|v'^{(i)}\right|\right| \in {}^*\mathbb{R}_{fin}$.*

*(iii) $\left|\left|v^{(i)} - v'^{(i)}\right|\right| \approx 0$.*

*Then there exist orthonormal sets $\{w^{(1)},\ldots,w^{(\gamma)}\}$ and $\{z^{(1)},\ldots,z^{(\gamma)}\}$ with the following properties:*

1. *For any $i \in \{1,\ldots,\gamma\}$, we have*

$$\begin{aligned} span(v^{(1)},\ldots,v^{(i)}) &= span(w^{(1)},\ldots,w^{(i)}),\\ \text{and } span(v'^{(1)},\ldots,v'^{(i)}) &= span(z^{(1)},\ldots,z^{(i)}). \end{aligned}$$

2. *For all $i \in \{1,\ldots,\gamma\}$, we have $\left|\left|w^{(i)} - z^{(i)}\right|\right| \approx 0$.*

*Proof.* Use the Gram-Schmidt algorithm on $\{v^{(1)},\ldots,v^{(\gamma)}\}$ and $\{v'^{(1)},\ldots,v'^{(\gamma)}\}$ to obtain

$\{w^{(1)}, \ldots, w^{(\gamma)}\}$ and $\{z^{(1)}, \ldots, z^{(\gamma)}\}$ respectively. We thus have:

$$
\begin{aligned}
w^{(1)} &:= \frac{v^{(1)}}{||v^{(1)}||}, \\
w^{(2)} &:= \frac{v^{(2)} - \langle v^{(2)}, w^{(1)} \rangle w^{(1)}}{||v^{(2)} - \langle v^{(2)}, w^{(1)} \rangle w^{(1)}||}, \\
w^{(3)} &:= \frac{v^{(3)} - \langle v^{(3)}, w^{(1)} \rangle w^{(1)} - \langle v^{(3)}, w^{(2)} \rangle w^{(2)}}{||v^{(3)} - \langle v^{(3)}, w^{(1)} \rangle w^{(1)} - \langle v^{(3)}, w^{(2)} \rangle w^{(2)}||}, \\
&\vdots \\
w^{(\gamma)} &:= \frac{v^{(\gamma)} - \langle v^{(\gamma)}, w^{(1)} \rangle w^{(1)} - \ldots - \langle v^{(\gamma)}, w^{(\gamma-1)} \rangle w^{(\gamma-1)}}{||v^{(\gamma)} - \langle v^{(\gamma)}, w^{(1)} \rangle w^{(1)} - \ldots - \langle v^{(\gamma)}, w^{(\gamma-1)} \rangle w^{(\gamma-1)}||},
\end{aligned}
\tag{C.3}
$$

and

$$
\begin{aligned}
z^{(1)} &:= \frac{v'^{(1)}}{||v'^{(1)}||}, \\
z^{(2)} &:= \frac{v'^{(2)} - \langle v'^{(2)}, z^{(1)} \rangle z^{(1)}}{||v'^{(2)} - \langle v'^{(2)}, z^{(1)} \rangle z^{(1)}||}, \\
z^{(3)} &:= \frac{v'^{(3)} - \langle v'^{(3)}, z^{(1)} \rangle z^{(1)} - \langle v'^{(3)}, z^{(2)} \rangle z^{(2)}}{||v'^{(3)} - \langle v'^{(3)}, z^{(1)} \rangle z^{(1)} - \langle v'^{(3)}, z^{(2)} \rangle z^{(2)}||}, \\
&\vdots \\
z^{(\gamma)} &:= \frac{v'^{(\gamma)} - \langle v'^{(\gamma)}, z^{(1)} \rangle z^{(1)} - \ldots - \langle v'^{(\gamma)}, z^{(\gamma-1)} \rangle z^{(\gamma-1)}}{||v'^{(\gamma)} - \langle v'^{(\gamma)}, z^{(1)} \rangle z^{(1)} - \ldots - \langle v'^{(\gamma)}, z^{(\gamma-1)} \rangle z^{(\gamma-1)}||}.
\end{aligned}
\tag{C.4}
$$

These sets $\{w^{(1)}, \ldots, w^{(\gamma)}\}$ and $\{z^{(1)}, \ldots, z^{(\gamma)}\}$ of internally orthonormal vectors satisfy (1) by construction. Therefore we need to only verify (2). The proof of (2) will be

done by induction on $i$. Observe that for $i = 1$, we have:

$$\begin{aligned}
\left|\left|w^{(1)} - z^{(1)}\right|\right| &= \left|\left|\frac{\left|\left|v'^{(1)}\right|\right| v^{(1)} - \left|\left|v^{(1)}\right|\right| v'^{(1)}}{||v^{(1)}|| \cdot ||v'^{(1)}||}\right|\right| \\
&= \frac{\left|\left|v^{(1)} - \frac{||v^{(1)}||}{||v'^{(1)}||} v'^{(1)}\right|\right|}{||v^{(1)}||} \\
&\approx \frac{\left|\left|v^{(1)} - 1 \cdot v'^{(1)}\right|\right|}{\mathbf{st}\left(||v^{(1)}||\right)} \\
&\approx 0. \qquad \text{(C.5)}
\end{aligned}$$

Similarly,

$$\begin{aligned}
\left|\left|w^{(2)} - z^{(2)}\right|\right| &= \left|\left|\frac{v^{(2)} - \langle v^{(2)}, w^{(1)}\rangle w^{(1)}}{||v^{(2)} - \langle v^{(2)}, w^{(1)}\rangle w^{(1)}||} - \frac{v'^{(2)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}}{||v'^{(2)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}||}\right|\right| \\
&= \frac{\left|\left|\alpha v^{(2)} - \alpha\left\langle v^{(2)}, w^{(1)}\right\rangle w^{(1)} - \beta v'^{(2)} + \beta\left\langle v'^{(2)}, z^{(1)}\right\rangle z^{(1)}\right|\right|}{\alpha\beta}, \qquad \text{(C.6)}
\end{aligned}$$

where $\alpha = \left|\left|v'^{(2)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}\right|\right|$ and $\beta = \left|\left|v^{(2)} - \langle v^{(2)}, w^{(1)}\rangle w^{(1)}\right|\right|$.

Geometrically, $\alpha$ (respectively $\beta$) represents the orthogonal projection of $v^{(2)}$ (respectively $v'^{(2)}$) onto the span of $v^{(1)}$ (respectively $v'^{(1)}$). Hence, by the SP condition, it follows that $\alpha\beta$ is non-infinitesimal.

By repeated uses of triangle inequality and Cauchy-Schwarz inequality, we have:

$$
\begin{aligned}
&|\alpha - \beta| \\
\leq &\left|\left|v^{(2)} - v'^{(2)}\right|\right| + \left|\left|\langle v^{(2)}, w^{(1)}\rangle w^{(1)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}\right|\right| \\
= &\left|\left|v^{(2)} - v'^{(2)}\right|\right| + \left|\left|\langle v^{(2)}, w^{(1)} - z^{(1)}\rangle w^{(1)} + \langle v^{(2)}, z^{(1)}\rangle w^{(1)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}\right|\right| \\
\leq &\left|\left|v^{(2)} - v'^{(2)}\right|\right| + \left|\langle v^{(2)}, w^{(1)} - z^{(1)}\rangle\right| \left|\left|w^{(1)}\right|\right| \\
&+ \left|\left|\langle v^{(2)}, z^{(1)}\rangle w^{(1)} - \langle v'^{(2)}, z^{(1)}\rangle z^{(1)}\right|\right| \\
\leq &\left|\left|v^{(2)} - v'^{(2)}\right|\right| + \left|\langle v^{(2)}, w^{(1)} - z^{(1)}\rangle\right| \left|\left|w^{(1)}\right|\right| \\
&+ \left|\left|\langle v^{(2)}, z^{(1)}\rangle \left(w^{(1)} - z^{(1)}\right) + \langle v^{(2)} - v'^{(2)}, z^{(1)}\rangle z^{(1)}\right|\right| \\
\leq &\left|\left|v^{(2)} - v'^{(2)}\right|\right| + \left|\left|v^{(2)}\right|\right| \left|\left|w^{(1)} - z^{(1)}\right|\right| \left|\left|w^{(1)}\right|\right| + \left|\left|v^{(2)}\right|\right| \left|\left|z^{(1)}\right|\right| \left|\left|w^{(1)} - z^{(1)}\right|\right| \\
&+ \left|\left|v^{(2)} - v'^{(2)}\right|\right| \left|\left|z^{(1)}\right|\right| \left|\left|z^{(1)}\right|\right|,
\end{aligned}
$$

which is infinitesimal by the hypothesis. Hence, we have

$$|\alpha - \beta| \approx 0. \tag{C.7}$$

Using triangle inequality and Cauchy-Schwarz inequality a few times, we have:

$$
\begin{aligned}
\left|\langle v^{(2)}, w^{(1)}\rangle - \langle v'^{(2)}, z^{(1)}\rangle\right| &= \left|\langle v^{(2)} - v'^{(2)}, w^{(1)}\rangle + \langle v'^{(2)}, w^{(1)} - z^{(1)}\rangle\right| \\
&\leq \left|\langle v^{(2)} - v'^{(2)}, w^{(1)}\rangle\right| + \left|\langle v'^{(2)}, w^{(1)} - z^{(1)}\rangle\right| \\
&\leq \left|\left|v^{(2)} - v'^{(2)}\right|\right| \left|\left|w^{(1)}\right|\right| + \left|\left|v'^{(2)}\right|\right| \left|\left|w^{(1)} - z^{(1)}\right|\right| \qquad \text{(C.8)}
\end{aligned}
$$

The right side of (C.8) is an infinitesimal by the hypothesis and (C.5). Since $\langle v^{(2)}, w^{(1)}\rangle, \langle v'^{(2)}, z^{(1)}\rangle$ are in ${}^*\mathbb{R}_{\text{fin}}$ (one can see this using Cauchy-Schwarz inequality), we thus get:

$$\langle v^{(2)}, w^{(1)}\rangle \approx \langle v'^{(2)}, z^{(1)}\rangle \tag{C.9}$$

Note that $\alpha, \beta \in {}^*\mathbb{R}_{\text{fin}}$ (one can see this by applying the triangle inequality and Cauchy-Schwarz inequality to the expressions for $\alpha$ and $\beta$). Using (C.7) and (C.9) in (C.6) (and using the fact that $\mathbf{st}\colon {}^*\mathbb{R}_{\text{fin}} \to \mathbb{R}$ is a ring homomorphism), we get

$$\left|\left|w^{(2)} - z^{(2)}\right|\right| \approx 0. \tag{C.10}$$

The proof of the case $i = 2$ from the case $i = 1$ clearly generalizes to show, by induction, that $\left|\left|w^{(i)} - z^{(i)}\right|\right| \approx 0$ for all $i \in \{1, \ldots, \gamma\}$. □

**Remark C.3.** Theorem C.2 shows that if two internal subspaces have bases of finite vectors satisfying SP such that they can be matched in pairs of infinitesimal distances, then the same is true for the orthonormalizations of these bases as well. This allows one to "rotate" one subspace to another through an orthogonal transformation of infinitesimal norm, as done in Section 3.

In all applications of this concept in the dissertation, the inner product space $V$ is taken to be ${}^*\mathbb{R}^N$ for some $N \in {}^*\mathbb{N}$ (usually taken to be hyperfinite). The vectors are usually hyperfinite truncations of an orthonormal collection of elements of $\ell^2(\mathbb{R})$. We next show that Theorem C.2 is applicable in that setting.

**Proposition C.4.** *Let $\{u^{(1)}, \ldots, u^{(\gamma)}\}$ be a finite collection of orthonormal vectors in $\ell^2(\mathbb{R})$.*

1. *For any $N > \mathbb{N}$, the collection $\{(u^{(1)})_{(N)}, \ldots, (u^{(\gamma)})_{(N)}\}$ satisfies the Separation Property.*
2. *For any $N > M > \mathbb{N}$, the collection of vectors $\{(u^{(1)})_{(M)}, \ldots, (u^{(\gamma)})_{(M)}\}$ (canonically viewed as vectors in ${}^*\mathbb{R}^N$) and $\{(u^{(1)})_{(N)}, \ldots, (u^{(\gamma)})_{(N)}\}$ satisfy the conditions in Theorem C.2.*

*Proof.* Let $\{u^{(1)}, \ldots, u^{(\gamma)}\}$ be as in the statement of the proposition. Let $N > \mathbb{N}$. By

Lemma B.1, $\{(u^{(1)})_{(N)}, \ldots, (u^{(\gamma)})_{(N)}\}$ is linearly independent. Therefore, we can apply the Gram-Schmidt orthonormalization to obtain the corresponding orthonormal set $\{z^{(1)}, \ldots, z^{(\gamma)}\}$.

Take $i \in \{1, \ldots, \gamma\}$, and let $\mathcal{V}' := \{(u^{(j_1)})_{(N)}, \ldots, (u^{(j_t)})_{(N)}\}$ be a subcollection not containing $(u^{(i)})_{(N)}$. Then we have:

$$\begin{aligned}\left|\left|(u^{(i)})_{(N)} - P_{\mathrm{span}(\mathcal{V}')}((u^{(i)})_{(N)})\right|\right| &= \left|\left|(u^{(i)})_{(N)} - \sum_{\theta=1}^{t}\langle (u^{(i)})_{(N)}, z^{(j_\theta)}\rangle z^{(j_\theta)}\right|\right| \\ &\geq \left|\left|(u^{(i)})_{(N)}\right|\right| - \left|\left|\sum_{\theta=1}^{t}\langle (u^{(i)})_{(N)}, z^{(j_\theta)}\rangle z^{(j_\theta)}\right|\right| \\ &\geq \left|\left|(u^{(i)})_{(N)}\right|\right| - \sum_{\theta=1}^{t}\left|\langle (u^{(i)})_{(N)}, z^{(j_\theta)}\rangle\right| \left|\left|z^{(j_\theta)}\right|\right| \\ &= \left|\left|(u^{(i)})_{(N)}\right|\right| - \sum_{\theta=1}^{t}\left|\langle (u^{(i)})_{(N)}, z^{(j_\theta)}\rangle\right|\end{aligned}$$

The second and third lines follow by triangle inequality, and the fourth line follows from the fact that $\left|\left|z^{(j)}\right|\right| = 1$ for all $j \in \{1, \ldots, \gamma\}$. Thus, to prove 1, it suffices to show the following claim:

**Claim C.5.** *We have* $\langle (u^{(i)})_{(N)}, z^{(j)}\rangle \approx 0$ *for all* $j \in \{1, \ldots, \gamma\}\backslash\{i\}$.

This is a straightforward consequence of the precise formulae for $z^{(i)}$ as per the

Gram-Schmidt orthonormalization procedure (see below):

$$
\begin{aligned}
z^{(1)} &:= \frac{(u^{(1)})_{(N)}}{\big|\big|(u^{(1)})_{(N)}\big|\big|}, \\
z^{(2)} &:= \frac{(u^{(2)})_{(N)} - \langle (u^{(2)})_{(N)}, z^{(1)}\rangle z^{(1)}}{\big|\big|(u^{(2)})_{(N)} - \langle (u^{(2)})_{(N)}, z^{(1)}\rangle z^{(1)}\big|\big|}, \\
z^{(3)} &:= \frac{(u^{(3)})_{(N)} - \langle (u^{(3)})_{(N)}, z^{(1)}\rangle z^{(1)} - \langle (u^{(3)})_{(N)}, z^{(2)}\rangle z^{(2)}}{\big|\big|(u^{(3)})_{(N)} - \langle (u^{(3)})_{(N)}, z^{(1)}\rangle z^{(1)} - \langle (u^{(3)})_{(N)}, z^{(2)}\rangle z^{(2)}\big|\big|}, \\
&\ \vdots \\
z^{(\gamma)} &:= \frac{(u^{(\gamma)})_{(N)} - \langle (u^{(\gamma)})_{(N)}, z^{(1)}\rangle z^{(1)} - \ldots - \langle (u^{(\gamma)})_{(N)}, z^{(\gamma-1)}\rangle z^{(\gamma-1)}}{\big|\big|(u^{(\gamma)})_{(N)} - \langle (u^{(\gamma)})_{(N)}, z^{(1)}\rangle z^{(1)} - \ldots - \langle (u^{(\gamma)})_{(N)}, z^{(\gamma-1)}\rangle z^{(\gamma-1)}\big|\big|}.
\end{aligned}
\tag{C.11}
$$

Indeed, the fact that $\langle u^{(i)}, u^{(j)}\rangle_{\ell^2(\mathbb{R})} = \lim_{n\to\infty} \langle (u^{(i)})_{(n)}, (u^{(j)})_{(n)}\rangle = \delta_{ij}$ implies (by the nonstandard characterization of limits) that $\langle (u^{(i)})_{(N)}, (u^{(j)})_{(N)}\rangle \approx 0$ for $i \neq j$, which proves the claim. This completes the proof of (1).

Now, let $N > M > \mathbb{N}$. By (1), both $\{(u^{(1)})_{(M)}, \ldots, (u^{(\gamma)})_{(M)}\}$ (viewed canonically as vectors in ${}^*\mathbb{R}^N$ and $\{(u^{(1)})_{(N)}, \ldots, (u^{(\gamma)})_{(N)}\}$ satisfy SP. Also, by a similar argument as in the proof of Claim C.5, we have

$$
\begin{aligned}
\big|\big|(u^{(i)})_{(N)}\big|\big| \approx \big|\big|(u^{(i)})_{(M)}\big|\big| \approx 1 &\text{ for all } i \in \{1, \ldots, \gamma\}, \\
\text{and } \big|\big|(u^{(i)})_{(N)} - (u^{(i)})_{(M)}\big|\big| \approx 0 &\text{ for all } i \in \{1, \ldots, \gamma\}.
\end{aligned}
$$

This completes the proof of (2). □

## Appendix D. Concluding the Theorem of Hewitt and Savage from the Theorem of Ressel

In this appendix, we prove that the theorem of Ressel showing Radon presentability of completely regular Hausdorff spaces ([83, Theorem 3, p. 906]) implies the theorem of Hewitt and Savage on the presentability of the Baire sigma algebra of compact Hausdorff spaces ([51, Theorem 7.2, p. 483]). Since we will have occasion to talk about the presentability of Baire sigma algebras and Radon presentability in the same context, it is desirable to reduce the risk of confusion by introducing more precise notation for the relevant sigma algebras.

**Notation D.1.** For a Hausdorff space $S$, let $\mathcal{B}_{\mathrm{a}}(\mathrm{S})$ denote its Baire sigma algebra, the smallest sigma algebra with respect to which all continuous functions $f\colon S \to \mathbb{R}$ are measurable). Let $\mathcal{B}(S)$ denote its Borel sigma algebra, the smallest sigma algebra containing all open subsets of $S$ (it is clear that $\mathcal{B}_{\mathrm{a}}(\mathrm{S}) \subseteq \mathcal{B}(S)$). Let $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ denote the set of all Radon probability measures on $S$, and let $\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})$ denote the set of all Baire probability measures on $S$. Let $\mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ be the smallest sigma algebra on $\mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ that makes all maps of the form $\mu \mapsto \mu(B)$ measurable, where $B \in \mathcal{B}(S)$. Let $\mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$ be the smallest sigma algebra on $\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})$ that makes all maps of the form $\mu \mapsto \mu(A)$ measurable, where $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$.

Note that any compact Hausdorff space is normal (see, for example, Kelley [60, Theorem 9, chapter 5]), and in particular completely regular. The key idea in going from Ressel's result to that of Hewitt–Savage is that on any completely regular Hausdorff space, a tight Baire measure has a unique extension to a Radon measure (see Bogachev [19, Theorem 7.3.3, p. 81, vol. 2]). In particular, since every Baire measure on a $\sigma$-compact space is tight, it follows that every Baire measure on a completely regular $\sigma$-compact Hausdorff

space admits a unique extension to a Radon measure on that space. See Bogachev [19, Corollary 7.3.4, p. 81, vol. 2] for this result. Bogachev also has a formula for this unique extension on [19, p. 78, vol. 2]. We record these facts as a lemma.

**Lemma D.2.** *Let $S$ be a completely regular $\sigma$-compact Hausdorff space. For a subset $A \subseteq S$, let $\tau_A(S)$ denote the collection of those open subsets of $S$ that contain $A$. For every $\mu \in \mathfrak{P}_{Ba}(S)$, there is a unique element $\hat{\mu} \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ such that $\hat{\mu}(A) = \mu(A)$ for all $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$. Furthermore, $\hat{\mu}$ is precisely given by the following formula:*

$$\hat{\mu}(B) = \inf_{U \in \tau_B(S)} \sup_{\substack{A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S}) \\ A \subseteq U}} \mu(A) \text{ for all } B \in \mathcal{B}(S). \tag{D.1}$$

As a consequence, we obtain the following lemma.

**Lemma D.3.** *Let $S$ be a completely regular $\sigma$-compact Hausdorff space. Consider the map $\hat{}\colon \mathfrak{P}_{Ba}(\mathrm{S}) \to \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ defined by $\hat{}(\mu) = \hat{\mu}$ for all $\mu \in \mathfrak{P}_{Ba}(\mathrm{S})$ (where $\hat{\mu}$ is as in (D.1)). Then $\hat{}$ is a bijection.*

*Furthermore, for a set $\mathcal{A} \in \mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S}))$, define $\hat{\mathcal{A}}$ to be its image under $\hat{}$ (thus $\hat{\mathcal{A}} := \{\hat{\mu} : \mu \in \mathcal{A}\}$). Then $\hat{\mathcal{A}} \in \mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ for all $\mathcal{A} \in \mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S}))$.*

*Proof.* If $\mu$ and $\nu$ are distinct elements of $\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})$, then there exists an $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$ such that $\mu(A) \neq \nu(A)$, which implies $\hat{\mu}(A) \neq \hat{\nu}(A)$, so that $\hat{\mu} \neq \hat{\nu}$. Thus $\hat{}$ is an injection. That it is also a surjection follows from the fact that for any $\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, its restriction $\mu\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S})}$ to the Baire sigma algebra is a Baire measure that has a unique Radon extension by Lemma D.2, so that it must be the case that

$$\mu = \widehat{\mu\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S})}} \text{ for all } \mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S})\,. \tag{D.2}$$

Consider the collection $\mathfrak{G}$ of sets $\mathcal{A} \in \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$ for which $\hat{\mathcal{A}}$ is an element of $\mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$, that is,

$$\mathfrak{G} := \{\mathcal{A} \in \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})) : \hat{\mathcal{A}} \in \mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))\}. \tag{D.3}$$

We want to show that $\mathfrak{G}$ equals $\mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$. It is not very difficult to see that for any collection $(A_n)_{n\in\mathbb{N}} \subseteq \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$, we have the following:

$$\widehat{\cup_{n\in\mathbb{N}}\mathcal{A}_n} = \cup_{n\in\mathbb{N}}\hat{\mathcal{A}}_n.$$

Hence, by the fact that $\mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ is a sigma algebra, it follows that $\mathfrak{G}$ is closed under countable unions. Furthermore, if $\mathcal{A} \in \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$, then we have the following (the inclusion from left to right follows from the injectivity of $\hat{\ }$, while the inclusion from right to left follows from the fact that $\hat{\ }$ is a bijection):

$$\widehat{\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}) \setminus \mathcal{A}} = \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) \setminus \hat{\mathcal{A}}. \tag{D.4}$$

This shows that $\mathfrak{G}$ is closed under complements as well. Since $\emptyset \in \mathfrak{G}$, it thus follows that $\mathfrak{G}$ is a sigma algebra. Thus by Dynkin's $\pi$-$\lambda$ theorem, it suffices to show that $\mathfrak{G}$ contains a $\pi$-system (that is, a collection of sets that is closed under finite intersections) that generates $\mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$. A convenient $\pi$-system of that type is the following (that this is a $\pi$-system is trivial, and the fact that the smallest sigma algebra containing it coincides with $\mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$ follows from the fact that any map on $\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})$ of the type $\mu \mapsto \mu(A)$ for some $A \in \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})$ is measurable on the former sigma algebra):

$$\mathfrak{A} := \{\mathfrak{A}^{A_1,\ldots,A_n}_{C_1,\ldots,C_n} : n \in \mathbb{N}, A_1, \ldots, A_n \in \mathcal{B}_{\mathrm{a}}(\mathrm{S}) \text{ and } C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R})\}, \tag{D.5}$$

where for any $n \in \mathbb{N}$, $A_1, \ldots, A_n \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$ and $C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R})$, the set $\mathfrak{A}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n}$ is defined as follows:

$$\mathfrak{A}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n} := \{\mu \in \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}) : \mu(A_1) \in C_1, \ldots, \mu(A_n) \in C_n\}. \tag{D.6}$$

For $n \in \mathbb{N}$, consider the sets $A_1, \ldots, A_n \in \mathcal{B}(S)$ and $C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R})$. Define the collection $\mathfrak{B}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n}$ as follows:

$$\mathfrak{B}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n} := \{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \mu(A_1) \in C_1, \ldots, \mu(A_n) \in C_n\} \in \mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})). \tag{D.7}$$

It thus suffices to show the following claim.

**Claim D.4.** *We have* $\widehat{\mathfrak{A}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n}} = \mathfrak{B}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n}$ *for all* $A_1, \ldots, A_n \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$ *and* $C_1, \ldots, C_n \in \mathcal{B}(\mathbb{R})$*.*

*Proof of Claim D.4.* Note that for any $\mathcal{A}, \mathcal{B} \in \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$, we have the following (the inclusion from left to right is trivial, while the inclusion from right to left follows from the injectivity of the map $\hat{}$):

$$\widehat{\mathcal{A} \cap \mathcal{B}} = \hat{\mathcal{A}} \cap \hat{\mathcal{B}}.$$

Since $\mathfrak{A}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n} = \cap_{i\in[n]} \mathfrak{A}_{C_i}^{A_i}$ and $\mathfrak{B}_{C_1,\ldots,C_n}^{A_1,\ldots,A_n} = \cap_{i\in[n]} \mathfrak{B}_{C_i}^{A_i}$, it suffices to show the following set equality:

$$\widehat{\mathfrak{A}_C^A} = \mathfrak{B}_C^A \text{ for any } C \in \mathcal{B}(\mathbb{R}) \text{ and } A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})\,. \tag{D.8}$$

Toward that end, let $C \in \mathcal{B}(\mathbb{R})$ and $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})$. If $\mu \in \mathfrak{A}_C^A$, then we have $\hat{\mu}(A) = \mu(A) \in C$, so that $\hat{\mu} \in \mathfrak{B}_C^A$. Thus the left side of (D.8) is contained in the right side of (D.8). Conversely, if $\mu \in \mathfrak{B}_C^A$, then $\mu = \widehat{\mu\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S})}}$, where $\mu\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S})} \in \mathfrak{A}_C^A$, completing the proof.

□

As a corollary, we now have a way to define a natural measure on $\mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}))$ corresponding to any measure on $\mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$ in the case when $S$ is completely regular, Hausdorff, and $\sigma$-compact.

**Corollary D.5.** *Let $S$ be a completely regular $\sigma$-compact Hausdorff space. Let $\hat{}: \mathfrak{P}_{Ba}(\mathrm{S}) \to \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$ be as in Lemma D.3. Suppose $\mathscr{P}$ is a probability measure on $\mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}))$. Define a map $\check{\mathscr{P}}: \mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S})) \to [0,1]$ as follows:*

$$\check{\mathscr{P}}(\mathcal{A}) := \mathscr{P}(\hat{\mathcal{A}}) \text{ for all } \mathcal{A} \in \mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S})). \tag{D.9}$$

*Then $\check{\mathscr{P}}$ is a probability measure on $\mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S}))$.*

*Proof.* The fact that $\check{\mathscr{P}}$ is well-defined follows from Lemma D.3. Its countable additivity follows from that of $\mathscr{P}$ and the fact that the map $\hat{}$ is injective. Finally, the fact that $\check{\mathscr{P}}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})) = 1$ follows from the surjectivity of the map $\hat{}$ (as we have $\widehat{\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})} = \mathfrak{P}_{\mathrm{r}}(\mathrm{S})$, whose measure with respect to $\mathscr{P}$ is one). □

We are now able to show that the main result in Hewitt–Savage [51] is a direct consequence of the theorem of Ressel on the Radon presentability of completely regular Hausdorff spaces.

**Theorem D.6** (Hewitt–Savage [51, Theorem 7.2, p. 483])**.** *Suppose all completely regular spaces are Radon presentable as in Definition 5.3.3. Let $S$ be a compact Hausdorff space equipped with its Baire sigma algebra $\mathcal{B}_{\mathrm{a}}(\mathrm{S})$. Suppose $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and let $(X_n)_{n\in\mathbb{N}}$ be a sequence of exchangeable random variables (with respect to the Baire*

*sigma algebra $\mathcal{B}_{\mathrm{a}}(\mathrm{S})$). In other words, suppose the following holds:*

$$\mathbb{P}(X_1 \in A_1, \ldots, X_k \in A_k) = \mathbb{P}(X_{\sigma(1)} \in A_1, \ldots, X_{\sigma(k)} \in A_k)$$
$$\text{for all } k \in \mathbb{N}, \sigma \in S_k, \text{ and } A_1, \ldots, A_k \in \mathcal{B}_{\mathrm{a}}(\mathrm{S})\,. \tag{D.10}$$

*Then there is a unique probability measure $\mathscr{Q}$ on $\mathcal{C}(\mathfrak{P}_{Ba}(\mathrm{S}))$ such that*

$$\mathbb{P}(X_1 \in A_1, \ldots, X_k \in A_k) = \int\limits_{\mathfrak{P}_{Ba}(\mathrm{S})} \mu(A_1) \cdot \ldots \cdot \mu(A_k) d\mathscr{Q}(\mu)$$
$$\text{for all } A_1, \ldots, A_k \in \mathcal{B}_a(S). \tag{D.11}$$

*Proof.* We will only prove the existence of a probability measure $\mathscr{Q}$ on $\mathcal{C}(\mathcal{B}_{\mathrm{a}}(\mathrm{S}))$ satisfying (D.11), with uniqueness following more elementarily from Hewitt–Savage [51, Theorem 9.4, p. 489].

Since $S$ is compact Hausdorff, so is the countable product $S^\infty$ under the product topology (this follows from Tychonoff's theorem). Furthermore, Bogachev [19, Lemma 6.4.2 (iii), p. 14, vol. 2] implies the following:

$$\mathcal{B}_a(S^\infty) = \bigotimes \mathcal{B}_a(S), \tag{D.12}$$

where $\bigotimes \mathcal{B}_a(S)$ denotes the product sigma algebra on $S^\infty$ induced by the Baire sigma algebra $S$ (thus $\bigotimes \mathcal{B}_a(S)$ is the smallest sigma algebra on $S^\infty$ that makes the projection $\pi_i \colon S^\infty \to S$ Baire measurable for each $i \in \mathbb{N}$). Let $\nu \in \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}^\infty)$ be the distribution of the $S^\infty$-valued Baire measurable random variable $(X_n)_{n\in\mathbb{N}}$ (the Baire measurability of this random variable follows from the Baire measurability of the $X_i$ together with (D.12)).

Let $\hat{}\colon \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}^\infty) \to \mathfrak{P}_{\mathrm{r}}(\mathrm{S}^\infty)$ be as in Lemma D.3. Consider $\hat{\nu} \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}^\infty)$. We show in the next claim that the Baire exchangeability of the sequence $(X_n)_{n\in\mathbb{N}}$ implies the ex-

changeability of the measure $\hat{\nu}$. In particular, let $\Omega' := S^\infty$, $\mathcal{F}' := \mathcal{B}(S^\infty)$, and $\mathbb{P}' := \hat{\nu}$. Consider the sequence of Borel measurable $S$-valued random variables $(Y_n)_{n\in\mathbb{N}}$ where, for each $n \in \mathbb{N}$, the map $Y_n \colon \Omega' \to S$ is the projection onto the $n^{\text{th}}$ coordinate. Then we have the following claim:

**Claim D.7.** *The sequence $(Y_n)_{n\in\mathbb{N}}$ is a jointly Radon distributed sequence of exchangeable random variables taking values in a completely regular Hausdorff space.*

*Proof of Claim D.7.* The fact that $(Y_n)_{n\in\mathbb{N}}$ is a jointly Radon distributed sequence is immediate from the construction. Thus we only need to check the exchangeability of the $(Y_n)_{n\in\mathbb{N}}$ as Borel measurable random variables.

To that end, suppose $k \in \mathbb{N}$ and $B \in \mathcal{B}(\mathbb{R}^k)$. Let $\psi \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S^k})$ be the Borel distribution of $(Y_1, \ldots, Y_k)$. That is, $\psi$ is the measure on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ given by the pushforward $\mathbb{P}' \circ (Y_1, \ldots, Y_k)^{-1}$ (which is Radon, being the marginal of a Radon distribution on $S^\infty$). Let $\psi'$ be its restriction to the Baire sigma algebra on $S^k$—that is, $\psi' := \psi\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S^k})}$. Let $\sigma \in S_k$, and let $\psi_\sigma$ be the pushforward $\mathbb{P}' \circ (Y_{\sigma(1)}, \ldots, Y_{\sigma(k)}) \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S^k})$ induced by the permuted random vector $(Y_{\sigma(1)}, \ldots, Y_{\sigma(k)})$, with $\psi'_\sigma := \psi_\sigma\restriction_{\mathcal{B}_{\mathrm{a}}(\mathrm{S^k})}$ being its restriction to the Baire sigma algebra on $S^k$. It suffices to show that $\psi = \psi_\sigma$.

Note that for any $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S}^{\mathrm{k}})$, we have the following chain of equalities:

$$\begin{aligned}
\psi'(A) &= \mathbb{P}'((Y_1, \ldots, Y_k) \in A) \\
&= \hat{\nu}(A) \\
&= \nu(A) \\
&= \mathbb{P}((X_1, \ldots, X_k) \in A) \\
&= \mathbb{P}((X_{\sigma(1)}, \ldots, X_{\sigma(k)}) \in A) \qquad \text{(D.13)} \\
&= \mathbb{P}'((Y_{\sigma(1)}, \ldots, Y_{\sigma(k)}) \in A), \\
&= \psi_\sigma(A) \\
&= \psi'_\sigma(A). \qquad \text{(D.14)}
\end{aligned}$$

In the above, equation (D.13) follows from the Baire-exchangeability of $(X_1, \ldots, X_k)$, while the other lines follow from the fact that $A \in \mathcal{B}_{\mathrm{a}}(\mathrm{S}^{\mathrm{k}})$.

Note that by Lemma D.2, we have $\psi = \hat{\psi}'$ and $\psi_\sigma = \hat{\psi}'_\sigma$. By (D.1), we thus have the following for any $B \in \mathcal{B}(S^k)$ (where we use (D.14) in the third line):

$$\begin{aligned}
\psi(B) &= \hat{\psi}'(B) \\
&= \inf_{U \in \tau_B(S^k)} \sup_{\substack{A \in \mathcal{B}_a(S^k) \\ A \subseteq U}} \psi'(A) \\
&= \inf_{U \in \tau_B(S^k)} \sup_{\substack{A \in \mathcal{B}_a(S^k) \\ A \subseteq U}} \psi'_\sigma(A) \\
&= \hat{\psi}'_\sigma(A) \\
&= \psi_\sigma(B) \text{ for all } B \in \mathbb{R}^k,
\end{aligned}$$

which completes the proof of the claim.

Since completely regular Hausdorff spaces are Radon presentable, we obtain a unique Radon measure $\mathscr{P}$ on $(\mathfrak{P}_{\mathrm{r}}(\mathrm{S}), \mathcal{C}(\mathfrak{P}_{\mathrm{r}}(\mathrm{S})))$ such that the following holds:

$$\mathbb{P}'(Y_1 \in B_1, \ldots, Y_k \in B_k) = \int\limits_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(B_1) \cdot \ldots \cdot \mu(B_k) d\mathscr{P}(\mu)$$

$$\text{for all } B_1, \ldots, B_k \in \mathcal{B}(S). \tag{D.15}$$

Define $\mathscr{Q} := \check{\mathscr{P}} \colon \mathcal{C}(\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}^\infty)) \to [0,1]$ as in Lemma D.5. We claim that $\mathscr{Q}$ satisfies (D.11). Indeed, if $k \in \mathbb{N}$ and $A_1, \ldots, A_k \in \mathcal{B}_a(S)$, then we have:

$$\begin{aligned}
\mathbb{P}(X_1 \in A_1, \ldots, X_k \in A_k) &= \nu(A_1 \times \ldots \times A_k) \\
&= \hat{\nu}(A_1 \times \ldots \times A_k) \\
&= \mathbb{P}'(Y_1 \in A_1, \ldots, Y_k \in A_k) \\
&= \int\limits_{\mathfrak{P}_{\mathrm{r}}(\mathrm{S})} \mu(A_1) \cdot \ldots \cdot \mu(A_k) d\mathscr{P}(\mu) \\
&= \int_{[0,1]} \mathscr{P}(\{\mu \in \mathfrak{P}_{\mathrm{r}}(\mathrm{S}) : \mu(A_1) \cdot \ldots \cdot \mu(A_k) > y\}) d\lambda(y) \\
&= \int_{[0,1]} \mathscr{P}(\widehat{\mathfrak{A}_y}) d\lambda(y),
\end{aligned}$$

where $\mathfrak{A}_y := \{\mu \in \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}) : \mu(A_1) \cdot \ldots \cdot \mu(A_k) > y\}$.

As a consequence, we have the following:

$$\begin{aligned}
\mathbb{P}(X_1 \in A_1, \ldots, X_k \in A_k) &= \int_{[0,1]} \check{\mathscr{P}}(\mathfrak{A}_y) d\lambda(y) \\
&= \int_{[0,1]} \mathscr{Q}(\{\mu \in \mathfrak{P}_{\mathrm{Ba}}(\mathrm{S}) : \mu(A_1) \cdot \ldots \cdot \mu(A_k) > y\}) d\lambda(y) \\
&= \int\limits_{\mathfrak{P}_{\mathrm{Ba}}(\mathrm{S})} \mu(A_1) \cdot \ldots \cdot \mu(A_k) d\mathscr{Q}(\mu),
\end{aligned}$$

which completes the proof. ∎

## Appendix E. A Proof of Theorem 7.3.1 Using Internal Bayes' Theorem

In this appendix, we will carry out an alternative proof of Theorem 7.3.1, which was the key ingredient in our proof of the generalization of de Finetti–Hewitt–Savage theorem. The proof that we will present here is a refinement of the Bayes' theorem-based idea from [5]. We restate Theorem 7.3.1 for convenience.

**Theorem 4.1.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of $S$-valued exchangeable random variables, where $(S, \mathfrak{S})$ is some measurable space. For each $N > \mathbb{N}$ and $\omega \in {}^*\Omega$, define the internal probability measure $\mu_{\omega,N}$ as follows:*

$$\mu_{\omega,N}(B) := \frac{\#\{i \in [N] : X_i(\omega) \in B\}}{n} \text{ for all } B \in {}^*\mathfrak{S}. \tag{E.1}$$

*Then we have:*

$$\begin{aligned} {}^*\mathbb{P}(X_1 \in B_1, \ldots, X_k \in B_k) &\approx {}^*\!\!\int_{{}^*\Omega} \mu_{\omega,N}(B_1)\cdots\mu_{\omega,N}(B_k)d^*\mathbb{P}(\omega) \\ &\text{for all } k \in \mathbb{N} \text{ and } B_1, \ldots, B_k \in {}^*\mathfrak{S}. \end{aligned} \tag{E.2}$$

It turns out that one difficulty in a direct generalization of the method in [5] is that the sets $B_i$ were all either $\{0\}$ or $\{1\}$ in [5], while they may have intersections in (E.2). We get around this difficulty by observing that it suffices to prove (E.2) for tuples $(B_1, \ldots, B_k)$ such that $B_i$ and $B_J$ are either disjoint or equal for all $i, j \in [k]$.

**Definition E.1.** Call a finite tuple $(B_1, \ldots, B_k)$ of sets *disjointified* if for all $i, j \in [k]$, we have $B_i \cap B_j = \emptyset$ or $B_i \cap B_j = B_i = B_j$. In the setting of Theorem 4.1, call an event *disjointified* if it is of the type $\{X_1 \in B_1, \ldots, X_k \in B_k\}$ for some disjointified tuple $(B_1, \ldots, B_k)$.

**Lemma E.2.** *Let $N > \mathbb{N}$. In the setting of Theorem 4.1, suppose that*

$$^*\mathbb{P}(X_1 \in A_1, \ldots, X_k \in A_k) \approx {}^*\!\!\int_{^*\Omega} \mu_{\omega,N}(A_1) \cdots \mu_{\omega,N}(A_k) d^*\mathbb{P}(\omega) \tag{E.3}$$

*for all $k \in \mathbb{N}$ and $A_1, \ldots, A_k \in {}^*\mathfrak{S}$ such that $(A_1, \ldots, A_k)$ is disjointified.*

*Then* (E.2) *holds.*

*Proof.* Suppose (E.3) holds. Let $B_1, \ldots, B_k \in {}^*\mathfrak{S}$ be fixed. We can write the event $\{X_1 \in B_1, \ldots, X_k \in B_k\}$ as a disjoint union of disjointified events. Indeed, for $d \in \{0, 1\}$ and a set $B \subseteq S$, let $B^d$ be equal to $B$ if $d = 1$, and let it be equal to the complement $S \backslash B$ if $d = 0$. For a tuple $a = (a_1, \ldots, a_k) \in \{0, 1\}^k$ of zeros and ones, define the following set:

$$[B_1, \ldots, B_k]^a := \bigcap_{i \in [k]} {B_i}^{a_i}. \tag{E.4}$$

Being a finite intersection of $^*$-measurable sets, the set $[B_1, \ldots, B_k]^a$ is $^*$-measurable for all $a \in \{0, 1\}^k$. For $i \in [k]$, define $\mathfrak{D}_i := \{(a_1, \ldots, a_k) \in \{0, 1\}^k : a_i = 1\}$. For a tuple $\tilde{a} = (\tilde{a}_1, \ldots, \tilde{a}_k) \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k$ of $k$-tuples, we define

$$[B_1, \ldots, B_k]^{\tilde{a}} := \{X_1 \in [B_1, \ldots, B_k]^{\tilde{a}_1}, \ldots, X_k \in [B_1, \ldots, B_k]^{\tilde{a}_k}\}. \tag{E.5}$$

It is clear that the event $[B_1, \ldots, B_k]^{\tilde{a}}$ is disjointified for each $\tilde{a} \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k$, and that

$$[B_1, \ldots, B_k]^{\tilde{a}} \cap [B_1, \ldots, B_k]^{\tilde{b}} = \emptyset \text{ if } \tilde{a}, \tilde{b} \text{ are distinct elements of } \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k.$$

We thus have the following representation as a disjoint union of disjointified events:

$$\{X_1 \in B_1, \ldots, X_k \in B_k\} = \bigsqcup_{\tilde{a} \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k} [B_1, \ldots, B_k]^{\tilde{a}}. \tag{E.6}$$

For any internal probability measure $\mu$ on $({}^*S, {}^*\mathfrak{S})$, its finite additivity yields the following for all $k \in \mathbb{N}$:

$$\mu(B_i) = \mu\left(\bigsqcup_{\tilde{a}_i \in \mathfrak{D}_i} [B_1, \ldots, B_k]^{\tilde{a}_i}\right) = \sum_{\tilde{a}_i \in \mathfrak{D}_i} \mu\left([B_1, \ldots, B_k]^{\tilde{a}_i}\right) \text{ for each } i \in [k]. \tag{E.7}$$

Taking the product of the terms in (E.7) as $i$ varies over $[k]$, and switching the order of $\sum$ and $\prod$ using distributivity of multiplication over addition (which is a legal move since these are finite sums and products), we have the following observation for any internal probability measure $\mu$ on $({}^*S, {}^*\mathfrak{S})$:

$$\prod_{i \in [k]} \mu(B_i) = \sum_{\tilde{a}=(\tilde{a}_1,\ldots,\tilde{a}_k) \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k} \left[\prod_{i \in [k]} \mu\left([B_1, \ldots, B_k]^{\tilde{a}_i}\right)\right] \text{ for all } k \in \mathbb{N}. \tag{E.8}$$

Applying (E.8) to the internal measure $\mu_{\omega,N}$ for each $\omega \in \mathbb{N}$ and then (internal) integrating with respect to ${}^*\mathbb{P}$, we obtain the following by the (internal) linearity of the (internal) expectation:

$$\begin{aligned}
&{}^*\!\!\int_{{}^*\Omega} \left(\prod_{i \in [k]} \mu(B_i)\right) d^*\mathbb{P}(\omega) \\
&= \sum_{\tilde{a}=(\tilde{a}_1,\ldots,\tilde{a}_k) \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k} {}^*\!\!\int_{{}^*\Omega} \left(\prod_{i \in [k]} \mu_{\cdot,N}\left([B_1, \ldots, B_k]^{\tilde{a}_i}\right)\right) d^*\mathbb{P}(\omega) \\
&= \sum_{\tilde{a}=(\tilde{a}_1,\ldots,\tilde{a}_k) \in \mathfrak{D}_1 \times \ldots \times \mathfrak{D}_k} {}^*\mathbb{P}\left(X_1 \in [B_1, \ldots, B_k]^{\tilde{a}_1}, \ldots, X_k \in [B_1, \ldots, B_k]^{\tilde{a}_k}\right),
\end{aligned}$$

where the last line follows from the hypothesis of the theorem. The proof is now completed by (E.5) and (E.6). □

For the rest of the appendix, we fix the following set-up. Let $N > \mathbb{N}$. We have established in Lemma E.2 that it suffices to show (E.3). Toward that end, let $A_1, \ldots, A_k \in {}^*\mathfrak{S}$ be such that the tuple $(A_1, \ldots, A_k)$ is disjointified. For some $n \in \mathbb{N}$, let $C_1, \ldots, C_n$ be

the distinct (disjoint) sets appearing in the tuple $(A_1, \ldots, A_k)$. For each $i \in [n]$, let $C_i$ appear in $(A_1, \ldots, A_k)$ with a frequency $k_i$. Note that this necessarily implies that $k_1 + \ldots + k_n = k$.

For each $i \in [n]$, let $Y_i \colon {}^*\Omega \to [N]$ be defined as follows:

$$Y_i(\omega) := \#\{j \in [N] : X_j(\omega) \in C_i\} = \sum_{j \in [N]} \mathbb{1}_{C_i}(X_j(\omega)) \text{ for all } \omega \in {}^*\Omega. \tag{E.9}$$

Thus $\mu_{\omega,N}(C_i) = \dfrac{Y_i(\omega)}{N}$ for all $\omega \in {}^*\Omega$.

Let $\vec{A}$, $\vec{X}$, and $\vec{Y}$ denote the tuples $(A_1, \ldots, A_k)$, $(X_1, \ldots, X_k)$, and $(Y_1, \ldots, Y_n)$ respectively. The following lemma follows from elementary combinatorial arguments.

**Lemma E.3.** *Suppose that $t_i \in {}^*\mathbb{N}$ are such that $t_i \geq k_i$ for all $i \in [n]$, and such that ${}^*\mathbb{P}(\vec{Y} = (t_1, \ldots, t_n)) > 0$. Then we have:*

$${}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) = \frac{1}{N(N-1)\ldots(N-(k-1))} \cdot \frac{t_1! \ldots t_n!}{(t_1 - k_1)! \ldots (t_n - k_n)!}. \tag{E.10}$$

*Proof.* Let $t_1, \ldots, t_n$ be as in the statement of the lemma. Define the following event:

$$\begin{aligned} E_{t_1,\ldots,t_n} := \{ & X_1, \ldots, X_{t_1} \in C_1; \\ & X_{t_1+1}, \ldots, X_{t_1+t_2} \in C_2; \\ & \ldots; \\ & X_{t_1+\ldots+t_{n-1}+1}, \ldots, X_{t_1+\ldots+t_n} \in C_n; \\ & X_i \in S \backslash C_1 \sqcup \ldots \sqcup C_n \text{ for all other } i \in [N]\}. \end{aligned}$$

By exchangeability and the fact that the $C_i$ are disjoint, we have the following:

$${}^*\mathbb{P}(\vec{Y} = (t_1, \ldots, t_n)) = N_1 {}^*\mathbb{P}\left(E_{t_1,\ldots,t_n}\right), \tag{E.11}$$

$$\text{and } {}^*\mathbb{P}(\vec{X} \in \vec{A} \text{ and } \vec{Y} = (t_1, \ldots, t_n)) = N_2 {}^*\mathbb{P}\left(E_{t_1,\ldots,t_n}\right), \tag{E.12}$$

where

$$N_1 = \text{Number of ways to choose } t_i \text{ spots of the } i^{\text{th}} \text{ kind in } [N] \text{ as } i \text{ varies over } [n]$$
$$= \binom{N}{t_1}\binom{N-t_1}{t_2}\cdot\ldots\cdot\binom{N-t_1-\ldots-t_{n-1}}{t_n}, \tag{E.13}$$

and

$$N_2 = \text{Number of ways to choose } (t_i-k_i) \text{ spots of the } i^{\text{th}} \text{ kind in } [N] \text{ as } i \text{ varies over } [n]$$
$$= \binom{N-k}{t_1-k_1}\binom{N-k-(t_1-k_1)}{t_2-k_2}\cdot\ldots\cdot\binom{N-k-(t_1+\ldots+t_{n-1}-k_1\ldots-k_{n-1})}{t_n-k_n}. \tag{E.14}$$

Since it is given that ${}^*\mathbb{P}(\vec{Y}=(t_1,\ldots,t_n))>0$, we thus have ${}^*\mathbb{P}\left(E_{t_1,\ldots,t_n}\right)>0$ by (E.11). By (E.11), (E.12), (E.13), and (E.14), we therefore obtain (E.10) after simplification. □

**Corollary E.4.** *Suppose that $t_i \in {}^*\mathbb{N}$ such that ${}^*\mathbb{P}(\vec{Y}=(t_1,\ldots,t_n))>0$. Then we have:*

$${}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n)) \approx \left(\frac{t_1}{N}\right)^{k_1}\cdot\ldots\cdot\left(\frac{t_n}{N}\right)^{k_n} \text{ for all } (t_1,\ldots,t_n)\in[N]^n. \tag{E.15}$$

*Proof.* Suppose that the $t_i\in[N]$ are such that ${}^*\mathbb{P}(\vec{Y}=(t_1,\ldots,t_n))>0$. If $t_i\geq k_i$ for all $i\in[n]$. Then by Lemma E.3, we obtain the following:

$$\frac{{}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n))}{\left(\frac{t_1}{N}\right)^{k_1}\cdot\ldots\cdot\left(\frac{t_n}{N}\right)^{k_n}} = \frac{1}{1-\frac{1}{N}}\cdots\frac{1}{1-\frac{k-1}{N}}\cdot\prod_{i\in[n]}\left(\prod_{j\in[k_i-1]}\left(1-\frac{j}{t_i}\right)\right) \tag{E.16}$$
$$< \frac{1}{1-\frac{1}{N}}\cdots\frac{1}{1-\frac{k-1}{N}} \approx 1. \tag{E.17}$$

Note that if $t_i > \mathbb{N}$ for all $i\in\mathbb{N}$, then both $\frac{1}{1-\frac{1}{N}}\cdots\frac{1}{1-\frac{k-1}{N}} \approx 1$ and $\prod_{i\in[n]}\left(\prod_{j\in[k_i-1]}\left(1-\frac{j}{t_i}\right)\right)\approx 1$, so that (E.16) implies that

$$\frac{{}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n))}{\left(\frac{t_1}{N}\right)^{k_1}\cdot\ldots\cdot\left(\frac{t_n}{N}\right)^{k_n}} \approx 1 \text{ if } t_1,\ldots,t_n>\mathbb{N}, \tag{E.18}$$

which, in particular, implies (E.15) in this case.

Now, if $t_j$ is in $\mathbb{N}$ for some $j \in [n]$ but such that $t_i \geq k$ for all $i \in [n]$ and ${}^*\mathbb{P}(\vec{Y} = (t_1, \ldots, t_n)) > 0$, then the inequality in (E.17) implies that

$$ {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) < 2\left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} < 2\left(\frac{t_j}{N}\right)^{k_j} \approx 0, $$

so that

$$ {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \approx 0 \approx \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n}, $$

proving (E.15) in that case as well.

Finally, if $t_i < k_i$ for any $i \in [n]$, then ${}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) = 0$, while $\left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} \approx 0$ in that case as well. This completes the proof. □

We record (E.18) in the proof of Corollary E.4 as its own result.

**Corollary E.5.** *Suppose that $t_i > \mathbb{N}$ such that ${}^*\mathbb{P}(\vec{Y} = (t_1, \ldots, t_n)) > 0$. Then we have the following approximate equality:*

$$ \frac{{}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n))}{\left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n}} \approx 1 \text{ if } t_1, \ldots, t_n > \mathbb{N}. $$

By (E.17) and underflow applied to Corollary E.5, we obtain the following.

**Corollary E.6.** *Given $\epsilon \in \mathbb{R}_{>0}$, there is an $m_\epsilon$ satisfying the following.*

$$ 1 - \epsilon < \frac{{}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n))}{\left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n}} < 1 + \epsilon $$

*if $t_1, \ldots, t_n > m_\epsilon$ are such that ${}^*\mathbb{P}(\vec{Y} = (t_1, \ldots, t_n)) > 0$.*

The proof of Corollary E.4 also leads to the following observation.

**Corollary E.7.** *For each $m \in {}^*\mathbb{N}$, define the set*

$$L_m := \{(t_1, \ldots, t_n) \in [N]^n : \text{ there is } j \in [n] \text{ such that } t_j \leq m\}. \tag{E.19}$$

*Then, we have the following for all $m \in \mathbb{N}$:*

$$\begin{aligned}
0 &\approx \sum_{(t_1,\ldots,t_n)\in L_m} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right) \\
&\approx \sum_{(t_1,\ldots,t_n)\in L_m} \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right).
\end{aligned}$$

*Proof.* Let $m \in \mathbb{N}$ and $L_m$ be as in the statement of the corollary. Noting that the event $\left\{\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right\}$ is the same as the event $\{\vec{Y} = (t_1, \ldots, t_n)\}$, we obtain the following from (E.17) (we also use the fact that if $t_i < k_i$ for any $i \in [n]$, then ${}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) = 0$):

$$\begin{aligned}
&\sum_{(t_1,\ldots,t_n)\in L_m} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}({}^*C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}({}^*C_n) = \frac{t_n}{N}\right) \\
&\leq 2 \sum_{(t_1,\ldots,t_n)\in L_m} \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}({}^*C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}({}^*C_n) = \frac{t_n}{N}\right) \\
&\leq 2 \sum_{j\in[n]} \left( \sum_{r\in[m]} \left[ \sum_{\substack{(t_1,\ldots,t_n)\in[N]^n \\ t_j = r}} \left(\frac{t_j}{N}\right)^{k_j} {}^*\mathbb{P}\left(\mu_{\cdot,N}({}^*C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}({}^*C_n) = \frac{t_n}{N}\right)\right]\right) \\
&\leq 2 \sum_{j\in[n]} \left( \left[ \sum_{\substack{(t_1,\ldots,t_n)\in[N]^n \\ t_j \leq m}} \frac{m}{N} {}^*\mathbb{P}\left(\mu_{\cdot,N}({}^*C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}({}^*C_n) = \frac{t_n}{N}\right)\right]\right) \\
&= \frac{2m}{N} \sum_{j\in[n]} {}^*\mathbb{P}\left(\mu_{\cdot,N}({}^*C_j) \leq \frac{m}{N}\right) \\
&\leq \frac{2mn}{N} \\
&\approx 0,
\end{aligned}$$

completing the proof. $\square$

We now have all the ingredients for our proof of Theorem 4.1.

*Proof of Theorem 4.1.* Conditioning on the various possible values of $Y_i$ as $i$ varies in $[n]$, and noting that the event $\left\{\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right\}$ is the same as the event $\{\vec{Y} = (t_1, \ldots, t_n)\}$, we obtain:

$$\begin{aligned}
&{}^*\mathbb{P}((X_1, \ldots, X_k) \in \vec{A}) \\
&= \sum_{(t_1,\ldots,t_n)\in[N]^n} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right)
\end{aligned} \tag{E.20}$$

Now, by the definition of expected values, we have the following equality:

$$\begin{aligned}
&{}^*\!\!\int_{{}^*\Omega} \mu_{\omega,N}(A_1) \cdots \mu_{\omega,N}(A_k) d^*\mathbb{P}(\omega) \\
&= \sum_{(t_1,\ldots,t_n)\in[N]^n} \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right).
\end{aligned} \tag{E.21}$$

Let $\epsilon \in \mathbb{R}_{>0}$ and let $m_\epsilon \in \mathbb{N}$ be as in Corollary E.6. By that corollary, we obtain:

$$\begin{aligned}
&\sum_{(t_1,\ldots,t_n)\in[N]^n} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right) \\
&> \sum_{(t_1,\ldots,t_n)\in L_{m_\epsilon}} {}^*\mathbb{P}(\vec{X} \in \vec{A} | \vec{Y} = (t_1, \ldots, t_n)) \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right) \\
&\quad + (1-\epsilon) \sum_{\substack{(t_1,\ldots,t_n)\in[N]^n \\ t_1,\ldots,t_n > m_\epsilon}} \left(\frac{t_1}{N}\right)^{k_1} \cdot \ldots \cdot \left(\frac{t_n}{N}\right)^{k_n} \cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1) = \frac{t_1}{N}, \ldots, \mu_{\cdot,N}(C_n) = \frac{t_n}{N}\right).
\end{aligned}$$

By taking standard parts and using Corollary E.7, the above yields the following

inequality:

$$\begin{aligned}
&\mathbf{st}\left[\sum_{(t_1,\ldots,t_n)\in[N]^n} {}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n))\cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1)=\frac{t_1}{N},\ldots,\mu_{\cdot,N}(C_n)=\frac{t_n}{N}\right)\right]\\
\geq&(1-\epsilon)\,\mathbf{st}\left[\sum_{(t_1,\ldots,t_n)\in[N]^n}\left(\frac{t_1}{N}\right)^{k_1}\cdot\ldots\cdot\left(\frac{t_n}{N}\right)^{k_n}\cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1)=\frac{t_1}{N},\ldots,\mu_{\cdot,N}(C_n)=\frac{t_n}{N}\right)\right].
\end{aligned}$$

Since $\epsilon\in\mathbb{R}_{>0}$ is arbitrary, we thus obtain:

$$\begin{aligned}
&\mathbf{st}\left[\sum_{(t_1,\ldots,t_n)\in[N]^n} {}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n))\cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1)=\frac{t_1}{N},\ldots,\mu_{\cdot,N}(C_n)=\frac{t_n}{N}\right)\right]\\
\geq&\,\mathbf{st}\left[\sum_{(t_1,\ldots,t_n)\in[N]^n}\left(\frac{t_1}{N}\right)^{k_1}\cdot\ldots\cdot\left(\frac{t_n}{N}\right)^{k_n}\cdot {}^*\mathbb{P}\left(\mu_{\cdot,N}(C_1)=\frac{t_1}{N},\ldots,\mu_{\cdot,N}(C_n)=\frac{t_n}{N}\right)\right].
\end{aligned} \tag{E.22}$$

But the reverse inequality to (E.22) is also true because of (E.17) and the fact that ${}^*\mathbb{P}(\vec{X}\in\vec{A}|\vec{Y}=(t_1,\ldots,t_n))=0$ if $t_i<k_i$ for any $i\in[n]$. This completes the proof by (E.20) and (E.21). □

## Appendix F. Permissions

The work in Chapter 2 is adopted from the author's published paper [4], which was published under a Creative Commons Attribution 3.0 License, a copy of which has been attached. The work in Chapter 4 (and some part of Chapter 1) is adopted from the author's published paper [5], whose author agreement form has been attached. These and other works being used in the dissertation (namely [4, 2, 5, 3]) are also available on arXiv under the arXiv.org perpetual, non-exclusive license available at `https://arxiv.org/licenses/nonexclusive-distrib/1.0/license.html`. All of these publishers allow using these works in a dissertation with proper citation.

# Creative Commons Legal Code

**Attribution 3.0 Unported**

*CREATIVE COMMONS CORPORATION IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS LICENSE DOES NOT CREATE AN ATTORNEY-CLIENT RELATIONSHIP. CREATIVE COMMONS PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. CREATIVE COMMONS MAKES NO WARRANTIES REGARDING THE INFORMATION PROVIDED, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM ITS USE.*

## *License*

THE WORK (AS DEFINED BELOW) IS PROVIDED UNDER THE TERMS OF THIS CREATIVE COMMONS PUBLIC LICENSE ("CCPL" OR "LICENSE"). THE WORK IS PROTECTED BY COPYRIGHT AND/OR OTHER APPLICABLE LAW. ANY USE OF THE WORK OTHER THAN AS AUTHORIZED UNDER THIS LICENSE OR COPYRIGHT LAW IS PROHIBITED.

BY EXERCISING ANY RIGHTS TO THE WORK PROVIDED HERE, YOU ACCEPT AND AGREE TO BE BOUND BY THE TERMS OF THIS LICENSE. TO THE EXTENT THIS LICENSE MAY BE CONSIDERED TO BE A CONTRACT, THE LICENSOR GRANTS YOU THE RIGHTS CONTAINED HERE IN CONSIDERATION OF YOUR ACCEPTANCE OF SUCH TERMS AND CONDITIONS.

**1. Definitions**

a. **"Adaptation"** means a work based upon the Work, or upon the Work and other pre-existing works, such as a translation, adaptation, derivative work, arrangement of music or other alterations of a literary or artistic work, or phonogram or performance and includes cinematographic adaptations or any other form in which the Work may be recast, transformed, or adapted including in any form recognizably derived from the original, except that a work that constitutes a Collection will not be considered an Adaptation for the purpose of this License. For the avoidance of doubt, where the Work is a musical work, performance or phonogram, the synchronization of the Work in timed-relation with a moving image ("synching") will be considered an Adaptation for the purpose of this License.
b. **"Collection"** means a collection of literary or artistic works, such as encyclopedias and anthologies, or performances, phonograms or broadcasts, or other works or subject matter other than works listed in Section 1(f) below, which, by reason of the selection and arrangement of their contents, constitute intellectual creations, in which the Work is included in its entirety in unmodified form along with one or more other contributions, each constituting separate and independent works in themselves, which together are assembled into a collective whole. A work that constitutes a Collection will not be considered an Adaptation (as defined above) for the purposes of this License.
c. **"Distribute"** means to make available to the public the original and copies of the Work or Adaptation, as appropriate, through sale or other transfer of ownership.
d. **"Licensor"** means the individual, individuals, entity or entities that offer(s) the Work under the terms of this License.
e. **"Original Author"** means, in the case of a literary or artistic work, the individual, individuals, entity or entities who created the Work or if no individual or entity can be identified, the publisher; and in addition (i) in the case of a performance the actors, singers, musicians, dancers, and other persons who act, sing, deliver, declaim, play in, interpret or otherwise perform literary or artistic works or expressions of folklore; (ii) in the case of a phonogram the producer being the person or legal entity who first fixes the sounds of a performance or other sounds; and, (iii) in the case of broadcasts, the organization that transmits the broadcast.
f. **"Work"** means the literary and/or artistic work offered under the terms of this License including without limitation any production in the literary, scientific and artistic domain, whatever may be the

mode or form of its expression including digital form, such as a book, pamphlet and other writing; a lecture, address, sermon or other work of the same nature; a dramatic or dramatico-musical work; a choreographic work or entertainment in dumb show; a musical composition with or without words; a cinematographic work to which are assimilated works expressed by a process analogous to cinematography; a work of drawing, painting, architecture, sculpture, engraving or lithography; a photographic work to which are assimilated works expressed by a process analogous to photography; a work of applied art; an illustration, map, plan, sketch or three-dimensional work relative to geography, topography, architecture or science; a performance; a broadcast; a phonogram; a compilation of data to the extent it is protected as a copyrightable work; or a work performed by a variety or circus performer to the extent it is not otherwise considered a literary or artistic work.

g. **"You"** means an individual or entity exercising rights under this License who has not previously violated the terms of this License with respect to the Work, or who has received express permission from the Licensor to exercise rights under this License despite a previous violation.
h. **"Publicly Perform"** means to perform public recitations of the Work and to communicate to the public those public recitations, by any means or process, including by wire or wireless means or public digital performances; to make available to the public Works in such a way that members of the public may access these Works from a place and at a place individually chosen by them; to perform the Work to the public by any means or process and the communication to the public of the performances of the Work, including by public digital performance; to broadcast and rebroadcast the Work by any means including signs, sounds or images.
i. **"Reproduce"** means to make copies of the Work by any means including without limitation by sound or visual recordings and the right of fixation and reproducing fixations of the Work, including storage of a protected performance or phonogram in digital form or other electronic medium.

**2. Fair Dealing Rights.** Nothing in this License is intended to reduce, limit, or restrict any uses free from copyright or rights arising from limitations or exceptions that are provided for in connection with the copyright protection under copyright law or other applicable laws.

**3. License Grant.** Subject to the terms and conditions of this License, Licensor hereby grants You a worldwide, royalty-free, non-exclusive, perpetual (for the duration of the applicable copyright) license to exercise the rights in the Work as stated below:

a. to Reproduce the Work, to incorporate the Work into one or more Collections, and to Reproduce the Work as incorporated in the Collections;
b. to create and Reproduce Adaptations provided that any such Adaptation, including any translation in any medium, takes reasonable steps to clearly label, demarcate or otherwise identify that changes were made to the original Work. For example, a translation could be marked "The original work was translated from English to Spanish," or a modification could indicate "The original work has been modified.";
c. to Distribute and Publicly Perform the Work including as incorporated in Collections; and,
d. to Distribute and Publicly Perform Adaptations.

e. For the avoidance of doubt:

    i. **Non-waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme cannot be waived, the Licensor reserves the exclusive right to collect such royalties for any exercise by You of the rights granted under this License;
    ii. **Waivable Compulsory License Schemes**. In those jurisdictions in which the right to collect royalties through any statutory or compulsory licensing scheme can be waived, the Licensor waives the exclusive right to collect such royalties for any exercise by You of the rights granted under this License; and,
    iii. **Voluntary License Schemes**. The Licensor waives the right to collect royalties, whether individually or, in the event that the Licensor is a member of a collecting society that administers voluntary licensing schemes, via that society, from any exercise by You of the rights granted under this License.

The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. Subject to Section 8(f), all rights not expressly granted by Licensor are hereby reserved.

**4. Restrictions.** The license granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a. You may Distribute or Publicly Perform the Work only under the terms of this License. You must include a copy of, or the Uniform Resource Identifier (URI) for, this License with every copy of the Work You Distribute or Publicly Perform. You may not offer or impose any terms on the Work that restrict the terms of this License or the ability of the recipient of the Work to exercise the rights granted to that recipient under the terms of the License. You may not sublicense the Work. You must keep intact all notices that refer to this License and to the disclaimer of warranties with every copy of the Work You Distribute or Publicly Perform. When You Distribute or Publicly Perform the Work, You may not impose any effective technological measures on the Work that restrict the ability of a recipient of the Work from You to exercise the rights granted to that recipient under the terms of the License. This Section 4(a) applies to the Work as incorporated in a Collection, but this does not require the Collection apart from the Work itself to be made subject to the terms of this License. If You create a Collection, upon notice from any Licensor You must, to the extent practicable, remove from the Collection any credit as required by Section 4(b), as requested. If You create an Adaptation, upon notice from any Licensor You must, to the extent practicable, remove from the Adaptation any credit as required by Section 4(b), as requested.
b. If You Distribute, or Publicly Perform the Work or any Adaptations or Collections, You must, unless a request has been made pursuant to Section 4(a), keep intact all copyright notices for the Work and provide, reasonable to the medium or means You are utilizing: (i) the name of the Original Author (or pseudonym, if applicable) if supplied, and/or if the Original Author and/or Licensor designate another party or parties (e.g., a sponsor institute, publishing entity, journal) for attribution ("Attribution Parties") in Licensor's copyright notice, terms of service or by other reasonable means, the name of such party or parties; (ii) the title of the Work if supplied; (iii) to the extent reasonably practicable, the URI, if any, that Licensor specifies to be associated with the Work, unless such URI does not refer to the copyright notice or licensing information for the Work; and (iv) , consistent with Section 3(b), in the case of an Adaptation, a credit identifying the use of the Work in the Adaptation (e.g., "French translation of the Work by Original Author," or "Screenplay based on original Work by Original Author"). The credit required by this Section 4 (b) may be implemented in any reasonable manner; provided, however, that in the case of a Adaptation or Collection, at a minimum such credit will appear, if a credit for all contributing authors of the Adaptation or Collection appears, then as part of these credits and in a manner at least as prominent as the credits for the other contributing authors. For the avoidance of doubt, You may only use the credit required by this Section for the purpose of attribution in the manner set out above and, by exercising Your rights under this License, You may not implicitly or explicitly assert or imply any connection with, sponsorship or endorsement by the Original Author, Licensor and/or Attribution Parties, as appropriate, of You or Your use of the Work, without the separate, express prior written permission of the Original Author, Licensor and/or Attribution Parties.
c. Except as otherwise agreed in writing by the Licensor or as may be otherwise permitted by applicable law, if You Reproduce, Distribute or Publicly Perform the Work either by itself or as part of any Adaptations or Collections, You must not distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation. Licensor agrees that in those jurisdictions (e.g. Japan), in which any exercise of the right granted in Section 3(b) of this License (the right to make Adaptations) would be deemed to be a distortion, mutilation, modification or other derogatory action prejudicial to the Original Author's honor and reputation, the Licensor will waive or not assert, as appropriate, this Section, to the fullest extent permitted by the applicable national law, to enable You to reasonably exercise Your right under Section 3(b) of this License (right to make Adaptations) but not otherwise.

**5. Representations, Warranties and Disclaimer**

UNLESS OTHERWISE MUTUALLY AGREED TO BY THE PARTIES IN WRITING, LICENSOR OFFERS THE WORK AS-IS AND MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND CONCERNING THE WORK, EXPRESS, IMPLIED, STATUTORY OR OTHERWISE, INCLUDING, WITHOUT LIMITATION, WARRANTIES OF TITLE, MERCHANTIBILITY, FITNESS FOR A PARTICULAR PURPOSE, NONINFRINGEMENT, OR THE ABSENCE OF LATENT OR OTHER DEFECTS, ACCURACY, OR THE PRESENCE OF ABSENCE OF ERRORS, WHETHER OR NOT DISCOVERABLE. SOME JURISDICTIONS DO NOT ALLOW THE EXCLUSION OF IMPLIED WARRANTIES, SO SUCH EXCLUSION MAY NOT APPLY TO YOU.

**6. Limitation on Liability.** EXCEPT TO THE EXTENT REQUIRED BY APPLICABLE LAW, IN NO EVENT WILL LICENSOR BE LIABLE TO YOU ON ANY LEGAL THEORY FOR ANY SPECIAL, INCIDENTAL, CONSEQUENTIAL, PUNITIVE OR EXEMPLARY DAMAGES ARISING OUT OF THIS LICENSE OR THE USE OF THE WORK, EVEN IF LICENSOR HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

**7. Termination**

a. This License and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this License. Individuals or entities who have received Adaptations or Collections from You under this License, however, will not have their licenses terminated provided such individuals or entities remain in full compliance with those licenses. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this License.
b. Subject to the above terms and conditions, the license granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under different license terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this License (or any other license that has been, or is required to be, granted under the terms of this License), and this License will continue in full force and effect unless terminated as stated above.

**8. Miscellaneous**

a. Each time You Distribute or Publicly Perform the Work or a Collection, the Licensor offers to the recipient a license to the Work on the same terms and conditions as the license granted to You under this License.
b. Each time You Distribute or Publicly Perform an Adaptation, Licensor offers to the recipient a license to the original Work on the same terms and conditions as the license granted to You under this License.
c. If any provision of this License is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this License, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.
d. No term or provision of this License shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.
e. This License constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This License may not be modified without the mutual written agreement of the Licensor and You.
f. The rights granted under, and the subject matter referenced, in this License were drafted utilizing the terminology of the Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979), the Rome Convention of 1961, the WIPO Copyright Treaty of 1996, the WIPO Performances and Phonograms Treaty of 1996 and the Universal Copyright Convention (as revised on July 24, 1971). These rights and subject matter take effect in the relevant jurisdiction in which the License terms are sought to be enforced according to the corresponding provisions of the implementation of those treaty provisions in the applicable national law. If the standard suite of rights granted under applicable copyright law includes additional rights not granted under this License, such additional rights are deemed to be included in the License; this License is not intended to restrict the license of any rights under applicable law.

## *Creative Commons Notice*

*Creative Commons is not a party to this License, and makes no warranty whatsoever in connection with the Work. Creative Commons will not be liable to You or any party on any legal theory for any damages whatsoever, including without limitation any general, special, incidental or consequential damages arising in connection to this license. Notwithstanding the foregoing two (2) sentences, if Creative Commons has expressly identified itself as the Licensor hereunder, it shall have all rights and obligations of Licensor.*

## Journal of Stochastic Analysis Author Agreement

**Return signed form to: kuo@math.lsu.edu**

**Journal of Stochastic Analysis**

Author's Name: IRFAN ALAM
Author's Address: DEPARTMENT OF MATHEMATICS, LSU, BATON ROUGE-70803
Author's Phone: +1 (225) 239 0855
Author's Email Address: IRFANALAMISI @GMAIL.COM
Work (Title of Manuscript): A NONSTANDARD PROOF OF DE FINETTI'S THEOREM FOR BERNOULLI RANDOM VARIABLES.

The Editors and the Author agree as follows:

1. In consideration of the Editors' agreement to publish the Work, Author hereby grants and assigns to Editors and to Louisiana State University Libraries the non-exclusive right to print, publish, reproduce, or distribute the Work throughout the world in all means of expression by any method now known or hereafter developed, including electronic formats, and to market or sell the Work or any part of it as it sees fit. Author further grants the Editors the right to use Author's name in association with the Work in published form and in advertising and promotional materials. Copyright of the Work remains in Author's name. The Author cannot terminate this license.

2. Author agrees not to publish the Work in print form prior to publication of the Work by the Editors. The author agrees to cite the original Journal of Stochastic Analysis publication when publishing the Work elsewhere.

3. The Author represents and warrants that the Work: (a) is the Author's original work and that Author has full power to enter into this Agreement; (b) does not infringe the copyright or property of another; (c) contains no material which is obscene, libelous, defamatory or previously published, in whole or in part. Author shall indemnify and hold Editors harmless against loss of expenses arising from breach of any such warranties.

Author signature: Alam

Date: 12/14/2020

# Bibliography


[1] L. Accardi and Y. G. Lu, *A continuous version of de Finetti's theorem*, Ann. Probab. **21** (1993), no. 3, 1478–1493. MR 1235425

[2] Irfan Alam, *Limiting spherical integrals of bounded continuous functions*, arXiv e-prints (2019), arXiv:1909.11075, recommended for publication in Journal of Theoretical Probability (2021).

[3] ———, *Generalizing the de Finetti–Hewitt–Savage theorem*, arXiv e-prints (2020), arXiv:2008.08754.

[4] Irfan Alam, *Limiting probability measures*, J. Log. Anal. **12** (2020), Paper No. 1, 35. MR 4195193

[5] ———, *A nonstandard proof of de Finetti's theorem for Bernoulli random variables*, J. Stoch. Anal. **1** (2020), no. 4, Art. 15, 18. MR 4188596

[6] Sergio Albeverio, Raphael Høegh-Krohn, Jens Erik Fenstad, and Tom Lindstrøm, *Nonstandard methods in stochastic analysis and mathematical physics*, Pure and Applied Mathematics, vol. 122, Academic Press, Inc., Orlando, FL, 1986. MR 859372

[7] J. M. Aldaz, *A characterization of universal Loeb measurability for completely regular Hausdorff spaces*, Canad. J. Math. **44** (1992), no. 4, 673–690. MR 1178563

[8] David J. Aldous, *Representations for partially exchangeable arrays of random variables*, J. Multivariate Anal. **11** (1981), no. 4, 581–598. MR 637937

[9] ———, *Exchangeability and related topics*, École d'été de probabilités de Saint-Flour, XIII—1983, Lecture Notes in Math., vol. 1117, Springer, Berlin, 1985, pp. 1–198. MR 883646

[10] A. D. Alexandroff, *Additive set-functions in abstract spaces*, Rec. Math. [Mat. Sbornik] N. S. **8 (50)** (1940), 307–348. MR 0004078

[11] Robert M. Anderson, *A non-standard representation for Brownian motion and Itô integration*, Israel J. Math. **25** (1976), no. 1-2, 15–46. MR 464380

[12] ———, *STAR-FINITE PROBABILITY THEORY*, ProQuest LLC, Ann Arbor, MI, 1977, Thesis (Ph.D.)–Yale University. MR 2627217

[13] ———, *Star-finite representations of measure spaces*, Trans. Amer. Math. Soc. **271** (1982), no. 2, 667–687. MR 654856

[14] ______, *Strong core theorems with nonconvex preferences*, Econometrica **53** (1985), no. 6, 1283–1294. MR 809911

[15] Robert M. Anderson and Salim Rashid, *A nonstandard characterization of weak convergence*, Proc. Amer. Math. Soc. **69** (1978), no. 2, 327–332. MR 480925

[16] Tim Austin, *On exchangeable random variables and the statistics of large graphs and hypergraphs*, Probab. Surv. **5** (2008), 80–145. MR 2426176

[17] Vieri Benci, Marco Forti, and Mauro Di Nasso, *The eightfold path to nonstandard analysis*, Nonstandard methods and applications in mathematics, Lect. Notes Log., vol. 25, Assoc. Symbol. Logic, La Jolla, CA, 2006, pp. 3–44. MR 2209073

[18] J. H. Blau, *The space of measures on a given set*, Fund. Math. **38** (1951), 23–34. MR 47117

[19] Vladimir I. Bogachev, *Measure theory. Vol. I, II*, Springer-Verlag, Berlin, 2007. MR 2267655

[20] ______, *Weak convergence of measures*, Mathematical Surveys and Monographs, vol. 234, American Mathematical Society, Providence, RI, 2018. MR 3837546

[21] Ludwig Boltzmann, *Studien über das Gleichgewicht der lebenden Kraft*, Wissenschafiliche Abhandlungen **1** (1868), 49–96.

[22] Hans Bühlmann, *Austauschbare stochastische Variablen und ihre Grenzwertsaetze*, Univ. California Publ. Statist. **3** (1960), 1–35 (1960). MR 117779

[23] François Caron and Emily B. Fox, *Sparse graphs using exchangeable random measures*, J. R. Stat. Soc. Ser. B. Stat. Methodol. **79** (2017), no. 5, 1295–1366. MR 3731666

[24] C. C. Chang and H. J. Keisler, *Model theory*, third ed., Studies in Logic and the Foundations of Mathematics, vol. 73, North-Holland Publishing Co., Amsterdam, 1990. MR 1059055

[25] Nigel Cutland and Siu-Ah Ng, *The Wiener sphere and Wiener measure*, Ann. Probab. **21** (1993), no. 1, 1–13. MR 1207212

[26] Nigel J. Cutland, *Nonstandard real analysis*, Nonstandard analysis (Edinburgh, 1996), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 493, Kluwer Acad. Publ., Dordrecht, 1997, pp. 51–76. MR 1603229

[27] D. Dacunha-Castelle, *A survey on exchangeable random variables in normed spaces*, Exchangeability in probability and statistics (Rome, 1981), North-Holland,

Amsterdam-New York, 1982, pp. 47–60. MR 675964

[28] Bruno de Finetti, *Funzione caratteristica di un fenomeno aleatorio*, Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928, 1929, pp. 179–190.

[29] ______, *La prévision : ses lois logiques, ses sources subjectives*, Ann. Inst. H. Poincaré **7** (1937), no. 1, 1–68. MR 1508036

[30] Claude Dellacherie and Paul-André Meyer, *Probabilities and potential*, North-Holland Mathematics Studies, vol. 29, North-Holland Publishing Co., Amsterdam-New York; North-Holland Publishing Co., Amsterdam-New York, 1978. MR 521810

[31] P. Diaconis and D. Freedman, *de Finetti's theorem for Markov chains*, Ann. Probab. **8** (1980), no. 1, 115–130. MR 556418

[32] ______, *Finite exchangeable sequences*, Ann. Probab. **8** (1980), no. 4, 745–764. MR 577313

[33] Persi Diaconis and Svante Janson, *Graph limits and exchangeable random graphs*, Rend. Mat. Appl. (7) **28** (2008), no. 1, 33–61. MR 2463439

[34] Lester E. Dubins, *Some exchangeable probabilities are singular with respect to all presentable probabilities*, Z. Wahrsch. Verw. Gebiete **64** (1983), no. 1, 1–5. MR 710644

[35] Lester E. Dubins and David A. Freedman, *Exchangeable processes need not be mixtures of independent, identically distributed random variables*, Z. Wahrsch. Verw. Gebiete **48** (1979), no. 2, 115–132. MR 534840

[36] E. B. Dynkin, *Classes of equivalent random quantities*, Uspehi Matem. Nauk (N.S.) **8** (1953), no. 2(54), 125–130. MR 0055601

[37] William Feller, *An introduction to probability theory and its applications. Vol. II*, Second edition, John Wiley & Sons, Inc., New York-London-Sydney, 1971. MR 0270403

[38] Thomas S. Ferguson, *A Bayesian analysis of some nonparametric problems*, Ann. Statist. **1** (1973), 209–230. MR 350949

[39] Xavier Fernique, *Processus linéaires, processus généralisés*, Ann. Inst. Fourier (Grenoble) **17** (1967), no. fasc., fasc. 1, 1–92. MR 221576

[40] Gerald B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999. MR 1681462

[41] David A. Freedman, *Invariants under mixing which generalize de Finetti's theorem: Continuous time parameter*, Ann. Math. Statist. **34** (1963), 1194–1216. MR 189111

[42] D. H. Fremlin, *Measure theory. Vol. 4*, Torres Fremlin, Colchester, 2006, Topological measure spaces. Part I, II, Corrected second printing of the 2003 original. MR 2462372

[43] P. M. Gartside and E. A. Reznichenko, *Near metric properties of function spaces*, Fund. Math. **164** (2000), no. 2, 97–114. MR 1784703

[44] Paul Gartside, *e-11 - generalized metric spaces, part i*, Encyclopedia of General Topology (Klaas Pieter Hart, Jun iti Nagata, Jerry E. Vaughan, Vitaly V. Fedorchuk, Gary Gruenhage, Heikki J.K. Junnila, Krystyna M. Kuperberg, Jan van Mill, Tsugunori Nogura, Haruto Ohta, Akihiro Okuyama, Roman Pol, and Stephen Watson, eds.), Elsevier, Amsterdam, 2003, pp. 273 – 275.

[45] Marie Gaudard and Donald Hadwin, *Sigma-algebras on spaces of probability measures*, Scand. J. Statist. **16** (1989), no. 2, 169–175. MR 1028976

[46] C. Ward Henson, *Analytic sets, Baire sets and the standard part map*, Canadian J. Math. **31** (1979), no. 3, 663–672. MR 536371

[47] Horst Herrlich, *Wann sind alle stetigen Abbildungen in $Y$ konstant?*, Math. Z. **90** (1965), 152–154. MR 185565

[48] Alexander Hertle, *Gaussian plane and spherical means in separable Hilbert spaces*, Measure theory, Oberwolfach 1981 (Oberwolfach, 1981), Lecture Notes in Math., vol. 945, Springer, Berlin-New York, 1982, pp. 314–335. MR 675296

[49] ______, *On the asymptotic behaviour of Gaussian spherical integrals*, Probability in Banach spaces, IV (Oberwolfach, 1982), Lecture Notes in Math., vol. 990, Springer, Berlin, 1983, pp. 221–234. MR 707521

[50] Edwin Hewitt, *On two problems of Urysohn*, Ann. of Math. (2) **47** (1946), 503–509. MR 17527

[51] Edwin Hewitt and Leonard J. Savage, *Symmetric measures on Cartesian products*, Trans. Amer. Math. Soc. **80** (1955), 470–501. MR 76206

[52] Irina Holmes and Ambar N. Sengupta, *A Gaussian Radon transform for Banach spaces*, J. Funct. Anal. **263** (2012), no. 11, 3689–3706. MR 2984081

[53] D. N. Hoover, *Row-column exchangeability and a generalized model for probability*, Exchangeability in probability and statistics (Rome, 1981), North-Holland, Amsterdam-New York, 1982, pp. 281–291. MR 675982

[54] Douglas N Hoover, *Relations on probability spaces and arrays of random variables*, Preprint, Institute for Advanced Study, Princeton, NJ **2** (1979).

[55] Olav Kallenberg, *On the representation theorem for exchangeable arrays*, J. Multivariate Anal. **30** (1989), no. 1, 137–154. MR 1003713

[56] ______, *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York, 1997. MR 1464694

[57] ______, *Probabilistic symmetries and invariance principles*, Probability and its Applications (New York), Springer, New York, 2005. MR 2161313

[58] Gopinath Kallianpur, *The topology of weak convergence of probability measures*, J. Math. Mech. **10** (1961), 947–969. MR 0132143

[59] H. Jerome Keisler, *Hyperfinite model theory*, Logic Colloquium 76 (Oxford, 1976), 1977, pp. 5–110. Studies in Logic and Found. Math., Vol. 87. MR 0491155

[60] John L. Kelley, *General topology*, Springer-Verlag, New York-Berlin, 1975, Reprint of the 1955 edition [Van Nostrand, Toronto, Ont.], Graduate Texts in Mathematics, No. 27. MR 0370454

[61] J. F. C. Kingman, *Uses of exchangeability*, Ann. Probability **6** (1978), no. 2, 183–197. MR 494344

[62] Werner Kirsch, *An elementary proof of de Finetti's theorem*, Statist. Probab. Lett. **151** (2019), 84–88. MR 3938896

[63] D. Landers and L. Rogge, *Universal Loeb-measurability of sets and of the standard part map with applications*, Trans. Amer. Math. Soc. **304** (1987), no. 1, 229–243. MR 906814

[64] Paul Lévy, *Leçons d'analyse fonctionnelle*, Gauthier-Villars Paris, 1922.

[65] J. E. Littlewood, *Lectures on the Theory of Functions*, Oxford University Press, 1944. MR 0012121

[66] Ambrose Lo, *Demystifying the integrated tail probability expectation formula*, The American Statistician **73** (2019), no. 4, 367–374.

[67] Peter A. Loeb, *Conversion from nonstandard to standard measure spaces and applications in probability theory*, Trans. Amer. Math. Soc. **211** (1975), 113–122. MR 390154

[68] ______, *Applications of nonstandard analysis to ideal boundaries in potential theory*,

Israel J. Math. **25** (1976), no. 1-2, 154–187. MR 457757

[69] ______, *Weak limits of measures and the standard part map*, Proc. Amer. Math. Soc. **77** (1979), no. 1, 128–135. MR 539645

[70] ______, *An introduction to general nonstandard analysis*, Nonstandard analysis for the working mathematician, Springer, Dordrecht, 2015, pp. 37–78. MR 3409513

[71] Albert T. Lundell and Stephen Weingram, *The topology of CW complexes*, The University Series in Higher Mathematics, Van Nostrand Reinhold Co., New York, 1969. MR 3822092

[72] George W. Mackey, *Borel structure in groups and their duals*, Trans. Amer. Math. Soc. **85** (1957), 134–165. MR 89999

[73] Pertti Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, Cambridge, 1995. MR 1333890

[74] James Clerk Maxwell, *V. illustrations of the dynamical theory of gases.—part i. on the motions and collisions of perfectly elastic spheres*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **19** (1860), no. 124, 19–32.

[75] H. P. McKean, *Geometry of differential space*, Ann. Probability **1** (1973), 197–206. MR 0353471

[76] Peter Orbanz and Daniel M Roy, *Bayesian models of graphs, arrays and other exchangeable random structures*, IEEE transactions on pattern analysis and machine intelligence **37** (2014), no. 2, 437–461.

[77] W. Pauli and C.P. Enz, *Thermodynamics and the kinetic theory of gases*, Dover books on physics, Dover Publications, 2000.

[78] Amy Peterson and Ambar N. Sengupta, *Limiting means for spherical slices*, Commun. Stoch. Anal. **12** (2018), no. 3, Art. 4, 271–281. MR 3939456

[79] ______, *The Gaussian limit for high-dimensional spherical means*, J. Funct. Anal. **276** (2019), no. 3, 815–866. MR 3906293

[80] Henri Poincaré, *Calcul des probabilités*, Gauthier-Villars, 1912.

[81] David Preiss, *Metric spaces in which Prohorov's theorem is not valid*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **27** (1973), 109–116. MR 360979

[82] Yu. V. Prokhorov, *Convergence of random processes and limit theorems in probability*

*theory*, Teor. Veroyatnost. i Primenen. **1** (1956), 177–238. MR 0084896

[83] Paul Ressel, *De Finetti-type theorems: an analytical approach*, Ann. Probab. **13** (1985), no. 3, 898–922. MR 799427

[84] David A. Ross, *Loeb measure and probability*, Nonstandard analysis (Edinburgh, 1996), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., vol. 493, Kluwer Acad. Publ., Dordrecht, 1997, pp. 91–120. MR 1603231

[85] ———, *Pushing down infinite Loeb measures*, Math. Scand. **104** (2009), no. 1, 108–116. MR 2498374

[86] Walter Rudin, *Principles of mathematical analysis*, third ed., McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. MR 0385023

[87] ———, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987. MR 924157

[88] Leonard J. Savage, *The foundations of statistics*, revised ed., Dover Publications, Inc., New York, 1972. MR 0348870

[89] Laurent Schwartz, *Radon measures on arbitrary topological spaces and cylindrical measures*, Published for the Tata Institute of Fundamental Research, Bombay by Oxford University Press, London, 1973, Tata Institute of Fundamental Research Studies in Mathematics, No. 6. MR 0426084

[90] Ambar N. Sengupta, *The Gaussian Radon transform as a limit of spherical transforms*, J. Funct. Anal. **271** (2016), no. 11, 3242–3268. MR 3554705

[91] F. G. Slaughter, Jr., *The closed image of a metrizable space is $M_1$*, Proc. Amer. Math. Soc. **37** (1973), 309–314. MR 310832

[92] Flemming Topsøe, *Compactness in spaces of measures*, Studia Math. **36** (1970), 195–222. (errata insert). MR 268347

[93] ———, *Topology and measure*, Lecture Notes in Mathematics, Vol. 133, Springer-Verlag, Berlin-New York, 1970. MR 0422560

[94] ———, *Compactness and tightness in a space of measures with the topology of weak convergence*, Math. Scand. **34** (1974), 187–210. MR 388484

[95] Paul Urysohn, *Über die Mächtigkeit der zusammenhängenden Mengen*, Math. Ann. **94** (1925), no. 1, 262–295. MR 1512258

[96] N. N. Vakhania, V. I. Tarieladze, and S. A. Chobanyan, *Probability distributions on*

*Banach spaces*, Mathematics and its Applications (Soviet Series), vol. 14, D. Reidel Publishing Co., Dordrecht, 1987, Translated from the Russian and with a preface by Wojbor A. Woyczynski. MR 1435288

[97] V. S. Varadarajan, *Measures on topological spaces*, Mat. Sb. (N.S.) **55 (97)** (1961), 35–100. MR 0148838

[98] ______, *Groups of automorphisms of Borel spaces*, Trans. Amer. Math. Soc. **109** (1963), 191–220. MR 159923

[99] Victor Veitch and Daniel M Roy, *The class of random graphs arising from exchangeable random measures*, arXiv preprint arXiv:1512.03099 (2015).

[100] Norbert Wiener, *Differential-space*, Journal of Mathematics and Physics **2** (1923), no. 1-4, 131–174.

[101] Gerhard Winkler, *Simplexes of measures with closed extreme boundary and presentability of Hausdorff spaces*, Math. Nachr. **146** (1990), 47–56. MR 1069046

## Vita

Irfan Alam was born in 1992 in Patna, India. He moved to Bangalore in 2009 to pursue the Bachelor of Mathematics (B.Math.) degree from the Indian Statistical Institute. After completing B.Math. (with honours) in 2012, he moved to the Kolkata branch of the same institute for a Master of Mathematics (M.Math.) degree, which he was awarded in 2014. He then moved to the US where he has been a graduate student at Louisiana State University (LSU) in Baton Rouge. He earned a Master of Science (M.S.) degree in mathematics from LSU in 2016. He is currently a candidate for the degrees of Master of Arts in philosophy as well as Doctor of Philosophy in mathematics, both of which will be awarded in August, 2021. In Fall 2021, he will start a three-year position as Hans Rademacher Instructor in the department of mathematics at the University of Pennsylvania.