Communications on Stochastic Analysis

Volume 4 | Number 1 Article 3

3-1-2010

Commutativity properties of conditional distributions and Palm measures

Olav Kallenberg

Communications on Stochastic Analysis
Vol. 4, No. 1 (2010) 21-34

Serials Publications
www.serialspublications.com

COMMUTATIVITY PROPERTIES OF CONDITIONAL DISTRIBUTIONS AND PALM MEASURES

OLAV KALLENBERG

Abstract. Given a probability space $(\Omega, \mathcal{A}, P)$ and some σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$, we consider iterated conditional distributions of the form $(P[\cdot\,|\,\mathcal{F}])[\cdot\,|\,\mathcal{G}]$. Thus, in the second step, we form the conditional distribution with respect to $\mathcal{G}$, using $P[\cdot\,|\mathcal{F}]$ instead of P as the underlying probability measure. Under suitable regularity conditions, we show that conditioning with respect to $\mathcal{F}$ and $\mathcal{G}$ commute. The commutativity property remains valid for the operations of conditioning on a σ-field $\mathcal{F}$ and of forming the Palm distributions with respect to a random measure ξ. This enables us to construct the Palm measures of ξ via conditioning on a suitable σ-field.

1. Introduction

Given a probability space $(\Omega, \mathcal{A}, P)$ and some σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$, we may define iterated conditioning with respect to $\mathcal{F}$ and $\mathcal{G}$ in two entirely different ways. The most familiar way is to form the product of the conditional expectations $E_{\mathcal{F}} = E[\cdot\,|\mathcal{F}]$ and $E_{\mathcal{G}} = E[\cdot\,|\mathcal{G}]$, regarded as linear contraction operators on $L^1(P)$. As part of Theorem 5.1, we show that $E_{\mathcal{F}}$ and $E_{\mathcal{G}}$ commute iff $\mathcal{F}$ and $\mathcal{G}$ are conditionally independent given $\mathcal{F} \cap \mathcal{G}$.

A much more subtle way, of main concern in this paper, is to iterate the construction of regular conditional distributions. Then recall that, when Ω is a Borel space with a sub-σ-field $\mathcal{F}$, we may choose versions of the conditional probabilities $P_{\mathcal{F}}A = P[A|\mathcal{F}]$ combining into a probability kernel on Ω. In other words, $P_{\mathcal{F}}(\omega, A)$ is then an $\mathcal{F}$-measurable function of $\omega \in \Omega$ for fixed A and a probability measure in $A \in \mathcal{A}$ for fixed ω. Using $P_{\mathcal{F}}$ instead of P as the underlying probability measure on Ω, we may repeat the procedure for a second σ-field $\mathcal{G}$ to form the iterated conditional distribution $(P_{\mathcal{F}})_{\mathcal{G}} = (P[\cdot\,|\,\mathcal{F}])[\cdot\,|\,\mathcal{G}]$.

The latter construction involves some obvious technical difficulties. Since each step gives rise to a measurable function on Ω, the iterated conditioning yields a function on Ω^2. Choosing the latter to be product measurable (in itself a highly non-trivial step), we consider the restriction to the main diagonal (where $\omega_1 = \omega_2$), so that $(P_{\mathcal{F}})_{\mathcal{G}}$ and $(P_{\mathcal{G}})_{\mathcal{F}}$ again become probability kernels on Ω. Thus, we define

$$(P_{\mathcal{F}})_{\mathcal{G}}(\omega) = (P[\cdot\,|\mathcal{F}]_{\omega})[\cdot\,|\mathcal{G}]_{\omega}, \quad \omega \in \Omega.$$

Received 2009-6-22; Communicated by the editors.
2000 *Mathematics Subject Classification.* 60A10, 60G57.
Key words and phrases. Kernels and disintegration, iterated conditioning, random measures, Palm measures and distributions.

Taking diagonal values may seem absurd in this context, since regular conditional distributions are defined only up to a null set. This odd and somewhat surprising feature is justified by the disintegration Theorem 6.4 in [6], which may be stated most strikingly in the "diagonal" form

$$E[f(\xi,\eta)\,|\,\xi] = E[f(s,\eta)\,|\,\xi]_{s=\xi} \quad \text{a.s.} \tag{1.1}$$

With the indicated hurdles overcome, we prove in Theorem 4.1 that $(P_{\mathcal{F}})_{\mathcal{G}} = (P_{\mathcal{G}})_{\mathcal{F}} = P_{\mathcal{F}\vee\mathcal{G}}$ a.s. In other words, we can form $P_{\mathcal{F}\vee\mathcal{G}}$ by conditioning on one σ-field at a time. As an immediate consequence, we see that $P_{\mathcal{F}} = E_{\mathcal{F}}(P_{\mathcal{G}})_{\mathcal{F}}$ and $P_{\mathcal{G}} = E_{\mathcal{G}}(P_{\mathcal{F}})_{\mathcal{G}}$ a.s., which amounts to successive changes in the conditioning σ-field:

$$P_{\mathcal{G}} \rightsquigarrow P_{\mathcal{F}\vee\mathcal{G}} \rightsquigarrow P_{\mathcal{F}}, \qquad P_{\mathcal{F}} \rightsquigarrow P_{\mathcal{F}\vee\mathcal{G}} \rightsquigarrow P_{\mathcal{G}}.$$

Kernel versions of those results require appropriate regularity conditions on Ω. For comparison, we give in Proposition 5.4 some more elementary, set-wise versions, valid for arbitrary Ω, $\mathcal{F}$, and $\mathcal{G}$.

Conditional distributions may be regarded as special Palm distributions, where the underlying random measure ξ is of the form δ_τ for some random element τ in S. For general random measures ξ on S with σ-finite intensity measure $E\xi$ and for any random element η in T, we may choose a random element τ in S with $P[\tau\in\cdot\,|\,\xi,\eta] = \xi$ a.s., and then define the Palm distributions of η with respect to ξ by ordinary conditioning, as $P[\eta\in\cdot\,|\,\tau]_s$ a.e. $E\xi$. For this intuitively appealing construction—equivalent to the classical Ryll-Nardzewski approach in [5, 10]—to make sense, we must allow the probability measures to be σ-finite, which is justified by Lemma 3.3 below. We shall often use the shorthand notation $P[\,\cdot\,\|\,\xi]_s$ for the Palm measure at s with respect to ξ.

The connection between Palm distributions and classical conditioning suggests that the previously mentioned commutativity properties for conditional distributions have counterparts for general Palm measures, which is indeed true. Thus, we show in Theorem 4.2 that Palm conditioning commutes with ordinary conditioning, so that a.e.

$$P_{\mathcal{F}}[\,\cdot\,\|\,\xi] = (P[\,\cdot\,\|\,\xi])_{\mathcal{F}}, \qquad P[\,\cdot\,\|\,\xi] = E[P_{\mathcal{F}}[\,\cdot\,\|\,\xi]\,\|\,\xi].$$

This allows us to obtain $P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$ directly from $P[\,\cdot\,\|\,\xi]$, which is useful when the Palm measures constitute our primary objects of study. It also enables us to construct $P[\,\cdot\,\|\,\xi]$ from P via conditioning on a suitable σ-field:

$$P \rightsquigarrow P_{\mathcal{F}} \rightsquigarrow P_{\mathcal{F}}[\,\cdot\,\|\,\xi] \rightsquigarrow P[\,\cdot\,\|\,\xi].$$

The latter device has numerous applications, to be considered elsewhere.

Our main results are given in Section 4, where we present the basic commutativity properties for conditional distributions and Palm measures. Various related iteration properties are given in Section 5, along with some general commutativity criteria for conditional expectation operators. Some basic results on kernels and disintegration appear in the preliminary Section 2, and in Section 3 we explain the precise meanings of Palm measures and iterated conditioning, and discuss the indicated conditioning approach to Palm distributions. Some general background information on random measures and their Palm distributions may be gathered from [1, 5, 7].

We conclude this section with some remarks on notation. Given a measure μ and a measurable function f on the same space, we write $\mu f = \int \mu(dx)f(x)$ and let $f \cdot \mu$ denote the measure with μ-density f. In formulas involving integrals, the integrand is tacitly assumed to be measurable and nonnegative. By $\mu \sim \nu$ we mean that $\mu \ll \nu$ and $\nu \ll \mu$. The total variation of a (signed) measure μ is denoted by $\|\mu\|$. For any probability measure P, the corresponding expectation is denoted by E, with the same subscript as P if any, and we write $E[\xi; A] = \int_A \xi dP$. The distribution of a random element ξ is often written as $\mathcal{L}(\xi)$. Given two σ-fields $\mathcal{F}$ and $\mathcal{G}$, we write $\mathcal{F} \vee \mathcal{G}$ for the smallest σ-field containing both. For any measurable spaces $S, T, \ldots$, the associated σ-fields are denoted by $\mathcal{S}, \mathcal{T}, \ldots$.

2. Kernels and Disintegration

Given two measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, we define a *kernel* from S to T as a function $\mu \geq 0$ on $S \times \mathcal{T}$ such that $\mu(s, B) = \mu_s B$ is $\mathcal{S}$-measurable in s for fixed B and a measure in B for fixed s. For any kernels ν from S to T and μ from $S \times T$ to U, their *composition* $\nu \otimes \mu$ is the kernel from S to $T \times U$ given by

$$(\nu \otimes \mu)_s f = \int \nu_s(dt) \int \mu_{s,t}(du)\, f(t, u), \tag{2.1}$$

for any measurable function $f \geq 0$ on $T \times U$. Some simple properties of kernels and their composition are given in [6], pp. 20–21. The following basic fact is implicit in [6].

Lemma 2.1. *Composition of kernels is associative.*

Proof. Consider any kernels ν from S to T, μ from $S \times T$ to U, and ρ from $S \times T \times U$ to V. Fixing an $s \in S$ and a measurable function $f \geq 0$ on $T \times U \times V$, we get by (2.1)

$$\begin{aligned}
(\nu \otimes (\mu \otimes \rho))_s f &= \int \nu_s(dt) \iint (\mu \otimes \rho)_{s,t}(du\, dv)\, f(t, u, v) \\
&= \int \nu_s(dt) \int \mu_{s,t}(du) \int \rho_{s,t,u}(dv)\, f(t, u, v) \\
&= \iint (\nu \otimes \mu)_s(dt\, du) \int \rho_{s,t,u}(dv)\, f(t, u, v) \\
&= ((\nu \otimes \mu) \otimes \rho)_s f,
\end{aligned}$$

which shows that $\nu \otimes (\mu \otimes \rho) = (\nu \otimes \mu) \otimes \rho$. □

A kernel μ from S to T is said to be *σ-finite*, if there exists a measurable function $f > 0$ on $S \times T$ such that $\mu_s f(s, \cdot) < \infty$ for all $s \in S$. It is called a *probability kernel* if $\|\mu_s\| = 1$ for all $s \in S$.

Lemma 2.2. *Let M be a measure on $S \times T$, where T is Borel. Then M is σ-finite iff $M = \nu \otimes \mu$ for some σ-finite measure ν on S and a σ-finite kernel μ from S to T. The measures μ_s are a.e. unique up to normalizations, depending reciprocally on the choice of ν, and we may choose μ to be a probability kernel iff $M(\cdot \times T)$ is σ-finite.*

Proof. When $\|M\| = 1$, Theorems 6.3 and 6.4 in [6] yield $M = \nu \otimes \mu$ with $\nu = M(\cdot \times T)$ for some probability kernel μ from S to T. Next let M be σ-finite and $\neq 0$, and choose a measurable function $g > 0$ on $S \times T$ with $Mg = 1$. Then $g \cdot M = \nu \otimes \mu$ with $\nu = (g \cdot M)(\cdot \times T)$ for some probability kernel μ from S to T, and so $M = \nu \otimes (h \cdot \mu)$ with $h = 1/g$. Here $h \cdot \mu$ is σ-finite since $g \cdot (h \cdot \mu) = gh \cdot \mu = \mu$.

Conversely, let $M = \nu \otimes \mu \neq 0$ with σ-finite ν and μ. Then choose some measurable functions $f > 0$ on S and $g > 0$ on $S \times T$ with $\nu f = 1$ and $\mu_s g(s, \cdot) < \infty$ for all $s \in S$. Modifying μ on a ν-null set to get $\mu_s g(s, \cdot) > 0$ and noting that $\mu_s g(s, \cdot)$ is measurable by Lemma 1.41 in [6], we may normalize g such that $\mu_s g(s, \cdot) = 1$ for all s. Then (2.1) yields $M(fg) = 1$ with $(fg)(s,t) = f(s)\, g(s,t) > 0$.

To prove the stated uniqueness, we may assume that $\|M\| = 1$, in which case the assertion follows from the a.s. uniqueness of conditional distributions. If $\nu = M(\cdot \times T)$ is σ-finite and $\neq 0$, we may choose a measurable function $f > 0$ on S with $\nu f = 1$. Then $\|f \cdot M\| = 1$, and so $f \cdot M = (f \cdot \nu) \otimes \mu$ for some probability kernel μ, which implies $M = \nu \otimes \mu$. Conversely, suppose that $M = \nu \otimes \mu$ for some σ-finite measure ν and a probability kernel μ. Then (2.1) yields $M(\cdot \times T) = \nu$, and so $M(\cdot \times T)$ is σ-finite. $\square$

The relation $M = \nu \otimes \mu$ is called a *disintegration* of M. If M is σ-finite, we may choose ν to be any σ-finite measure $\nu \sim M(\cdot \times T)$. The measures μ_s are a.e. unique up to normalizations, depending on the normalization of ν. If even S is Borel, there is also a *dual* disintegration $\tilde{M} = \nu' \otimes \mu'$ in terms of a σ-finite measure ν' on T and a σ-finite kernel μ' from T to S, where $\tilde{M} f = M \tilde{f}$ with $\tilde{f}(t,s) = f(s,t)$, and we may write $\nu \otimes \mu \mathrel{\tilde{=}} \nu' \otimes \mu'$.

The disintegration of kernels is more subtle.

Theorem 2.3. *Let ρ be a σ-finite kernel from S to $T \times U$, where U is Borel.*

- *(i) If $\mathcal{T}$ is countably generated, there exist some σ-finite kernels ν from S to T and μ from $S \times T$ to U such that $\rho = \nu \otimes \mu$.*
- *(ii) For any σ-finite measure P on S, there exist some kernels ν and μ as in (i) such that $P \otimes \rho = P \otimes \nu \otimes \mu$.*

Both assertions remain true for any fixed, σ-finite kernel ν from S to T such that $\nu_s \sim \rho_s(\cdot \times U)$ for all $s \in S$.

Part (i) fails without some regularity conditions on T (cf. [8]), and we have only the weaker statement (ii). Here the conclusion is equivalent to $\rho f = (\nu \otimes \mu) f$ a.s. P, for any measurable function $f \geq 0$ on $S \times T \times U$.

Proof. (i) First let ρ be a probability kernel from S to $T \times U$, so that $\nu_s = \rho_s(\cdot \times U)$ is a probability kernel from S to T. Since $\mathcal{T}$ is countably generated and U is Borel, Proposition 7.26 in [6] (cf. [2], p. 57) yields a probability kernel μ from $S \times T$ to U satisfying

$$\mu_{s,t} B = \frac{\rho_s(dt \times B)}{\nu_s(dt)}, \quad t \in T \text{ a.e. } \nu_s,\ s \in S,\ B \in \mathcal{U},$$

which implies $\rho = \nu \otimes \mu$ by Theorem 6.4 in [6].

For a general σ-finite kernel ρ, there exists a measurable function $g > 0$ on $S \times T \times U$ such that $\rho_s g(s, \cdot) = 1$ on the set $S' = \{s \in S;\ \rho_s \neq 0\}$. Then $g \cdot \rho$ is a probability kernel from S' to $T \times U$, and so $g \cdot \rho = \nu \otimes \mu$ for some probability kernels ν from S' to T and μ from $S' \times T$ to U. Writing $h = 1/g$, we get $\rho = \nu \otimes (h \cdot \mu)$ on S', where $h \cdot \mu$ is clearly a σ-finite kernel from $S' \times T$ to U. We may finally choose $\nu_s = \mu_{s,t} = 0$ for $s \in S \setminus S'$.

Now fix any σ-finite kernels ν from S to T and ρ from S to $T \times U$ such that $\nu_s \sim \rho_s(\cdot \times U)$ for all $s \in S$. As before, we may reduce to the case where $\|\nu_s\| = \|\rho_s\| = 1$ on S'. Putting $\nu'_s = \rho_s(\cdot \times U)$, we get as before a disintegration $\rho = (\nu + \nu') \otimes \mu$ for some kernel μ from $S \times T$ to U. In particular, $\nu' = h \cdot (\nu + \nu')$ with $h(s,t) = \mu_{s,t} U$, and so $\nu = (1 - h) \cdot (\nu + \nu')$. Since $\nu \sim \nu'$, we obtain $\nu + \nu' = h' \cdot \nu$ with $h' = (1-h)^{-1} 1\{h < 1\}$, which implies $\rho = \nu \otimes \mu'$ with $\mu' = h'\mu$.

(ii) Again we may assume that $\|\rho_s\| \equiv 1$. For any probability measure P on S, we may form a probability measure $M = P \otimes \rho$ on $S \times T \times U$. Projecting both sides onto $S \times T$ yields $\bar{M} = P \otimes \nu$, where $\bar{M} = M(\cdot \times U)$ and $\nu = \rho(\cdot \times U)$. Furthermore, Lemma 2.2 yields $M = \bar{M} \otimes \mu$ for some kernel μ from $S \times T$ to U. Hence, by Lemma 2.1,

$$P \otimes \rho = M = \bar{M} \otimes \mu = (P \otimes \nu) \otimes \mu = P \otimes (\nu \otimes \mu).$$

If even ν is fixed with $\nu_s \sim \rho_s(\cdot \times U)$, then $P \otimes \nu \sim (P \otimes \rho)(\cdot \times U)$, and we may choose μ to be a kernel from $S \times T$ to U satisfying $P \otimes \rho = (P \otimes \nu) \otimes \mu$. The assertion now follows as before. $\square$

3. Palm Measures and Iteration

The prime examples of disintegration kernels are of course the regular conditional distributions. Here ξ and η are random elements in some measurable spaces S and T with T Borel, and we consider disintegrations of the form $\mathcal{L}(\xi, \eta) = \mathcal{L}(\xi) \otimes \mu$, where μ is a probability kernel from S to T satisfying $\mu_\xi = P[\eta \in \cdot \mid \xi]$ a.s. If the basic probability space $(\Omega, \mathcal{A}, P)$ is itself Borel, then for any σ-field $\mathcal{F} \subset \mathcal{A}$ we can choose ξ and η to be the identity maps of Ω into $(\Omega, \mathcal{F})$ and $(\Omega, \mathcal{A})$, respectively, which gives the conditional probability $P[\cdot|\mathcal{F}]$ as a kernel $P_\mathcal{F}$ from $(\Omega, \mathcal{F})$ to $(\Omega, \mathcal{A})$.

A *random measure* on a measurable space S is defined as a kernel ξ from Ω to S. We shall always assume that ξ is *uniformly σ-finite*, in the sense that there exists a measurable partition $B_1, B_2, \ldots$ of S such that $\xi B_k < \infty$ a.s. for every k. In that case, ξ may also be regarded as a random element in the measure space $\mathcal{M}_S$, consisting of all measures μ on S with $\mu B_k < \infty$ for all k. Note that if S is a Borel space, then so is $\mathcal{M}_S$. The *intensity measure* $E\xi$ of ξ is given by $(E\xi)f = E(\xi f)$.

Now consider a random measure ξ on S and a random element η in T, where both S and T are Borel. The associated *Campbell measure* $C_{\xi,\eta}$ on $S \times T$ is defined by

$$C_{\xi,\eta} f \equiv \iint C_{\xi,\eta}(ds\, dt)\, f(s,t) = E \int \xi(ds)\, f(s, \eta), \quad f \geq 0,$$

where f is understood to be measurable. If $C_{\xi,\eta}$ is σ-finite, which holds automatically when ξ is η-measurable, there exists a disintegration $C_{\xi,\eta} = \lambda \otimes Q$ in terms of a σ-finite measure λ on S and a σ-finite kernel Q from S to T. Here we call λ a *supporting measure* of ξ, while the Q_s are known as the associated *Palm measures* of η with respect to ξ at the points $s \in S$. From Lemma 2.2 we see that the Q_s are a.e. finite iff $E\xi$ is σ-finite, in which case we may take $\lambda = E\xi$ and choose the Q_s to be probability measures on T, then referred to as *Palm distributions*.

When Ω itself is Borel, we may choose η to be the identity mapping on Ω, so that the associated Campbell measure becomes $C_\xi = P \otimes \xi \overset{\sim}{=} \nu \otimes Q$, and the Palm measures or distributions Q_s are defined directly on Ω. In that case, we often write $Q_s = P[\,\cdot\,\|\,\xi]_s$. (For general η we have instead $C_{\xi,\eta} = \mathcal{L}(\eta) \otimes E_\eta \xi$.) To obtain the n-th order Palm measures, we simply replace ξ by the n-fold product measure $\xi^{\otimes n}$ on S^n. Finally, given a point process $\xi = \sum_{i \in J} \delta_{\tau_i}$ on S, we define the n-th order *reduced Palm measures* of ξ by disintegration of the *compound Campbell measure*

$$C_\xi^{(n)} f = E \sum_{m \le \xi} f(\xi - m, m), \quad f \ge 0,$$

where the summation extends over all measures $m \in \mathcal{N}_S$ of the form $\sum_{i \in I} \delta_{\tau_i}$ with finite $I \subset J$.

We turn to the construction of iterated conditional distributions. Then consider a Borel probability space $(\Omega, \mathcal{A}, P)$ with two σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$. Letting ξ, η, and ι denote the identity maps of Ω into $(\Omega, \mathcal{F})$, $(\Omega, \mathcal{G})$, and $(\Omega, \mathcal{A})$, respectively, we first form the disintegration $\mathcal{L}(\xi, \iota) = \mathcal{L}(\xi) \otimes P_\mathcal{F}$, which gives $P_\mathcal{F}$ as an $\mathcal{F}$-measurable kernel on Ω. Next repeat the procedure with P and ξ replaced by $P_\mathcal{F}$ and η, as in

$$P_\mathcal{F}\{(\eta, \iota) \in \cdot\} = P_\mathcal{F}\{\eta \in \cdot\} \otimes P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}],$$

to form a measure-valued function $P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}]$ on Ω^2. Finally, take diagonal values

$$(P_\mathcal{F})_\mathcal{G}(\omega) = P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}](\omega, \omega), \quad \omega \in \Omega,$$

to obtain a measure-valued function $(P_\mathcal{F})_\mathcal{G}$ on Ω. It is crucial for our purposes to choose suitably measurable versions of $P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}]$ and $(P_\mathcal{F})_\mathcal{G}$.

Lemma 3.1. *Consider a Borel probability space $(\Omega, \mathcal{A}, P)$ and some σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$.*

- *(i) If $\mathcal{G}$ is countably generated, then $P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}]$ has an $(\mathcal{F} \otimes \mathcal{G})$-measurable version for any choice of $P_\mathcal{F}$.*
- *(ii) For general $\mathcal{G}$, we can choose a version of the kernel $P_\mathcal{F}\{(\eta, \iota) \in \cdot\}$ above such that $P_\mathcal{F}[\,\cdot\,|\,\mathcal{G}]$ has an $(\mathcal{F} \otimes \mathcal{G})$-measurable version.*

In either case, $(P_\mathcal{F})_\mathcal{G}$ becomes $(\mathcal{F} \vee \mathcal{G})$-measurable.

Proof. Claim (i) follows immediately from Proposition 2.3 (i). In (ii), Proposition 2.3 (ii) yields some probability kernels ν from $(\Omega, \mathcal{F})$ to $(\Omega, \mathcal{G})$ and μ from $(\Omega^2, \mathcal{F} \otimes \mathcal{G})$ to $(\Omega, \mathcal{A})$ satisfying

$$\mathcal{L}(\xi, \eta, \iota) = \mathcal{L}(\xi) \otimes (\nu \otimes \mu),$$

which gives $\rho = \nu \otimes \mu$ as a version of $P_\mathcal{F}\{(\eta, \iota) \in \cdot\}$. The last assertion holds since the pair (ξ, η) is $(\mathcal{F} \vee \mathcal{G})$-measurable. □

In (ii) it is not clear whether we can choose a version of $P_\mathcal{F}$ itself with the stated property. This is because our new version $\nu \otimes \mu$ of the kernel $P_\mathcal{F}\{(\eta, \zeta) \in \cdot\}$ may not be of the same form.

We turn to the iteration of conditioning and Palm disintegration. To define $P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$, we first form $P_{\mathcal{F}}$ as before as an $\mathcal{F}$-measurable probability kernel on Ω. Since ξ is $\mathcal{A}$-measurable, we may next form the associated Campbell measure $P_{\mathcal{F}} \otimes \xi$ as a kernel from Ω to $\Omega \times S$, and then construct $P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$ by suitable disintegration. To form $(P[\,\cdot\,\|\,\xi])_{\mathcal{F}}$, we start from $P[\,\cdot\,\|\,\xi]$, regarded as a kernel from S to Ω. Introducing the identity maps η and ι from Ω to $(\Omega, \mathcal{F})$ and $(\Omega, \mathcal{A})$, respectively, we form the kernel $P[\,\cdot\,\|\,\xi] \circ (\eta, \iota)^{-1}$ from S to $(\Omega^2, \mathcal{F} \otimes \mathcal{A})$. It remains to construct $(P[\,\cdot\,\|\,\xi])_{\mathcal{F}}$ as before by suitable disintegration.

Lemma 3.2. *Consider a Borel probability space $(\Omega, \mathcal{A}, P)$, a σ-field $\mathcal{F} \subset \mathcal{A}$, and a random measure ξ on a Borel space S such that $P \otimes \xi$ is σ-finite on $\mathcal{F} \otimes \mathcal{S}$.*

(i) The kernel $P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$ has an $(\mathcal{F} \otimes \mathcal{S})$-measurable version.
(ii) If $\mathcal{F}$ is countably generated, then $(P[\,\cdot\,\|\,\xi])_{\mathcal{F}}$ has an $(\mathcal{F} \otimes \mathcal{S})$-measurable version for any choice of $P[\,\cdot\,\|\,\xi]$.
(iii) For general $\mathcal{F}$, we can choose a version of the kernel $P[\,\cdot\,\|\,\xi] \circ (\eta, \iota)^{-1}$ such that $(P[\,\cdot\,\|\,\xi])_{\mathcal{F}}$ has an $(\mathcal{F} \otimes \mathcal{S})$-measurable version.

Proof. Since S is Borel, the σ-field $\mathcal{S}$ is countably generated, and (i) follows. Claims (ii) and (iii) can be proved as in case of Lemma 3.1. □

The previous iteration requires us to consider conditional distributions based on possibly infinite "probability" measures. Say that P is a *pseudo-probability measure* on Ω if it is σ-finite and $\neq 0$.

Lemma 3.3. *Consider a pseudo-probability space $(\Omega, \mathcal{A}, P)$ and a σ-field $\mathcal{F} \subset \mathcal{A}$ such that P remains σ-finite on $\mathcal{F}$. Choose an $\mathcal{F}$-measurable random variable $\rho > 0$ with $E\rho = 1$, and put $\hat{P} = \rho \cdot P$. Then the random variables $E[\,\xi\,|\mathcal{F}] = \hat{E}[\,\xi\,|\mathcal{F}]$ are a.s. independent of the choice of ρ, and they obey all properties of conditional expectations and distributions listed in Theorems 6.1 and 6.3–4 of* [6].

Proof. The chain rule $f \cdot (g \cdot \mu) = (fg) \cdot \mu$ ensures that, if μ and ν are σ-finite measures on a measurable space S with $\mu \ll \nu$ and if $\rho > 0$ is a measurable function on S, then $d\mu/d\nu = d(\rho \cdot \mu)/d(\rho \cdot \nu)$ a.e. ν. This shows that $\hat{E}[\,\xi\,|\mathcal{F}]$ is a.s. independent of ρ. Since $L^1(P)$ may not be contained in $L^1(\hat{P})$, we must allow random variables $\xi \geq 0$ with $\hat{E}\xi = \infty$. Defining $E[\,\xi\,|\mathcal{F}] = \hat{E}[\,\xi\,|\mathcal{F}]$ for a fixed $\rho > 0$ with $E\rho = 1$, we can easily verify all properties involving a single σ-field $\mathcal{F}$. Only the chain rule (tower property) involves two different σ-fields $\mathcal{F} \subset \mathcal{G}$, and we need to define ρ in terms of the smallest one, $\mathcal{F}$. □

Allowing conditioning based on pseudo-probability measures, we may express the Palm distributions $P[\,\cdot\,\|\,\xi]$ in terms of ordinary conditioning. Informally, they agree with the conditional distributions, given a random element τ in S, sampled according to the conditional pseudo-distribution ξ. To develop this idea, consider first a single unit mass $\xi = \delta_\tau$, located at a random point τ in S. Here clearly $P[\eta \in \cdot\,\|\,\xi]_s = P[\eta \in \cdot\,|\,\tau]_s$ a.s. This remains true when $\|\xi\| = c$ a.s. for some constant $c \in (0, \infty)$, provided that we choose τ to satisfy $c\,P[\tau \in \cdot\,|\,\xi, \eta] = \xi$ a.s. The next step is to assume the intensity measure $E\xi$ to be σ-finite. The general case requires iterated conditioning and will not be considered until Theorem 5.2.

Lemma 3.4. *Consider a random measure ξ on S and a random element η in T, where S and T are Borel and $E\xi$ is σ-finite. Introduce a pseudo-probability space $(\tilde{\Omega}, \tilde{P})$ with some random elements τ in S and $\tilde{\eta}$ in T such that*

$$\tilde{E}f(\tau, \tilde{\eta}) = E \int \xi(ds)\, f(s, \eta), \quad f \geq 0. \tag{3.1}$$

Then

$$P[\eta \in \cdot \,\|\, \xi]_s = \tilde{P}[\tilde{\eta} \in \cdot \,|\, \tau]_s, \quad s \in S \text{ a.e. } E\xi.$$

Informally, we are choosing τ to satisfy $P[\tau \in \cdot \,|\, \xi, \eta] = \xi$ a.s. Similar results hold for multivariate and reduced Palm measures.

Proof. For measurable sets $B \subset S$ and functions $f \geq 0$ on T, we have

$$\begin{aligned} \tilde{E}[E[f(\eta) \,\|\, \xi]_\tau;\, \tau \in B] &= E \int_B E[f(\eta) \,\|\, \xi]_s\, \xi(ds) \\ &= \int_B E[f(\eta) \,\|\, \xi]_s\, E\xi(ds) \\ &= Ef(\eta)\, \xi B = \tilde{E}[f(\tilde{\eta});\, \tau \in B], \end{aligned}$$

where the first and last steps hold by (3.1), the second step holds by the definition of $E\xi$, and the third step holds by Palm disintegration. □

4. Main Iteration Principles

Here we establish the basic commutativity properties for conditional distributions and Palm measures. Recall that $P_{\mathcal{F}} = P[\cdot \,|\, \mathcal{F}]$ and $E_{\mathcal{F}} = E[\cdot \,|\, \mathcal{F}]$. The iterated conditional distributions $(P_{\mathcal{F}})_{\mathcal{G}}$ and $(P_{\mathcal{G}})_{\mathcal{F}}$ are defined as probability kernels from $(\Omega, \mathcal{F} \vee \mathcal{G})$ to $(\Omega, \mathcal{A})$, in the sense of Lemma 3.1 (i) or (ii).

Theorem 4.1. *For any Borel probability space $(\Omega, \mathcal{A}, P)$ and σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$, we have a.s.*

(i) $(P_{\mathcal{F}})_{\mathcal{G}} = (P_{\mathcal{G}})_{\mathcal{F}} = P_{\mathcal{F} \vee \mathcal{G}}$,
(ii) $P_{\mathcal{F}} = E_{\mathcal{F}}(P_{\mathcal{G}})_{\mathcal{F}}, \quad P_{\mathcal{G}} = E_{\mathcal{G}}(P_{\mathcal{F}})_{\mathcal{G}}$.

Proof. Let ξ, η, and ι denote the identity mappings of Ω into the measurable spaces $(\Omega, \mathcal{F})$, $(\Omega, \mathcal{G})$, and $(\Omega, \mathcal{A})$, respectively, and define the kernels ρ from $(\Omega, \mathcal{F})$ to $(\Omega, \mathcal{A})$, μ from $(\Omega, \mathcal{F})$ to $(\Omega^2, \mathcal{G} \otimes \mathcal{A})$, and ν from $(\Omega^2, \mathcal{F} \otimes \mathcal{G})$ to $(\Omega, \mathcal{A})$ by

$$\rho_\xi = P[\cdot \,|\, \xi], \quad \mu_s = \rho_s\{(\eta, \iota) \in \cdot\}, \quad \nu_{s,\eta} = \rho_s[\cdot \,|\, \eta],$$

where ν is such as in Lemma 3.1. Note that $\nu_s = (\nu_{s,t};\, t \in \Omega)$ is then a $\mathcal{G}$-measurable kernel on Ω for every fixed s. For any $(\mathcal{F} \otimes \mathcal{G} \otimes \mathcal{A})$-measurable function $f \geq 0$ on Ω^3, we now apply Theorem 6.4 in [6] in turn to the functions

$$f(s,t,u), \quad g_s(t,u) = f(s,t,u), \quad h(s,t) = \int \nu_{s,t}(du)\, f(s,t,u),$$

with associated disintegrations

$$\mathcal{L}(\xi, \eta, \iota) = \mathcal{L}(\xi) \otimes \mu, \quad \mu_s = \mu_s(\cdot \times \Omega) \otimes \nu_s, \quad \mathcal{L}(\xi, \eta) = \mathcal{L}(\xi) \otimes \mu(\cdot \times \Omega),$$

to obtain

$$\begin{aligned} Ef(\xi,\eta,\iota) &= E\int\!\!\int \mu_\xi(dt\,du)\,f(\xi,t,u) \\ &= E\int \mu_\xi(dt\times\Omega)\int \nu_{\xi,t}(du)\,f(\xi,t,u) \\ &= E\int \nu_{\xi,\eta}(du)\,f(\xi,\eta,u). \end{aligned}$$

Here the uniqueness of the disintegration yields $\nu_{\xi,\eta} = P[\,\cdot\,|\,\xi,\eta]$ a.s., which means that $(P_\mathcal{F})_\mathcal{G} = P_{\mathcal{F}\vee\mathcal{G}}$ a.s. Similarly, $(P_\mathcal{G})_\mathcal{F} = P_{\mathcal{F}\vee\mathcal{G}}$ a.s., and (i) follows. To prove (ii), we note that

$$P_\mathcal{F} = E_\mathcal{F} P_{\mathcal{F}\vee\mathcal{G}} = E_\mathcal{F}(P_\mathcal{G})_\mathcal{F}, \qquad P_\mathcal{G} = E_\mathcal{G} P_{\mathcal{F}\vee\mathcal{G}} = E_\mathcal{G}(P_\mathcal{F})_\mathcal{G},$$

by (i) along with the tower property of conditional expectations. □

The commutativity relations in Theorem 4.1 have the following counterparts for Palm measures $P[\,\cdot\,\|\,\xi]$ and their conditional versions $P_\mathcal{F}[\,\cdot\,\|\,\xi]$ and $(P[\,\cdot\,\|\,\xi])_\mathcal{F}$, where the iteration is now defined in the sense of Lemma 3.2. As before, any product-measurable function on S^2 or Ω^2 is evaluated along the main diagonal.

Theorem 4.2. *Given a Borel probability space $(\Omega,\mathcal{A},P)$ and a σ-field $\mathcal{F}\subset\mathcal{A}$, let ξ be a random measure on a Borel space S such that $P\otimes\xi$ is σ-finite on $\mathcal{F}\otimes\mathcal{S}$. Then*

(i) $P_\mathcal{F}[\,\cdot\,\|\,\xi] = (P[\,\cdot\,\|\,\xi])_\mathcal{F} = (P\otimes\xi)_{\mathcal{F}\otimes\mathcal{S}}$ *a.e.* $P\otimes\xi$,

(ii) $P[\,\cdot\,\|\,\xi] = E[P_\mathcal{F}[\,\cdot\,\|\,\xi]\,\|\,\xi]$ *a.e.* $E\xi$,

where all three members of (i) can be chosen to be probability kernels from $(\Omega\times S,\ \mathcal{F}\otimes\mathcal{S})$ to $(\Omega,\mathcal{A})$.

Here $(P\otimes\xi)_{\mathcal{F}\otimes\mathcal{S}}$ denotes the set of conditional distributions, under the Campbell measure $P\otimes\xi$ and with respect to the σ-field $\mathcal{F}\otimes\mathcal{S}$, defined in the sense of Lemma 3.3 as a probability kernel from $\Omega\times S$ to Ω. Part (ii) may be regarded as a version of the formula $P = EP_\mathcal{F}$, applied to the Palm measures $P[\,\cdot\,\|\,\xi]_s$. Detailed justifications are provided below, as part of our proofs.

We begin with some technical prerequisites.

Lemma 4.3. *Given P, $\mathcal{F}$, and ξ as in Theorem 4.2, we have for any $(\mathcal{F}\otimes\mathcal{S})$-measurable process $Y\ge 0$ on S*

(i) $E_\mathcal{F}\int Y_s\,\xi(ds) = \int Y_s E_\mathcal{F}\xi(ds)$ *a.s. P,*

(ii) $E_\mathcal{F}[Y_s; A\,\|\,\xi]_s = Y_s P_\mathcal{F}[A\,\|\,\xi]_s$ *a.e.* $P\otimes\xi$.

Proof. (i) We may write $Y_s = f(s,\eta)$, where η denotes the identity mapping from Ω to $(\Omega,\mathcal{F})$. Then $\int \xi(ds)\,f(s,t)$ is an $(\mathcal{A}\otimes\mathcal{F})$-measurable function on Ω^2 (cf. [6], p. 21), and so the disintegration theorem, in the form of (1.1), yields

$$E_\mathcal{F}\int Y_s\,\xi(ds) = \Big(E_\mathcal{F}\int f(s,t)\,\xi(ds)\Big)_{t=\eta} = \int f(s,\eta)\,E_\mathcal{F}\xi(ds),$$

where the second step holds by the definition of $E_\mathcal{F}\xi$ in terms of $P_\mathcal{F}$.

(ii) Regarding $\mu = P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$ as a kernel from $S \times T$ to Ω, where T denotes the measurable space $(\Omega, \mathcal{F})$, we obtain the asserted formula by writing

$$\int \mu_{s,t}(du)\, f(s,t)\, g(u) = f(s,t) \int \mu_{s,t}(du)\, g(u),$$

for any measurable functions $f \geq 0$ on $S \times T$ and $g \geq 0$ on Ω. Our qualification is needed since each side is determined only up to a $(P \otimes \xi)$-null set in $S \times T$. $\square$

Proof of Theorem 4.2: (i) Since $P \otimes \xi$ is σ-finite on $\mathcal{F} \otimes \mathcal{S}$, Lemma 3.3 ensures the existence of $(P \otimes \xi)_{\mathcal{F} \otimes \mathcal{S}}$ as an $(\mathcal{F} \otimes \mathcal{S})$-measurable probability kernel from $\Omega \times S$ to Ω. Choosing an $(\mathcal{F} \otimes \mathcal{S})$-measurable process $Y > 0$ with $(P \otimes \xi)Y < \infty$, we get by Lemma 4.3 (i)

$$E \int Y_s E_{\mathcal{F}} \xi(ds) = E E_{\mathcal{F}} \int Y_s\, \xi(ds) = (P \otimes \xi) Y < \infty,$$

which shows that $E_{\mathcal{F}}\xi$ is a σ-finite kernel from Ω to S. Choosing $E_{\mathcal{F}}\xi$ as our supporting kernel for ξ under $P_{\mathcal{F}}$ and invoking Lemma 3.2 (i), we obtain $P_{\mathcal{F}}[\,\cdot\,\|\,\xi]$ as a probability kernel from $\Omega \times S$ to Ω. For any disintegration $P \otimes \xi \cong \nu \otimes P[\,\cdot\,\|\,\xi]$ we get $E[Y_s \,\|\, \xi]_s < \infty$ a.e. ν, which shows that the Palm measures $P[\,\cdot\,\|\,\xi]_s$ are a.e. σ-finite on $\mathcal{F}$. Hence, by Lemmas 3.2 (ii)–(iii) and 3.3, even $(P[\,\cdot\,\|\,\xi])_{\mathcal{F}}$ can be chosen to form a probability kernel from $\Omega \times S$ to Ω.

Now write $P_\xi^s = P[\,\cdot\,\|\,\xi]_s$, for convenience, and fix any supporting measure ν of ξ. Then for any set $A \in \mathcal{A}$ and $(\mathcal{F} \otimes \mathcal{S})$-measurable process $Y \geq 0$ on S, we have

$$\begin{aligned} E\int \xi(ds)\, Y_s P_\xi^s[A|\mathcal{F}] &= \int \nu(ds)\, E_\xi^s(\, Y_s P_\xi^s[A|\mathcal{F}]\,) \\ &= \int \nu(ds)\, E_\xi^s[\, Y_s; A] = E[\int \xi(ds)\, Y_s; A] \\ &= E\int E_{\mathcal{F}}\xi(ds)\, E_{\mathcal{F}}[\, Y_s; A \,\|\, \xi]_s \\ &= E\int \xi(ds)\, Y_s P_{\mathcal{F}}[A \,\|\, \xi]_s. \end{aligned}$$

Here the first and third steps hold by Palm disintegration. In the second step, we apply the formula $E(\alpha P[A|\mathcal{F}]) = E[\alpha; A]$, valid for $\mathcal{F}$-measurable $\alpha \geq 0$, to the measures P_ξ^s and random variables Y_s. The fourth step holds by Palm disintegration with respect to the probability measures $P_{\mathcal{F}}$. The last step holds by Lemma 4.3 (i)–(ii), with (i) applied to the process $E_{\mathcal{F}}[Y_s; A \,\|\, \xi]_s$. The assertion now follows since Y was arbitrary.

(ii) Applying the formula $P = EP_{\mathcal{F}}$ to the Palm measures P_ξ^s and using (i), we get

$$P_\xi^s = E_\xi^s (P_\xi^s)_{\mathcal{F}} = E_\xi^s (P_{\mathcal{F}})_\xi^s, \quad s \in S \text{ a.e. } E\xi.$$

To justify the second equality, we may write

$$\begin{aligned} \int \nu(ds)\, \big\| E_\xi^s (P_\xi^s)_{\mathcal{F}} - E_\xi^s (P_{\mathcal{F}})_\xi^s \big\| &\leq \int \nu(ds)\, E_\xi^s \big\| (P_\xi^s)_{\mathcal{F}} - (P_{\mathcal{F}})_\xi^s \big\| \\ &= E\int \xi(ds)\, \big\| (P_\xi^s)_{\mathcal{F}} - (P_{\mathcal{F}})_\xi^s \big\| = 0, \end{aligned}$$

where the first step holds by the triangle inequality, the second step holds by Palm disintegration, and the last step holds by (i). Hence, the norm on the left vanishes a.e. $\nu \sim E\xi$. □

5. Further Commutativity Properties

The iterated conditioning in the last section should be distinguished from repeated applications of the conditional expectations $E_{\mathcal{F}} = E[\,\cdot\,|\mathcal{F}]$, regarded as linear contraction operators on $L^1(P)$. For any σ-fields $\mathcal{F}$ and $\mathcal{G}$, the product $E_{\mathcal{F}}E_{\mathcal{G}}$ is again an operator of the same type, given for any $\xi \in L^1$ by

$$E_{\mathcal{F}}E_{\mathcal{G}}\,\xi = E[E[\,\xi\,|\,\mathcal{G}]\,|\,\mathcal{F}] \text{ a.s.}$$

We give some general criteria for commutativity. By $\mathcal{F} \perp\!\!\!\perp_{\mathcal{H}} \mathcal{G}$ we mean that $\mathcal{F}$ and $\mathcal{G}$ are conditionally independent, given a third σ-field $\mathcal{H}$. For any subspaces $F, G \subset L^2$, we write $F^{\perp}$ for the orthogonal complement of F and $F \vee G$ for the closed linear subspace spanned by F and G.

Theorem 5.1. *For any probability space $(\Omega, \mathcal{A}, P)$ with σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$ and generated subspaces $F, G \subset L^2(P)$, these conditions are equivalent:*

(i) $E_{\mathcal{F}}E_{\mathcal{G}} = E_{\mathcal{G}}E_{\mathcal{F}}$,
(ii) $\mathcal{F} \perp\!\!\!\perp_{\mathcal{F}\cap\mathcal{G}} \mathcal{G}$ *a.s.*,
(iii) $F \perp G$ *on* $(F \cap G)^{\perp}$,
(iv) $F^{\perp} \perp G^{\perp}$ *on* $F \vee G$,

in which case $E_{\mathcal{F}}E_{\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}}$. More generally, $\mathcal{F} \perp\!\!\!\perp_{\mathcal{H}} \mathcal{G}$ for some σ-fields $\mathcal{F}, \mathcal{G}, \mathcal{H} \subset \mathcal{A}$ iff $E_{\mathcal{F}\vee\mathcal{H}}$ and $E_{\mathcal{G}\vee\mathcal{H}}$ commute with product $E_{\mathcal{H}}$.

Condition (i) is obvious when $\mathcal{F} \subset \mathcal{G}$, and when $\mathcal{F} \perp\!\!\!\perp \mathcal{G}$ it follows from Fubini's theorem. It is also well-known for σ-fields of the form $\mathcal{F}_\sigma$ and $\mathcal{F}_\tau$, where σ and τ are optional times with respect to a filtration $\mathcal{F} = (\mathcal{F}_t)$ satisfying the "usual" conditions of right-continuity and completeness (cf. [6], pp. 127, 135). Here the standard proof uses martingales (cf. [2], pp. 15, 82, or [9], pp. 161, 191). In all those cases, (ii) is easily verified directly.

Condition (ii) may seem like a very special case of conditional independence. However, the general relation $\mathcal{F} \perp\!\!\!\perp_{\mathcal{H}} \mathcal{G}$ is equivalent to $\mathcal{F}\vee\mathcal{H} \perp\!\!\!\perp_{\mathcal{H}} \mathcal{G}\vee\mathcal{H}$ by Corollary 6.7 (i) in [6], and so we may assume that $\mathcal{H} \subset \mathcal{F} \cap \mathcal{G}$. But then $\mathcal{H} = \mathcal{F} \cap \mathcal{G}$ a.s. by part (ii) of the same result.

Commutativity criteria for Hilbert space projections appear in many standard texts, such as in [3], p. 514, and [4], p. 211. However, I was unable to find a reference for the geometric condition (iii) and its dual (iv), familiar to every calculus student for planes in three-dimensional space.

Proof of Theorem 5.1: Assuming (ii), we get by Proposition 6.6 in [6]

$$E_{\mathcal{F}}E_{\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}}E_{\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}}E_{\mathcal{F}} = E_{\mathcal{G}}E_{\mathcal{F}},$$

which proves (i). Conversely, (i) yields $E_{\mathcal{F}}E_{\mathcal{G}}\xi \in L^1(\mathcal{F} \cap \mathcal{G})$ for any $\xi \in L^1$. Since also

$$E[E_{\mathcal{F}}E_{\mathcal{G}}\xi; A] = E[E_{\mathcal{G}}\xi; A] = E[\xi; A], \quad A \in \mathcal{F} \cap \mathcal{G},$$

we get $E_{\mathcal{F}}E_{\mathcal{G}}\xi = E_{\mathcal{F}\cap\mathcal{G}}\xi$ a.s. Hence, $E_{\mathcal{F}}E_{\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}} = E_{\mathcal{F}\cap\mathcal{G}}E_{\mathcal{G}}$, and (ii) follows by Proposition 6.6 in [6].

Under (i) the L^2-projections $P_F = E_{\mathcal{F}}$ and $P_G = E_{\mathcal{G}}$ commute, and as before the product equals P_K with $K = F \cap G$. Hence, for any $x \in F \cap K^\perp$ we have $P_G x = P_G P_F x = P_K x = 0$, and so $x \perp G$. This shows that $F \cap K^\perp \perp G$, and (iii) follows. Conversely, (iii) implies

$$L^2 = K \oplus (F \cap K^\perp) \oplus (G \cap K^\perp) \oplus (F^\perp \cap G^\perp). \tag{5.1}$$

In particular, every $x \in L^2$ is a linear combination of elements in K, $F \cap K^\perp$, $G \cap K^\perp$, and $F^\perp \cap G^\perp$, and by linearity we may prove (i) by checking the relation $P_F P_G = P_K$ for each of them. This is an elementary exercise, completing the proof that (i)–(iii) are equivalent.

Using the decomposition (5.1), we may easily check that (iii) implies (iv). Since $(F \vee G)^\perp = F^\perp \cap G^\perp$ and therefore $F \vee G = (F^\perp \cap G^\perp)^\perp$, condition (iv) agrees with (iii), applied to the sets $F^\perp$ and $G^\perp$. Hence, even (iv) implies (iii), and so the two conditions are equivalent. This argument relies on the fact that $(K^\perp)^\perp = K$ for any closed linear subspace K.

To prove the last statement, put $\hat{\mathcal{F}} = \mathcal{F} \vee \mathcal{H}$ and $\hat{\mathcal{G}} = \mathcal{G} \vee \mathcal{H}$, and note that $\mathcal{F} \perp\!\!\!\perp_{\mathcal{H}} \mathcal{G}$ iff $\hat{\mathcal{F}} \perp\!\!\!\perp_{\mathcal{H}} \hat{\mathcal{G}}$ by Corollary 6.7 in [6]. By Proposition 6.6 in [6] this is equivalent to $E_{\hat{\mathcal{F}}} E_{\hat{\mathcal{G}}} = E_{\mathcal{H}} E_{\hat{\mathcal{G}}} = E_{\mathcal{H}}$, and by symmetry it is also equivalent to $E_{\hat{\mathcal{G}}} E_{\hat{\mathcal{F}}} = E_{\mathcal{H}}$. The assertion now follows by combination. □

We proceed with an interpretation of Palm measures in terms of conditional distributions, extending the elementary Lemma 3.4. Since the normalization of infinite Palm measures is arbitrary, we need to express the general relationship in terms of elementary conditioning, which is how the iterated conditioning again comes in.

Theorem 5.2. *Consider a random measure ξ on S and a random element η on T, where S and T are Borel and the Campbell measure $C_{\xi,\eta}$ is σ-finite. Fix a set $B \in \mathcal{T}$ such that $E[\xi;\, \eta \in B]$ is σ-finite, and put $A = \eta^{-1}B$. Then for τ and $\tilde{\eta}$ as in* (3.1)*, we have*

$$P[\,\cdot \,\|\, \xi]_s[\eta \in \cdot \,|A] = \tilde{P}[\tilde{\eta} \in \cdot \,|\, \tau, A]_s, \quad s \in S \ \text{ a.e. } E[\xi; A]. \tag{5.2}$$

The right-hand side of (5.2) can be interpreted in two different ways:

$$\tilde{P}[\tilde{\eta} \in \cdot \,|\, \tau, A] = \tilde{P}[\,\cdot \,|\, \tau][\tilde{\eta} \in \cdot \,|\tilde{A}] = \tilde{P}[\,\cdot \,|\tilde{A}][\tilde{\eta} \in \cdot \,|\, \tau], \tag{5.3}$$

where $\tilde{A} = \tilde{\eta}^{-1}B$. Since we are not assuming $\tilde{P} \circ \tau^{-1} = E\xi$ to be σ-finite, the measures $\tilde{P}[\,\cdot \,|\, \tau]_s$ may be infinite, and we need to define the kernel $\tilde{P}[\,\cdot \,|\, \tau]$ by general disintegration of the measure $\tilde{P} \circ (\tau, \tilde{\eta})^{-1} = C_{\xi,\eta}$. This makes $\tilde{P}[\,\cdot \,|\, \tau]$ a version of $P[\,\cdot \,\|\, \xi]$, and (5.2) becomes trivial under the first interpretation.

The second interpretation in (5.3) is more substantial and may be less obvious. The following result shows that the two versions are indeed equivalent. It also justifies our conditioning on A in (5.2) and (5.3).

Lemma 5.3. *Consider a measure ν on S and a kernel μ from S to T, where T is Borel and ν and μ are σ-finite, and put $M = \nu \otimes \mu$. Fix an $A \in \mathcal{T}$ such that $\hat{\nu} = M(\cdot \times A)$ is σ-finite, let $\hat{M}$ denote the restriction of M to $S \times A$, and choose a kernel $\hat{\mu}$ from S to A with $\hat{M} = \hat{\nu} \otimes \hat{\mu}$. Then*

$$\hat{\mu}_s = \mu_s[\,\cdot \,|A], \quad s \in S \ \text{ a.e. } \hat{\nu}. \tag{5.4}$$

Proof. Writing

$$\hat{\nu}B = M(B \times A) = \int_B \nu(ds)\,\mu_s A, \quad B \in \mathcal{S}, \tag{5.5}$$

we see that $\mu_s A > 0$ for $s \in S$ a.e. $\hat{\nu}$. Since $\hat{\nu}$ is σ-finite, we also have $\mu_s A < \infty$ a.s. ν, and this remains true for $\hat{\nu}$ since $\hat{\nu} \ll \nu$. Hence, $\mu_s A \in (0, \infty)$ for $s \in S$ a.e. $\hat{\nu}$, which shows that (5.4) makes sense.

Now define $\hat{S} = \{s \in S;\, \mu_s A \in (0, \infty)\}$, and conclude from (5.5) that $\nu = g \cdot \hat{\nu}$ on $\hat{S}$ with $g(s) = 1/\mu_s A$. Hence, on $\hat{S} \times A$ we have

$$\hat{\nu} \otimes \hat{\mu} = \hat{M} = M = \nu \otimes \mu = (g \cdot \hat{\nu}) \otimes \mu = \hat{\nu} \otimes g\mu,$$

and so the uniqueness in Lemma 2.2 yields

$$\hat{\mu}_s = g(s)\,\mu_s \text{ on } A, \quad s \in \hat{S} \text{ a.e. } \hat{\nu},$$

which is equivalent to (5.4). □

We conclude with some elementary, set-wise versions of Theorems 4.1 and 4.2, valid for arbitrary $\mathcal{F}$, $\mathcal{G}$, and Ω.

Proposition 5.4. *Given a probability space $(\Omega, \mathcal{A}, P)$ and some σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{A}$, fix any $A \in \mathcal{A}$ and $\alpha \in L^1(\mathcal{F} \vee \mathcal{G})$. Then these conditions are equivalent:*

(i) $\alpha = P_{\mathcal{F}\vee\mathcal{G}}(A)$ *a.s.,*
(ii) $E_{\mathcal{F}}[\alpha; G] = P_{\mathcal{F}}(A \cap G)$ *a.s.,* $G \in \mathcal{G}$,
(iii) $E_{\mathcal{G}}[\alpha; F] = P_{\mathcal{G}}(A \cap F)$ *a.s.,* $F \in \mathcal{F}$.

Note that (ii) and (iii) are weak versions of the statements $\alpha = (P_{\mathcal{F}})_{\mathcal{G}}(A)$ and $\alpha = (P_{\mathcal{G}})_{\mathcal{F}}(A)$, respectively. Thus, the equivalence of (i) − (iii) corresponds to Theorem 4.1 (i).

Proof. By symmetry it is enough to prove that (i) ⇔ (ii). Assuming (i) and letting $G \in \mathcal{G}$ be arbitrary, so that also $G \in \mathcal{F} \vee \mathcal{G}$, we get

$$E_{\mathcal{F}}[\alpha; G] = E_{\mathcal{F}} P_{\mathcal{F}\vee\mathcal{G}}(A \cap G) = P_{\mathcal{F}}(A \cap G) \text{ a.s.},$$

which proves (ii). To prove the converse, we need to show that the solution (i) to equation (ii) is a.s. unique. Thus, let α be such that $E_{\mathcal{F}}[\alpha; G] = 0$ a.s. for all $G \in \mathcal{G}$. Then clearly $E[\alpha;\, F \cap G] = 0$ for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, and so a monotone-class argument yields $E[\alpha; A] = 0$ for all $A \in \mathcal{F} \vee \mathcal{G}$. Since $\alpha \in L^1(\mathcal{F} \vee \mathcal{G})$, we obtain $\alpha = 0$ a.s. □

We turn to a set-wise version of Theorem 4.2, where we assume for simplicity that $E\xi$ is σ-finite.

Proposition 5.5. *Given a probability space $(\Omega, \mathcal{A}, P)$ and σ-field $\mathcal{F} \subset \mathcal{A}$, let ξ be a random measure on a Borel space S such that $E\xi$ is σ-finite, and fix any $A \in \mathcal{A}$ and $Y \in L^1(\mathcal{F} \otimes \mathcal{S},\, P \otimes \xi)$. Then these conditions are equivalent:*

(i) $Y = (P \otimes \xi)[A \times S \mid \mathcal{F} \otimes \mathcal{S}]$ *a.e.* $P \otimes \xi$,
(ii) $E[Y_s; F \,\|\, \xi]_s = P[A \cap F \,\|\, \xi]_s$ *a.e.* $E\xi$, $F \in \mathcal{F}$,
(iii) $\int_B E_{\mathcal{F}}\xi(ds)\, Y_s = E_{\mathcal{F}}[\xi B; A]$ *a.s.* P, $B \in \mathcal{S}$.

Here (i) means that $Y = (P \otimes \xi)_{\mathcal{F}\otimes\mathcal{S}}(A)$ a.e., while (ii) and (iii) are weak versions of the statements $Y = (P[\,\cdot\,\|\,\xi])_{\mathcal{F}}(A)$ and $Y = P_{\mathcal{F}}[A\,\|\,\xi]$, respectively. Hence, the equivalence of (i) – (iii) corresponds to Theorem 4.2 (i).

Proof. Condition (ii) means that

$$E\int_B \xi(ds)\, Y_s 1_F = E[\,\xi B;\, A\cap F\,], \quad B\in\mathcal{S},\ F\in\mathcal{F}. \tag{5.6}$$

From Lemma 4.3 (i) we see that (5.6) is equivalent to (iii). Furthermore, (5.6) extends by a monotone-class argument to

$$E\int \xi(ds)\, Y_s Z_s = E\int \xi(ds)\, Z_s 1_A,$$

for any indicator function Z on $\mathcal{F}\otimes\mathcal{S}$. Thus, (5.6) is also equivalent to (i). □

References

1. Daley, D. J. and Vere-Jones, D.: *An Introduction to the Theory of Point Processes II*, 2nd ed. Springer, New York, 2008.
2. Dellacherie, C. and Meyer, P. A.: *Probabilités et Potentiel II, Théorie des Martingales*, Hermann, Paris, 1980.
3. Dunford, N. and Schwartz, J. T.: *Linear Operators I, General Theory*, Wiley, New York, 1988.
4. Friedman, A. *Foundations of Modern Analysis*, Dover, New York, 1982.
5. Kallenberg, O.: *Random Measures*, 4th ed. Akademie-Verlag and Academic Press, Berlin and London, 1986.
6. Kallenberg, O.: *Foundations of Modern Probability*, 2nd ed. Springer, New York, 2002.
7. Matthes, K., Kerstan, J., and Mecke, J.: *Infinitely Divisible Point Processes.* Wiley, Chichester, 1978.
8. Meyer, P. A. and Yor, M.: Sur la théorie de la prédiction, et le problème de décomposition des tribus $\mathcal{F}^\circ_{t+}$, *Séminaire de Probabilités X*, *Lect. Notes in Math.* **511** pp. 104–117. Springer, Berlin, 1976.
9. Rogers, L. C. G. and Williams, D.: *Diffusions, Markov Processes, and Martingales I*, 2nd ed. Wiley, Chichester, 1994.
10. Ryll-Nardzewski, C.: Remarks on processes of calls. *Proc. 4th Berkeley Symp. Math. Statist. Probab.* **2**, (1961) 455–465, University of California Press, Berkeley.

Olav Kallenberg: Mathematics & Statistics Department, Auburn University, Auburn, Alabama 36849, USA

E-mail address: `kalleoh@auburn.edu`