

INDIAN STATISTICAL INSTITUTE, KOLKATA
MASTER OF STATISTICS



Multivariate Analysis

Final Report — Group G

**Analysis of multivariate heavy-tailed
distributions using Hill estimation**

11 NOVEMBER 2025

Adityarup Laha, Deeptendra Banerjee, Swapnamoy Samadder

Advisor/Instructor

DR. MINERVA MUKHOPADHYAY
ISRU, INDIAN STATISTICAL INSTITUTE, KOLKATA



Contents

1	Introduction	1
2	Prerequisites	1
2.1	Radon Measures	1
2.2	Vague Topology	2
2.3	The Poisson Process	2
2.4	The Laplace functional	3
2.5	Laplace functional of PRM	4
3	Heavy-tailed Measures and The Hill Estimator	4
3.1	Regular Variation	4
3.2	The Hill Estimator	6
3.3	Choosing the optimal k_n	7
4	Hill Estimation in Multivariate Contexts	7
4.1	Generalising heavy-tailed measures	8
4.2	Estimation for the standard case: common tail parameter	9
4.3	Estimation for the non-standard case: differing tail parameters	10
5	Concluding Remarks	11



1 Introduction

Heavy-tailed distributions are prevalent in a wide range of real-world phenomena, such as finance, telecommunications, natural disasters, and social sciences. These distributions are characterized by a slower decay of their tails, leading to higher probabilities of extreme values compared to their light-tailed counterparts. Understanding and estimating the tail behaviour is critical for accurately modelling rare events, assessing risks, and designing robust systems. For example, in risk management, capturing the heavy-tail behaviour is essential for metrics like value-at-risk or extreme loss probabilities.

A common method for estimating the tail index, a parameter quantifying the tail's heaviness, is the Hill estimator. Introduced within the framework of extreme value theory, the Hill estimator is designed to work with heavy-tailed distributions, specifically those in the domain of attraction of a Pareto-like tail. It relies on the asymptotic behaviour of order statistics, focusing on the largest observations in the dataset. Despite its simplicity, the Hill estimator requires careful tuning, particularly in selecting the threshold that separates the tail from the bulk of the distribution.

Before delving into the specifics of Hill estimation, it is essential to first understand several foundational concepts. The following section will provide the necessary theoretical background to frame and justify the application of the Hill estimator.

2 Prerequisites

2.1 Radon Measures

Definition 2.1. Let μ be a measure on the σ -algebra of Borel sets of a Hausdorff topological space (any two distinct points can have disjoint neighbourhoods) S .

- The measure μ is called **inner regular** or **tight** if, for every open set U ,

$$\mu(U) = \sup\{\mu(K) \mid K \text{ is a compact subset of } U\}$$

- The measure μ is called **outer regular** if, for every Borel set B ,

$$\mu(B) = \inf\{\mu(U) \mid U \text{ is an open set containing } B\}.$$

- The measure μ is called **locally finite** if every point of X has a neighborhood U for which $\mu(U)$ is finite.

If μ is locally finite, then it follows that μ is finite on compact sets, and for locally compact Hausdorff spaces, the converse holds, too. Thus, in this case, local finiteness may be equivalently replaced by finiteness on compact subsets.

Definition 2.2 (Radon measure). The measure μ is called a **Radon measure** if it is inner regular and locally finite.

In many situations, such as finite measures on locally compact spaces, this also implies outer regularity. We shall be working with non-negative Radon measures, in particular, counting measures.



Take a locally compact, separable topological space \mathcal{S} . Define $M_+(\mathcal{S})$, the set of Radon measures on \mathcal{S} ; and $M_p(\mathcal{S})$, the set of counting measures on \mathcal{S} .

For a nonnegative, bounded measurable function $f : \mathcal{S} \rightarrow \mathbb{R}_+$ and $\mu \in M_+(\mathcal{S})$, we use the notation

$$\underline{\mu}(f) = \int_{x \in \mathcal{S}} f(x) \mu(dx).$$

and for $m \in M_p(\mathcal{S})$; $m(\cdot) = \sum_i \mathbf{1}\{\cdot \ni x_i\}$,

$$\underline{m}(f) = \sum_i f(x_i).$$

2.2 Vague Topology

Let $C_K^+(\mathcal{S})$ be the set of all compactly supported non-negative continuous functions on \mathcal{S} .

Definition 2.3 (Vague Metric on the set of Radon Measures). *Let $\mu_1, \mu_2 \in M_+(\mathcal{S})$ and $\{f_n\}_{n \in \mathbb{N}}$ be a sequence in $C_K(\mathcal{S})$ such that $\forall i \ \|f_i\|_{sup} \leq 1$ and $\forall i \neq j, \text{int}(\text{support}(f_i)) \cap \text{int}(\text{support}(f_j)) = \emptyset$. Then*

$$d_v(\mu_1, \mu_2) = \sup_{\{f_n\}} \sum_{i=1}^{\infty} \frac{1}{2^i} \min \left\{ \|\underline{\mu}_1(f_i) - \underline{\mu}_2(f_i)\|, 1 \right\}$$

Then, we define $\mathcal{M}_+(\mathcal{S})$, the family of Borel sets of $M_+(\mathcal{S})$ generated by the vague metric; and $\mathcal{M}_p(\mathcal{S})$, the family of Borel sets of $M_p(\mathcal{S})$ inherited from the vague metric.

An open set inherited from vague metric looks like

$$\{\mu \in M_+(\mathcal{S}) : \underline{\mu}(f_i) \in (a_i, b_i), 0 \leq a_i < b_i \in \mathbb{R}, \forall f_i, i \in \mathbb{N}\}$$

See also: https://en.wikipedia.org/wiki/Vague_topology

Definition 2.4 (Vague Convergence). *A sequence of Radon measures $\{\mu_n\}_{n \in \mathbb{N}}$ is said to vaguely converge to μ if $\lim_{n \rightarrow \infty} d_v(\mu_n, \mu) \rightarrow 0$.*

2.3 The Poisson Process

A point process on \mathcal{S} is any measurable function of the form $N : (\Omega, \mathcal{A}) \rightarrow (M_p(\mathcal{S}), \mathcal{M}_p(\mathcal{S}))$

Definition 2.5 (Poisson Process). *N is called a Poisson process with mean measure μ , or equivalently, a Poisson random measure $\text{PRM}(\mu)$, if the following conditions hold:*

1. For $A \in \mathcal{B}(\mathcal{S})$ (viz. the Borel subsets of \mathcal{S}),

$$\mathbb{P}[N(A) = k] = \begin{cases} \frac{e^{-\mu(A)} (\mu(A))^k}{k!} & , \text{ if } \mu(A) < \infty \\ 0 & , \text{ if } \mu(A) = \infty \end{cases}$$

2. If A_1, \dots, A_k are disjoint subsets of \mathcal{S} in $\mathcal{B}(\mathcal{S})$, then $N(A_1), \dots, N(A_k)$ are independent random variables.



2.4 The Laplace functional

Suppose \mathcal{B}_+ denotes the set of nonnegative, bounded, measurable functions from $\mathcal{S} \rightarrow \mathbb{R}_+$, and let

$$M : (\Omega, \mathcal{A}, \mathbb{P}) \mapsto (M_+(\mathcal{S}), \mathcal{M}_+(\mathcal{S}))$$

be a random Radon measure (i.e., a random element of $M_+(\mathcal{S})$).

Definition 2.6 (Laplace Functional). *The Laplace functional of M is*

$$\begin{aligned} \Psi_M(f) &= \mathbb{E} \exp\{-\underline{M}(f)\} = \int_{\Omega} \exp\{-\underline{M}(\omega, f)\} d\mathbb{P}(\omega) \\ &= \int_{M_+(\mathcal{S})} \exp\{-\mu(f)\} \mathbb{P} \circ M^{-1}(d\mu) \\ &= \int_{\mathbb{R}^+} \exp\{-z\} d\mathbb{P}(\{\underline{M}(f) \leq z\}). \end{aligned}$$

Note that if P is a probability measure on $\mathcal{M}_+(\mathcal{S})$, its Laplace functional is given by

$$\int_{M_+(\mathcal{S})} \exp\{-\underline{\mu}(f)\} P(d\mu), \quad f \in \mathcal{B}_+.$$

Theorem 2.1. *If M is a random measure on \mathcal{S} , the Laplace functional $\Psi_M(f)$, $f \in C_K^+(\mathcal{S})$, uniquely determines the distribution of M .*

Proof. The distribution of M is the measure $\mathbb{P} \circ M^{-1}$ on $\mathcal{M}_+(\mathcal{S})$, the Borel σ -algebra generated by the open subsets of $M_+(\mathcal{S})$. Recall that $\mathcal{G}(M_+(\mathcal{S}))$, the class of open subsets of $M_+(\mathcal{S})$, is generated by the class \mathcal{C} of basis sets, which has a typical member

$$\{\mu \in M_+(\mathcal{S}) : \mu(f_i) \in (a_i, b_i), i = 1, \dots, d\},$$

for $f_i \in C_K^+(\mathcal{S})$, $i = 1, \dots, d$. The class \mathcal{C} is closed under finite intersections and hence is a Π -system generating the Borel σ -algebra. By Dynkin's π - λ theorem, it suffices to show the Laplace functional uniquely determines probabilities on \mathcal{C} .

Since

$$\mathbb{P} \circ M^{-1} \{\mu \in M_+(\mathcal{S}) : \mu(f_i) \in (a_i, b_i), i = 1, \dots, d\} = \mathbb{P} [\underline{M}(f_i) \in (a_i, b_i), i = 1, \dots, d],$$

it suffices to show $\Psi_M(\cdot)$ determines the joint distribution of

$$(\underline{M}(f_1), \dots, \underline{M}(f_d)).$$

This joint distribution is determined by its Laplace transform

$$\mathbb{E} \exp \left\{ - \sum_{i=1}^d \lambda_i \underline{M}(f_i) \right\}, \quad \lambda_i > 0, \quad i = 1, \dots, d.$$

However, this is equal to

$$\mathbb{E} \exp \left\{ - \underline{M} \left(\sum_{i=1}^d \lambda_i f_i \right) \right\} = \Psi_M \left(\sum_{i=1}^d \lambda_i f_i \right).$$



Since $\sum_{i=1}^d \lambda_i f_i \in C_K^+(\mathcal{S})$, knowledge of $\{\Psi_M(f); f \in C_K^+(\mathcal{S})\}$ is determining. ■

2.5 Laplace functional of PRM

The distribution of $\text{PRM}(\mu)$ is uniquely determined by conditions 1 and 2 in Definition 2.5. Furthermore, the point process N is $\text{PRM}(\mu)$ if and only if its Laplace functional is of the form

$$\Psi_N(f) = \exp \left\{ - \int_{\mathcal{S}} \left(1 - e^{-f(x)} \right) \mu(dx) \right\}, \quad f \in \mathcal{B}_+.$$

Theorem 2.2 (Convergence Criterion). *Let $\{\eta_n, n \geq 0\}$ be random elements of $M_+(\mathcal{S})$. Then*

$$\eta_n \Rightarrow \eta_0 \quad \text{in } M_p(\mathcal{S})$$

if and only if

$$\Psi_{\eta_n}(f) = \mathbb{E} e^{-\eta_n(f)} \rightarrow \mathbb{E} e^{-\eta_0(f)} = \Psi_{\eta_0}(f), \quad \forall f \in C_K^+(\mathcal{S}).$$

Thus, weak convergence is characterized by convergence of Laplace functionals on $C_K^+(\mathcal{S})$.

Theorem 2.3. *Suppose that for $n \geq 1$ we have $X_{n,j}, j \geq 1$ is a sequence of iid random elements of $(\mathcal{S}, \mathcal{B})$. Let γ be $\text{PRM}(\mu)$ on $M_p(\mathcal{S})$. We have*

$$\sum_{i=1}^n \mathbf{1}\{\cdot \ni X_{n,i}\} \Longrightarrow \gamma$$

on $M_p(\mathcal{S})$ iff

$$n\mathbb{P}[X_{n,1} \in \cdot] \xrightarrow{v} \mu$$

Suppose additionally that $0 < a_n < \infty$. Then for measure $\mu \in M_+(\mathcal{S})$ we have

$$\frac{1}{a_n} \sum_{j=1}^n \mathbf{1}\{\cdot \ni X_{n,j}\} \Longrightarrow \mu$$

on $M_+(\mathcal{S})$ iff

$$\frac{n}{a_n} \mathbb{P}[X_{n,1} \in \cdot] \xrightarrow{v} \mu$$

3 Heavy-tailed Measures and The Hill Estimator

3.1 Regular Variation

Definition 3.1 (Regularly varying function). *A measurable function $g : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ (written $g \in \text{RV}_\rho$) if for $x > 0$,*

$$\lim_{t \rightarrow \infty} \frac{g(tx)}{g(t)} = x^\rho$$

We call ρ the exponent of variation.

Definition 3.2 (Heavy-tailed distribution). *Suppose $X \sim F$, a CDF. We say X is heavy-tailed with tail parameter $\alpha > 0$, if $1 - F$ is a regularly varying function with exponent of variation $-\alpha$, i.e. $1 - F \in \text{RV}_{-\alpha}$.*



Theorem 3.1. Suppose X is a nonnegative random variable with CDF F . Set $\bar{F} = 1 - F$. The following are equivalent:

(a) $\bar{F} \in RV_{-\alpha}$, $\alpha > 0$.

(b) There exists a sequence $\{b_n\}$ with $b_n \rightarrow \infty$ such that

$$\lim_{n \rightarrow \infty} n\bar{F}(b_n x) = x^{-\alpha}, \quad x > 0.$$

(c) There exists a sequence $\{b_n\}$ with $b_n \rightarrow \infty$ such that

$$\mu_n(\cdot) := n\mathbb{P}\left[\frac{X}{b_n} \in \cdot\right] \xrightarrow{v} \nu_\alpha(\cdot)$$

in $M_+(0, \infty]$, where $\nu_\alpha((x, \infty]) = x^{-\alpha}$.

Proof. The equivalence of (a) and (b) is straightforward and can be easily verified. To prove (b) \implies (c), let $f \in C_K^+((0, \infty])$. We need to show that

$$\mu_n(f) := n\mathbb{E}f\left(\frac{X_1}{b_n}\right) = \int f(x)n\mathbb{P}\left[\frac{X_1}{b_n} \in dx\right] \rightarrow \nu_\alpha(f).$$

Since f has compact support, there exists $\delta > 0$ such that the support of f is contained in $(\delta, \infty]$. By (b), we know

$$\mu_n(x, \infty] \rightarrow x^{-\alpha} = \nu_\alpha(x, \infty] \quad \forall x > 0.$$

On $(\delta, \infty]$, define

$$P_n(\cdot) = \frac{\mu_n}{\mu_n(\delta, \infty]}$$

so that P_n is a probability measure on $(\delta, \infty]$. Then for $y \in (\delta, \infty]$,

$$P_n(y, \infty] \rightarrow P(y, \infty] = \frac{y^{-\alpha}}{\delta^{-\alpha}}.$$

Thus, $\{P_n\}$ converges weakly to P . Since f is bounded and continuous on $(\delta, \infty]$, we get from weak convergence that

$$P_n(f) \rightarrow P(f)$$

that is,

$$\frac{\mu_n(f)}{\mu_n(\delta, \infty]} \rightarrow \frac{\nu_\alpha(f)}{\delta^{-\alpha}}.$$

which implies

$$\mu_n(f) \rightarrow \nu_\alpha(f),$$

Now to show (c) \implies (b) we observe that $(x, \infty]$ is relatively compact and $\nu_\alpha(\partial(x, \infty]) = \nu_\alpha(\{x\}) = 0$. Since $\mu_n \xrightarrow{v} \nu_\alpha$, we have $\mu_n(x, \infty] \rightarrow \nu_\alpha(x, \infty]$ ■



3.2 The Hill Estimator

Theorem 3.2 (Convergence of regular-varying tails). *Suppose that $\{X_j\}_{j \in \mathbb{N}}$ are iid, nonnegative random variables whose common distribution has a regularly varying tail with tail parameter α , which implies that*

$$\frac{n}{k} \mathbb{P} \left[\frac{X_1}{b(n/k)} \in \cdot \right] \xrightarrow{v} \nu_\alpha(\cdot)$$

in $M_+(0, \infty]$ as $n \rightarrow \infty$ and $k = k(n) \rightarrow \infty$ with $n/k \rightarrow \infty$, where, $b(t) = F^{-1}(1 - t^{-1})$. Then in $M_+(0, \infty]$,

$$\nu_{k,n} \Rightarrow \nu_\alpha \quad \text{as } n \rightarrow \infty$$

where

$$\nu_\alpha((x, \infty]) = x^{-\alpha}, \quad \nu_{k,n}(\cdot) := \frac{1}{k} \sum \mathbf{1} \left\{ \cdot \ni \frac{X_i}{b(n/k)} \right\}$$

Proof. It suffices to show for a sequence $h_j \in \mathcal{C}_K^+(0, \infty]$ that in \mathbb{R}^∞ ,

$$(\nu_{k,n}(h_j), j \geq 1) \Rightarrow (\nu_\alpha(h_j), j \geq 1) \quad \text{as } n \rightarrow \infty$$

Convergence in \mathbb{R}^∞ reduces to convergence in \mathbb{R}^d for any d , so it suffices to show that

$$(\nu_{k,n}(h_j), 1 \leq j \leq d) \Rightarrow (\nu_\alpha(h_j), 1 \leq j \leq d) \quad \text{as } n \rightarrow \infty$$

To show this, we can show the joint Laplace transforms converge, so we assume $\lambda_j > 0$, $j = 1, \dots, d$, and show that

$$\mathbb{E} \left[\exp \left(- \sum_{j=1}^d \lambda_j \nu_{k,n}(h_j) \right) \right] \rightarrow \mathbb{E} \left[\exp \left(- \sum_{j=1}^d \lambda_j \nu_\alpha(h_j) \right) \right].$$

However,

$$\sum_{j=1}^d \lambda_j \nu_{k,n}(h_j) = \nu_{k,n} \left(\sum_{j=1}^d \lambda_j h_j \right);$$

similarly for ν_α substituted for $\nu_{k,n}$. Since $\sum_{j=1}^d \lambda_j h_j \in \mathcal{C}_K^+(0, \infty]$, it suffices to show for any $h \in \mathcal{C}_K^+(0, \infty]$ that

$$\mathbb{E} [\exp(-\nu_{k,n}(h))] \rightarrow \mathbb{E} [\exp(-\nu_\alpha(h))],$$

which is a reduction of the original task to a one-dimensional chore. The left side of the above equation is

$$\begin{aligned} \mathbb{E} \left[\exp \left(- \frac{1}{k} \sum_{j=1}^n h \left(\frac{X_j}{b(n/k)} \right) \right) \right] &= \left(\mathbb{E} \left[\exp \left(- \frac{1}{k} h \left(\frac{X_1}{b(n/k)} \right) \right) \right] \right)^n \\ &= \left(1 - \frac{\int_{(0, \infty]} (1 - \exp(-\frac{1}{k} h(x))) n \mathbb{P} \left[\frac{X}{b(n/k)} \in dx \right]}{n} \right)^n \end{aligned}$$

This converges to $\exp(-\nu_\alpha(h))$ since

$$\int_{(0, \infty]} \left(1 - \exp \left(- \frac{1}{k} h(x) \right) \right) n \mathbb{P} \left[\frac{X_1}{b(n/k)} \in dx \right] \approx \int_{(0, \infty]} h(x) \frac{n}{k} P \left[\frac{X_1}{b(n/k)} \in dx \right] \rightarrow \nu_\alpha(h),$$



where the approximate equivalence in the previous line can be justified by writing upper and lower bounds resulting from expanding the term $(1 - \exp(-\frac{1}{k}h(x)))$ to get upper and lower bounds. ■

Definition 3.3 (Hill Estimator). *Let $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \dots, \mathbf{Z}_n$ be iid samples from a heavy-tailed distribution with heavy tail parameter α . Then, the Hill estimator of α^{-1} based on $k = k(n)$ upper-order statistics is defined as*

$$H_{k,n} = \frac{1}{k} \sum_{i=n-k+1}^n \ln \frac{\mathbf{Z}_{(i)}}{\mathbf{Z}_{(n-k)}}$$

Theorem 3.3 (Consistency of Hill Estimator). *Let $\nu_{k,n} \Rightarrow \nu_\alpha$, and let $\{k_n\}_{n \in \mathbb{N}}$ be a sequence such that $k_n \rightarrow \infty$ and $\frac{k}{n} \rightarrow 0$ as $n \rightarrow \infty$, then*

$$H_{k,n} \xrightarrow{p} \frac{1}{\alpha}$$

Proof. See Theorem 4.2 of Resnick, 2007. ■

3.3 Choosing the optimal k_n

While the above consistency result holds for any $\{k_n\}$ satisfying those constraints, the rate of convergence varies significantly over the choice of sequence taken. Several methods have been proposed to best optimise this. Common approaches include:

1. Fixed sequences such as $k_n = \sqrt{n} \log(n)$ are commonly used as they are very simple to implement and work well in most scenarios.
2. A bootstrap based approach seeking to minimise the asymptotic variance of the estimate was proposed in Danielsson et al., 2001.
3. Among the data-driven approaches, perhaps the most popular method is the Hill plot (Drees et al., 2000) and its variants (smooHill, altHill, PERHILL, etc.)

4 Hill Estimation in Multivariate Contexts

So far, we have talked about Hill estimation in the univariate context. In particular, we have looked at regular variation at $+\infty$ for univariate CDFs. One can analogously look at regular variation at $-\infty$ without issue.

To successfully apply the above ideas in a multivariate context, we need to answer a few key questions —

- Q1. What does it mean to be “heavy-tailed” in the multivariate context?
- Q2. How do we account for the directionality of tails, and how do we interpret them?
- Q3. Are all tail parameters necessarily the same? If not, how do we model that?
- Q4. How does one perform Hill estimation in the multivariate context?



4.1 Generalising heavy-tailed measures

While attempting to generalise Hill estimation to multivariate, the first problem we encounter is there is no directionality, hence no obvious way to define regularly varying distributions. Fortunately, an analogue of Theorem 3.1(c) is still well-defined —

Definition 4.1 (Multivariate regular-varying distribution). *Let \mathbf{Z} be a random vector in some finite-dimensional Euclidean space \mathcal{S} . We say that \mathbf{Z} has a regularly varying tail if $\exists b_n \rightarrow \infty$ as $n \rightarrow \infty$.*

$$n\mathbb{P}\left[\frac{1}{b_n}\mathbf{Z} \in \cdot\right] \xrightarrow{v} \nu_*(\cdot)$$

In the above, $\nu_* \in M_+(\mathcal{S})$ is the limit measure, analogous to ν_α in Theorem 3.1(c). Its marginals capture the tail behaviour of the respective component of \mathbf{Z} . This addresses (Q1).

If we assume $Z^{(j)}$ is regular varying with tail parameter α_j ; for $j = 1, \dots, d$, then of course, it is possible to estimate α_j 's using univariate Hill estimation. However, this doesn't tell us anything about the joint behaviour of the tails. Modelling ν_* is more complicated than simply estimating a finite number of α 's — in particular, the phenomenon of interest is **tail dependence**, i.e. when occurrence of extreme events in one component is linked to the occurrence of extreme events in another.

Observe that for $\mathbf{Z}' = Q\mathbf{Z}$ for Q orthogonal,

$$\nu'_*(\cdot) = \nu_*({Q^{-1}\mathbf{x} : x \in \cdot})$$

i.e., ν'_* is simply a rotated version of ν_* .

Obviously, a rotation doesn't meaningfully affect the tail characteristics of the data. Thus, intuitively, writing $\mathbf{Z} \mapsto (R, \Theta) = (\|\mathbf{Z}\|, \frac{\mathbf{Z}}{\|\mathbf{Z}\|})$,

$$\nu_* \cong (\text{something dependent on } R) \times (\text{something dependent on } \Theta)$$

If we further assume that all components of \mathbf{Z} have the same marginal tail parameter, i.e.

$$\frac{\mathbb{P}[Z^{(i)} > x]}{\mathbb{P}[Z^{(j)} > x]} \rightarrow r_{ij} \in (0, \infty) \forall i, j = 1, \dots, \dim(\mathbf{Z})$$

then it can be shown that $\nu_* \cong c\nu_\alpha \times S$, for some $c > 0$, some ν_α as the *radial measure*, and some S as the *angular measure*. We are interested in estimating α , the common tail parameter, and S , the angular measure, which captures the joint behaviour, including tail dependence.

Hereforth, $S(\cdot)$ is a probability measure on \mathbb{N}_+ , which is the unit ball with respect to norm $\|\cdot\|$ centred at $\mathbf{0}$ in \mathcal{S} .

Definition 4.2 (Multivariate heavy-tailed distribution). *Let \mathbf{Z} be a random vector in some finite-dimensional Euclidean space \mathcal{S} . We say that \mathbf{Z} is regularly varying with the tail parameter α if $\exists b(t) \rightarrow \infty$ as $t \rightarrow \infty, c > 0$.*

$(R, \Theta) = (\|\mathbf{Z}\|, \frac{\mathbf{Z}}{\|\mathbf{Z}\|})$, we have

$$t\mathbb{P}\left[\left(\frac{R}{b(t)}, \Theta\right) \in \cdot\right] \xrightarrow{v} c\nu_\alpha \times S$$

in $M_+((0, \infty] \times \mathbb{N}_+)$ where $\nu_\alpha((x, \infty]) = x^{-\alpha}$



Theorem 4.1 (Key theorem for multivariate heavy-tailed distributions). *Let $\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \dots$ be iid random samples from a multivariate heavy-tailed distribution with (common) heavy-tailed parameter α . After transformation to polar coordinates, let the sequence be $\{(R_1, \boldsymbol{\Theta}_1), (R_2, \boldsymbol{\Theta}_2), \dots\}$. Then the following are equivalent*

1. *There exists a sequence $b_n \rightarrow \infty$ such that*

$$\sum_{i=1}^n \mathbf{1} \left\{ \cdot \ni \left(\frac{R_i}{b_n}, \boldsymbol{\Theta}_i \right) \right\} \Rightarrow PRM(c\nu_\alpha \times S)$$

in $M_p((0, \infty] \times \mathbb{N}_+)$.

2. *In $M_+((0, \infty] \times \mathbb{N}_+)$,*

$$\frac{1}{k} \sum_{i=1}^n \mathbf{1} \left\{ \cdot \ni \left(\frac{R_i}{b}, \boldsymbol{\Theta}_i \right) \right\} \Rightarrow c\nu_\alpha \times S,$$

3. *\mathbf{Z} is a random vector with a regularly varying tail with tail measure α .*

Proof. Refer Theorem 6.1 and 6.2 of Resnick, 2007. Note that the bridge between the lists of equivalences in Theorems 6.1 and 6.2 is Theorem 5.3. ■

4.2 Estimation for the standard case: common tail parameter

By consequence of the above theorem, assuming the components of \mathbf{Z} are tail equivalent, the problem of estimating α can easily be reduced to the univariate case by taking the norm $R = \|\mathbf{Z}\|$, and doing univariate Hill estimation on R .

This approach has been discussed in detail for elliptically distributed regular-varying distributions in Dominicy et al., 2017, under *angular estimators*. Notice that this method is clearly not scale invariant. A location-scale invariant estimator was also proposed therein that instead uses ellipsoids to contour the “outer” observations. Details are available in Dominicy et al., 2017 and Heikkilä et al., 2019.

Once we have the estimated tail parameter, we aim to assess *tail dependence*. The literature traditionally considers two tail dependence coefficients: upper and lower.

Definition 4.3. *The upper and lower tail dependence coefficients for a random vector \mathbf{X} having components $X_i \sim F_i$ are*

$$\delta_{ij}^u = \lim_{t \rightarrow 1} \mathbb{P}(X_i \geq F_i^{-1}(t) \mid X_j \geq F_j^{-1}(t))$$

$$\delta_{ij}^l = \lim_{t \rightarrow 1} \mathbb{P}(X_i \leq F_i^{-1}(t) \mid X_j \leq F_j^{-1}(t))$$

Courtesy of Schmidt, 2002, we know under ellipticity, the two expressions are equal, thus

$$\delta_{ij}^u = \delta_{ij}^l =: \delta_{ij} = \frac{\int_0^\zeta s^\alpha (s^2 - 1)^{-1/2} ds}{\int_0^1 s^\alpha (s^2 - 1)^{-1/2} ds}$$

where $\zeta = \frac{\sqrt{1+\rho}}{2}$, ρ_{ij} being the linear correlation coefficient. α can be replaced by its estimate, and ρ_{ij} by $\hat{\rho}_{ij,n} = \sin(\frac{\pi}{2} \hat{\tau}_{ij,n})$ where $\hat{\tau}_{ij,n}$ is the estimate of Kendall’s correlation.

This semi-parametric estimate of tail dependence allows us to assess the probability of extreme events in one component given another, which is particularly important in fields such as finance and econometrics. Indeed,



Dominicy et al., 2017 concludes with a discussion on the tail dependence of stock markets in windowed sub-samples over time. Crucially, the geographic regional structure presents itself in this tail dependence heatmap (Figure 3). More generally, in the standard case, it is possible to estimate the angular measure $S(\cdot)$, since

$$\frac{\sum_{i=1}^n \mathbf{1} \left\{ [1, \infty] \times \cdot \ni \left(\frac{R_i}{b}, \Theta_i \right) \right\}}{\sum_{i=1}^n \mathbf{1} \left\{ [1, \infty] \ni \left(\frac{R_i}{b} \right) \right\}} \Rightarrow S(\cdot)$$

We interpret this as the empirical probability measure of the points whose norms exceed 1 approximates S . Apart from normalization, if we consider the points $\{\Theta_i \mid R_i > 1\}$ and make a density plot, we should get a visualisation of the density of $S(\cdot)$. A notable mode in the density around $\frac{\pi}{4}$ would be indicative of dependence, while one near 0 or $\frac{\pi}{2}$ would indicate the lack thereof. For further details, see Section 9.2.(2-3) of Resnick, 2007.

Thus, Hill estimation opens avenues into analysis of extreme values and events in a way not possible through classical non-parametric methods such as the bootstrap because of the dubiousness of the convergence of these methods in many heavy-tailed contexts.

However, all our methods so far have a crucial drawback: they require the assumption of tail equivalence, i.e., a common tail parameter α . In most realistic scenarios, this assumption is only approximately true at best. We now try to relax this assumption.

4.3 Estimation for the non-standard case: differing tail parameters

We have already seen earlier that while estimating the marginal α_j 's is easy, it is not very useful by itself. We need to capture the joint behaviour, which is only possible by estimating the angular measure $S(\cdot)$.

A method based on ranks overcomes some of the drawbacks of the previous multivariate methods. The rank method does not require estimation of the marginal α_j 's, yet it achieves transformation to the standard case, allowing estimation of the angular measure. It is simple to implement; however, the transformation destroys the iid property of the sample, and asymptotic analysis is sophisticated. For a given k , the transform is —

$$(\mathbf{Z}_1, \dots, \mathbf{Z}_n) \mapsto \left\{ \left(\frac{k}{r_i^{(j)}}; j = 1, \dots, d \right); i = 1, \dots, n \right\}$$

where $r_i^{(j)} = \sum_{l=1}^n \mathbf{1} Z_l^{(j)} \geq Z_i^{(j)}$ is the rank of the i^{th} observation when sorted in the decreasing order of the j^{th} component.

The details of this method are quite involved; details are available in Section 9.2.3 of Resnick, 2007. Importantly, it turns out that after the transformation, we fall back into the standard case, allowing us to estimate $S(\cdot)$ using the usual polar transform on top of the rank transform.

Unfortunately, there is no neat analogue of Schmidt, 2002 for the non-standard case. We were unable to find results that would allow us to estimate tail dependence when there are differing α_j 's present. However, on a purely *ad hoc* basis, we use something like

$$\delta_{ij} = \frac{\int_0^\zeta (s^\alpha + s^\beta)(s^2 - 1)^{-1/2} ds}{\int_0^1 (s^\alpha + s^\beta)(s^2 - 1)^{-1/2} ds}$$



where α, β are the two tail parameters. The supposed advantage would be that this coefficient depends only on the tails of X_i and X_j , whereas earlier the common tail parameter α would be influenced by tails of **all** the components, not just X_i and X_j . However, the validity of this is doubtful, to say the least. Further research is needed in this area to derive estimators for the tail dependence for the non-standard case.

5 Concluding Remarks

In this project, we've attempted to showcase the process of multivariate Hill estimation – its rich literature, variety of methods, typical assumptions, and drawbacks. We began from relatively well understood univariate heavy-tailed distributions, looked at generalisations into the multivariate context, and ended close to the cutting edge of the field as it stands to date, with varying tail parameters. On the way, we found specific areas in need of further research and theoretical developments.

However, no discussion on this topic would be complete without mentioning that Hill estimation is not the only method for estimating the tail parameter. Other methods, such as Pickands estimator, (multivariate) peaks-over-threshold (POT) etc. exist and are often used. Some recent novel developments in this field include Németh and Zempléni, 2020 and Sasaki and Wang, 2024.

Hill estimation features prominently in econometric, actuarial, and financial research, as we have seen with Dominicy et al., 2017, Dominicy et al., 2020, Danish re-insurance, etc. Overall, Hill estimation is a powerful and versatile tool in dealing with a large class of heavy-tailed distributions and finds applications in the analysis of extreme values in several contexts.

References

- Drees, H., Resnick, S., & Haan, L. d. (2000). How to make a Hill plot. *The Annals of Statistics*, 28(1), 254–274. <https://doi.org/10.1214/aos/1016120372>
- Danielsson, J., de Haan, L., Peng, L., & de Vries, C. G. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis*, 76(2), 226–248. <https://doi.org/10.1006/jmva.2000.1903>
- Schmidt, R. (2002). Tail dependence for elliptically contoured distributions. *Mathematical Methods of Operations Research*, 55(2), 301–327. <https://doi.org/10.1007/s001860200191>
- Resnick, S. I. (2007). *Heavy-Tail Phenomena*. Springer. <https://doi.org/10.1007/978-0-387-45024-7>
- Dominicy, Y., Ilmonen, P., & Veredas, D. (2017). Multivariate Hill Estimators. *International Statistical Review*, 85(1), 108–142. <https://doi.org/10.1111/insr.12120>
- Heikkilä, M., Dominicy, Y., & Ilmonen, P. (2019). On multivariate separating Hill estimator under estimated location and scatter [Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/02331888.2018.1548016>]. *Statistics*, 53(2), 301–320. <https://doi.org/10.1080/02331888.2018.1548016>
- Dominicy, Y., Heikkilä, M., Ilmonen, P., & Veredas, D. (2020). Flexible multivariate Hill estimators. *Journal of Econometrics*, 217(2), 398–410. <https://doi.org/10.1016/j.jeconom.2019.12.010>
- Németh, L., & Zempléni, A. (2020). Regression Estimator for the Tail Index. *Journal of Statistical Theory and Practice*, 14(3), 48. <https://doi.org/10.1007/s42519-020-00114-7>
- Sasaki, Y., & Wang, Y. (2024). On uniform confidence intervals for the tail index and the extreme quantile. *Journal of Econometrics*, 244(1), 105865. <https://doi.org/10.1016/j.jeconom.2024.105865>