

March 2021

Rate of Convergence in the Central Limit Theorem for iid Pareto Variables

Claas Becker

Hochschule RheinMain, 65022 Wiesbaden, Germany, claas.becker@hs-rm.de

Manuel Bohnet

Hochschule RheinMain, 65022 Wiesbaden, Germany, manuel_bohnet@web.de

Sarah Kummert

Hochschule RheinMain, 65022 Wiesbaden, Germany, sarah.kummert@gmx.de

Follow this and additional works at: <https://digitalcommons.lsu.edu/josa>



Part of the [Analysis Commons](#), and the [Other Mathematics Commons](#)

Recommended Citation

Becker, Claas; Bohnet, Manuel; and Kummert, Sarah (2021) "Rate of Convergence in the Central Limit Theorem for iid Pareto Variables," *Journal of Stochastic Analysis*: Vol. 2 : No. 1 , Article 3.

DOI: 10.31390/josa.2.1.03

Available at: <https://digitalcommons.lsu.edu/josa/vol2/iss1/3>

RATE OF CONVERGENCE IN THE CENTRAL LIMIT THEOREM FOR IID PARETO VARIABLES

CLAAS BECKER*, MANUEL BOHNET, AND SARAH KUMMERT

ABSTRACT. We estimate the rate of convergence in the central limit theorem for a sequence of iid Pareto variables X_k with shape parameter r . If $r \leq 4$, $E(|X_1|^3) = \infty$ and the Berry-Esseen theorem cannot be applied. In these cases the rate of convergence is very slow and can be expressed as a function of r .

1. Introduction

Let $(X_k)_{k \in \mathbb{N}}$ be a sequence of iid random variables with mean μ and variance σ^2 . By the central limit theorem, the standardized random sums

$$Z_n = \frac{1}{\sqrt{n\sigma^2}} \sum_{k=1}^n (X_k - \mu)$$

converge weakly to the standard normal distribution. The rate of convergence can be measured in terms of

$$\|F_n - \Phi\|_\infty = \sup_{x \in \mathbb{R}} \{|F_n(x) - \Phi(x)|\}$$

where F_n is the cumulative distribution function of Z_n and Φ is the cumulative distribution function of the standardized normal distribution. It is well known that without further assumptions on the distribution of the X_k , the rate of convergence cannot be faster than $1/\sqrt{n}$, see [3], p.448.

If $E(|X_1|^3) < \infty$, then, by the Berry-Esseen theorem,

$$\|F_n - \Phi\|_\infty \leq \frac{CE(|X_1|^3)}{\sigma^3 \sqrt{n}}.$$

In recent years, significant progress has been made to obtain sharp estimates for the constant C , see [2], [4], [5] and the references therein.

In this paper, we focus on iid random variables with a Pareto distribution. Their probability density function is given by

$$f(x) = \begin{cases} (r-1) x_{\min}^{r-1} x^{-r}, & x \geq x_{\min} > 0 \\ 0, & \text{otherwise} \end{cases} \quad (1.1)$$

Received 2020-11-29; Accepted 2021-1-21; Communicated by the editors.

2010 *Mathematics Subject Classification.* Primary 60F05, Secondary 60E99.

Key words and phrases. Pareto distribution, central limit theorem, Berry-Esseen theorem.

* Corresponding author.

with shape parameter $r > 1$ and location parameter x_{\min} . We are interested in tail behaviour and therefore, for simplicity, put $x_{\min} = 1$.

If $r \leq 4$, $E(|X_1|^3) = \infty$ and the Berry-Esseen theorem cannot be applied. For the same reason, an Edgeworth expansion does not work. Instead, we rely on the technique of truncated moments.

2. The Truncation Method

For each X_k , we take any $\hat{\tau}_k, \tau_k > 0$ and define the truncated random variables

$$\bar{X}_k = X_k \cdot 1_{]-\hat{\tau}_k, \tau_k[} \quad \text{and} \quad X'_k = X_k - \bar{X}_k .$$

The following theorem is due to Feller [1].

Theorem 2.1. *Let $(X_k)_{k \in \mathbb{N}}$ be a sequence of mutually independent random variables. We assume $E(X_k) = 0$ and $E(X_k^2) = \sigma_k^2 < \infty$ for all $k \in \mathbb{N}$. Put*

$$\begin{aligned} \beta'_k &= E(X_k'^2) & b' &= \beta'_1 + \dots + \beta'_n \\ \gamma_k &= E(|\bar{X}_k|^3) & c &= \gamma_1 + \dots + \gamma_n \\ s^2 &= \sigma_1^2 + \dots + \sigma_n^2 . \end{aligned}$$

Then

$$\|F_n - \Phi\|_\infty \leq 6 \left(\frac{c}{s^3} + \frac{b'}{s^2} \right) \quad (2.1)$$

where F_n is the cumulative distribution function of the standardized sum

$$\frac{1}{\sqrt{s^2}} \sum_{k=1}^n X_k .$$

Note that the dependency on n is hidden in the constants b' , c and s^2 .

3. Rate of Convergence

Theorem 3.1. *Assume that $(Y_k)_{k \in \mathbb{N}}$ are iid Pareto with shape parameter $r > 3$ and location parameter $x_{\min} = 1$. We put $\mu = E(Y_n)$, $X_n = Y_n - \mu$ and $\sigma^2 = E(X_n^2)$. Let F_n denote the cumulative distribution function of*

$$Z_n = \frac{1}{\sqrt{n\sigma^2}} \sum_{k=1}^n X_k .$$

Then there exist constants $c_1, c_2, c_3 > 0$ such that the asymptotic behaviour¹ of the bound in inequality (2.1) is given by

$$\|F_n - \Phi\|_\infty \leq 6 \left(\frac{c}{s^3} + \frac{b'}{s^2} \right) \sim \begin{cases} c_1 n^{\frac{1}{2}(3-r)}, & 3 < r < 4 \\ c_2 n^{-\frac{1}{2}} \ln n, & r = 4 \\ c_3 n^{-\frac{1}{2}}, & r > 4. \end{cases}$$

¹ $a_n \sim b_n$ iff $\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = 1$

Proof. We have

$$\mu = E(Y_k) = \frac{r-1}{r-2} \quad \text{and} \quad E(Y_k^2) = \frac{r-1}{r-3}.$$

The probability density function of X_k is

$$f(x) = \begin{cases} (r-1)(x+\mu)^{-r}, & x \geq 1-\mu \\ 0, & \text{otherwise.} \end{cases}$$

Since the shifted Pareto distributions have support in the interval $[1-\mu, \infty[$, we can use symmetric truncation $\bar{X}_k = X_k \cdot 1_{] -\tau_k, \tau_k[}$. We try the ansatz

$$\tau_k = k^\alpha \quad \text{with} \quad \alpha > 0.$$

The main challenge is to choose α in such a way that both terms $\frac{c}{s^3}$ and $\frac{b'}{s^2}$ in inequality (2.1) decay sufficiently fast.

(1) *The case $3 < r < 4$:*

We compute β'_k and γ_k . A simple computation shows that

$$\begin{aligned} \beta'_k = E(X_k'^2) &= (r-1) \int_{\tau_k}^{\infty} x^2 (x+\mu)^{-r} dx \\ &= (r-1) \int_{\tau_k+\mu}^{\infty} (x-\mu)^2 x^{-r} dx \\ &= -(r-1) \left(\frac{\tilde{\tau}_k^{3-r}}{3-r} - 2\mu \frac{\tilde{\tau}_k^{2-r}}{2-r} + \mu^2 \frac{\tilde{\tau}_k^{1-r}}{1-r} \right) \end{aligned} \tag{3.1}$$

with the abbreviation $\tilde{\tau}_k = \tau_k + \mu$. Therefore

$$b' = \sum_{k=1}^n \beta'_k \leq -(r-1) \sum_{k=1}^n \left(\frac{\tilde{\tau}_k^{3-r}}{3-r} + \mu^2 \frac{\tilde{\tau}_k^{1-r}}{1-r} \right).$$

$$\begin{aligned} \gamma_k &= E(|\bar{X}_k|^3) \\ &= -2(r-1) \left(\frac{\mu^{4-r}}{4-r} - 3\mu \frac{\mu^{3-r}}{3-r} + 3\mu^2 \frac{\mu^{2-r}}{2-r} - \mu^3 \frac{\mu^{1-r}}{1-r} \right) \\ &\quad + (r-1) \left(\frac{\tilde{\tau}_k^{4-r} + 1}{4-r} - 3\mu \frac{\tilde{\tau}_k^{3-r} + 1}{3-r} + 3\mu^2 \frac{\tilde{\tau}_k^{2-r} + 1}{2-r} - \mu^3 \frac{\tilde{\tau}_k^{1-r} + 1}{1-r} \right) \end{aligned} \tag{3.2}$$

Therefore the following inequality holds

$$c = \sum_{k=1}^n \gamma_k \leq (r-1) \sum_{k=1}^n \left(\frac{\tilde{\tau}_k^{4-r}}{4-r} - 3\mu \frac{\tilde{\tau}_k^{3-r}}{3-r} - \mu^3 \frac{\tilde{\tau}_k^{1-r}}{1-r} \right) + n(r-1)(\theta_1 - 2\theta_2)$$

with constants

$$\theta_1 = \frac{1}{4-r} - 3\mu \frac{1}{3-r} - \mu^3 \frac{1}{1-r} \quad \text{and} \quad \theta_2 = 3\mu^2 \frac{\mu^{2-r}}{2-r}.$$

Using $\tilde{\tau}_k = \tau_k + \mu = k^\alpha + \mu$, we arrive at the inequalities

$$\tilde{\tau}_k^{4-r} \leq 3k^{\alpha(4-r)} \quad \tilde{\tau}_k^{3-r} \leq k^{\alpha(3-r)} \quad \tilde{\tau}_k^{1-r} \leq k^{\alpha(1-r)} .$$

Therefore

$$\begin{aligned} c + sb' &\leq (r-1) \left(\frac{3}{4-r} \sum_{k=1}^n k^{\alpha(4-r)} - \frac{3\mu}{3-r} \sum_{k=1}^n k^{\alpha(3-r)} - \frac{\mu^3}{1-r} \sum_{k=1}^n k^{\alpha(1-r)} \right) \\ &\quad + n(r-1)(\theta_1 - 2\theta_2) \\ &\quad - s(r-1) \left(\frac{1}{3-r} \sum_{k=1}^n k^{\alpha(3-r)} + \frac{\mu^2}{1-r} \sum_{k=1}^n k^{\alpha(1-r)} \right) . \end{aligned}$$

Note that s is not a constant since $s^2 = \sum_{k=1}^n \sigma^2 = n\sigma^2$. Approximating integrals by Riemann sums, we have

$$\sum_{k=1}^{n-1} k^{\alpha(4-r)} \leq \int_1^n x^{\alpha(4-r)} dx = \frac{n^{\alpha(4-r)+1} - 1}{\alpha(4-r) + 1} .$$

If $\alpha \neq -\frac{1}{3-r}$, we have

$$\sum_{k=2}^n k^{\alpha(3-r)} \leq \int_1^n x^{\alpha(3-r)} dx = \frac{n^{\alpha(3-r)+1} - 1}{\alpha(3-r) + 1} .$$

Analogously, if $\alpha \neq -\frac{1}{1-r}$, we obtain

$$\sum_{k=2}^n k^{\alpha(1-r)} \leq \int_1^n x^{\alpha(1-r)} dx = \frac{n^{\alpha(1-r)+1} - 1}{\alpha(1-r) + 1} .$$

Combining these inequalities, we arrive at

$$\begin{aligned} 6 \frac{c + sb'}{s^3} &\leq \frac{6\beta_1}{\sigma^3} \left(3 \frac{n^{\alpha(4-r)-\frac{1}{2}} - n^{-\frac{3}{2}}}{\beta_2} + 3 \frac{n^{\alpha(4-r)-\frac{3}{2}}}{4-r} - 3\mu \frac{n^{\alpha(3-r)-\frac{1}{2}} - n^{-\frac{3}{2}}}{\beta_3} \right. \\ &\quad \left. - n^{-\frac{3}{2}} \frac{3\mu}{3-r} - \mu^3 \frac{n^{\alpha(1-r)-\frac{1}{2}} - n^{-\frac{3}{2}}}{\beta_4} - n^{-\frac{3}{2}} \frac{\mu^3}{1-r} + n^{-\frac{1}{2}}(\theta_1 - 2\theta_2) \right. \\ &\quad \left. - \sigma \frac{n^{\alpha(3-r)} - n^{-1}}{\beta_3} - \sigma \frac{n^{-1}}{3-r} - \sigma \mu^2 \frac{n^{\alpha(1-r)} - n^{-1}}{\beta_4} - \sigma \frac{n^{-1}\mu^2}{1-r} \right) \end{aligned}$$

with constants

$$\begin{aligned} \beta_1 &= r-1 & \beta_2 &= \alpha(4-r)^2 + 4-r \\ \beta_3 &= \alpha(3-r)^2 + 3-r & \beta_4 &= \alpha(1-r)^2 + 1-r . \end{aligned}$$

It turns out that, as $n \rightarrow \infty$, only the two terms containing $n^{\alpha(4-r)-\frac{1}{2}}$ and $n^{\alpha(3-r)}$ ultimately determine the rate of convergence. For fast convergence of the first term, it would be best to put $\alpha = 0$. But for $\alpha = 0$ the second term does not converge at all. Comparing the exponents in both terms, we find that $\alpha = \frac{1}{2}$ is optimal. This choice for α is consistent with the restrictions $\alpha \neq -\frac{1}{3-r}$ and $\alpha \neq -\frac{1}{1-r}$ which we made before.

A review of our previous computations in the cases $\alpha = -\frac{1}{3-r}$ and $\alpha = -\frac{1}{1-r}$ shows that the term containing $n^{\alpha(4-r)-\frac{1}{2}}$ does not vanish. Therefore these choices for α do not improve the asymptotic estimate.

(2) *The case $r > 4$:*

Equations (3.1) and (3.2) also hold in this case. We arrive at the inequalities

$$b' = \sum_{k=1}^n \beta'_k \leq -(r-1) \sum_{k=1}^n \left(\frac{\tilde{\tau}_k^{3-r}}{3-r} + \mu^2 \frac{\tilde{\tau}_k^{1-r}}{1-r} \right)$$

$$c = \sum_{k=1}^n \gamma_k \leq (r-1) \sum_{k=1}^n \left(-3\mu \frac{\tilde{\tau}_k^{3-r}}{3-r} - \mu^3 \frac{\tilde{\tau}_k^{1-r}}{1-r} \right) + n(r-1)(\bar{\theta}_1 - 2\bar{\theta}_2)$$

with constants

$$\bar{\theta}_1 = \frac{3\mu}{r-3} + \frac{\mu^3}{r-1}$$

$$\bar{\theta}_2 = \frac{\mu^{4-r}}{4-r} + 3\mu^2 \frac{\mu^{2-r}}{2-r}.$$

With the inequalities

$$\tilde{\tau}_k^{3-r} \leq k^{\alpha(3-r)} \quad \text{and} \quad \tilde{\tau}_k^{1-r} \leq k^{\alpha(1-r)}$$

we arrive at

$$c + sb' \leq (r-1) \left(-\frac{3\mu}{3-r} \sum_{k=1}^n k^{\alpha(3-r)} - \frac{\mu^3}{1-r} \sum_{k=1}^n k^{\alpha(1-r)} \right)$$

$$+ n(r-1)(\bar{\theta}_1 - 2\bar{\theta}_2)$$

$$- s(r-1) \left(\frac{1}{3-r} \sum_{k=1}^n k^{\alpha(3-r)} + \frac{\mu^2}{1-r} \sum_{k=1}^n k^{\alpha(1-r)} \right).$$

Again, approximating integrals by Riemann sums, we obtain, if $\alpha \neq -\frac{1}{3-r}$ and $\alpha \neq -\frac{1}{1-r}$

$$\sum_{k=2}^n k^{\alpha(3-r)} \leq \int_1^n x^{\alpha(3-r)} dx = \frac{n^{\alpha(3-r)+1} - 1}{\alpha(3-r) + 1}$$

$$\sum_{k=2}^n k^{\alpha(1-r)} \leq \int_1^n x^{\alpha(1-r)} dx = \frac{n^{\alpha(1-r)+1} - 1}{\alpha(1-r) + 1}.$$

This leads to the inequality

$$6 \frac{c + sb'}{s^3} \leq \frac{6\beta_1}{\sigma^3} \left(-3\mu \frac{n^{\alpha(3-r)-\frac{1}{2}} - n^{-\frac{3}{2}}}{\beta_3} - \frac{3\mu n^{-\frac{3}{2}}}{3-r} - \mu^3 \frac{n^{\alpha(1-r)-\frac{1}{2}} - n^{-\frac{3}{2}}}{\beta_4} \right. \\ \left. - \frac{\mu^3 n^{-\frac{3}{2}}}{1-r} + n^{-\frac{1}{2}}(\bar{\theta}_1 - 2\bar{\theta}_2) \right. \\ \left. - \sigma \frac{n^{\alpha(3-r)} - n^{-1}}{\beta_3} - \sigma \frac{n^{-1}}{3-r} - \sigma \mu^2 \frac{n^{\alpha(1-r)} - n^{-1}}{\beta_4} - \sigma \frac{n^{-1}\mu^2}{1-r} \right)$$

with constants β_1, β_3 , and β_4 as before.

It is the term containing $n^{-\frac{1}{2}}$ that determines the rate of convergence. We can, for instance, put $\alpha = 1$ which is in line with the conditions $\alpha \neq -\frac{1}{3-r}$ and $\alpha \neq -\frac{1}{1-r}$.

A review of our previous computations in the cases $\alpha = -\frac{1}{3-r}$ and $\alpha = -\frac{1}{1-r}$ shows that the term containing $n^{-\frac{1}{2}}$ does not vanish. Therefore these choices for α do not improve the asymptotic estimate.

(3) *The case $r = 4$:*

In this case we have

$$\beta'_k = 3 \left(\tilde{\tau}_k^{-1} - \mu \tilde{\tau}_k^{-2} + \frac{\mu^2}{3} \tilde{\tau}_k^{-3} \right) \leq 3 \left(\tilde{\tau}_k^{-1} + \frac{\mu^2}{3} \tilde{\tau}_k^{-3} \right)$$

with $\tilde{\tau}_k = \tau_k + \mu$ as before.

$$\gamma_k = \int_{-\infty}^{\infty} |y|^3 3 (y + \mu)^{-4} 1_{[1-\mu, \tau_k](y)} dy \leq 9\mu + \mu^3 + 3 \left(\ln \tilde{\tau}_k + 3\mu \tilde{\tau}_k^{-1} + \frac{\mu^3}{3} \tilde{\tau}_k^{-3} \right)$$

where no attempt at sharp estimates for constants has been made. With the inequalities

$$\ln \tilde{\tau}_k \leq \alpha \ln k + \ln 3 \quad \tilde{\tau}_k^{-1} \leq k^{-\alpha} \quad \tilde{\tau}_k^{-3} \leq k^{-3\alpha}$$

we obtain

$$c + sb' \leq 3\alpha \sum_{k=1}^n \ln k + 9\mu \sum_{k=1}^n k^{-\alpha} + \mu^3 \sum_{k=1}^n k^{-3\alpha} + n(3 \ln 3 + \theta) \\ + 3s \sum_{k=1}^n k^{-\alpha} + s\mu^2 \sum_{k=1}^n k^{-3\alpha}$$

with the constant $\theta = 9\mu + \mu^3$. It is the sum containing $\ln k$ that makes the case $r = 4$ different. Approximating integrals by Riemann sums, we have, if $\alpha \neq 1$ and

$\alpha \neq \frac{1}{3}$,

$$\begin{aligned} \sum_{k=1}^{n-1} \ln k &\leq \int_1^n \ln x \, dx = n \ln n - n + 1 \\ \sum_{k=2}^n k^{-\alpha} &\leq \int_1^n x^{-\alpha} \, dx = \frac{n^{1-\alpha} - 1}{1-\alpha} \\ \sum_{k=2}^n k^{-3\alpha} &\leq \int_1^n x^{-3\alpha} \, dx = \frac{n^{1-3\alpha} - 1}{1-3\alpha}. \end{aligned}$$

Using the equation $s^2 = \sum_{k=1}^n \sigma^2 = n\sigma^2$, we obtain the inequality

$$\begin{aligned} &6 \frac{c + sb'}{s^3} \\ &\leq \frac{6}{\sigma^3} \left(3\alpha (\ln n (n^{-\frac{1}{2}} + n^{-\frac{3}{2}})) - n^{-\frac{1}{2}} + n^{-\frac{3}{2}} \right) + 9\mu \left(\frac{n^{-\frac{1}{2}-\alpha} - n^{-\frac{3}{2}}}{1-\alpha} + n^{-\frac{3}{2}} \right) \\ &\quad + \mu^3 \left(\frac{n^{-\frac{1}{2}-3\alpha} - n^{-\frac{3}{2}}}{1-3\alpha} + n^{-\frac{3}{2}} \right) + n^{-\frac{1}{2}} (3 \ln 3 + \theta) \\ &\quad + 3\sigma \left(\frac{n^{-\alpha} - n^{-1}}{1-\alpha} + n^{-1} \right) + \sigma\mu^2 \left(\frac{n^{-3\alpha} - n^{-1}}{1-3\alpha} + n^{-1} \right). \end{aligned}$$

If we want the term containing $\ln n n^{-\frac{1}{2}}$ to vanish, we have to put $\alpha = 0$. However, in this case the term $n^{-\alpha}$ does not converge any more. Therefore the choice $\alpha = \frac{1}{2}$ is optimal.

A review of our computations shows that in the cases $\alpha = 1$ and $\alpha = \frac{1}{3}$ the term containing $\ln n n^{-\frac{1}{2}}$ does not vanish so that the estimate is not improved. \square

4. Concluding Remarks

If $r \leq 3$, the variance of the Pareto distribution is infinite and the formulation of the central limit theorem in terms of standardized sums is not possible.

In the case $3 < r < 4$, $E(|X_1|^3) = \infty$ and convergence is slower than in the Berry-Esseen theorem. Note that the bound in theorem 3.1 is a monotonously decreasing function of r . As r approaches 3 from above, the rate of convergence can be arbitrarily slow. For instance, in the case $r = 3.5$, the rate of convergence is proportional to $n^{-\frac{1}{4}}$. Of course, using this methodology, we have just obtained upper bounds. However, we have done extensive computer simulations that indicate that the rate of convergence in this case is indeed proportional to $n^{-\frac{1}{4}}$. Though we have taken great care of numerical inaccuracies, computer simulations should always be taken with caution.

The case $r = 4$ is borderline because for $r \leq 4$ $E(|X_1|^3) = \infty$, but for $r > 4$ $E(|X_1|^3) < \infty$. For $r > 4$ but close to 4, the rate of convergence is faster than for $r = 4$. Conversely, for $r < 4$ but close to 4, the rate of convergence is slower than

for $r = 4$. This is true because for large n

$$n^{-\frac{1}{2}} \ln n \leq n^{-\frac{1}{2}+\epsilon}$$

since $\lim_{n \rightarrow \infty} \frac{\ln n}{n^\epsilon} = 0$, which can be seen by applying l'Hôpital's rule.

If $r > 4$, $E(|X_1|^3) < \infty$, and our estimate in theorem 3.1 gives the same rate of convergence as the Berry-Esseen theorem.

References

1. Feller, W.: On the Berry-Esseen Theorem, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **10** (1968), 261–268.
2. Shevtsova, I.: On the absolute constants in the Berry-Esseen type inequalities for identically distributed summands, arXiv:1111.6554v1.
3. Shiryaev, A. N.: *Probability*, Vol. 1, 3rd ed., Springer, 2016.
4. Tyurin, I. S.: On the accuracy of the Gaussian approximation, *Dokl. Math.* **80** (2009), no. 3, 840–843, <https://doi.org/10.1134/S1064562409060155>.
5. Tyurin, I. S.: Refinement of the upper bounds of the constants in Lyapunov's theorem, *Russ. Math. Surv.* **65** (2010), no. 3, 586–588.

CLAAS BECKER: HOCHSCHULE RHEINMAIN, 65022 WIESBADEN, GERMANY
Email address: `claas.becker@hs-rm.de`

MANUEL BOHNET: HOCHSCHULE RHEINMAIN, 65022 WIESBADEN, GERMANY
Email address: `manuel_bohnet@web.de`

SARAH KUMMERT: HOCHSCHULE RHEINMAIN, 65022 WIESBADEN, GERMANY
Email address: `sarah.kummert@gmx.de`