HEAVY TAILED ANALYSIS
EURANDOM
SUMMER 2005

SIDNEY RESNICK

School of Operations Research and Industrial Engineering
Cornell University
Ithaca, NY 14853 USA
sir1@cornell.edu
http://www.orie.cornell.edu/∼sid
&
Eurandom

1. Course Abstract

This is a survey of some of the mathematical, probabilistic and statistical tools used in heavy tail analysis. Heavy tails are characteristic of phenomena where the probability of a huge value is relatively big. Record breaking insurance losses, financial log-returns, file sizes stored on a server, transmission rates of files are all examples of heavy tailed phenomena. The modeling and statistics of such phenomena are tail dependent and much different than classical modeling and statistical analysis which give primacy to central moments, averages and the normal density, which has a wimpy, light tail. An organizing theme is that many limit relations giving approximations can be viewed as applications of almost surely continuous maps.

2. Introduction

Heavy tail analysis is an interesting and useful blend of mathematical analysis, probability and stochastic processes and statistics. Heavy tail analysis is the study of systems whose behavior is governed by large values which shock the system periodically. This is in contrast to many stable systems whose behavior is determined largely by an averaging effect. In heavy tailed analysis, typically the asymptotic behavior of descriptor variables is determined by the large values or merely a single large value.

Roughly speaking, a random variable X has a heavy (right) tail if there exists a positive parameter $\alpha > 0$ such that

$$P[X > x] \sim x^{-\alpha}, \quad x \to \infty. \tag{2.1}$$

(Note here we use the notation

$$f(x) \sim g(x), \quad x \to \infty$$

as shorthand for

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1,$$

for two real functions f, g.) Examples of such random variables are those with Cauchy, Pareto, t, F or stable distributions. Stationary stochastic processes, such as the ARCH, GARCH, EGARCH etc, which have been proposed as models for financial returns have marginal distributions satisfying (2.1). It turns out that (2.1) is not quite the right mathematical setting for discussing heavy tails (that pride of place belongs to regular variation of real functions) but we will get to that in due course.

Note the elementary observation that a heavy tailed random variable has a relatively large probability of exhibiting a really large value, compared to random variables which have exponentially bounded tails such as normal, Weibull, exponential or gamma random variables. For a $N(0,1)$ normal random variable N, with density $n(x)$, we have by Mill's ratio that

$$P[N > x] \sim \frac{n(x)}{x} \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad x \to \infty,$$

which has much weaker tail weight than suggested by (2.1).

There is a tendency to sometimes confuse the concept of a heavy tail distribution with the concept of a distribution with infinite right support. (For a probability distribution F, the support is the smallest closed set C such that $F(C) = 1$. For the exponential distribution with no translation, the support is $[0, \infty)$ and for the normal distribution, the support is $\mathbb{R}$.) The distinction is simple and exemplified by comparing a normally distributed random variable with one whose distribution is Pareto. Both have positive probability of achieving a value bigger than any pre-assigned threshold. However, the Pareto random variable has, for large thresholds, a much bigger probability of exceeding the threshold. One cannot rule out heavy tailed distributions by using the argument that everything in the world is bounded unless one agrees to rule out all distributions with unbounded support.

Much of classical statistics is often based on averages and moments. Try to imagine a statistical world where you do not rely on moments since if (2.1) holds, moments above the α-th do not exist! This follows since

$$\int_0^\infty x^{\beta-1} P[X > x] dx \approx \int_0^\infty x^{\beta-1} x^{-\alpha} dx \begin{cases} < \infty, & \text{if } \beta < \alpha, \\ = \infty, & \text{if } \beta \geq \alpha, \end{cases}$$

where (in this case)

$$\int f \approx \int g$$

means both integrals either converge or diverge together. Much stability theory in stochastic modeling is expressed in terms of mean drifts but what if the means do not exist. Descriptor variables in queueing theory are often in terms of means such as mean waiting time, mean queue lengths and so on. What if such expectations are infinite?

Consider the following scenarios where heavy tailed analysis is used.

(i) Finance. It is empirically observed that "returns" possess several notable features, sometimes called *stylized facts*. What is a "return"? Suppose $\{S_i\}$ is the stochastic process representing the price of a speculative asset (stock, currency, derivative, commodity (corn, coffee, etc)) at the ith measurement time. The return process is

$$\tilde{R}_i := (S_i - S_{i-1})/S_{i-1};$$

that is, the process giving the relative difference of prices. If the returns are small then the differenced log-Price process approximates the return process

$$\begin{aligned} R_i := \log S_i - \log S_{i-1} &= \log \frac{S_i}{S_{i-1}} = \log\left(1 + \left(\frac{S_i}{S_{i-1}} - 1\right)\right) \\ &\sim \frac{S_i}{S_{i-1}} - 1 = \tilde{R}_i \end{aligned}$$

since for $|x|$ small,

$$\log(1 + x) \sim x$$

by, say, L'Hospital's rule. So instead of studying the returns $\{\tilde{R}_i\}$, the differenced log-Price process $\{R_i\}$ is studied and henceforth we refer to $\{R_i\}$ as the *returns*.

Empirically, either process is often seen to exhibit notable properties:

(1) Heavy tailed marginal distributions (but usually $2 < \alpha$ so the mean and variance exist);
(2) Little or no correlation. However by squaring or taking absolute values of the process one gets a highly correlated, even long range dependent process.
(3) The process is dependent. (If the random variables were independent, so would the squares be independent but squares are typically correlated.)

Hence one needs to model the data with a process which is stationary, has heavy tailed marginal distributions and a dependence structure. This leads to the study of specialized models in economics with lots of acronyms like ARCH and GARCH. Estimation of, say, the marginal distribution's shape parameter α are made more complex due to the fact that the observations are not independent.

Classical Extreme Value Theory which subsumes heavy tail analysis uses techniques to estimate *value-at-risk* or (VaR), which is an extreme quantile of the profit/loss density, once the density is estimated.

Note, that given S_0, there is a one-to-one correspondence between

$$\{S_0, S_1, \dots, S_T\} \text{ and } \{S_0, R_1, \dots, R_T\}$$

since

$$\begin{aligned}\sum_{t=1}^{T} R_t =&(\log S_1 - \log S_0) + (\log S_2 - \log S_1)\\ &\qquad + \cdots + (\log S_T - \log S_{T-1})\\ =&\log S_T - \log S_0 = \log \frac{S_T}{S_0},\end{aligned}$$

so that

$$S_T = S_0 e^{\sum_{t=1}^{T} R_t}. \tag{2.2}$$

Why deal with returns rather than the price process?

(1) The returns are scale free and thus independent of the size of the investment.
(2) Returns have more attractive statistical properties than prices such as stationarity. Econometric models sometimes yield non-stationary price models but stationary returns.

To convince you this might make a difference to somebody, note that from 1970-1995, the two worst losses world wide were Hurrricane Andrew (my wife's cousin's yacht in Miami wound up on somebody's roof 30 miles to the north) and the Northridge earthquake in California. Losses in 1992 dollars were \$16,000 and \$11,838 million dollars respectively. (Note the unit is "millions of dollars".)

Why deal with log-returns rather than returns?

(1) Log returns are nicely additive over time. It is easier to construct models for additive phenomena than multiplicative ones (such as $1 + \tilde{R}_t = S_t/S_{t-1}$). One

can recover S_T from log-returns by what is essentially an additive formula (2.2). (Additive is good!) Also, the T-day return process

$$R_T - R_1 = \log S_T - \log S_0$$

is additive. (Additive is good!)

(2) Daily returns satisfy

$$\frac{S_t}{S_{t-1}} - 1 \geq -1,$$

and for statistical modeling, it is a bit unnatural to have the variable bounded below by -1. For instance one could not model such a process using a normal or two-sided stable density.

(3) Certain economic facts are easily expressed by means of log-returns. For example, if S_t is the exchange rate of the US dollar against the British pound and $R_t = \log(S_t/S_{t-1})$, then $1/S_t$ is the exchange rate of pounds to dollars and the return from the point of view of the British investor is

$$\log \frac{1/S_t}{1/S_{t-1}} = \log \frac{S_{t-1}}{S_t} = -\log \frac{S_t}{S_{t-1}}$$

which is minus the return for the American investor.

(4) The operations of taking logarithms and differencing are standard time series tools for coercing a data set into looking stationary. Both operations, as indicated, are easily undone. So there is a high degree of comfort with these operations.

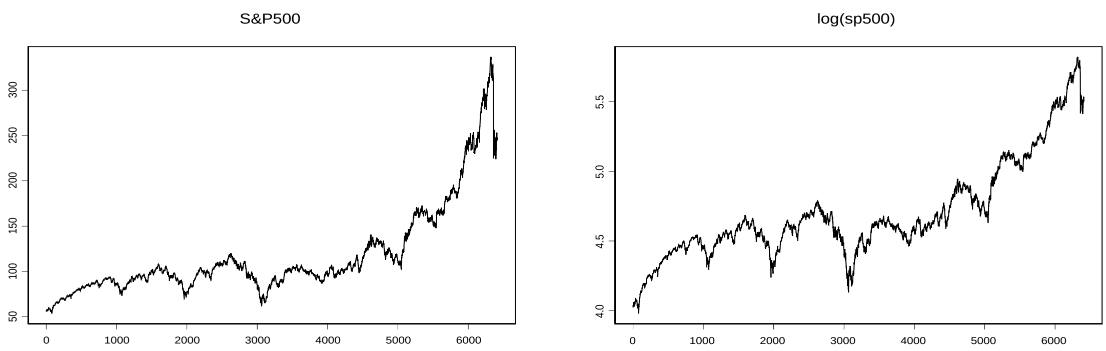

FIGURE 1. Time series plot of S&P 500 data (left) and log(S&P500) (right).

Example 1 (Standard & Poors 500)**.** We consider the data set *fm-poors.dat* in the package Xtremes which gives the Standard & Poors 500 stock market index. The data is daily data from July 1962 to December 1987 but of course does not include days when the market is closed. In Figure 1 we display the time series plots of the actual data for the index and the log of the data. Only a lunatic would conclude these two series were stationary. In the left side of Figure 2 we exhibit the 6410 returns $\{R_t\}$ of the data by differencing at lag 1 the log(S&P) data. On the right side is the sample autocorrelation function. There is a biggish lag 1 correlation but otherwise few spikes are outside the magic window.

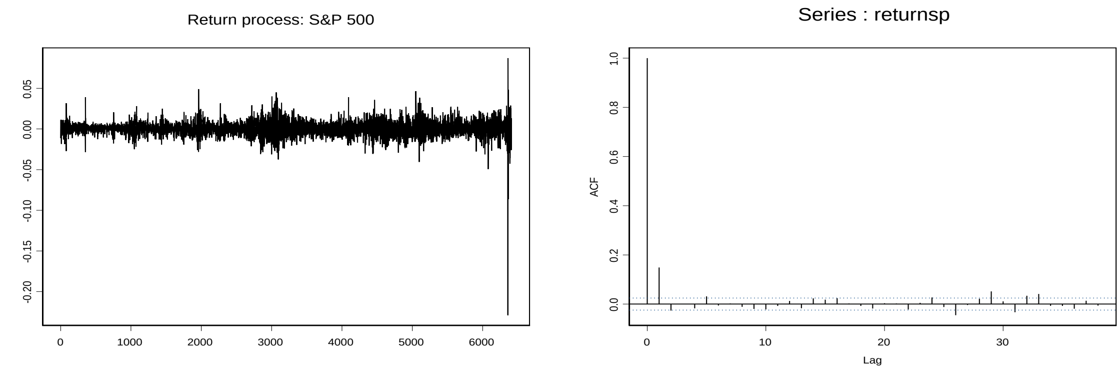

FIGURE 2. Time series plot of S&P 500 return data (left) and the autocorrelation function (right).

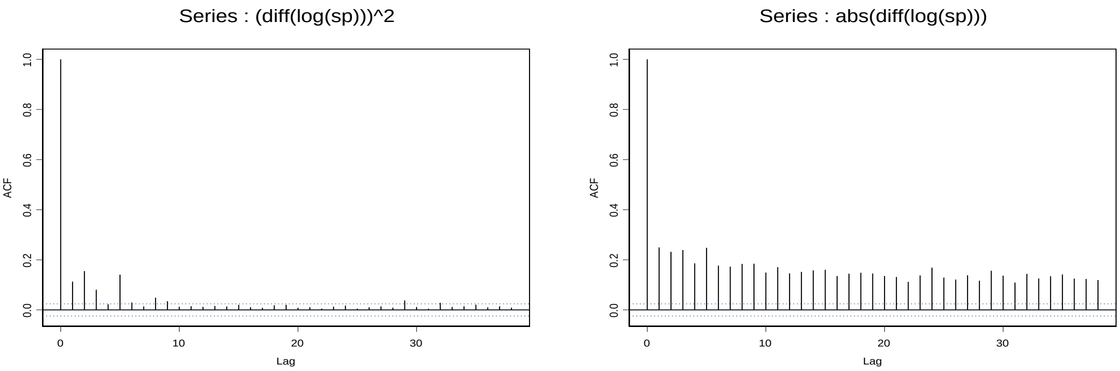

FIGURE 3. i) The autocorrelation function of the squared returns (left). (ii) The autocorrelation function of the absolute values of the returns. (right)

For a view of the *stylized facts* about these data, and to indicate the complexities of the dependence structure, we exhibit the autocorrelation function of the squared returns in Figure 3 (left) and on the right the autocorrelation function for the absolute value of the returns. Though there is little correlation in the original series, the iid hypothesis is obviously false.

One can compare the heaviness of the right and left tail of the marginal distribution of the process $\{R_t\}$ even if we do not believe that the process is iid. A reasonable assumption seems to be that the data can be modelled by a stationary, uncorrelated process and we hope the standard exploratory extreme value and heavy tailed methods developed for iid processes still apply. We apply the QQ-plotting technique to the data. After playing a bit with the number of upper order statistics used, we settled on $k = 200$ order statistics for the positive values (upper tail) which gives the slope estimate of $\hat{\alpha} = 3.61$. This is shown in the left side of Figure 4. On the right side of Figure 4 is the comparable plot for the left tail; here we applied the routine to abs(returns[returns¡0]); that is, to the absolute value of the negative data points in the log-return sample. After some experimentation, we obtained an estimate $\hat{\alpha} = 3.138$ using $k = 150$. Are the two tails symmetric which is a common theoretical assumption? Unlikely!

(ii) Insurance and reinsurance. The general theme here is to model insurance claim sizes and frequencies so that premium rates may be set intelligently and risk to the insurance company quantified.

Smaller insurance companies sometimes pay for reinsurance or excess-of-loss (XL) insurance to a bigger company like Lloyd's of London. The excess claims over a certain contractually agreed threshhold is covered by the big insurance company. Such excess claims are by definition very large so heavy tail analysis is a natural tool to apply. What premium should the big insurance company charge to cover potential losses?

As an example of data you might encounter, consider the Danish data on large fire insurance losses McNeil (1997), Resnick (1997). Figure 5 gives a time series plot of the 2156 Danish data consisting of losses over one million Danish Krone (DKK) and the right hand plot is the QQ plot of this data yielding a remarkably straight plot. The straight line plot indicates the appropriateness of heavy tail analysis.

(iii) Data networks. A popular idealized data transmission model of a source destination pair is an *on/off* model where constant rate transmissions alternate with *off* periods. The *on* periods are random in length with a *heavy tailed distribution* and this leads to occasional large transmission lengths. The model offers an explanation of perceived *long range dependence* in measured traffic rates. A competing model which is marginally more elegant in our eyes is the infinite source Poisson model to be discussed later along with all its warts.

Example 2. The Boston University study (Crovella and Bestavros (1995), Crovella and Bestavros (1996), Cunha et al. (1995)) suggests self-similarity of web traffic stems from heavy tailed file sizes. This means that we treat files as being randomly selected from a population and if X represents a randomly selected file size then the heavy tail hypothesis

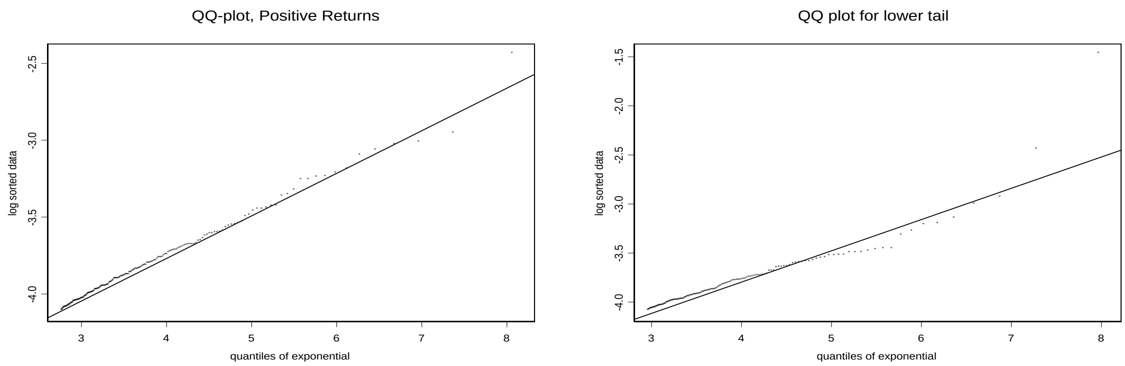

FIGURE 4. Left: QQ-plot and parfit estimate of α for the right tail using $k = 200$ upper order statistics. Right: QQ-plot and parfit estimate of α for the left tail using the absolute value of the negative values in the log-returns.

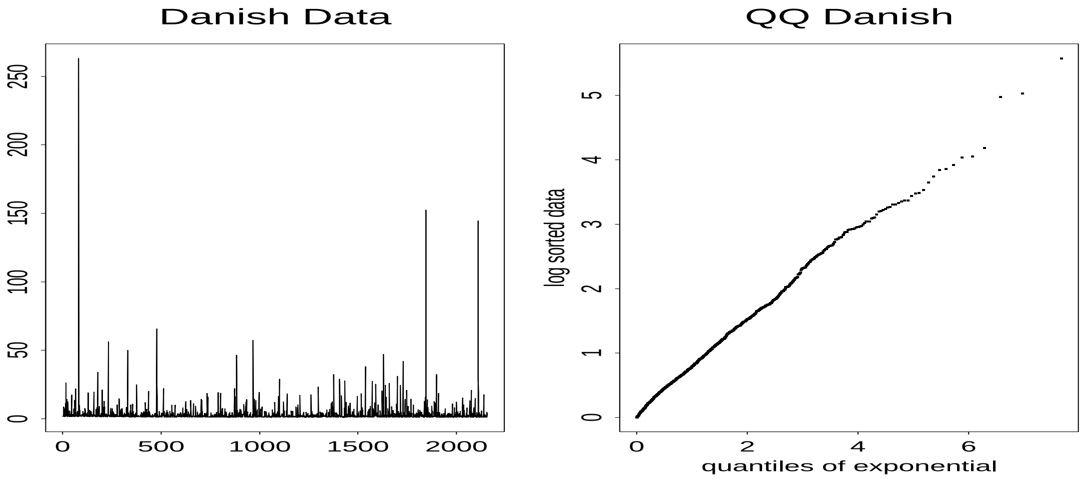

FIGURE 5. Danish Data (left) and QQ-plot.

means for large $x > 0$

$$P[X > x] \sim x^{-\alpha}, \quad \alpha > 0, \tag{2.3}$$

where α is a shape parameter that must be statistically estimated. The BU study reports an overall estimate for a five month measurement period (see Cunha et al. (1995)) of $\alpha = 1.05$. However, there is considerable month-to-month variation in these estimates and, for instance, the estimate for November 1994 in room 272 places α in the neighborhood of 0.66. Figure 6 gives the QQ and Hill plots (Beirlant et al. (1996), Hill (1975), Kratz and Resnick (1996), Resnick and Stărică (1997)) of the file size data for the month of November in the Boston University study. These are two graphical methods for estimating α and will be discussed in more detail later.

Extensive traffic measurements of *on* periods are reported in Willinger et al. (1995) where measured values of α were usually in the interval $(1, 2)$. Studies of sizes of files accessed on various servers by the Calgary study (Arlitt and Williamson (1996)), report estimates of α from 0.4 to 0.6. So accumulating evidence already exists which suggests values of α outside the range $(1, 2)$ should be considered. Also, as user demands on the web grow and access speeds increase, there may be a drift toward heavier file size distribution tails. However, this is a hypothesis that is currently untested.

REFERENCES

M. Arlitt and C. Williamson. Web server workload characterization: The search for invariants (extended version). In *Proceedings of the ACM Sigmetrics Conference, Philadelphia, Pa*, 1996. Available from {mfa126,carey}@cs.usask.ca.

J. Beirlant, P. Vynckier, and J. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, 91(436):1659–1667, 1996. ISSN 0162-1459.

M. Crovella and A. Bestavros. Explaining world wide web traffic self–similarity. Preprint available as TR-95-015 from {crovella,best}@cs.bu.edu, 1995.

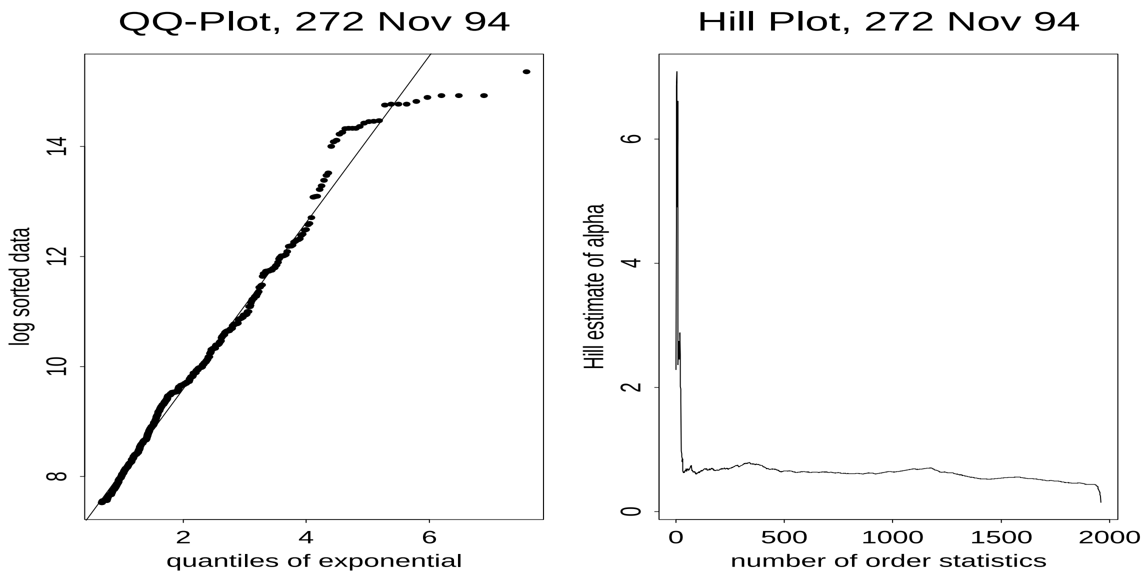

FIGURE 6. QQ and Hill plots of November 1994 file lengths.

M. Crovella and A. Bestavros. Self-similarity in world wide web traffic: evidence and possible causes. In *Proceedings of the ACM SIGMETRICS '96 International Conference on Measurement and Modeling of Computer Systems*, volume 24, pages 160–169, 1996.

C. Cunha, A. Bestavros, and M. Crovella. Characteristics of www client–based traces. Preprint available as BU-CS-95-010 from {crovella,best}@cs.bu.edu, 1995.

B. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

M. Kratz and S. Resnick. The qq–estimator and heavy tails. *Stochastic Models*, 12:699–724, 1996.

A. McNeil. Estimating the tails of loss severity distributions using extreme value theory. *Astin Bulletin*, 27:117–137, 1997.

S. Resnick. Discussion of the danish data on large fire insurance losses. *Astin Bulletin*, 27: 139–151, 1997.

S. Resnick and C. Stărică. Smoothing the Hill estimator. *Adv. Applied Probability*, 29: 271–293, 1997.

W. Willinger, M. Taqqu, M. Leland, and D. Wilson. Self–similarity in high–speed packet traffic: analysis and modelling of ethernet traffic measurements. *Statistical Science*, 10: 67–85, 1995.

3. A Crash Course on Regular Variation

The theory of regularly varying functions is the appropriate mathematical analysis tool for proper discussion of heavy tail phenomena. We begin by reviewing some results from analysis starting with uniform convergence.

3.1. **Preliminaries from analysis.**

3.1.1. *Uniform convergence.* If $\{f_n, n \geq 0\}$ are real valued functions on $\mathbb{R}$ (or, in fact, any metric space) then f_n converges uniformly on $A \subset \mathbb{R}$ to f if

$$\sup_A |f_0(x) - f_n(x)| \to 0 \tag{3.1}$$

as $n \to \infty$. The definition would still make sense of the range of $f_n, n \geq 0$ were a metric space but then $|f_0(x) - f_n(x)|$ would need to be replaced by $d(f_0, f_n)$, where $d(\cdot, \cdot)$ is the metric. For functions on $\mathbb{R}$, the phrase *local uniform convergence* means that (3.1) holds for any compact interval A.

If $U_n, n \geq 0$ are non-decreasing real valued functions on $\mathbb{R}$, then a useful fact is that if U_0 is continuous and $U_n(x) \to U_0(x)$ as $n \to \infty$ for all x, then $U_n \to U$ locally uniformly; i.e. for any $a < b$

$$\sup_{x \epsilon [a,b]} |U_n(x) - U_0(x)| \to 0.$$

(See (Resnick, 1987, page 1).) One proof of this fact is outlined as follows: If U_0 is continuous on $[a, b]$, then it is uniformly continuous. From the uniform convergence, for any x, there is an interval-neighborhood O_x on which $U_0(\cdot)$ oscillates by less than a given ϵ. This gives an open cover of $[a, b]$. Compactness of $[a, b]$ allows us to prune $\{O_x,\ x \in [a, b]\}$ to obtain a finite subcover $\{(a_i, b_i), i = 1, \ldots, K\}$. Using this finite collection and the monotonicity of the functions leads to the result: Given $\epsilon > 0$, there exists some large N such that if $n \geq N$ then

$$\max_{1 \leq i \leq N} |U_n(a_i) - U_0(a_i)| \bigvee U_n(b_i) - U_0(b_i)| < \epsilon,$$

(by pointwise convergence). Observe that

$$\sup_{x \in [a,b]} |U_n(x) - U_0(x)| \leq \max_{1 \leq i \leq N} \sup_{[a_i, b_i]} |U_n(x) - U_0(x)|. \tag{3.2}$$

For any $x \in [a_i, b_i]$, we have by monotonicity

$$\begin{aligned} U_n(x) - U_0(x) \leq & U_n(b_i) - U_0(a_i) \\ \leq & U_0(b_i) + \epsilon - U_0(a_i), \quad \text{(by (3.2))} \\ \leq & 2\epsilon, \end{aligned}$$

with a similar lower bound. This is true for all i and hence we get uniform convergence on $[a, b]$.

3.1.2. *Inverses of monotone functions.* Suppose $H : \mathbb{R} \mapsto (a, b)$ is a non-decreasing function on $\mathbb{R}$ with range (a, b), $-\infty \le a < b \le \infty$. With the convention that the infimum of an empty set is $+\infty$, we define the (left continuous) inverse $H^{\leftarrow} : (a, b) \mapsto \mathbb{R}$ of H as

$$H^{\leftarrow}(y) = \inf\{s : H(s) \ge y\}.$$

In case the function H is right continuous we have the following interesting properties:

$$A(y) := \{s : H(s) \ge y\} \text{ is closed,} \tag{3.3}$$

$$H(H^{\leftarrow}(y)) \ge y \tag{3.4}$$

$$H^{\leftarrow}(y) \le t \text{ iff } y \le H(t). \tag{3.5}$$

For (3.3), observe that if $s_n \in A(y)$ and $s_n \downarrow s$, then $y \le H(s_n) \downarrow H(s)$ so $H(s) \ge y$ and $s \in A(y)$. If $s_n \uparrow s$ and $s_n \in A(y)$, then $y \le H(s_n) \uparrow H(s-) \le H(s)$ and $H(s) \ge y$ so $s \in A(y)$ again and $A(y)$ is closed. Since $A(y)$ is closed, $\inf A(y) \in A(y)$; that is, $H^{\leftarrow}(y) \in A(y)$ which means $H(H^{\leftarrow}(y)) \ge y$. This gives (3.4). Lastly, (3.5) follows from the definition of $H^{\leftarrow}$.

3.1.3. *Convergence of monotone functions.* For any function H denote

$$\mathcal{C}(H) = \{x \in \mathbb{R} : H \text{ is finite and continuous at } x\}.$$

A sequence $\{H_n, n \ge 0\}$ of non-decreasing functions on $\mathbb{R}$ converges weakly to H_0 if as $n \to \infty$ we have

$$H_n(x) \to H_0(x),$$

for all $x \in \mathcal{C}(H_0)$. We will denote this by $H_n \to H_0$ and no other form of convergence for monotone functions will be relevant. If $F_n, n \ge 0$ are non-defective distributions, then a myriad of names give equivalent concepts: complete convergence, vague convergence, weak* convergence, narrow convergence. If $X_n, n \ge 0$ are random variables and X_n has distribution function $F_n, n \ge 0$, then $X_n \Rightarrow X_0$ means $F_n \to F_0$. For the proof of the following, see (Billingsley, 1986, page 343), (Resnick, 1987, page 5), (Resnick, 1998, page 259).

Proposition 1. *If $H_n, n \ge 0$ are non-decreasing functions on $\mathbb{R}$ with range (a, b) and $H_n \to H_0$, then $H_n^{\leftarrow} \to H_0^{\leftarrow}$ in the sense that for $t \in (a, b) \cap \mathcal{C}(H_0^{\leftarrow})$*

$$H_n^{\leftarrow}(t) \to H_0^{\leftarrow}(t).$$

3.1.4. *Cauchy's functional equation.* Let $k(x), x \in \mathbb{R}$ be a function which satisfies

$$k(x + y) = k(x) + k(y), x, y \in R.$$

If k is measurable and bounded on a set of positive measure, then $k(x) = cx$ for some $c \in \mathbb{R}$. (See Seneta (1976), (Bingham et al., 1987, page 4).)

3.2. **Regular variation: definition and first properties.** An essential analytical tool for dealing with heavy tails, long range dependence and domains of attraction is the theory of regularly varying functions. This theory provides the correct mathematical framework for considering things like Pareto tails and algebraic decay.

Roughly speaking, *regularly varying functions* are those functions which behave asymptotically like power functions. We will deal currently only with real functions of a real variable. Consideration of multivariate cases and probability concepts suggests recasting definitions in terms of vague convergence of measures but we will consider this reformulation later.

Definition 1. A measurable function $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$ (written $U \in RV_\rho$) if for $x > 0$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho.$$

We call ρ the *exponent of variation*.

If $\rho = 0$ we call U *slowly varying*. Slowly varying functions are generically denoted by $L(x)$. If $U \in RV_\rho$, then $U(x)/x^\rho \in RV_0$ and setting $L(x) = U(x)/x^\rho$ we see it is always possible to represent a ρ-varying function as $x^\rho L(x)$.

Examples. The canonical ρ-varying function is x^ρ. The functions $\log(1+x)$, $\log\log(e+x)$ are slowly varying, as is $\exp\{(\log x)^\alpha\}, 0 < \alpha < 1$. Any function U such that $\lim_{x\to\infty} U(x) =: U(\infty)$ exists finite is slowly varying. The following functions are not regularly varying: $e^x, \sin(x+2)$. Note $[\log x]$ is slowly varying, but $\exp\{[\log x]\}$ is not regularly varying.

In probability applications we are concerned with distributions whose tails are regularly varying. Examples are

$$1 - F(x) = x^{-\alpha},\ x \geq 1,\ \alpha > 0,$$

and the extreme value distribution

$$\Phi_\alpha(x) = \exp\{-x^{-\alpha}\},\ x \geq 0.$$

$\Phi_\alpha(x)$ has the property

$$1 - \Phi_\alpha(x) \sim x^{-\alpha} \text{ as } x \to \infty.$$

A stable law (to be discussed later) with index $\alpha, 0 < \alpha < 2$ has the property

$$1 - G(x) \sim cx^{-\alpha},\ x \to \infty,\ c > 0.$$

The Cauchy density $f(x) = (\pi(1+x^2))^{-1}$ has a distribution function F with the property

$$1 - F(x) \sim (\pi x)^{-1}.$$

If $N(x)$ is the standard normal df then $1 - N(x)$ is not regularly varying nor is the tail of the Gumbel extreme value distribution $1 - \exp\{-e^{-x}\}$.

The definition of regular variation can be weakened slightly (cf de Haan (1970), Feller (1971), Resnick (1987)).

Proposition 2. *(i) A measurable function $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ varies regularly if there exists a function h such that for all $x > 0$*

$$\lim_{t\to\infty} U(tx)/U(t) = h(x).$$

In this case $h(x) = x^\rho$ *for some* $\rho \in \mathbb{R}$ *and* $U \in RV_\rho$.

(ii) A monotone function $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ *varies regularly provided there are two sequences* $\{\lambda_n\}, \{a_n\}$ *of positive numbers satisfying*

$$a_n \to \infty, \qquad \lambda_n \sim \lambda_{n+1}, \quad n \to \infty, \tag{3.6}$$

and for all $x > 0$

$$\lim_{n\to\infty} \lambda_n U(a_n x) =: \chi(x) \text{ exists positive and finite.} \tag{3.7}$$

In this case $\chi(x)/\chi(1) = x^\rho$ *and* $U \in RV_\rho$ *for some* $\rho \in \mathbb{R}$.

We frequently refer to (3.7) as the *sequential form of regular variation*. For probability purposes, it is the most useful. Typically U is a distribution tail, $\lambda_n = n$ and a_n is a distribution quantile.

Proof. (i) The function h is measurable since it is a limit of measurable functions. Then for $x > 0, y > 0$

$$\frac{U(txy)}{U(t)} = \frac{U(txy)}{U(tx)} \cdot \frac{U(tx)}{U(t)}$$

and letting $t \to \infty$ gives

$$h(xy) = h(y)h(x).$$

So h satisfies the Hamel equation, which by change of variable can be converted to the Cauchy equation. Therefore, the form of h is $h(x) = x^\rho$ for some $\rho \in \mathbb{R}$.

(ii) For concreteness assume U is nondecreasing. Assume (3.6) and (3.7) and we show regular variation. Since $a_n \to \infty$, for each t there is a finite $n(t)$ defined by

$$n(t) = \inf\{m : a_{m+1} > t\}$$

so that

$$a_{n(t)} \le t < a_{n(t)+1}.$$

Therefore by monotonicity for $x > 0$

$$\left(\frac{\lambda_{n(t)+1}}{\lambda_{n(t)}}\right)\left(\frac{\lambda_{n(t)}U(a_{n(t)}x)}{\lambda_{n(t)+1}U(a_{n(t)+1})}\right) \le \frac{U(tx)}{U(t)} \le \left(\frac{\lambda_{n(t)}}{\lambda_{n(t)+1}}\right)\left(\frac{\lambda_{n(t)+1}U(a_{n(t)+1}x)}{\lambda_{n(t)}U(a_{n(t)})}\right).$$

Now let $t \to \infty$ and use (3.6) and (3.7) to get $\lim_{t\to\infty} \frac{U(tx)}{U(t)} = 1\frac{\chi(x)}{\chi(1)}$. Regular variation follows from part (i). $\square$

Remark 1. Proposition 2 (ii) remains true if we only assume (3.7) holds on a dense set. This is relevant to the case where U is nondecreasing and $\lambda_n U(a_n x)$ converges weakly.

3.2.1. *A maximal domain of attraction.* Suppose $\{X_n, n \geq 1\}$ are iid with common distribution function $F(x)$. The extreme is

$$M_n = \bigvee_{i=1}^{n} X_i = \max\{X_1, \dots, X_n\}.$$

One of the extreme value distributions is

$$\Phi_\alpha(x) := \exp\{-x^{-\alpha}\}, \quad x > 0,\ \alpha > 0.$$

What are conditions on F, called *domain of attraction conditions*, so that there exists $a_n > 0$ such that

$$P[a_n^{-1} M_n \leq x] = F^n(a_n x) \to \Phi_\alpha(x) \tag{3.8}$$

weakly. How do you characterize the normalization sequence $\{a_n\}$?

Set $x_0 = \sup\{x : F(x) < 1\}$ which is called the right end point of F. We first check (3.8) implies $x_0 = \infty$. Otherwise if $x_0 < \infty$ we get from (3.8) that for $x > 0, a_n x \to x_0$; i.e. $a_n \to x_0 x^{-1}$. Since $x > 0$ is arbitrary we get $a_n \to 0$ whence $x_0 = 0$. But then for $x > 0, F^n(a_n x) = 1$, which violates (3.8). Hence $x_0 = \infty$.

Furthermore $a_n \to \infty$ since otherwise on a subsequence $n', a_{n'} \leq K$ for some $K < \infty$ and

$$0 < \Phi_\alpha(1) = \lim_{n' \to \infty} F^{n'}(a_{n'}) \leq \lim_{n' \to \infty} F^{n'}(K) = 0$$

since $F(K) < 1$ which is a contradiction.

In (3.8), take logarithms to get for $x > 0, \lim_{n \to \infty} n(-\log F(a_n x)) = x^{-\alpha}$. Now use the relation $-\log(1-z) \sim z$ as $z \to 0$ and (7) is equivalent to

$$\lim_{n \to \infty} n(1 - F(a_n x)) = x^{-\alpha}, \quad x > 0. \tag{3.9}$$

From (3.9) and Proposition 2 we get

$$1 - F(x) \sim x^{-\alpha} L(x), \quad x \to \infty, \tag{3.10}$$

for some $\alpha > 0$. To characterize $\{a_n\}$ set $U(x) = 1/(1 - F(x))$ and (3.9) is the same as

$$U(a_n x)/n \to x^\alpha, x > 0$$

and inverting we find via Proposition 1 that

$$\frac{U^{\leftarrow}(ny)}{a_n} \to y^{1/\alpha}, y > 0.$$

So $U^{\leftarrow}(n) = (1/(1-F))^{\leftarrow}(n) \sim a_n$ and this determines a_n by the convergence to types theorem.(See Feller (1971), Resnick (1987, 1998).)

Conversely if (3.10) holds, define $a_n = U^{\leftarrow}(n)$ as previously. Then

$$\lim_{n \to \infty} \frac{1 - F(a_n x)}{1 - F(a_n)} = x^{-\alpha}$$

and we recover (3.9) provided $1 - F(a_n) \sim n^{-1}$ or what is the same provided $U(a_n) \sim n$ i.e., $U(U^{\leftarrow}(n)) \sim n$. Recall from (3.5), that $z < U^{\leftarrow}(n)$ iff $U(z) < n$ and setting $z = U^{\leftarrow}(n)(1-\varepsilon)$ and then $z = U^{\leftarrow}(n)(1+\varepsilon)$ we get

$$\frac{U(U^{\leftarrow}(n))}{U(U^{\leftarrow}(n)(1+\varepsilon))} \le \frac{U(U^{\leftarrow}(n))}{n} \le \frac{U(U^{\leftarrow}(n))}{U(U^{\leftarrow}(n)(1-\varepsilon))}.$$

Let $n \to \infty$, remembering $U = 1/(1-F) \in RV_\alpha$. Then

$$(1+\varepsilon)^{-\alpha} \le \liminf_{n\to\infty} n^{-1}U(U^{\leftarrow}(n)) \le \limsup_{n\to\infty} U(U^{\leftarrow}(n)) \le (1-\varepsilon)^{-\alpha}$$

and since $\varepsilon > 0$ is arbitrary the desired result follows.

3.3. **Regular variation: Deeper Results; Karamata's Theorem.** There are several deeper results which give the theory power and utility: uniform convergence, Karamata's theorem which says a regularly varying function integrates the way you expect a power function to integrate, and finally the Karamata representation theorem.

3.3.1. *Uniform convergence.* The first useful result is the uniform convergence theorem.

Proposition 3. *If $U \in RV_\rho$ for $\rho \in \mathbb{R}$, then*

$$\lim_{t\to\infty} U(tx)/U(t) = x^\rho$$

locally uniformly in x on $(0,\infty)$. If $\rho < 0$, then uniform convergence holds on intervals of the form $(b,\infty), b > 0$. If $\rho > 0$ uniform convergence holds on intervals $(0,b]$ provided U is bounded on $(0,b]$ for all $b > 0$.

If U is monotone the result already follows from the discussion in Subsubsection 3.1.1, since we have a family of monotone functions converging to a continuous limit. For detailed discussion see Bingham et al. (1987), de Haan (1970), Geluk and de Haan (1987), Seneta (1976).

3.3.2. *Integration and Karamata's theorem.* The next set of results examines the integral properties of regularly varying functions. For purposes of integration, a ρ-varying function behaves roughly like x^ρ. We assume all functions are locally integrable and since we are interested in behavior at ∞ we assume integrability on intervals including 0 as well.

Theorem 1 (Karamata's Theorem)**.** *(a) Suppose $\rho \ge -1$ and $U \in RV_\rho$. Then $\int_0^x U(t)dt \in RV_{\rho+1}$ and*

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \rho + 1. \tag{3.11}$$

If $\rho < -1$ (or if $\rho = -1$ and $\int_x^\infty U(s)ds < \infty$) then $U \in RV_\rho$ implies $\int_x^\infty U(t)dt$ is finite, $\int_x^\infty U(t)dt \in RV_{\rho+1}$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = -\rho - 1. \tag{3.12}$$

(b) If U satisfies

$$\lim_{x\to\infty} \frac{xU(x)}{\int_0^x U(t)dt} = \lambda \in (0,\infty) \tag{3.13}$$

then $U \in RV_{\lambda-1}$. If $\int_x^\infty U(t)dt < \infty$ and

$$\lim_{x\to\infty} \frac{xU(x)}{\int_x^\infty U(t)dt} = \lambda \in (0,\infty) \tag{3.14}$$

then $U \in RV_{-\lambda-1}$.

What Theorem 1 emphasizes is that for the purposes of integration, the slowly varying function can be passed from inside to outside the integral. For example the way to remember and interpret (3.11) is to write $U(x) = x^\rho L(x)$ and then observe

$$\int_0^x U(t)dt = \int_0^x t^\rho L(t)dt$$

and pass the $L(t)$ in the integrand outside as a factor $L(x)$ to get

$$\begin{aligned}&\sim L(x)\int_0^x t^\rho dt = L(x)x^{\rho+1}/(\rho+1)\\ &= xx^\rho L(x)/(\rho+1) = xU(x)/(\rho+1),\end{aligned}$$

which is equivalent to the assertion (3.11).

Proof. (a). For certain values of ρ, uniform convergence suffices after writing say

$$\frac{\int_0^x U(s)ds}{xU(x)} = \int_0^x \frac{U(sx)}{U(x)}ds.$$

If we wish to proceed, using elementary concepts, consider the following approach, which follows de Haan (1970).

If $\rho > -1$ we show $\int_0^\infty U(t)dt = \infty$. From $U \in RV_\rho$ we have

$$\lim_{s\to\infty} U(2s)/U(s) = 2^\rho > 2^{-1}$$

since $\rho > -1$. Therefore there exists s_0 such that $s > s_0$ necessitates $U(2s) > 2^{-1}U(s)$. For n with $2^n > s_0$ we have

$$\int_{2^{n+1}}^{2^{n+2}} U(s)ds = 2\int_{2^n}^{2^{n+1}} U(2s)ds > \int_{2^n}^{2^{n+1}} U(s)ds$$

and so setting $n_0 = \inf\{n : 2^n > s_0\}$ gives

$$\int_{s_0}^\infty U(s)ds \ge \sum_{n:2^n>s_0} \int_{2^{n+1}}^{2^{n+2}} U(s)ds > \sum_{n\ge n_0} \int_{2^{n_0+1}}^{2^{n_0+2}} U(s)ds = \infty.$$

Thus for $\rho > -1, x > 0$, and any $N < \infty$ we have

$$\int_0^t U(sx)ds \sim \int_N^t U(sx)ds, t\to\infty,$$

since $U(sx)$ is a ρ-varying function of s. For fixed x and given ε, there exists N such that for $s > N$

$$(1-\varepsilon)x^\rho U(s) \leq U(sx) \leq (1+\varepsilon)x^\rho U(s)$$

and thus

$$\begin{aligned}\limsup_{t\to\infty} \frac{\int_0^{tx} U(s)ds}{\int_0^t U(s)ds} =& \limsup_{t\to\infty} \frac{x\int_0^t U(sx)ds}{\int_0^t U(s)ds}\\ =& \limsup_{t\to\infty} \frac{x\int_N^t U(sx)ds}{\int_N^t U(s)ds}\\ \leq& \limsup_{t\to\infty} x^{\rho+1}(1+\varepsilon)\frac{\int_N^t U(s)ds}{\int_N^t U(s)ds}\\ =&(1+\varepsilon)x^{\rho+1}.\end{aligned}$$

An analogous argument applies for lim inf and thus we have proved

$$\int_0^x U(s)ds \in RV_{\rho+1}$$

when $\rho > -1$.

In case $\rho = -1$ then either $\int_0^\infty U(s)ds < \infty$ in which case $\int_0^x U(s)ds \in RV_{-1+1} = RV_0$ or else $\int_0^\infty U(s)ds = \infty$ and the previous argument is applicable. So we have checked that for $\rho \geq -1, \int_0^x U(s)ds \in RV_{\rho+1}$.

We now focus on proving (3.11) when $U \in RV_\rho$, $\rho \geq -1$. As in the development leading to (3.22), set

$$b(x) = xU(x)/\int_0^x U(t)dt$$

so that integrating $b(x)/x$ leads to the representations

$$\begin{aligned}\int_0^x U(s)ds =& c\exp\left\{\int_1^x t^{-1}b(t)dt\right\}\\ U(x) =& cx^{-1}b(x)\exp\left\{\int_1^x t^{-1}b(t)dt\right\}.\end{aligned} \tag{3.15}$$

We must show $b(x) \to \rho+1$. Observe first that

$$\begin{aligned}\liminf_{x\to\infty} 1/b(x) =& \liminf_{x\to\infty} \frac{\int_0^x U(t)dt}{xU(x)}\\ =& \liminf_{x\to\infty} \int_0^1 \frac{U(sx)}{U(x)}ds.\end{aligned}$$

Now make a change of variable $s = x^{-1}t$ and and by Fatou's lemma this is

$$\geq \int_0^1 \liminf_{x\to\infty}(U(sx)/U(x))ds$$

$$= \int_0^1 s^\rho ds = \frac{1}{\rho+1}$$

and we conclude

$$\limsup_{x\to\infty} b(x) \le \rho + 1. \tag{3.16}$$

If $\rho = 1$ then $b(x) \to 0$ as desired, so now suppose $\rho > -1$.

We observe the following properties of $b(x)$:

(i) $b(x)$ is bounded on a semi-infinite neighborhood of ∞ (by (3.16)).

(ii) b is slowly varying since $xU(x) \in RV_{\rho+1}$ and $\int_0^x U(s)ds \in RV_{\rho+1}$.

(iii) We have

$$b(xt) - b(x) \to 0$$

boundedly as $x \to \infty$.

The last statement follows since by slow variation

$$\lim_{x\to\infty} (b(xt) - b(x))/b(x) = 0$$

and the denominator is ultimately bounded.

From (iii) and dominated convergence

$$\lim_{x\to\infty} \int_1^s t^{-1}(b(xt) - b(x))dt = 0$$

and the left side may be rewritten to obtain

$$\lim_{x\to\infty} \left\{ \int_1^s t^{-1} b(xt)dt - b(x) \log s \right\} = 0. \tag{3.17}$$

From (3.15)

$$c \exp\left\{ \int_1^x t^{-1} b(t)dt \right\} = \int_0^x U(s)ds \in RV_{\rho+1}$$

and from the regular variation property

$$\begin{aligned}(\rho+1)\log s &= \lim_{x\to\infty} \log \left\{ \frac{\int_0^{xs} U(t)dt}{\int_0^x U(t)dt} \right\} \\ &= \lim_{x\to\infty} \int_x^{xs} t^{-1} b(t)dt = \lim_{x\to\infty} \int_1^s t^{-1} b(xt)dt\end{aligned}$$

and combining this with (3.17) leads to the desired conclusion that $b(x) \to \rho + 1$.

(b). We suppose (3.13) holds and check $U \in RV_{\lambda-1}$. Set

$$b(x) = xU(x) / \int_0^x U(t)dt$$

so that $b(x) \to \lambda$. From (3.15)

$$U(x) = cx^{-1} b(x) \exp \int_1^x t^{-1} b(t)dt \Big\}$$

$$= cb(x) \exp \int_1^x t^{-1}(b(t) - 1)dt \Big\}$$

and since $b(t) - 1 \to \lambda - 1$ we see that U satisfies the representation of a $(\lambda - 1)$ varying function. □

3.3.3. *Karamata's representation.* Theorem 1 leads in a straightforward way to what has been called the *Karamata representation* of a regularly varying function.

Corollary 1 (The Karamata Representation)**.** *(i) The function L is slowly varying iff L can be represented as*

$$L(x) = c(x) \exp \left\{ \int_1^x t^{-1}\varepsilon(t)dt \right\}, \quad x > 0, \tag{3.18}$$

where $c : \mathbb{R}_+ \mapsto \mathbb{R}_+$, $\varepsilon : \mathbb{R}_+ \mapsto \mathbb{R}_+$ and

$$\lim_{x\to\infty} c(x) = c \in (0, \infty), \tag{3.19}$$

$$\lim_{t\to\infty} \varepsilon(t) = 0. \tag{3.20}$$

(ii) A function $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is regularly varying with index ρ iff U has the representation

$$U(x) = c(x) \exp \left\{ \int_1^x t^{-1}\rho(t)dt \right\} \tag{3.21}$$

where $c(\cdot)$ satisfies (3.19) *and* $\lim_{t\to\infty} \rho(t) = \rho$*. (This is obtained from (i) by writing $U(x) = x^\rho L(x)$ and using the representation for L.)*

Proof. If L has a representation (3.18) then it must be slowly varying since for $x > 1$

$$\lim_{t\to\infty} L(tx)/L(t) = \lim_{t\to\infty} (c(tx)/c(t)) \exp \left\{ \int_t^{tx} s^{-1}\varepsilon(s)ds \right\}.$$

Given ε, there exists t_0 by (3.20) such that

$$-\varepsilon < \varepsilon(t) < \varepsilon, \quad t \geq t_0,$$

so that

$$-\varepsilon \log x = -\varepsilon \int_t^{tx} s^{-1}ds \leq \int_t^{tx} s^{-1}\varepsilon(s)ds \leq \varepsilon \int_t^{tx} s^{-1}ds = \varepsilon \log x.$$

Therefore $\lim_{t\to\infty} \int_t^{tx} s^{-1}\varepsilon(s)ds = 0$ and $\lim_{t\to\infty} L(tx)/L(t) = 1$.

Conversely suppose $L \in RV_0$. By Karamata's theorem

$$b(x) := xL(x)/\int_0^x L(s)ds \to 1$$

and $x \to \infty$. Note

$$L(x) = x^{-1}b(x) \int_0^x L(s)ds.$$

Set $\varepsilon(x) = b(x) - 1$ so $\varepsilon(x) \to 0$ and

$$\int_1^x t^{-1}\varepsilon(t)dt = \int_1^x \left(L(t)/\int_0^t L(s)ds \right) dt - \log x$$

$$= \int_1^x d\left(\log \int_0^t L(s)ds\right) - \log x$$
$$= \log\left(x^{-1}\int_0^x L(s)ds \Big/ \int_0^1 L(s)ds\right)$$

whence

$$\exp\left\{\int_1^x t^{-1}\varepsilon(t)dt\right\} = x^{-1}\int_0^x L(s)ds\Big/\int_0^1 L(s)ds$$
$$= L(x)\Big/\left(b(x)\int_0^1 L(s)ds\right), \tag{3.22}$$

and the representation follows with

$$c(x) = b(x)\int_0^1 L(s)ds.$$

□

3.3.4. *Differentiation.* The previous results describe the asymptotic properties of the indefinite integral of a regularly varying function. We now describe what happens when a ρ-varying function is differentiated.

Proposition 4. *Suppose $U : R_+ \mapsto R_+$ is absolutely continuous with density u so that*

$$U(x) = \int_0^x u(t)dt.$$

(a) (Von Mises) If

$$\lim_{x\to\infty} xu(x)/U(x) = \rho, \tag{3.23}$$

then $U \in RV_\rho$.

(b) (Landau, 1916) See also (de Haan, 1970, page 23, 109), Seneta (1976),Resnick (1987). If $U \in RV_\rho, \rho \in \mathbb{R}$, and u is monotone then (3.23) *holds and if $\rho \neq 0$ then $|u|(x) \in RV_{\rho-1}$.*

Proof. (a) Set

$$b(x) = xu(x)/U(x)$$

and as before we find

$$U(x) = U(1)\exp\left\{\int_1^x t^{-1}b(t)dt\right\}$$

so that U satisfies the representation theorem for a ρ-varying function.

(b) Suppose u is nondecreasing. An analogous proof works in the case u is nonincreasing. Let $0 < a < b$ and observe

$$(U(xb) - U(xa))/U(x) = \int_{xa}^{xb} u(y)dy/U(x).$$

By monotonicity we get

$$u(xb)x(b-a)/U(x) \geq (U(xb) - U(xa))/U(x) \geq u(xa)x(b-a)/U(x). \tag{3.24}$$

From (22) and the fact that $U \in RV_\rho$ we conclude

$$\limsup_{x\to\infty} xu(xa)/U(x) \le (b^\rho - a^\rho)/(b-a) \tag{3.25}$$

for any $b > a > 0$. So let $b \downarrow a$, which is tantamount to taking a derivative. Then (3.25) becomes

$$\limsup_{x\to\infty} xu(xa)/U(x) \le \rho a^{\rho-1} \tag{3.26}$$

for any $a > 0$. Similarly from the left-hand equality in (3.24) after letting $a \uparrow b$ we get

$$\liminf_{x\to\infty} xu(xb)/U(x) \ge \rho b^{\rho-1} \tag{3.27}$$

for any $b > 0$. Then (3.23) results by setting $a = 1$ in (3.26) and $b = 1$ in (3.27). □

3.4. **Regular variation: Further properties.** For the following list of properties, it is convenient to define *rapid variation* or regular variation with index ∞. Say $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is regularly varying with index ∞ ($U \in RV_\infty$) if for every $x > 0$

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\infty := \begin{cases} 0, & \text{if } x < 1, \\ 1, & \text{if } x = 1, \\ \infty, & \text{if } x > 1. \end{cases}$$

Similarly $U \in RV_{-\infty}$ if

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^{-\infty} := \begin{cases} \infty, & \text{if } x < 1, \\ 1, & \text{if } x = 1, \\ 0, & \text{if } x > 1. \end{cases}$$

The following proposition collects useful properties of regularly varying functions. (See de Haan (1970).)

Proposition 5. *(i) If $U \in RV_\rho$, $-\infty \le \rho \le \infty$, then*

$$\lim_{x\to\infty} \log U(x)/\log x = \rho$$

so that

$$\lim_{x\to\infty} U(x) = \begin{cases} 0, & \text{if } \rho < 0, \\ \infty, & \text{if } \rho > 0. \end{cases}$$

(ii) (Potter bounds.) Suppose $U \in RV_\rho$, $\rho \in \mathbb{R}$. Take $\varepsilon > 0$. Then there exists t_0 such that for $x \ge 1$ and $t \ge t_0$

$$(1-\varepsilon)x^{\rho-\varepsilon} < \frac{U(tx)}{U(t)} < (1+\varepsilon)x^{\rho+\varepsilon}. \tag{3.28}$$

(iii) If $U \in RV_\rho$, $\rho \in \mathbb{R}$, and $\{a_n\}$, $\{a'_n\}$ satisfy, $0 < a_n \to \infty$, $0 < a_{n'} \to \infty$, and $a_n \sim c a_{n'}$, for $0 < c < \infty$, then $U(a_n) \sim c^\rho U(a_{n'})$. If $\rho \ne 0$ the result also holds for $c = 0$ or ∞. Analogous results hold with sequences replaced by functions.

(iv) If $U_1 \in RV_{\rho_1}$ and $U_2 \in RV_{\rho_2}$ and $\lim_{x\to\infty} U_2(x) = \infty$ then

$$U_1 \circ U_2 \in RV_{\rho_1\rho_2}.$$

(v) Suppose U is nondecreasing, $U(\infty)=\infty$, and $U\in RV_\rho, 0\le\rho\le\infty$. Then

$$U^{\leftarrow}\in RV_{\rho^{-1}}.$$

(vi) Suppose U_1,U_2 are nondecreasing and ρ-varying, $0<\rho<\infty$. Then for $0\le c\le\infty$

$$U_1(x)\sim cU_2(x),\quad x\to\infty$$

iff

$$U_1^{\leftarrow}(x)\sim c^{-\rho^{-1}}U_2^{\leftarrow}(x),\quad x\to\infty.$$

(vii) If $U\in RV_\rho$, $\rho\neq 0$, then there exists a function U^* which is absolutely continuous, strictly monotone, and

$$U(x)\sim U(x)^*,\quad x\to\infty.$$

Proof. (i) We give the proof for the case $0<\rho<\infty$. Suppose U has Karamata representation

$$U(x)=c(x)\exp\int_1^x t^{-1}\rho(t)dt\Big\}$$

where $c(x)\to c>0$ and $\rho(t)\to\rho$. Then

$$\log U(x)/\log x=o(1)+\int_1^x t^{-1}\rho(t)dt/\int_1^x t^{-1}dt\to\rho.$$

(ii) Using the Karamata representation

$$U(tx)/U(t)=(c(tx)/c(t))\exp\left\{\int_1^x s^{-1}\rho(ts)ds\right\}$$

and the result is apparent since we may pick t_0 so that $t>t_0$ implies $\rho-\varepsilon<\rho(ts)<\rho+\varepsilon$ for $s>1$.

(iii) If $c>0$ then from the uniform convergence property in Proposition 3

$$\lim_{n\to\infty}\frac{U(a_n)}{U(a_{n'})}=\lim_{n\to\infty}\frac{U(a_{n'}(a_n/a_{n'}))}{U(a_{n'})}=\lim_{t\to\infty}\frac{U(tc)}{U(t)}=c^\rho.$$

(iv) Again by uniform convergence, for $x>0$

$$\begin{aligned}\lim_{t\to\infty}\frac{U_1(U_2(tx))}{U_1(U_2(t))}&=\lim_{t\to\infty}\frac{U_1(U_2(t)(U_2(tx)/U_2(t)))}{U_1(U_2(t))}\\&=\lim_{y\to\infty}\frac{U_1(yx^{\rho_2})}{U_1(y)}=x^{\rho_2\rho_1}.\end{aligned}$$

(v) Let $U_t(x)=U(tx)/U(t)$ so that if $U\in RV_\rho$ and U is nondecreasing then $(0<\rho<\infty)$

$$U_t(x)\to x^\rho,\quad t\to\infty,$$

which implies by Proposition 1

$$U_t^{\leftarrow}(x)\to x^{\rho^{-1}},\quad t\to\infty;$$

that it,

$$\lim_{t\to\infty}U^{\leftarrow}(xU(t))/t=x^{\rho^{-1}}.$$

Therefore

$$\lim_{t\to\infty} U^{\leftarrow}(xU(U^{\leftarrow}(t)))/U^{\leftarrow}(t) = x^{\rho^{-1}}.$$

This limit holds locally uniformly since monotone functions are converging to a continuous limit. Now $U \circ U^{\leftarrow}(t) \sim t$ as $t \to \infty$, and if we replace x by $xt/U \circ U^{\leftarrow}(t)$ and use uniform convergence we get

$$\begin{aligned}\lim_{t\to\infty} \frac{U^{\leftarrow}(tx)}{U^{\leftarrow}(t)} &= \lim_{t\to\infty} \frac{U^{\leftarrow}((xt/U \circ U^{\leftarrow}(t))U \circ U^{\leftarrow}(t))}{U^{\leftarrow}(t)} \\ &= \lim_{t\to\infty} \frac{U^{\leftarrow}(xU \circ U^{\leftarrow}(t))}{U^{\leftarrow}(t)} = x^{\rho^{-1}}\end{aligned}$$

which makes $U^{\leftarrow} \in RV_{\rho^{-1}}$.

(vi) If $c > 0, 0 < \rho < \infty$ we have for $x > 0$

$$\lim_{t\to\infty} \frac{U_1(tx)}{U_2(t)} = \lim_{t\to\infty} \frac{U_1(tx)U_2(tx)}{U_2(tx)U_2(t)} = cx^{\rho}.$$

Inverting we find for $y > 0$

$$\lim_{t\to\infty} U_1^{\leftarrow}(yU_2(t))/t = (c^{-1}y)^{\rho^{-1}}$$

and so

$$\lim_{t\to\infty} U_1^{\leftarrow}(yU_2 \circ U_2^{\leftarrow}(t))/U_2^{\leftarrow}(t) = (c^{-1}y)^{\rho^{-1}}$$

and since $U_2 \circ U_2^{\leftarrow}(t) \sim t$

$$\lim_{t\to\infty} U_1^{\leftarrow}(yt)/U_2^{\leftarrow}(t) = (c^{-1}y)^{\rho^{-1}}.$$

Set $y = 1$ to obtain the result.

(vii) For instance if $U \in RV_{\rho}, \rho > 0$ define

$$U^*(t) = \int_1^t s^{-1}U(s)ds.$$

Then $s^{-1}U(s) \in RV_{\rho-1}$ and by Karamata's theorem

$$U(x)/U^*(x) \to \rho.$$

U^* is absolutely continuous and since $U(x) \to \infty$ when $\rho > 0, U^*$ is ultimately strictly increasing. □

References

P. Billingsley. *Probability and Measure.* Wiley, New York, 2nd edition, 1986.

N. Bingham, C. Goldie, and J. Teugels. *Regular Variation.* Cambridge University Press, 1987.

L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes.* Mathematisch Centrum Amsterdam, 1970.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, New York, 2nd edition, 1971.

J. L. Geluk and L. de Haan. *Regular Variation, Extensions and Tauberian Theorems*, volume 40 of *CWI Tract*. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1987. ISBN 90-6196-324-9.

S. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

S. Resnick. *A Probability Path*. Birkhäuser, Boston, 1998.

E. Seneta. *Regularly Varying Functions*. Springer-Verlag, New York, 1976. Lecture Notes in Mathematics, 508.

4. A Crash Course in Weak Convergence.

Many asymptotic properties of statistics in heavy tailed analysis are clearly understood with a fairly high level interpretation which comes from the modern theory of weak convergence of probability measures on metric spaces as originally promoted in Billingsley (1968) and updated in Billingsley (1999).

4.1. **Definitions.** Let $\mathbb{S}$ be a complete, separable metric space with metric d and let $\mathcal{S}$ be the Borel σ- algebra of subsets of $\mathbb{S}$ generated by open sets. Suppose $(\Omega, \mathcal{A}, \mathbb{P})$ is a probability space. A random element X in $\mathbb{S}$ is a measurable map from such a space $(\Omega, \mathcal{A})$ into $(\mathbb{S}, \mathcal{S})$.

With a random variable, a point $\omega \in \Omega$ is mapped into a real valued member of $\mathbb{R}$. With a random element, a point $\omega \in \Omega$ is mapped into a an element of the metric space $\mathbb{S}$. Here are some common examples of this paradigm.

Metric space $\mathbb{S}$	**Random element X is a:**
$\mathbb{R}$	random variable
$\mathbb{R}^d$	random vector
$C[0,\infty)$, the space of real valued, continuous functions on $[0,\infty)$	random process with continuous paths
$D[0,\infty)$, the space of real valued, right continuous functions on $[0,\infty)$ with finite left limits existing on $(0,\infty)$	right continuous random process with jump discontinuities
$M_p(\mathbb{E})$, the space of point measures on a nice space $\mathbb{E}$	stochastic point process on $\mathbb{E}$
$M_+(\mathbb{E})$, the space of Radon measures on a nice space $\mathbb{E}$.	random measure on $\mathbb{E}$

Table 1. Various metric spaces and random elements.

Given a sequence $\{X_n, n \geq 0\}$ of random elements of $\mathbb{S}$, there is a corresponding sequence of distributions on $\mathcal{S}$,

$$P_n = \mathbb{P} \circ X_n^{-1} = \mathbb{P}[X_n \in \cdot], \quad n \geq 0.$$

P_n is called the *distribution* of X_n. Then X_n converges weakly to X_0 (written $X_n \Rightarrow X_0$ or $P_n \Rightarrow P_0$) if whenever $f \in C(\mathbb{S})$, the class of bounded, continuous real valued functions on $\mathbb{S}$, we have

$$\mathbf{E}f(X_n) = \int_{\mathbb{S}} f(x) P_n(dx) \to \mathbf{E}f(X_0) = \int_{\mathbb{S}} f(x) P_0(dx).$$

Recall that the definition of weak convergence of random variables in $\mathbb{R}$ is given in terms of one dimensional distribution functions which does not generalize nicely to higher dimensions. The definition in terms of integrals of test functions $f \in C(\mathbb{S})$ is very flexible and well defined for any metric space $\mathbb{S}$.

4.2. **Basic properties of weak convergence.**

4.2.1. *Portmanteau Theorem.* The basic *Portmanteau Theorem* ((Billingsley, 1968, page 11), Billingsley (1999)) says the following are equivalent:

$$X_n \Rightarrow X_0. \tag{4.1}$$

$$\lim_{n\to\infty} \mathbb{P}[X_n \in A] = \mathbb{P}[X_0 \in A], \ \forall A \in \mathcal{S} \text{ such that } \mathbb{P}[X_0 \in \partial A] = 0. \tag{4.2}$$

Here ∂A denotes the boundary of the set A.

$$\limsup_{n\to\infty} \mathbb{P}[X_n \in F] \leq \mathbb{P}[X_0 \in F], \ \forall \text{ closed } F \in \mathcal{S}. \tag{4.3}$$

$$\liminf_{n\to\infty} \mathbb{P}[X_n \in G] \geq \mathbb{P}[X_0 \in G], \ \forall \text{ open } G \in \mathcal{S}. \tag{4.4}$$

$$\mathbf{E}f(X_n) \to \mathbf{E}f(X_0), \quad \text{for all } f \text{ which are bounded and } \underline{\text{uniformly}} \text{ continuous.} \tag{4.5}$$

Although it may seem comfortable to express weak convergence of probability measures in terms of sets, it is mathematically simplist to rely on integrals with respect to test functions as given, for instance, in (4.5).

4.2.2. *Skorohod's theorem.* A nice way to think about weak convergence is using *Skorohod's theorem* ((Billingsley, 1971, Proposition 0.2)) which, for certain purposes, allows one to replace convergence in distribution with almost sure convergence. In a theory which relies heavily on continuity, this is a big advantage. Almost sure convergence, being pointwise, is very well suited to continuity arguments.

Let $\{X_n, n \geq 0\}$ be random elements of the metric space $(\mathbb{S}, \mathcal{S})$ and suppose the domain of each X_n is $(\Omega, \mathcal{A}, \mathbb{P})$. Let

$$([0,1], \mathcal{B}[0,1], \mathbb{LEB}(\cdot))$$

be the usual probability space on $[0,1]$, where $\mathbb{LEB}(\cdot)$ is Lebesgue measure or length. Skorohod's theorem says that $X_n \Rightarrow X_0$ iff there exists random elements $\{X_n^*, n \geq 0\}$ in $\mathbb{S}$ defined on the uniform probability space such that

$$X_n \stackrel{d}{=} X_n^*, \quad \text{for each } n \geq 0,$$

and

$$X_n^* \to X_0^* \text{ a.s.}$$

The second statement means

$$\mathbb{LEB}\Big\{t \in [0,1] : \lim_{n\to\infty} d(X_n^*(t), X_0^*(t)) = 0\Big\} = 1.$$

Almost sure convergence always implies convergence in distribution so Skorohod's theorem provides a partial converse. To see why almost sure convergence implies weak convergence is easy. With $d(\cdot,\cdot)$ as the metric on $\mathbb{S}$ we have $d(X_n, X_0) \to 0$ almost surely and for any $f \in C(\mathbb{S})$ we get by continuity that $f(X_n) \to f(X_0)$, almost surely. Since f is bounded, by dominated convergence we get $\mathbf{E}f(X_n) \to \mathbf{E}f(X_0)$.

Recall that in one dimension, Skorohod's theorem has an easy proof. If $X_n \Rightarrow X_0$ and X_n has distribution function F_n then

$$F_n \to F_0, \quad n \to \infty.$$

Thus, by Proposition 1, $F_n^{\leftarrow} \to F_0^{\leftarrow}$. Then with U, the identity function on $[0,1]$, (so that U is uniformly distributed)

$$X_n \stackrel{d}{=} F_n^{\leftarrow}(U) =: X_n^*, \quad n \geq 0,$$

and

$$\begin{aligned}\mathbb{LEB}[X_n^* \to X_0^*] =& \mathbb{LEB}\{t \in [0,1] : F_n^{\leftarrow}(t) \to F_n^{\leftarrow}(t)]\\ \geq& \mathbb{LEB}\big(\mathcal{C}(F_0^{\leftarrow})\big) = 1,\end{aligned}$$

since the set of discontinuities of the monotone function $F_0^{\leftarrow}(\cdot)$ is countable, and hence has Lebesgue measure 0.

The power of weak convergence theory comes from the fact that once a basic convergence result has been proved, many corollaries emerge with little effort, often using only continuity. Suppose $(\mathbb{S}_i, d_i), i = 1, 2$, are two metric spaces and $h : \mathbb{S}_1 \mapsto \mathbb{S}_2$ is continuous. If $\{X_n, n \geq 0\}$ are random elements in $(\mathbb{S}_1, \mathcal{S}_1)$ and $X_n \Rightarrow X_0$ then $h(X_n) \Rightarrow h(X_0)$ as random elements in $(\mathbb{S}_2, \mathcal{S}_2)$.

To check this is easy: Let $f_2 \in C(\mathbb{S}_2)$ and we must show that $\mathbf{E}f_2(h(X_n)) \to \mathbf{E}f_2(h(X_0))$. But $f_2(h(X_n)) = f_2 \circ h(X_n)$ and since $f_2 \circ h \in C(\mathbb{S}_1)$, the result follows from the definition of $X_n \Rightarrow X_0$ in $\mathbb{S}_1$.

If $\{X_n\}$ are random variables which converge, then letting $h(x) = x^2$ or $\arctan x$ or $\dots$ yields additional convergences for free.

4.2.3. *Continuous mapping theorem.* In fact, the function h used in the previous paragraphs, need not be continuous everywhere and in fact, many of the maps h we will wish to use are definitely not continuous everywhere.

Theorem 2 (Continuous Mapping Theorem.). *Let $(\mathbb{S}_i, d_i), i = 1, 2$, be two metric spaces and suppose $\{X_n, n \geq 0\}$ are random elements of $(\mathbb{S}_1, \mathcal{S}_1)$ and $X_n \Rightarrow X_0$. For a function $h : S_1 \mapsto S_2$, define the discontinuity set of h as*

$$D_h := \{s_1 \in S_1 : h \textit{ is discontinuous at } s_1\}.$$

If h satisfies

$$\mathbb{P}[X_0 \in D_h] = \mathbb{P}[X_0 \in \{s_1 \in S_1 : h \textit{ is discontinuous at } s_1\}] = 0$$

then

$$h(X_n) \Rightarrow h(X_0)$$

in $\mathbb{S}_2$.

Proof. For a traditional proof, see (Billingsley, 1968, page 30). This result is an immediate consequence of Skorohod's theorem. If $X_n \Rightarrow X_0$ then there exist almost surely convergent random elements of $\mathbb{S}_1$ defined on the unit interval, denoted X_n^*, such that

$$X_n^* \stackrel{d}{=} X_n, \quad n \geq 0.$$

Then it follows that

$$\mathbb{LEB}[h(X_n^*) \to h(X_0^*)] \geq \mathbb{LEB}[X_0^* \notin D_h]$$

where we denote by $\mathrm{disc}(h)$ the discontinuity set of h; that is, the complement of $\mathcal{C}(h)$. Since $X_0 \stackrel{d}{=} X_0^*$ we get the previous probability equal to

$$=\mathbb{P}[X_0 \notin \mathrm{disc}(h)] = 1$$

and therefore $h(X_n^*) \to h(X_0^*)$ almost surely. Since almost sure convergence implies convergence in distribution $h(X_n^*) \Rightarrow h(X_0^*)$. Since $h(X_n) \stackrel{d}{=} h(X_n^*)$, $n \geq 0$, the result follows. □

4.2.4. *Subsequences and Prohorov's theorem.* Often to prove weak convergence, subsequence arguments are used and the following is useful. A family Π of probability measures on a complete, separable metric space is *relatively compact* if every sequence $\{P_n\} \subset \Pi$ contains a weakly convergent subsequence. Relative compactness is theoretically useful but hard to check in practice so we need a workable criterion. Call the family Π tight (and by abuse of language we will refer to the corresponding random elements also as a tight family) if for any ε, there exists a compact $K_\varepsilon \in \boldsymbol{\mathcal{S}}$ such that

$$P(K_\varepsilon) > 1 - \varepsilon, \quad \text{for all } P \in \Pi.$$

This is the sort of condition that precludes probability mass from escaping from the state space. Prohorov's theorem (Billingsley (1968)) assures us that when S is separable and complete, tightness of Π is the same as relative compactness. Tightness is checkable although it is seldom easy.

4.3. **Some useful metric spaces.** It pays to spend a bit of time remembering details of examples of metric spaces that will be useful. To standardize notation we set

$$\begin{aligned}\mathcal{F}(\mathbb{S}) &= \text{ closed subsets of } \mathbb{S},\\ \mathcal{G}(\mathbb{S}) &= \text{ open subsets of } \mathbb{S},\\ \mathcal{K}(\mathbb{S}) &= \text{ compact subsets of } \mathbb{S}.\end{aligned}$$

4.3.1. $\mathbb{R}^d$*, finite dimensional Euclidean space.* We set

$$\mathbb{R}^d := \{(x_1, \dots, x_d) : \ x_i \in \mathbb{R}, \ i = 1, \dots, d\} = \mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}.$$

The metric is defined by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2},$$

for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. Convergence of a sequence in this space is equivalent to componentwise convergence.

Define an interval

$$(\boldsymbol{a}, \boldsymbol{b}] = \{\boldsymbol{x} \in \mathbb{R}^d : a_i < x_i \leq b_i, \ i = 1, \dots, d\}$$

A probability measure P on $\mathbb{R}^d$ is determined by its distribution function

$$F(\boldsymbol{x}) := P(-\boldsymbol{\infty}, \boldsymbol{x}]$$

and a sequence of probability measures $\{P_n, n \geq 0\}$ on $\mathbb{R}^d$ converges to P_0 iff

$$F_n(\boldsymbol{x}) \to F_0(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathcal{C}(F_0).$$

Note this says that a sequence of random vectors converges in distribution iff their distribution functions converge weakly. While this is concrete, it is seldom useful since multivariate distribution functions are usually awkward to deal with in practice.

Also, recall $K \in \mathcal{K}(\mathbb{R}^d)$ iff K is closed and bounded.

4.3.2. $\mathbb{R}^\infty$*, sequence space.* Define

$$\mathbb{R}^\infty := \{(x_1, x_2, \dots) : x_i \in \mathbb{R},\ i \geq 1\} = \mathbb{R} \times \mathbb{R} \times \dots .$$

The metric can be defined by

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{\infty} (|x_i - y_i| \wedge 1) 2^{-i},$$

for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$. This gives a complete, separable metric space where convergence of a family of sequences means coordinatewise convergence which means

$$\boldsymbol{x}(n) \to \boldsymbol{x}(0) \text{ iff } x_i(n) \to x_i(0),\ \forall i \geq 1.$$

The topology $\mathcal{G}(\mathbb{R}^\infty)$ can be generated by basic neighborhoods of the form

$$N_k(\boldsymbol{x}) = \{\boldsymbol{y} : \bigvee_{i=1}^{d} |x_i - y_i| < \epsilon\},$$

as we vary d, the center $\boldsymbol{x}$ and ϵ.

A set $A \subset \mathbb{R}^\infty$ is relatively compact iff every one-dimensional section is bounded; that is iff for any $i \geq 1$

$$\{x_i : \boldsymbol{x} \in A\} \text{ is bounded.}$$

4.3.3. $C[0,1]$ *and* $C[0,\infty)$*, continuous functions.* The metric on $C[0,M]$, the space of real valued continuous functions with domain $[0,M]$ is the uniform metric

$$d_M(x(\cdot), y(\cdot)) = \sup_{0 \leq t \leq M} |x(t) - y(t)| =: \|x(\cdot) - y(\cdot)\|_M.$$

and the metric on $C[0,\infty)$ is

$$d(x(\cdot), y(\cdot)) = \sum_{n=1}^{\infty} \frac{d_n(x,y) \wedge 1}{2^n}$$

where we interpret $d_n(x,y)$ as the $C[0,n]$ distance of x and y restricted to $[0,n]$. The metric on $C[0,\infty)$ induces the topology of local uniform convergence.

For $C[0,1]$ (or $C[0,M]$), we have that every function is uniformly continuous since a continuous function on a compact set is always uniformly continuous. Uniform continuity can be expressed by the modulus of continuity which is defined for $x \in C[0,1]$ by

$$\omega_x(\delta) = \sup_{|t-s|<\delta} |x(t) - x(s)|, \quad 0 < \delta < 1.$$

Then, uniform continuity means

$$\lim_{\delta \to 0} \omega_x(\delta) = 0.$$

The Arzela-Ascoli theorem says a uniformly bounded equicontinuous family of functions in $C[0,1]$ has a uniformly convergent subsequence; that is, this family is relatively compact or has compact closure. Thus a set $A \subset C[0,1]$ is relatively compact iff

(i) A is uniformly bounded; that is,

$$\sup_{0\le t\le 1}\sup_{x\in K}|x(t)|<\infty, \tag{4.6}$$

and

(ii) A is equicontinuous; that is

$$\lim_{\delta\downarrow 0}\sup_{x\in K}\omega_x(\delta)=0.$$

Since the functions in a compact family vary in a controlled way, (4.6) can be replaced by

$$\sup_{x\in K}|x(0)|<\infty. \tag{4.7}$$

Compare this result with the compactness characterization in $\mathbb{R}^\infty$ where compactness meant each one-dimensional section was compact. Here, a continuous function is compact if each one dimensional section is compact in a uniform way AND equicontinuity is present.

4.3.4. $D[0,1]$ *and* $D[0,\infty)$. Start by considering $D[0,1]$, the space of right continuous functions on $[0,1)$ which have finite left limits on $(0,1]$. Minor changes allow us to consider $D[0,M]$ for any $M>0$.

In the uniform topology, two functions $x(\cdot)$ and $y(\cdot)$ are close if their graphs are uniformly close. In the Skorohod topology on $D[0,1]$, we consider x and y close if after deforming the time scale of one of them, say y, the resulting graphs are close. Consider the following simple example:

$$x_n(t)=1_{[0,\frac12+\frac1n]}(t),\quad x(t)=1_{[0,\frac12]}(t). \tag{4.8}$$

The uniform distance is always 1 but a time deformation allows us to consider the functions to be close. (Various metrics and their applications to functions with jumps are considered in detail in Whitt (2002).)

Define time deformations

$$\begin{aligned}\Lambda=\{\lambda :&[0,1]\mapsto[0,1]:\lambda(0)=0,\ \lambda(1)=1,\\ &\lambda(\cdot)\text{ is continuous, strictly increasing, 1-1, onto.}\}\end{aligned} \tag{4.9}$$

Let $e(t)\in\Lambda$ be the identity transformation and denote the uniform distance between x,y as

$$\|x-y\|:=\bigvee_{t=0}^{1}|x(t)-y(t)|.$$

The Skorohod metric $d(x,y)$ between two functions $x,y\in D[0,1]$ is

$$\begin{aligned}d(x,y)=&\inf\{\epsilon>0:\exists\lambda\in\Lambda,\ \text{such that }\|\lambda-e\|\vee\|x-y\circ\lambda\|\le\epsilon\},\\ =&\inf_{\lambda\in\Lambda}\|\lambda-e\|\vee\|x-y\circ\lambda\|.\end{aligned}$$

Simple consequences of the definitions:

(1) Given a sequence $\{x_n\}$ of functions in $D[0,1]$, we have $d(x_n, x_0) \to 0$, iff there exist $\lambda_n \in \Lambda$ and

$$\|\lambda - e\| \to 0, \quad \|x_n \circ \lambda_n - x_0\| \to 0. \tag{4.10}$$

(2) From the definition, we always have

$$d(x, y) \leq \|x - y\|, \quad x, y \in D[0,1]$$

since one choice of λ is the identity but this may not give the infimum. Therefore, uniform convergence always implies Skorohod convergence. The converse is very false; see (4.8).

(3) If $d(x_n, x_0) \to 0$ for $x_n, \in D[0,1]$, $n \geq 0$, then for all $t \in \mathcal{C}(x_0)$, we have pointwise convergence

$$x_n(t) \to x_0(t).$$

To see this suppose (4.10) holds. Then

$$\|\lambda_n - e\| = \|\lambda_n^{\leftarrow} - e\| \to 0.$$

Thus

$$\begin{aligned} |x_n(t) - x_0(t)| \leq & |x_n(t) - x_0 \circ \lambda_n^{\leftarrow}(t)| + |x_0 \circ \lambda_n^{\leftarrow}(t) - x_0(t)| \\ \leq & \|x_n \circ \lambda_n - x_0\| + o(1), \end{aligned}$$

since x is continuous at t and $\lambda_n^{\leftarrow} \to e$.

(4) If $d(x_n, x_0) \to 0$ and $x \in C[0,1]$, then uniform convergence holds.

If (4.10) holds then as in item 3 we have for each $t \in [0,1]$

$$|x_n(t) - x_0(t)| \leq \|x_n \circ \lambda_n - x_0\| + \|x_0 - x_0 \circ \lambda_n\| \to 0$$

and hence

$$\|x_n(t) - x_0(t)\| \to 0.$$

THE SPACE $D[0, \infty)$. Now we extend this metric to $D[0, \infty)$. For a function $x \in D[0, \infty)$ write

$$r_s x(t) = x(t), \quad 0 \leq t \leq s,$$

for the restriction of x to the interval $[0, s]$ and write

$$\|x\|_s = \bigvee_{t=0}^{s} |x(t)|.$$

Let d_s be the Skorohod metric on $D[0, s]$ and define d_∞, the Skorohod metric on $D[0, \infty)$ by

$$d_\infty(x, y) = \int_0^\infty e^{-s}\Big(d_s(r_s x, r_s y) \wedge 1\Big) ds.$$

The impact of this is that Skorohod convergence on $D[0, \infty)$ reduces to convergence on finite intervals since $d_\infty(x_n, x_0) \to 0$ iff for any $s \in \mathcal{C}(x_0)$ we have $d_s(r_s x_n, r_s x_0) \to 0$.

4.3.5. *Radon measures and point measures; vague convergence.* Suppose $\mathbb{E}$ is a nice space. The technical meaning of *nice* is that $\mathbb{E}$ should be a locally compact topological space with countable base but it is safe to think of $\mathbb{E}$ as a finite dimensional Euclidean space and we may think of $\mathbb{E}$ as a subset of $\mathbb{R}^d$. The case $d = 1$ is important but $d > 1$ is also very useful. When it comes time to construct point processes, $\mathbb{E}$ will be the space where our points live. We assume $\mathbb{E}$ comes with a σ-field $\mathcal{E}$ which can be the σ-field generated by the open sets or, equivalently, the rectangles of E. So the important sets in $\mathcal{E}$ are built up from rectangles.

How can we model a random distribution of points in $\mathbb{E}$? One way is to specify random elements in $\mathbb{E}$, say $\{X_n\}$, and then to say that a stochastic point process is the counting function whose value at the region $A \in \mathcal{E}$ is the number of random elements $\{X_n\}$ which fall in A. This is intuitively appealing but has some technical drawbacks and it mathematically preferable to focus on counting functions rather than points.

A measure $\mu : \mathcal{E} \mapsto [0, \infty]$ is an assignment of positive numbers to sets in $\mathcal{E}$ such that

(1) $\mu(\emptyset) = 0$ and $\mu(A) \geq 0$, for all $A \in \mathcal{E}$,
(2) If $\{A_n, n \geq 1\}$ are mutually disjoint sets in $\mathcal{E}$, then the σ-additivity property holds

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

The measure μ is called *Radon*, if

$$\mu(K) < \infty, \quad \forall K \in \mathcal{K}(\mathbb{E});$$

so that compact sets are known to have finite μ-mass. Knowing where the measure is required to be finite helps us to keep track of infinities in a useful way and prevents illegal operations like $\infty - \infty$.

Define

$$M_+(\mathbb{E}) = \{\mu : \mu \text{ is a non-negative measure on } \mathcal{E}\text{and } \mu \text{ is Radon.}\} \tag{4.11}$$

The space $M_+(\mathbb{E})$ can be made into a complete separable metric space under what is called the *vague* metric. For now, instead of describing the metric, we will describe the notion of convergence the metric engenders.

The way we defined convergence of *probability* measures was by means of test functions. We integrate a test function which is bounded and continuous on the metric space and if the resulting sequence of numbers converges, then we have weak convergence. However, with infinite measures in $M_+(\mathbb{E})$, we cannot just integrate a bounded function to get something finite. However, we know our measures are also Radon and this suggests using functions which vanish off of compact sets. So define

$$C_K^+(\mathbb{E}) = \{f : \mathbb{E} \mapsto \mathbb{R}_+ : f \text{ is continuous with compact support.}\}$$

For a function to have compact support means that it vanishes off a compact set.

The notion of convergence in $M_+(\mathbb{E})$: If $\mu_n \in M_+(\mathbb{E})$, for $n \geq 0$, then μ_n converges vaguely to μ_0, written $\mu_n \stackrel{v}{\to} \mu_0$, provided for all $f \in C_K^+(\mathbb{E})$ we have

$$\mu_n(f) := \int_{\mathbb{E}} f(x)\mu_n(dx) \to \mu_0(f) := \int_{\mathbb{E}} f(x)\mu_0(dx),$$

as $n \to \infty$.

Example 3 (Trivial but mildly illuminating example)**.** Suppose $\mathbb{E}$ is some finite dimensional Euclidean space and define for $\boldsymbol{x} \in \mathbb{E}$, and $A \in \mathcal{E}$

$$\epsilon_{\boldsymbol{x}}(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \in A^c. \end{cases}$$

Then

$$\mu_n := \epsilon_{\boldsymbol{x}_n} \xrightarrow{v} \mu_n := \epsilon_{\boldsymbol{x}_0},$$

in $M_+(\mathbb{E})$, iff

$$\boldsymbol{x}_n \to \boldsymbol{x}_0$$

in the metric on $\mathbb{E}$.

To see this, suppose that $\boldsymbol{x}_n \to \boldsymbol{x}_0$ and $f \in C_K^+(\mathbb{E})$. Then

$$\mu_n(f) = f(\boldsymbol{x}_n) \to f(\boldsymbol{x}_0) = \mu_0(f),$$

since f is continuous and the points are converging. Conversely, suppose $\boldsymbol{x}_n \not\to \boldsymbol{x}_0$. Define $\phi : \mathbb{R} \mapsto [0,1]$ by

$$\phi(t) = \begin{cases} 1, & \text{if } t < 0, \\ 1 - t, & \text{if } 0 \leq t \leq 1, \\ 0, & \text{if } t > 1. \end{cases}$$

There exists $\{n'\}$ such that $d(\boldsymbol{x}_{n'}, \boldsymbol{x}_0) > \epsilon$. Define

$$f(\boldsymbol{y}) = \phi(d(\boldsymbol{x}_0, \boldsymbol{y})/\epsilon)$$

so that $f \in C_K(\mathbb{E})$. Then

$$|f(\boldsymbol{x}_{n'}) - f(\boldsymbol{x}_0)| = |0 - 1| \not\to 0$$

and then we have $\mu_n(f) \not\to \mu_0(f)$.

Point measures. A point measure m is an element of $M_+(\mathbb{E})$ of the form

$$m = \sum_i \epsilon_{x_i}. \tag{4.12}$$

Built into this definition is the understanding that $m(\cdot)$ is Radon: $m(K) < \infty$, for $K \in \mathcal{K}(\mathbb{E})$. Think of $\{x_i\}$ as the atoms and m as the function which counts how many atoms fall in a set. The set $M_p(\mathbb{E})$ is the set of all point measures satisfying (4.12). This turns out to be a closed subset of $M_+(\mathbb{E})$.

More on $M_+(\mathbb{E})$ (and hence, more on $M_p(\mathbb{E})$): THE VAGUE TOPOLOGY ON $M_+(\mathbb{E})$, OPEN SETS: We can specify open sets, a topology (a system of open sets satisfying closure properties) and then a notion of "distance" in $M_+(\mathbb{E})$. Define a *basis* set to be a subset of $M_+(\mathbb{E})$ of the form

$$\{\mu \in M_+(\mathbb{E}) : \mu(f_i) \in (a_i, b_i), i = 1, \ldots, k\} \tag{4.13}$$

where $f_i \in C_K^+(\mathbb{E})$ and $0 \leq a_i \leq b_i$. Now imagine varying the choices of integer k, functions $f_1, \ldots, f_k$, and endpoints $a_1, \ldots, a_k; b_1, \ldots, b_k$. Unions of basis sets form the class of open sets constituting the vague topology.

The topology is metrizable as a complete, separable metric space and we can put a metric $d(\cdot,\cdot)$ on the space which yields the same open sets. The metric "d" can be specified as follows: There is some sequence of functions $f_i \in C_K^+(\mathbb{E})$ and for $\mu_1, \mu_2 \in M_+(\mathbb{E})$

$$d(\mu_1,\mu_2) = \sum_{i=1}^{\infty} \frac{|\mu_1(f_i) - \mu_2(f_i)| \wedge 1}{2^i}. \tag{4.14}$$

An interpretation: If $\mu \in M_+(\mathbb{E})$, μ is determined by knowledge of $\{\mu(f), f \in C_K^+(\mathbb{E}))$. This may seem reasonable and we will see why this is true shortly. Think of μ as an object with components $\{\mu(f), f \in C_K^+(\mathbb{E})$. Think of $\mu(f)$ as the f^{th}-component of μ. Then (4.14) indicates, in fact, it is enough to have a countable set of components to determine μ and we can think about μ being represented

$$\mu = \{\mu(f_i), i \geq 1\}. \tag{4.15}$$

So there is an analogy with $\mathbb{R}^\infty$.

This analogy makes plausible the following characterization of compactness: A subset $M \subset M_+(\mathbb{E})$ is vaguely relatively compact iff

$$\sup_{\mu \in M} \mu(f) < \infty, \quad \forall f \in C_K^+(\mathbb{E}). \tag{4.16}$$

One direction of the compactness characterization in (4.16) is easy to prove and helps us digest the concepts. Suppose M is relatively compact. For $f \in C_K^+(\mathbb{E})$, define $T_f : M_+(\mathbb{E}) \mapsto [0,\infty)$ by

$$T_f(\mu) = \mu(f).$$

Then T_f is continuous since $\mu_n \xrightarrow{v} \mu$ implies

$$T_f(\mu_n) = \mu_n(f) \to \mu(f) = T_f(\mu).$$

For fixed $f \in C_K^+(\mathbb{E})$, we note

$$\sup_{\mu\in M} \mu(f) = \sup_{\mu \in M} T_f(\mu) = \sup_{\mu \in M^-} T_f(\mu),$$

since the supremum of a continuous function on M must be the same as the supremum on the closure M^-.

If M is relatively compact, then the closure M^- is compact. Since T_f is continuous on $M_+(\mathbb{E})$, $T_f(M^-)$ is a compact subset of $[0,\infty)$. (Continuous images of compact sets are compact.) Compact sets in $[0,\infty)$ are bounded so

$$\infty > \sup T_f(M^-) = \sup\{T_f(\mu), \mu \in M^-\} = \sup_{\mu \in M^-} \{\mu(f)\}.$$

Why emphasize integrals of test functions rather than measures of sets? Proofs are a bit easier with this formulation and it is easier to capitilize on continuity arguments. One can always formulate parallel definitions and concepts with sets using a variant of Urysohn's lemma. See (Dudley, 1989, page 47), (Simmons, 1963, page 135), Kallenberg (1983), (Resnick, 1987, page 141).

Lemma 1. *(a) Suppose* $K \in \mathcal{K}(\mathbb{E})$*. There exists* $K_n \in \mathcal{K}(\mathbb{E})$*,* $K_n \downarrow K$ *and there exist* $f_n \in C_K^+(\mathbb{E})$*, with* $\{f_n\}$ *non-increasing such that*

$$1_K \le f_n \le 1_{K_n} \downarrow 1_K. \tag{4.17}$$

(b) Suppose $G \in \mathcal{G}(\mathbb{E})$*, and* G *is relatively compact. There exist open, relatively compact* $G_n \uparrow G$ *and* $f_n \in C_K^+(\mathbb{E})$*, with* $\{f_n\}$ *non-decreasing such that*

$$1_G \ge f_n \ge 1_{G_n} \uparrow 1_G \tag{4.18}$$

From Lemma 1, comes a Portmanteau Theorem.

Theorem 3. *Let* $\mu, \mu_n \in M_+(\mathbb{E})$*. The following are equivalent.*

(i) $\mu_n \xrightarrow{v} \mu$.
(ii) $\mu_n(B) \to \mu(B)$ *for all relatively compact* B *satisfying* $\mu(\partial B) = 0$.
(iii) *For all* $K \in \mathcal{K}(\mathbb{E})$ *we have*

$$\limsup_{n\to\infty} \mu_n(K) \le \mu(K)$$

and for all $G \in \mathcal{G}$ *which are relatively compact, we have*

$$\liminf_{n\to\infty} \mu_n(G) \ge \mu(G).$$

5. Vague Convergence, Regular Variation and the Multivariate Case.

Regular variation of distribution tails can be reformulated in terms of vague convergence and with this reformulation, the generalization to higher dimensions is effortless. We begin by discussing the reformulation in one dimension.

5.0.6. *Vague convergence on* $(0,\infty]$.

Theorem 4. *Suppose* X_1 *is a non-negative random variable with distribution function* $F(x)$*. Set* $\bar F = 1 - F$*. The following are equivalent:*

(i) $\bar F \in RV_{-\alpha},\ \alpha > 0$.
(ii) *There exists a sequence* $\{b_n\}$*, with* $b_n \to \infty$ *such that*

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

(iii) *There exists a sequence* $\{b_n\}$ *with* $b_n \to \infty$ *such that*

$$\nu_n(\cdot) := nP\left[\frac{X_1}{b_n} \in \cdot\right] \xrightarrow{v} \nu(\cdot) \tag{5.1}$$

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

Remark 2. (a) If any of (i), (ii) or (iii) is true we may always define

$$b(t) = \left(\frac{1}{1-F}\right)^{\leftarrow}(t) \tag{5.2}$$

and set $b_n = b(n)$. Note if (i) holds, then

$$\bar F \in RV_{-\alpha} \text{ implies } \frac{1}{1-F} \in RV_{\alpha} \text{ implies } b(\cdot) = \left(\frac{1}{1-F}\right)^{\leftarrow}(\cdot) \in RV_{1/\alpha}.$$

(b) Note in (iii) that the space $\mathbb{E} = (0, \infty]$ has 0 excluded and ∞ included. This is required since we need neighborhoods of ∞ to be relatively compact. Vague convergence only controls set wise convergence on relatively compact sets (with no mass on the boundary). With the usual topology on $[0, \infty)$ sets of the form (x, ∞) are not bounded; yet consideration of $n\bar{F}(b_n x) = nP[X_1/b_n > x]$ requires considering exactly such sets. We need some topology which makes semi-infinite intervals compact. More on this later. If it helps, think of $(0, \infty]$ as the homeomorphic stretching of $(0, 1]$ or as the homeomorphic image of $[0, \infty)$ under the map $x \mapsto 1/x$ which takes $[0, \infty) \mapsto (0, \infty]$. A convenient way to handle this is by using the *one point uncompactification* method to be discussed soon in Subsection 5.0.7.

(c) Preview of things to come: Note that if $\{X_j, j \geq 1\}$ is an iid sequence of non-negative random variables with common distribution F, then the measure ν_n defined in (5.1)

$$\nu_n(\cdot) = \mathbf{E}\left(\sum_{i=1}^{n} \epsilon_{X_i/b(n)}(\cdot)\right)$$

is the mean measure of the empirical measure of the scaled sample. The convergence of ν_n is equivalent to convergence of the sequence of empirical measures to a limiting Poisson process.

Proof. The equivalence of (i) and (ii) is Part (ii) of Proposition 2.

(ii) $\rightarrow$ *(iii)*. Let $f \in C_K^+((0, \infty])$ and we must show

$$\nu_n(f) := nEf\left(\frac{X_1}{b_n}\right) = \int f(x)nP[\frac{X_1}{b_n} \in dx] \rightarrow \nu(f).$$

Since f has compact support, the support of f is contained in $(\delta, \infty]$ for some $\delta > 0$. We know

$$\nu_n(x, \infty] \rightarrow x^{-\alpha} = \nu(x, \infty], \quad \forall x > 0. \tag{5.3}$$

On $(\delta, \infty]$ define

$$P_n(\cdot) = \frac{\nu_n}{\nu_n(\delta, \infty]} \tag{5.4}$$

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

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

In $\mathbb{R}$, convergence of distribution functions (or tails) is equivalent to weak convergence so $\{P_n\}$ converges weakly to P. Since f is bounded and continuous on $(\delta, \infty]$, we get from weak convergence:

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

that is,

$$\frac{\nu_n(f)}{\nu_n(\delta, \infty]} \rightarrow \frac{\nu(f)}{\delta^{-\alpha}}$$

and in light of (5.3), this implies

$$\nu_n(f) \rightarrow \nu(f)$$

as required.

$(iii) \to (ii)$. Since

$$\nu_n \xrightarrow{v} \nu,$$

we have

$$\nu_n(x,\infty] \to \nu(x,\infty], \quad \forall x > 0,$$

since $(x,\infty]$ is relatively compact and

$$\nu(\partial(x,\infty]) = \nu(\{x\}) = 0.$$

□

5.0.7. *Topological clarification: The one point uncompactification.* In reformulating the function theory concept of regularly varying functions into a measure theory concept, there is continual need to deal with sets which are bounded away from 0. Such sets need to be regarded as "bounded" in an appropriate topology so sequences of measures of such sets can converge non-trivially. A convenient way to think about this is by means of something we will call the one point un-compactification.

Let $(\mathbb{X}, \mathcal{T})$ be a nice topological space; $\mathbb{X}$ is the set and $\mathcal{T}$ is the topology, that is a collection of subsets of $\mathbb{X}$ designated as *open* satisfying

(i) Both $\emptyset \in \mathcal{T}$ and $\mathbb{X} \in \mathcal{T}$.
(ii) The collection $\mathcal{T}$ is closed under finite intersections and arbitrary unions.

(For our purposes, $\mathbb{X}$ would be a subset of Euclidean space.) Consider a subset $\mathbb{D} \subset \mathbb{X}$ and define

$$\mathbb{X}^{\#} = \mathbb{X} \setminus \mathbb{D}$$

and give $\mathbb{X}^{\#}$ the relative topology

$$\mathcal{T}^{\#} = \mathcal{T} \bigcap \mathbb{D}^c = \mathcal{T} \bigcap \mathbb{X}^{\#}.$$

So a set is open in $\mathbb{X}^{\#}$ if it is an open subset of $\mathbb{X}$ intersected with $\mathbb{X}^{\#}$.

What we need to understand is what are the compact sets of $\mathbb{X}^{\#}$.

Proposition 6. *Suppose, as usual, the compact subsets of $\mathbb{X}$ are denoted by $\mathcal{K}(\mathbb{X})$. Then*

$$\mathcal{K}(\mathbb{X}^{\#}) = \{K \in \mathcal{K}(\mathbb{X}) : K \bigcap \mathbb{D} = \emptyset\}$$

are the compact subsets of $\mathbb{X}^{\#}$.

So the compact sets of $\mathbb{X}^{\#}$ are the original compact sets of $\mathbb{X}$, provided they do not intersect the piece D chopped away from $\mathbb{X}$ to form $\mathbb{X}^{\#}$.

Specialize this to the *one point un-compactification*: Suppose $\mathbb{E}$ is a compact set and $e \in \mathbb{E}$. Give $\mathbb{E}\setminus\{e\}$ the relative topology consisting of sets in $\mathbb{E}\setminus\{e\}$ of the form $G\setminus\{e\}$, where $G \in \mathcal{G}(\mathbb{E})$. The compact sets of $\mathbb{E}\setminus\{e\}$ are those compact subsets $K \subset \mathbb{E}$ such that $e \notin K$.

So the *one point un-compactification* describes what are the compact sets when a compact space is punctured by the removal of a point.

Consider the following special cases:

(1) Suppose $\mathbb{E}$ is the compact set $[0,\infty]^d = [\mathbf{0},\boldsymbol{\infty}]$, which we may consider as the stretching of $[0,1]^d = [\mathbf{0},\mathbf{1}]$ onto $[\mathbf{0},\boldsymbol{\infty}]$. The compact sets of $[\mathbf{0},\boldsymbol{\infty}]$ consist of any closed set. The compact subsets of $[\mathbf{0},\boldsymbol{\infty}] \setminus \{\mathbf{0}\}$ are closed subsets of $[\mathbf{0},\boldsymbol{\infty}]$ bounded away from $\mathbf{0}$.

(2) Suppose $\mathbb{E}$ is the compact set $[-\boldsymbol{\infty},\boldsymbol{\infty}]$. The compact sets of $[-\boldsymbol{\infty},\boldsymbol{\infty}]$ consist of any closed set. The compact subsets of $[-\boldsymbol{\infty},\boldsymbol{\infty}]\setminus\{\mathbf{0}\}$ are closed subsets of $[-\boldsymbol{\infty},\boldsymbol{\infty}]$ bounded away from $\mathbf{0}$. This choice of $\mathbb{E}$, and the associated space of Radon measures $M_+(\mathbb{E})$ is needed for considering weak convergence of partial sums to multivariate Lévy processes and for analyzing multivariate problems related to *value at risk*.

(3) As a last example of the use of Proposition 6 suppose $\mathbb{E} = [\mathbf{0},\boldsymbol{\infty}] \setminus \{\mathbf{0}\}$ and define the cone $\mathbb{E}^0$ by

$$\mathbb{E}^0 := \{\boldsymbol{s} \in \mathbb{E} : \text{ For some } 1 \le i < j \le d, s^{(i)} \wedge s^{(j)} > 0\},$$

where we wrote the vector $\boldsymbol{s} = (s^{(1)},\dots,s^{(d)})$. Here is an alternative description of $\mathbb{E}^0$: For $i = 1,\dots,d$, define the basis vectors

$$\boldsymbol{e}_i = (0,\dots,0,1,0,\dots,0)$$

so that the axes originating at $\mathbf{0}$ are $\mathbb{L}_i := \{t\boldsymbol{e}_i, t > 0\}$, $i = 1,\dots,d$. Then we also have

$$\mathbb{E}^0 = \mathbb{E} \setminus \bigcup_{i=1}^d L_i.$$

If $d = 2$, we have $\mathbb{E}^0 = (0,\infty]^2$.

The relatively compact subsets of $\mathbb{E}^0$ are those sets bounded away from the axes originating at $\mathbf{0}$. So G is relatively compact in $\mathbb{E}^0$ if for some $\delta > 0$ we have that for every $\boldsymbol{x} \in G$, for some $1 \le i < j \le d$, that $x^{(i)} \wedge x^{(j)} > \delta$.

Such a space is useful in consideration of asymptotic independence.

Proof. (Proposition 6.) Begin by supposing

$$K \in \mathcal{K}(\mathbb{X}), \quad K \cap \mathcal{D} = \emptyset,$$

and we show $K \in \mathcal{K}(\mathbb{X}^\#)$. Let

$$\{G_\gamma^\# = G_\gamma \cap \mathbb{X}^\#, \gamma \in \Lambda\}$$

be some arbitrary cover of K by open subsets of $\mathbb{X}^\#$ where $G_\gamma \in \mathcal{G}(\mathbb{X})$ and Λ is some index set. So

$$K \subset \bigcup_{\gamma\in\Lambda} G_\gamma \bigcap \mathbb{X}^\# \subset \bigcup_{\gamma\in\Lambda} G_\gamma.$$

Since $K \in \mathcal{K}(\mathbb{X})$, there is a finite subcollection indexed by $\Lambda' \subset \Lambda$ such that $K \subset \bigcup_{\gamma\in\Lambda'} G_\gamma$. Since $K \cap \mathbb{D} = \emptyset$,

$$K \subset \bigcup_{\gamma\in\Lambda'} G_\gamma \bigcap \mathbb{X}^\#.$$

So, any cover of K by open subsets of $\mathbb{X}^\#$ has a finite subcover and thus K is compact in $\mathbb{X}^\#$. Thus

$$\{K \in \mathcal{K}(\mathbb{X}) : K \bigcap \mathbb{D} = \emptyset\} \subset \mathcal{K}(\mathbb{X}^\#).$$

The converse is quite similar. □

References

P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

P. Billingsley. *Weak convergence of measures: Applications in probability.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1971. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 5.

P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

R. Dudley. *Real Analysis and Probability.* Wadsworth and Brook/Cole, 1989.

O. Kallenberg. *Random Measures.* Akademie-Verlag, Berlin, third edition, 1983. ISBN 0-12-394960-2.

S. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York, 1987.

G. Simmons. *Topology and Modern Analysis.* McGraw-Hill, New York, 1963.

W. Whitt. *Stochastic Processs Limits: An Introduction to Stochastic-Process Limits And their Application to Queues.* Springer-Verlag, New York, 2002. ISBN 0-387-95358-2.