2

Scale Invariance, Power Laws, and Regular Variation

In our daily lives, many things that we come across have a size, or scale, that we associate with them. For example, people's heights and weights differ, but they are not *that* different – they rarely differ by more than a factor of two or three and do not differ much from the population average. In contrast, the incomes of people we encounter in our daily lives do not have a typical size or scale – they may differ by a factor of *100 or more* and can be very far from the population average! This contrast is a consequence of the fact that many heavy-tailed phenomena, such as incomes, are *scale invariant*, aka, *scale-free*, while light-tailed phenomena, such as heights and weights, are not.

Scale invariance is a property that feels particularly magical the first time you observe it. An object is scale invariant if it looks the same regardless of what scale you look at it. Perhaps the easiest way to understand scale invariance is using fractals, like the one shown in Figure 2.1. If you zoom in or out, the picture will look the same. It turns out that Pareto distributions have the same property (see Figure 2.2). But Pareto distributions are even more special than fractals. With fractals, you have to zoom in or out in specific, discrete steps for the picture to look the same; with Pareto distributions, the invariance holds across a continuum of scale changes.

Scale invariance is a particularly mysterious aspect of heavy-tailed phenomena. It is natural to think of the average of a distribution as a good predictor of what samples will occur; but for scale invariant distributions the average is actually a very poor predictor. This fact leads to many of the counterintuitive properties of heavy-tailed distributions. For example, consider the old economics joke: "If Bill Gates walks into a bar . . . on average, everybody in the bar is a millionaire."

Though initially mysterious and counterintuitive, scale invariance is a beautiful and widely observed phenomenon that has attracted attention well beyond mathematics and statistics (e.g., in physics, computer science, economics, and even art). For example, scale invariance is an important concept in both classical and quantum field theory as well as statistical mechanics. In fact, it is closely tied to the notion of "universality" in physics, which relates to the fact that widely different systems can be described by the same underlying theory. Further, in the context of network science, scale invariance has received considerable attention. Widely varying networks have been found to have scale invariant degree distributions (and are thus termed "scale-free networks" [25, 41]) and this observation has had dramatic implications on our understanding of the structural properties of networks. For a discussion of scale invariance broadly, we recommend [220]. Here, we focus on scale invariance in the context of probability and statistics.

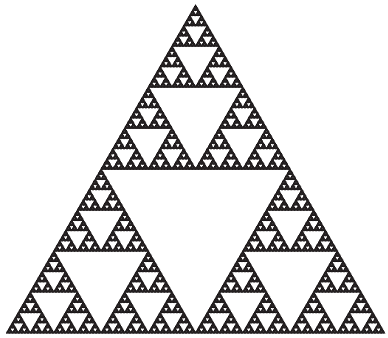


Figure 2.1 Illustration of the Sierpinski fractal [151].

In particular, in this chapter we explore the mathematics of the property of scale invariance and its connections with Pareto distributions and so-called *power law distributions*. Note that both “scale invariance” and “power law” are often (mis)used synonymously with “heavy-tailed,” and thus it is important to point out that not all heavy-tailed distributions are scale invariant or have a power law (though all scale invariant distributions are heavy-tailed, as are all power law distributions). In this chapter, we formalize and generalize the notions of scale invariance and power laws as a subclass of heavy-tailed distributions termed “regularly varying distributions.” In addition, we explain why the class of regularly varying distributions is particularly appealing from a mathematical perspective. The properties of this class shed light on many of the counterintuitive properties of heavy-tailed distributions, highlighting what properties of heavy-tailed distributions can be viewed as simple consequences of scale invariance. Further, to illustrate the usefulness of the class, we demonstrate how to apply properties of regular variation in order to analyze heavy-tailed phenomena more broadly. These examples show that it is not much more challenging to analyze the entire class of regularly varying distributions than it is to work with the specific case of the Pareto distribution.

2.1 Scale Invariance and Power Laws

To this point we have only introduced scale invariance informally as the property that something looks the “same” regardless of the scale at which it is observed. Given that our focus is on probability distributions, we can rephrase this idea as follows: if the scale (or units) with which the samples from the distribution are measured is changed, then the shape of the distribution is unchanged. This leads to the following formal definition.

Definition 2.1 A distribution function F is scale invariant if there exists an $x_0 > 0$ and a continuous positive function g such that

$$\bar{F}(\lambda x) = g(\lambda)\bar{F}(x),$$

for all x, λ satisfying $x, \lambda x \geq x_0$.

To interpret the definition of scale invariant, one can think of λ as the “change of scale” for the units being used. With this interpretation, the definition says that the shape of the

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

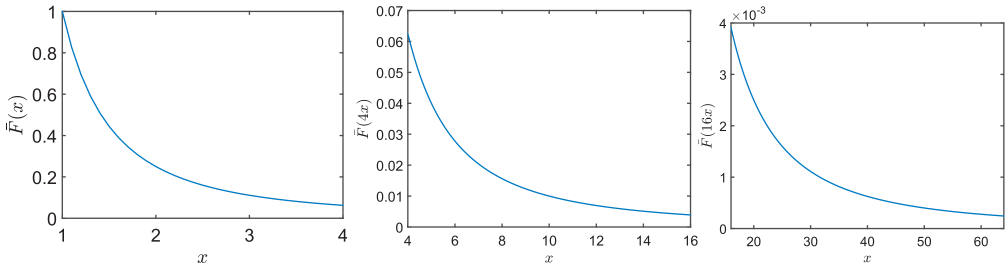


Figure 2.2 The complementary cumulative distribution function (c.c.d.f.) corresponding to the Pareto distribution ($\alpha = 2$, $x_m = 1$) plotted at different scales of the independent variable. Note that the shape of the curve is preserved up to a multiplicative scaling, consistent with scale invariance.

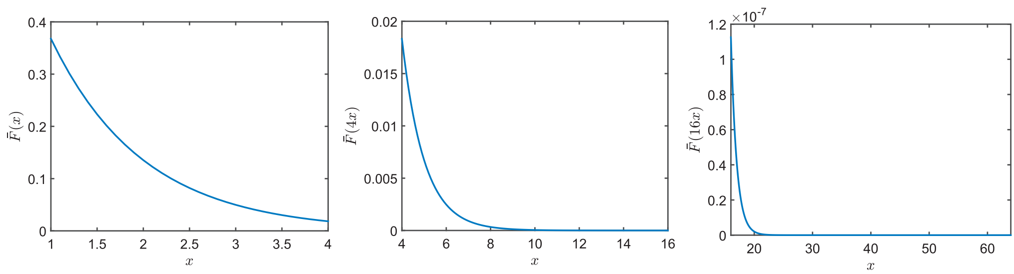


Figure 2.3 The complementary cumulative distribution function (c.c.d.f.) corresponding to the Exponential distribution (with mean 1) plotted at different scales of the independent variable. Note that the shape of the curve looks fundamentally altered at different scales.

distribution $\bar{F}$ remains unchanged up to a multiplicative factor $g(\lambda)$ if the measurements are scaled by λ. This is exactly what is shown in Figure 2.2 for the Pareto distribution.

More formally, to see that the Pareto is scale invariant, recall that a Pareto distribution has $\bar{F}(x) = (x/x_m)^{-\alpha}$ for $x > x_m$. Thus,

$$\bar{F}(\lambda x) = \left(\frac{\lambda x}{x_m}\right)^{-\alpha} = \bar{F}(x)\lambda^{-\alpha}, \text{ whenever } x, \lambda x > x_m.$$

Scale invariance is an elegant property, but it is also a fragile one. In particular, it does not hold for most probability distributions. For example, it is easy to see that the Exponential distribution is not scale invariant. Recall that an Exponential distribution has $\bar{F}(x) = e^{-\mu x}$ for $x \geq 0$. Therefore,

$$\bar{F}(\lambda x) = e^{-\mu \lambda x} = \bar{F}(x)e^{-\mu(\lambda - 1)x}.$$

Thus, there is not a choice for g that is independent of x. This is also illustrated in Figure 2.3.

One may initially think that the lack of scale invariance of the Exponential distribution is a consequence of its being a light-tailed distribution. But that is not the case. For example, let us generalize the Exponential distribution to the Weibull distribution, $\bar{F}(x) = e^{-\beta x^{\alpha}}$ for $x \geq 0$, which is equivalent to the Exponential distribution when the shape parameter $\alpha = 1$ and is heavy-tailed when $\alpha < 1$. As the following calculation shows, the Weibull is also not scale invariant:

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

$$\bar{F}(\lambda x) = e^{-\beta(\lambda x)^\alpha} = \bar{F}(x)e^{-\beta x^\alpha(\lambda^\alpha - 1)}.$$

These examples start to give some intuition about scale invariance, but they leave open a fundamental, natural question:

Which distributions are scale invariant?

From the examples, we know that there is at least one scale invariant distribution (the Pareto distribution), but we also know that not all common distributions are scale invariant – not even all common heavy-tailed distributions. Perhaps surprisingly, it turns out that scale invariance is an extremely special property: distributions with "power law tails," (i.e., tails that match the Pareto distribution up to a multiplicative constant) are the *only* scale invariant distributions. That is, "scale invariance" can be thought of interchangeably with "power law." The following theorem states this formally.

Theorem 2.2 *A distribution function F is scale invariant if and only if F has a power law tail, that is, there exists $x_0 > 0$, $c \geq 0$, and $\alpha > 0$ such that $\bar{F}(x) = cx^{-\alpha}$ for $x \geq x_0$.*

Proof Note that the case where $\bar{F}$ is identically zero over $[x_0, \infty)$ trivially satisfies the conditions of the lemma (this corresponds to the case $c = 0$).

Excluding the trivial case from consideration, it is easy to see that $\bar{F}(x)$ must be nonzero for all $x \geq x_0$. Indeed, if $\bar{F}(x') = 0$ for some $x' \geq x_0$, then for any $x \geq x_0$, $\bar{F}(x) = \bar{F}(x')g(x/x') = 0$.

Fix $x, y > 0$. We may then pick z large enough such that $z, zx, zxy \geq x_0$. From the scale invariant property of $\bar{F}$, $\bar{F}(xyz) = \bar{F}(z)g(xy)$. Of course, we may also write $\bar{F}(xyz) = \bar{F}(xz)g(y) = \bar{F}(z)g(x)g(y)$. Since $\bar{F}(z) \neq 0$, we can immediately see that the function g satisfies the following property:

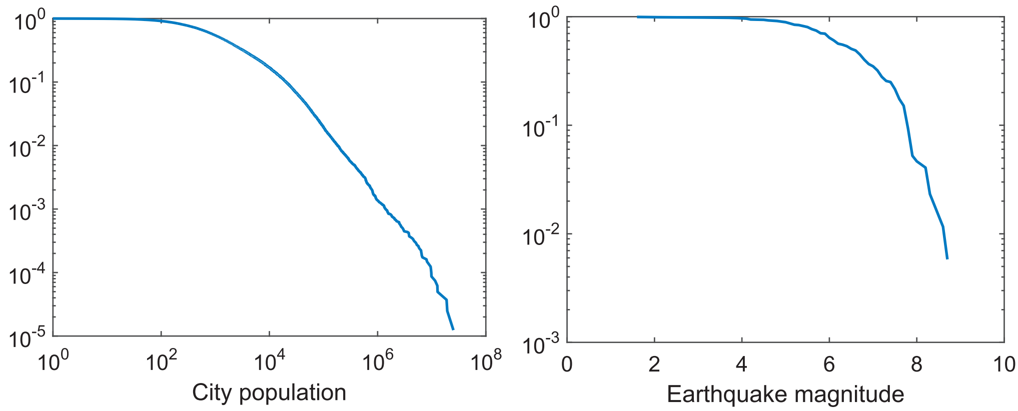
$$g(xy) = g(x)g(y) \quad \text{for all } x, y > 0. \tag{2.1}$$

This is a very special property, and the only continuous positive functions satisfying the condition in (2.1) are $g(x) = x^{-\alpha}$ for some $\alpha \in \mathbb{R}$.[1] Noting that $\bar{F}(x) = \bar{F}(x_0)g(x/x_0)$ for all $x \geq x_0$, we conclude that $\alpha > 0$ (since $\bar{F}$ must be monotonically decreasing, with $\lim_{x\to\infty} \bar{F}(x) = 0$). Therefore, $\bar{F}(x) = cx^{-\alpha}$ for $x \geq x_0$ for some $c, \alpha > 0$. □

2.2 Approximate Scale Invariance and Regular Variation

We have just seen that all scale invariant distributions are power law distributions, aka distributions with tails matching a Pareto distribution up to a multiplicative constant. This makes scale invariance a very special property, or a very fragile property depending on how you look at it. In fact, one interpretation of Theorem 2.2 is that we should not expect to see scale invariance in reality since it is so fragile.

[1] Defining $f(x) := \log g(e^x)$, the condition (2.1) is equivalent to $f(x + y) = f(x) + f(y)$ for $x, y \in \mathbb{R}$. This is known as *Cauchy's functional equation*. The stated claim now follows from the fact that the only continuous solutions of Cauchy's functional equation are of the form $f(x) = \alpha x$ for $\alpha \in \mathbb{R}$ (see, for example, [5]).



(a) Populations of US cities as per the 2010 census (data sourced from [2]).

(b) Intensities of earthquakes in the US between 1900 and 2017 (sourced from [1]). Earthquake intensity is measured on the Richter scale, which is inherently logarithmic. Thus, the values of the x-axis should be interpreted as being proportional to the logarithm of the intensity of the earthquake.

Figure 2.4 Empirical c.c.d.f. (aka rank plot) corresponding to two real-world datasets on a log-log scale.

In the strictest sense, that interpretation is correct. It is quite unusual for the distribution of an observed phenomenon to *exactly* match a power law distribution and thus be scale invariant. Instead, what tends to be observed in reality is that the *body* of the distribution is not an exact power law, but the *tail* of the distribution is *approximately* a power law. Consider the examples in Figure 2.4, which depicts the empirical c.c.d.f. corresponding to two real-world datasets on a log-log scale. Notice that the body of the empirical c.c.d.f. (on the log-log scale) does not look linear, which it would if the data were sampled from a power law distribution, but instead seems to approach a straight line *asymptotically*, which suggests that the c.c.d.f. behaves *asymptotically* like a power law.

Given that we should not expect to see precise scale invariance in real observations, it is natural to shift our focus from precise scale invariance to notions of approximate or asymptotic scale invariance; and it is natural not to focus on the whole distribution, but rather on just the tail of the distribution. In particular, the relevant formalism becomes *asymptotic scale invariance*, which we define here.

Definition 2.3 A distribution F is asymptotically scale invariant if there exists a continuous positive function g such that for any $\lambda > 0$,

$$\lim_{x\to\infty} \frac{\bar{F}(\lambda x)}{\bar{F}(x)} = g(\lambda).$$

The notion of asymptotic scale invariance almost exactly parallels the notion of scale invariance, except that it only requires the property to hold in the limit as $x \to \infty$, that

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

is, it only requires the property to approximately hold for the tail: $\bar{F}(\lambda x) \sim g(\lambda)\bar{F}(x)$ as $x \to \infty$.[2]

As a result, it is immediately clear that Pareto distributions are asymptotically scale invariant:

$$\bar{F}(\lambda x)/\bar{F}(x) = \lambda^{-\alpha}.$$

Similarly, it is easy to see that asymptotic scale invariance is still quite a special property that is not satisfied by most distributions. For example, the Weibull and Exponential (Weibull with $\alpha = 1$) distributions are not asymptotically scale invariant since, as $x \to \infty$,

$$\frac{\bar{F}(\lambda x)}{\bar{F}(x)} = e^{-\beta(\lambda^\alpha - 1)x^\alpha} \to \begin{cases} \infty, & \lambda < 1, \\ 1, & \lambda = 1, \\ 0, & \lambda > 1. \end{cases}$$

However, asymptotic scale invariance is significantly broader than scale invariance, and it is easy to see that other distributions besides power law distributions are asymptotically scale invariant. For example, it follows from Exercise 8 that the convolution of a Pareto and an Exponential distribution is asymptotically scale invariant (though it is clearly not scale invariant). Similarly, consider the Fréchet distribution, which we introduced in Section 1.2.4 and will appear in our analysis of extremal processes in Chapter 7. This distribution, which is supported over the nonnegative reals, is defined by the distribution function $F(x) = e^{-x^{-\alpha}}$ for $x \geq 0$, where the parameter $\alpha > 0$. While this distribution is clearly not a power law, it is not hard to see that $\bar{F}(x) \sim x^{-\alpha}$ (see Exercise 1). In other words, the Fréchet distribution has an asymptotically power law tail, which in turn implies asymptotic scale invariance:

$$\lim_{x \to \infty} \frac{\bar{F}(\lambda x)}{\bar{F}(x)} = \lambda^{-\alpha}.$$

In general, since asymptotic scale invariance focuses only on the tail of the distribution, the body of such a distribution may behave in an arbitrary manner as long as the tail is approximately scale invariant.

As in the case of scale invariance, given the examples above, the natural question becomes:

Which distributions are asymptotically scale invariant?

It is clear that the class of asymptotically scale invariant distributions includes a variety of heavy-tailed distributions beyond the Pareto distribution, but it is also clear that it does not include all heavy-tailed distributions since it does not include the Weibull distribution. However, the fact that "scale invariant" can be thought of equivalently to "power law," leads to the suggestion that "asymptotically scale invariant" should correspond to some notion of "approximately power law," and this turns out to be true. In particular, it turns out that asymptotically scale invariant distributions have tails that are approximately power law in a rigorous sense that can be formalized via the class of *regularly varying distributions*.

[2] Throughout the book we use $a(x) \sim b(x)$ as $x \to \infty$ to mean $\lim_{x\to\infty} a(x)/b(x) = 1$.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

Definition 2.4 A function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying of index $\rho \in \mathbb{R}$, denoted $f \in \mathcal{RV}(\rho)$, if for all $y > 0$,

$$\lim_{x\to\infty} \frac{f(xy)}{f(x)} = y^\rho.$$

Further, for $\rho \le 0$, a distribution F is regularly varying of index ρ, denoted as $F \in \mathcal{RV}(\rho)$, if $\bar{F}(x) = 1 - F(x)$ is a regularly varying function of index ρ.[3]

The form of the definition immediately makes clear that regularly varying distributions are asymptotically scale invariant. Further, since $\lim_{x\to\infty} \bar{F}(xy)/\bar{F}(x) = y^\rho$, they seem to mimic the behavior of power law distributions, such as the Pareto distribution, asymptotically. This intuitively suggests that all asymptotically scale invariant distributions are regularly varying distributions – which turns out to be true.

Theorem 2.5 *A distribution F is asymptotically scale invariant if and only if it is regularly varying.*

Proof It is immediately clear that regularly varying distributions are asymptotically scale invariant, and so we need only prove the other direction. Fix $x, y > 0$. The asymptotic scale-free property implies that

$$\lim_{z\to\infty} \frac{\bar{F}(xyz)}{\bar{F}(z)} = g(xy).$$

We can also compute the same limit by writing $\frac{\bar{F}(xyz)}{\bar{F}(z)} = \frac{\bar{F}(xyz)}{\bar{F}(xz)}\frac{\bar{F}(xz)}{\bar{F}(z)}$. Note that $\frac{\bar{F}(xyz)}{\bar{F}(xz)} \to g(y)$ and $\frac{\bar{F}(xz)}{\bar{F}(z)} \to g(x)$ as $z \to \infty$, implying that

$$\lim_{z\to\infty} \frac{\bar{F}(xyz)}{\bar{F}(z)} = g(x)g(y).$$

From this, we can conclude that the function g satisfies

$$g(xy) = g(x)g(y) \quad \text{for all } x, y > 0.$$

This is the same relationship we used in the proof of Theorem 2.2 and, as in that case, it follows that there exists $\theta \in \mathbb{R}$ such that $g(x) = x^\theta$. Of course, by definition, this means that $\bar{F}$ is a regularly varying function, and F is a regularly varying distribution. □

Theorem 2.5 shows that the class of regularly varying distributions characterizes precisely those distributions that are asymptotically scale invariant, and from it we can immediately

[3] It is common practice to write the domain of regularly varying functions as $\mathbb{R}_+$. That said, it is important to understand that regular variation is an *asymptotic* property of a function as its argument tends to ∞. Thus, for a function f to be regularly varying, we only require that its domain includes $[x_0, \infty)$ for some $x_0 > 0$. Similarly, for a distribution to be regularly varying, we only require that its support include $[x_0, \infty)$ for some $x_0 > 0$. Specifically, regularly varying distributions need not be supported on the nonnegative reals. For example, the Cauchy distribution is regularly varying (see Exercise 1). Finally, why the index ρ must be nonpositive for regularly varying distributions will become apparent soon (see Lemma 2.7).

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

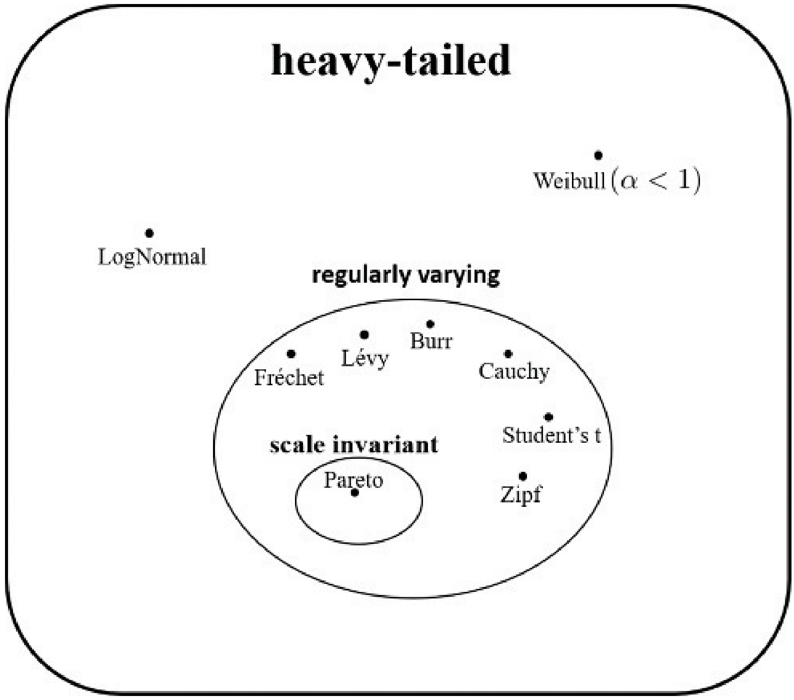


Figure 2.5 Scale invariant and regularly varying distributions.

see that a number of common heavy-tailed distributions are scale invariant. In particular, with a little effort, it is possible to verify that the Student's t-distribution, the Cauchy distribution, the Burr distribution, the Lévy, and also the Zipf distribution are all regularly varying and thus asymptotically scale invariant (see Exercise 1). This is summarized in Figure 2.5. We have not yet proven that regularly varying distributions are heavy-tailed; this follows from the analytic properties of regularly varying functions discussed next (see Lemma 2.9).

2.3 Analytic Properties of Regularly Varying Functions

The fact that regularly varying distributions are exactly those distributions that are asymptotically scale invariant suggests that they should be able to be analyzed (at least asymptotically) as if they are simply Pareto distributions. In fact, this intuition is correct and can be formalized explicitly, as we show in this section. Concretely, the properties we outline in this section provide the tools that enable regularly varying distributions to be analyzed "as if" they were polynomials, as far as the tail is concerned. This makes them remarkably easy to work with and shows that the added generality that comes from working with the class of regularly varying distributions, as opposed to working specifically with Pareto distributions, comes without too much added technical complexity.

To begin, it is important to formalize exactly what we mean when we say that regularly varying distributions have tails that are approximately power law. To do this, we need to first introduce the concept of a *slowly varying function*.

Definition 2.6 A function $L\colon \mathbb{R}_+ \to \mathbb{R}_+$ is said to be slowly varying if $\lim_{x\to\infty} \frac{L(xy)}{L(x)} = 1$ for all $y > 0$.

Slowly varying functions are simply regularly varying functions of index zero. So, intuitively, they can be thought of as functions that grow/decay asymptotically slower than any polynomial; for example, $\log x$, $\log\log x$, and so on. This can be formalized as follows.

Lemma 2.7 *If the function $L\colon \mathbb{R}_+ \to \mathbb{R}_+$ is slowly varying, then*

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

We prove this lemma later in this section using properties of regularly varying functions. But we state it now in order to highlight an equivalent definition of regularly varying distributions as distributions that are "asymptotically power law." The following representation theorem for regularly varying functions makes this precise.

Theorem 2.8 *A function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying with index ρ if and only if $f(x) = x^\rho L(x)$, where $L(x)$ is a slowly varying function.*

Proof We start by proving the "if" direction. Suppose that $f \in \mathcal{RV}(\rho)$. Define $L(x) = f(x)/x^\rho$. To prove the result, it is enough to show that L is slowly varying, which can be argued as follows:

$$\lim_{x\to\infty} \frac{L(xy)}{L(x)} = \lim_{x\to\infty} \frac{f(xy)}{f(x)} \frac{x^\rho}{(xy)^\rho} = 1.$$

To prove the other direction, we need to show that, given $f(x) = x^\rho L(x)$, where $L(x)$ is a slowly varying function, $f \in \mathcal{RV}(\rho)$. For $y > 0$,

$$\lim_{x\to\infty} \frac{f(xy)}{f(x)} = \lim_{x\to\infty} \frac{(xy)^\rho}{x^\rho} \frac{L(xy)}{L(x)} = y^\rho,$$

which implies, by definition, that $f \in \mathcal{RV}(\rho)$. □

It is important to remember when applying this theorem that regularly varying *functions* can have arbitrary index ρ; however, regularly varying *distributions* must have index $\rho \le 0$.[4]

The key implication of Theorem 2.8 in the context of heavy-tailed distributions is that regularly varying distributions can be thought of as distributions with approximately power law tails in a rigorous sense. That is, they differ from a power law tail only by a slowly varying function $L(x)$, which can intuitively be treated as a constant when doing analysis. This intuition leads to many of the analytic properties that we discuss in the remainder of this section.

However, before we move to the analytic properties of regularly varying distributions, it is useful to illustrate how powerful the representation theorem is by itself. To illustrate

[4] One peculiarity of this notational convention is that a Pareto distribution with tail index α, where $\alpha > 0$, is regularly varying with index $-\alpha$.

this, we use it in Lemma 2.9 to argue that regularly varying distributions are heavy-tailed. Of course, this is not a surprising result, given the tie to Pareto distributions, but it is an important foundational result and it provides a simple illustration of how to work with the representation theorem.

Lemma 2.9 *All regularly varying distributions are heavy-tailed.*

Proof Suppose that the distribution F is regularly varying. We know then that $\bar{F}(x) = x^{-\alpha}L(x)$, where $\alpha \geq 0$, and $L(x)$ is a slowly varying function. Consider $\mu > 0$ and $\beta > \alpha$. Now,

$$\frac{\bar{F}(x)}{e^{-\mu x}} = (x^{-\beta}e^{\mu x})(x^{\beta-\alpha}L(x)).$$

Since $x^{-\beta}e^{\mu x} \to \infty$ as $x \to \infty$, and $x^{\beta-\alpha}L(x) \to \infty$ as $x \to \infty$ (via Lemma 2.7), we can conclude that $\lim_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \infty$, which proves that F is heavy-tailed. □

The above provides an example of using the fact that regular varying distributions have tails that are approximately power law; however, this representation of regularly varying distributions has much broader implications as well. In particular, in the remainder of this section we illustrate a variety of analytic properties of regularly varying distributions that highlight how regularly varying distributions can be analyzed "as if" they were Pareto distributions, as far as the tail is concerned.

In the following, we focus on two crucial analytic properties: (i) integration/differentiation of regularly varying functions, which is formalized via Karamata's theorem, and (ii) inverting the integral transforms of regularly varying functions to understand properties of the tail of the distribution, which is formalized via Karamata's Tauberian theorem. In each case, we include a simple example application of the result. Then, in the following section (Section 2.4), we prove closure properties of regularly varying distributions with respect to various algebraic operations in order to provide further illustrations of how to apply these analytic properties.

2.3.1 Integration and Differentiation of Regularly Varying Distributions

Perhaps one of the most appealing aspects of working with power law and Pareto distributions is that when one needs to manipulate them to calculate moments, conditional probabilities, convolutions, and other such things, all that is required is the integration or differentiation of polynomials, which is quite straightforward. This is in stark contrast to distributions such as the Gaussian and LogNormal, which can be very difficult to work with in this way.

One of the nicest properties of regularly varying distributions is that, in a sense, you can treat them as if they were simply polynomials when integrating or differentiating them – as long as you only care about the tail – and so they are not much more difficult to work with than Pareto distributions. This is especially useful when calculating things like moments, conditional probabilities, convolutions, and so on, as we shall see repeatedly in the remainder of this chapter and throughout the book.

The foundational properties of regularly varying functions with respect to integration and differentiation are typically referred to as Karamata's theorem. This result provides the building block for working with regularly varying distributions.

Karamata's Theorem

Karamata's theorem is perhaps the most important result in the study of regular variation. We start our discussion of it by stating Karamata's theorem for integration of regularly varying functions, since its statement is a bit cleaner than that of differentiation.

It is useful to begin by anticipating what we should expect Karamata's theorem to say. To do this, begin by considering what would happen if $f(t) = t^\rho$. In that case,

$$\int_0^x f(t)dt = \frac{x^{\rho+1}}{\rho+1} = \frac{xf(x)}{\rho+1} \text{ if } \rho > -1, \text{ and}$$
$$\int_x^\infty f(t)dt = \frac{x^{\rho+1}}{-(\rho+1)} = \frac{xf(x)}{-(\rho+1)} \text{ if } \rho < -1.$$

Thus, we may expect that Karamata's theorem would say that, asymptotically, the integrals of regularly varying functions should behave as if the function were a polynomial as far as the tail is concerned (i.e., the $=$ above should be replaced by a $\sim$). In fact, this is exactly what Karamata's theorem states.

Theorem 2.10 (Karamata's Theorem)

(a) For $\rho > -1$, $f \in \mathcal{RV}(\rho)$ if and only if

$$\int_0^x f(t)dt \sim \frac{xf(x)}{\rho+1}.$$

(b) For $\rho < -1$, $f \in \mathcal{RV}(\rho)$ if and only if

$$\int_x^\infty f(t)dt \sim \frac{xf(x)}{-(\rho+1)}.$$

Not surprisingly, regularly varying distributions also asymptotically behave as if they were polynomials with respect to differentiation. In particular, if $f(x) = x^\alpha$, then $f'(x) = \alpha x^{\alpha-1}$, and so $\alpha f(x) = xf'(x)$. The following result, which is commonly referred to as the *monotone density theorem*, shows that exactly this relationship holds for regularly varying distributions with $=$ replaced by $\sim$, modulo some technical conditions.

Theorem 2.11 (Monotone Density Theorem) *Suppose that the function f is absolutely continuous with derivative f'. If $f \in \mathcal{RV}(\rho)$ and f' is eventually monotone, then $xf'(x) \sim \rho f(x)$. Moreover, if $\rho \neq 0$, then $|f'(x)| \in \mathcal{RV}(\rho - 1)$.*[5]

[5] A function f is said to be *absolutely continuous* if it has a derivative f' almost everywhere that is integrable, such that

$$f(x) = f(0) + \int_0^x f'(t)dt \quad \forall x.$$

A function g is *eventually monotone* if there exists $x_0 > 0$ such that g is monotone over $[x_0, \infty)$.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

In what follows, we give the proof of Theorem 2.11. The proof of Theorem 2.10 is a bit more cumbersome, and we refer the interested reader to [183, Section 2.3.2] for the proof.

Proof of Theorem 2.11 For simplicity, we assume that $f'(x)$ is nondecreasing over $x \geq x_0$ (the proof for the case of eventually nonincreasing f' is along similar lines). Fixing a, b such that $0 < a < b$, we may write

$$\int_{ax}^{bx} \frac{f'(t)}{f(x)} dt = \frac{f(bx) - f(ax)}{f(x)}.$$

For $x > x_0/a$, the monotonicity of f' implies that

$$\frac{f'(ax)x(b-a)}{f(x)} \leq \frac{f(bx) - f(ax)}{f(x)} \leq \frac{f'(bx)x(b-a)}{f(x)}. \tag{2.2}$$

Noting that $f \in \mathcal{RV}(\rho)$, the first inequality in (2.2) implies that

$$\limsup_{x\to\infty} \frac{f'(ax)x}{f(x)} \leq \frac{b^\rho - a^\rho}{b-a}.$$

Next, letting $b \downarrow a$ on the right side of the above inequality and noting that this corresponds to taking the derivative of the function x^ρ at $x = a$, we obtain

$$\limsup_{x\to\infty} \frac{f'(ax)x}{f(x)} \leq \rho a^{\rho-1}. \tag{2.3}$$

Similarly, using the second inequality in (2.2) and letting $a \uparrow b$, we obtain

$$\liminf_{x\to\infty} \frac{f'(bx)x}{f(x)} \geq \rho b^{\rho-1}. \tag{2.4}$$

Setting $a = 1$ in (2.3) and $b = 1$ in (2.4), we conclude that $xf'(x) \sim \rho f(x)$. Finally, when $\rho \neq 0$, it is easy to see that $f'(x) \sim \rho\left(\frac{f(x)}{x}\right)$ implies that $|f'(x)| \in \mathcal{RV}(\rho - 1)$. □

Hopefully, it is already clear that Karamata's theorem is a particularly appealing and powerful property of regularly varying functions. But, it is worth considering a few examples in order to emphasize this further. Perhaps the most powerful example of the use of Karamata's theorem is in deriving the so-called "Karamata representation theorem" for regularly varying functions.

Karamata's Representation Theorem

We have already discussed one representation theorem for regularly varying functions (Theorem 2.8), which allows us to write any regularly varying function as $\bar{F}(x) = x^\rho L(x)$ for some slowly varying function $L(x)$. This is a particularly nice form since it highlights the view of regularly varying distributions as asymptotically power law; however, it is also a fairly implicit view of regularly varying functions since the form of $L(x)$ is not defined. Using Karamata's theorem, we can derive a much more precise representation theorem for regularly varying functions.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

Theorem 2.12 (Karamata's Representation Theorem) *$f \in \mathcal{RV}(\rho)$ if and only if it can be represented as*

$$f(x) = c(x)\exp\left\{\int_1^x \frac{\beta(t)}{t}dt\right\} \tag{2.5}$$

for $x > 0$, where $\lim_{x\to\infty} c(x) = c \in (0, \infty)$ and $\lim_{x\to\infty} \beta(x) = \rho$.

The representation of regularly varying distributions in Karamata's representation theorem may initially seem surprising since it does not superficially look like a power law. However, note that if one treats $\beta(t)$ as if it is a constant ρ (which it converges to in the limit), then the exponent becomes $\rho \log x$ and so the power law form appears (since $e^{\rho \log x} = x^\rho$).

Proof To begin, let us first check that if a function f can be represented via (2.5), then $f \in \mathcal{RV}(\rho)$. Note that for $y > 0$,

$$\frac{f(xy)}{f(x)} = \frac{c(xy)}{c(x)}\exp\left\{\int_x^{xy} \frac{\beta(t)}{t}dt\right\}.$$

Now, since $\beta(x) \to \rho$ as $x \to \infty$, given $\epsilon > 0$, there exists $x_0 > 0$ such that $\rho - \epsilon < \beta(x) < \rho + \epsilon$ for all $x > x_0$. Therefore, for x large enough so that $x, xy > x_0$,

$$\frac{c(xy)}{c(x)}\exp\left\{\int_x^{xy} \frac{\rho-\epsilon}{t}dt\right\} < \frac{f(xy)}{f(x)} < \frac{c(xy)}{c(x)}\exp\left\{\int_x^{xy} \frac{\rho+\epsilon}{t}dt\right\},$$

which is equivalent to

$$\frac{c(xy)}{c(x)}y^{\rho-\epsilon} < \frac{f(xy)}{f(x)} < \frac{c(xy)}{c(x)}y^{\rho+\epsilon}.$$

Now, since $\frac{c(xy)}{c(x)} \to 1$ as $x \to \infty$, we obtain

$$y^{\rho-\epsilon} \le \liminf_{x\to\infty} \frac{f(xy)}{f(x)} \le \limsup_{x\to\infty} \frac{f(xy)}{f(x)} \le y^{\rho+\epsilon}.$$

Letting $\epsilon \downarrow 0$, we finally conclude that $\lim_{x\to\infty} \frac{f(xy)}{f(x)} = y^\rho$, which, of course, implies by definition that $f \in \mathcal{RV}(\rho)$.

Next, we prove that if $f \in \mathcal{RV}(\rho)$, then f has a representation of the form (2.5). We prove this first for the slowly varying case (i.e., $\rho = 0$), and then consider the case of general ρ.

Accordingly, suppose that $L \in \mathcal{RV}(0)$. Define $b(x) = \frac{xL(x)}{\int_0^x L(t)dt}$. Note that Karamata's theorem implies that $b(x) \to 1$ as $x \to \infty$. Let $\beta_L(x) = b(x) - 1$. Now,

$$\begin{aligned}\int_1^x \frac{\beta_L(t)}{t}dt &= \int_1^x \frac{L(t)}{\int_0^t L(y)dy}dt - \log(x)\\ &= \log\left(\int_0^x L(y)dy\right) - \log\left(\int_0^1 L(y)dy\right) - \log(x).\end{aligned}$$

Now, using $\int_0^x L(y)dy = \frac{xL(x)}{b(x)}$, we obtain

$$\int_1^x \frac{\beta_L(t)}{t}dt = \log\left(\frac{L(x)}{b(x)\int_0^1 L(y)dy}\right),$$

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

which finally gives us

$$L(x) = c_L(x)\exp\left\{\int_1^x \frac{\beta_L(t)}{t}dt\right\}, \tag{2.6}$$

where $c_L(x) = b(x)\int_0^1 L(y)dy$. Noting that $c_L(x) \to \int_0^1 L(y)dy$ and $\beta_L(x) \to 0$ as $x \to \infty$, we have proved that L has the postulated representation.

Finally, moving to the case of general ρ, suppose that $f \in \mathcal{RV}(\rho)$. We know then that $f(x) = x^\rho L(x)$, where $L(x)$ is a slowly varying function. We have already established that $L(x)$ can be represented as (2.6), with $\beta_L(x) \to 0$ and $c_L(x) \to c \in (0, \infty)$ as $x \to \infty$. It then follows immediately that

$$f(x) = c_L(x)\exp\left\{\int_1^x \frac{(\beta_L(t) + \rho)}{t}dt\right\},$$

which gives us the desired representation. □

Karamata's representation theorem is an extremely powerful tool for working with regularly varying distributions. To see this, note that it is straightforward to prove Lemma 2.7 using Karamata's representation theorem (see Exercise 4). Additionally, Karamata's representation theorem can be used to show a number of other properties of regularly varying distributions that connect them to power law and Pareto distributions. We illustrate two of these here: (i) the observation that regularly varying distributions appear approximately linear on a log-log plot, and (ii) properties of the moments of regularly varying distributions.

Let us start by considering the behavior of regularly varying distributions in logarithmic scale. We have seen earlier that one distinguishing property of Pareto distributions is that they are exactly linear when viewed on a log-log scale. Specifically, recall that for Pareto distributions $\bar{F}(x) = (x/x_m)^{-\alpha}$ for $x > x_m$, and so

$$\log \bar{F}(x) = -\alpha \log(x) + \alpha \log(x_m).$$

Thus, $\log \bar{F}(x)$ is exactly linear in terms of $\log x$, with slope α. This is a property that allows for easy preliminary identification of them in data, as we have seen in Chapter 1 and explore in detail in Chapter 8. Note that one must be very cautious using this approach for estimation, as we illustrate in Chapter 8.

Using Karamata's representation theorem, we can easily obtain the corresponding property for the tail of regularly varying distributions. In particular, we have the following result, which shows that the tail of regularly varying distributions with index α is asymptotically linear with slope α when viewed on a log-log plot.

Lemma 2.13 *If f is a regularly varying function with index ρ, then*

$$\lim_{x\to\infty} \frac{\log f(x)}{\log(x)} = \rho.$$

Proof From Karamata's representation theorem, we know that

$$f(x) = c(x)\exp\left\{\int_1^x \frac{\beta(t)}{t}dt\right\},$$

where $\lim_{x\to\infty} c(x) = c \in (0, \infty)$ and $\lim_{x\to\infty} \beta(x) = \rho$.

Given $\epsilon > 0$, there exists $x_0 > 1$ such that for all $x \geq x_0$, $\rho - \epsilon < \beta(x) < \rho + \epsilon$. Therefore, for $x > x_0$,

$$\begin{aligned}\log f(x) &\leq \log c(x) + \int_1^{x_0} \frac{\beta(t)}{t} dt + \int_{x_0}^{x} \frac{\rho + \epsilon}{t} dt \\ &= \log c(x) + \int_1^{x_0} \frac{\beta(t)}{t} dt + (\rho + \epsilon)(\log(x) - \log(x_0)).\end{aligned}$$

From the preceding inequality, it follows that

$$\limsup_{x \to \infty} \frac{\log f(x)}{\log(x)} \leq \rho + \epsilon.$$

Using similar arguments, it can be shown that

$$\liminf_{x \to \infty} \frac{\log f(x)}{\log(x)} \geq \rho - \epsilon.$$

Letting ϵ approach zero completes the proof. □

Next, let us move to studying the moments of regularly varying distributions. Recall that the moments of Pareto distributions are a bit peculiar: for Pareto(x_m, ρ) distributions, the ith moment is finite if $i < \rho$ and infinite if $i > \rho$. The fact that moments can be infinite is, as we have seen in Chapter 1, not just of theoretical interest. Data from a variety of situations has been shown to exhibit power law tails with ρ around 1.2–2.1, and so is often well approximated by distributions with infinite variance.

Using the above result, it is not hard to show that regularly varying distributions have moments that parallel those of Pareto distributions. In particular, we have the following result.

Theorem 2.14 *Suppose that a nonnegative random variable X is regularly varying of index $-\rho$. Then*

$$\begin{aligned}\mathbb{E}\left[X^i\right] &< \infty \text{ for } 0 \leq i < \rho, \\ \mathbb{E}\left[X^i\right] &= \infty \text{ for } i > \rho.\end{aligned}$$

The moment conditions in the theorem above should not be particularly surprising at this point since computing moments has to do with integration and the finiteness of moments has to do mainly with the tail, which means that Karamata's theorem should ensure that regularly varying distributions behave like power laws. This intuition serves as a good guide for the proof, which makes use of Lemma 2.13, which was a consequence of Karamata's theorem. We leave the proof of the result as an exercise for the reader (see Exercise 5).

2.3.2 *Integral Transforms of Regularly Varying Distributions*

Integral transforms like the moment generating function, the characteristic function, and the Laplace–Stieltjes transform are of fundamental importance in probability, as well as in many physical problems in applied mathematics. It is often easier to study probabilistic and stochastic models using transforms than it is to study them directly as a result of the ease of computing convolutions, moments, time scalings, performing integration of the distribution, and so on. Thus, one can typically complete the analysis in "transform space" and then invert the transform to understand the distribution itself, taking advantage of the uniqueness of the representation.

In the context of this book, we have already seen the importance of transforms in the definition of heavy-tailed distributions. Recall that the definition of heavy-tailed distributions explicitly uses the moment generating function (m.g.f.) and defines heavy-tailed distributions as those distributions for which $M_X(t) := \mathbb{E}\left[e^{tX}\right] = \infty$ for all $t > 0$. This means that, while moment generating functions are often a powerful analytic tool, working with the m.g.f. of heavy-tailed distributions is problematic. Thus, one needs to consider other integral transforms in the case of heavy-tailed distributions. This section provides the tools for working with transforms in the heavy-tailed setting.

Though the m.g.f. is not appropriate for heavy-tailed distributions, one can instead use other transforms. When the distribution corresponds to a nonnegative random variable, the Laplace–Stieltjes transform (LST) is appropriate and, more generally, the characteristic function is the appropriate tool. The *Laplace–Stieltjes transform* of a function f is defined as

$$\psi_f(s) := \int_{-\infty}^{\infty} e^{-sx} df(x).$$

Specializing to probability distributions, given a random variable X following distribution F, the *Laplace–Stieltjes transform* of F (or X) is defined as

$$\psi_X(s) := \int_{-\infty}^{\infty} e^{-sx} dF(x) = \mathbb{E}\left[e^{-sX}\right].$$

Notice that the LST of a distribution is related to the m.g.f. via a change of variable: we replace the argument t in the definition of $M(t)$ by $-s$. Similarly, the characteristic function can be obtained by replacing t with it, where i is the imaginary unit. So, given a random variable X following distribution F, the *characteristic function* of F (or X) is defined as

$$\phi_X(t) := \mathbb{E}\left[e^{itX}\right] = \mathbb{E}\left[\cos(tX) + i\sin(tX)\right].$$

Note that if X is nonnegative, the LST $\psi_X(s)$ is well defined and finite for $s \geq 0$. On the other hand, the characteristic function $\phi(t)$ associated with any distribution is well defined and finite for all t.

To get a feel for the behavior of transforms in the case of heavy-tailed distributions it is useful to look at the specific case of a power-law function. To keep things simple, we consider power laws of the form

$$f(x) = \begin{cases} x^\rho & x \geq 0, \\ 0 & x < 0, \end{cases}$$

where $\rho > 0$. Note that with $\rho > 0$, f cannot capture a probability distribution, but limiting our attention to positive indices makes things simpler, so we do that for now and then extend

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

the analysis to probability distributions later in the section. We do this because the case of $\rho > 0$ turns out to be quite instructive. In this case, the LST of f can be written as follows:

$$\begin{aligned}\psi_f(s) &= \int_0^\infty e^{-sx} df(x) \\ &= \rho \int_0^\infty e^{-sx} x^{\rho-1} dx \\ &= \rho s^{-\rho} \int_0^\infty e^{-sx} (sx)^{\rho-1} d(sx) \\ &= \rho s^{-\rho} \Gamma(\rho) = s^{-\rho} \Gamma(\rho+1),\end{aligned}$$

where the last line uses the Gamma function Γ, which is a continuous extension of the factorial function to the real numbers defined as $\Gamma(z) := \int_0^\infty e^{-t} t^{z-1} dt$ and satisfying $\Gamma(n) = (n-1)!$ for $n \in \mathbb{Z}$ and $z\Gamma(z) = \Gamma(z+1)$.

The calculation in this example reveals something exciting. The LST of a function f with a power law (as $x \to \infty$) also behaves like a power law (as $s \downarrow 0$). This is exciting because it suggests that one can potentially understand properties of the tail of a distribution by studying properties of the LST near zero. Of course, one could potentially obtain this information by inverting the LST, but that is typically very involved and cannot be done in closed form apart from a few special cases. In contrast, in this example, the tail behavior of f can be obtained with a simple observation about ψ.

However, before getting too excited, it is important to remember that, so far, we have only seen this behavior in the case of specific f following a power law with a positive index. Thus, the question becomes:

Do regularly varying distributions have regularly varying transforms?

If so, it would be quite powerful since it is often the case that one can derive the LST in situations where it is not tractable to work directly with the distribution.

Fortunately. the answer is "yes." Results of this form are called *Abelian* and *Tauberian* theorems, and there is a wide variety of these theorems for the LST and the characteristic function. As a first example, we present Karamata's Tauberian theorem, which is the direct extension of the power-law example above for the case of increasing functions.

Theorem 2.15 (Karamata's Tauberian Theorem) *Let f be a nonnegative right-continuous increasing function such that $f(x) = 0$ for $x < 0$, and let $\rho \geq 0$. Then, for slowly varying $L(x)$, the following are equivalent:*

$$f(x) \sim L(x)x^\rho \qquad (x \to \infty), \tag{2.7}$$

$$\psi_f(s) \sim \Gamma(\rho+1)L(1/s)s^{-\rho} \qquad (s \downarrow 0). \tag{2.8}$$

This result says something very powerful. Informally, it says that if the behavior of the LST as $s \downarrow 0$ is approximately a power law, then the corresponding function is also a power law *with the same index*. Further, given the representation of regularly varying functions in Theorem 2.8, this can be interpreted in a different light as well. In particular, since (2.7)

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

characterizes regularly varying functions, the theorem states that regularly varying functions are exactly those that have LSTs that are regularly varying around zero.

Though Theorem 2.15 is commonly called a Tauberian theorem, it actually includes both a Tauberian theorem and an Abelian theorem. In particular, the direction showing that (2.7) implies (2.8) is called an Abelian theorem, and the reverse direction is a Tauberian theorem. The Tauberian direction is typically harder to prove, which is why such theorems are typically referred to as Tauberian theorems. We omit the proof of Theorem 2.15 here; the interested reader is referred to Theorem 1.7.1 in [31].

Theorem 2.15 is powerful but does not yet give us exactly what we would like since it still assumes that the index ρ is positive. Thus, it does not apply to regularly varying *distributions* directly. However, it is possible to remedy this. In particular, the following is a more general version of Karamata's Tauberian theorem that uses a Taylor expansion of the LST in terms of the moments of the distribution.

Theorem 2.16 *Consider a nonnegative random variable X with distribution F. For $n \in \mathbb{Z}_+$, suppose that $\mathbb{E}[X^n] < \infty$. Then for slowly varying $L(x)$ and $\alpha = n + \beta$ where $\beta \in (0,1)$, the following are equivalent:*

$$\bar{F}(x) \sim \frac{(-1)^n}{\Gamma(1-\alpha)} x^{-\alpha} L(x) \qquad (x \to \infty), \tag{2.9}$$

$$(-1)^{n+1} \left[\psi_X(s) - \sum_{i=0}^{n} \frac{\mathbb{E}[X^i](-s)^i}{i!} \right] \sim s^\alpha L(1/s) \qquad (s \downarrow 0). \tag{2.10}$$

To interpret the statement of Theorem 2.16, note that if a nonnegative random variable X satisfies $\mathbb{E}[X^n] < \infty$, then its LST can be expressed via a Taylor expansion as follows:

$$\psi_X(s) = \sum_{i=0}^{n} \frac{\mathbb{E}[X^i](-s)^i}{i!} + o(s^n). \tag{2.11}$$

The Abelian component of Theorem 2.16 states that if X is regularly varying with index $-\alpha$, where $\alpha \in (n, n+1)$, then the $o(s^n)$ correction term in (2.11) is of the order of s^α. The Tauberian component makes the converse implication.

To understand this connection better, let us first consider the case $n = 0$. For this case, Theorem 2.16 states that for $\alpha \in (0,1)$,

$$\bar{F}(x) \sim \frac{1}{\Gamma(1-\alpha)} x^{-\alpha} L(x) \iff \psi_X(s) - 1 \sim -s^\alpha L(1/s).$$

Clearly, this statement applies to distributions with infinite mean. Consider, for example, the Lévy distribution, which is parameterized by $c > 0$ and has LST $\psi(s) = e^{-\sqrt{2sc}}$ for $s \geq 0$ [120]. Noting that as $s \downarrow 0$, $1 - \psi(s) \sim -\sqrt{2cs}$, Theorem 2.16 (specifically, the Tauberian part) implies that the Lévy tail satisfies $\bar{F}(x) \sim \frac{\sqrt{2c}}{\Gamma(1/2)} x^{-1/2}$, that is, $\bar{F}(x) \sim \sqrt{\frac{2c}{\pi}}\, x^{-1/2}$. The same conclusion can be arrived at by applying Karamata's theorem to the Lévy density function; see Exercise 1.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

Next, consider the case $n = 1$. In this case, Theorem 2.16 states that for $\alpha \in (1, 2)$,

$$\bar{F}(x) \sim \frac{-1}{\Gamma(1-\alpha)} x^{-\alpha} L(x) \iff \psi_X(s) - 1 + \mathbb{E}[X]\, s \sim s^{\alpha} L(1/s).$$

This statement in turn is applicable to distributions with a finite first moment, but an infinite second moment.

Proof sketch of Theorem 2.16 We present here the proof of Theorem 2.16 for the case $n = 0$ to illustrate how Theorem 2.16 actually follows from Theorem 2.15. The case $n \geq 1$ is slightly more cumbersome, but follows along similar lines (see Exercise 10).

Let us first consider the Abelian direction. Accordingly, suppose that $\bar{F}(x) \sim \frac{1}{\Gamma(1-\alpha)} x^{-\alpha} L(x)$ for $\alpha \in (0, 1)$. Consider now the function $g(x) = \int_0^x \bar{F}(y)dy$, which has LST $\psi_g(s) = \frac{1-\psi_X(s)}{s}$ (checking this claim is left as an exercise for the reader). From Karamata's theorem (Theorem 2.10), note that

$$g(x) \sim \frac{1}{(1-\alpha)\Gamma(1-\alpha)} x^{1-\alpha} L(x).$$

Since $1 - \alpha > 0$, we can now invoke Theorem 2.15 (specifically, the Abelian part) to conclude that

$$\psi_g(s) = \frac{1-\psi_X(s)}{s} \sim \frac{\Gamma(2-\alpha)}{(1-\alpha)\Gamma(1-\alpha)} s^{\alpha-1} L(1/s),$$

which in turn implies that $\psi_X(s) - 1 \sim -s^{\alpha} L(1/s)$ (note that $\Gamma(2-\alpha) = (1-\alpha)\Gamma(1-\alpha)$).

Next, consider the Tauberian direction, that is, suppose that $\psi_X(s) - 1 \sim -s^{\alpha} L(1/s)$, which is equivalent to $\psi_g(s) \sim s^{\alpha-1} L(1/s)$. Invoking Theorem 2.15 (specifically, the Tauberian part), we conclude that $g(x) \sim \frac{1}{\Gamma(2-\alpha)} x^{1-\alpha} L(x)$. Finally, the monotone density theorem (Theorem 2.11) implies that

$$\bar{F}(x) \sim \frac{1-\alpha}{\Gamma(2-\alpha)} x^{-\alpha} L(x) = \frac{1}{\Gamma(1-\alpha)} x^{-\alpha} L(x).$$

This completes the proof.

The proof for general n follows along similar lines; the Abelian direction involves applying Karamata's theorem $n + 1$ times, while the Tauberian direction involves applying the monotone density theorem $n + 1$ times (see Exercise 10). □

Karamata's Tauberian theorem is only one of many Tauberian theorems that are useful when studying heavy-tailed distributions. In particular, given that it relies on the LST, the versions we have stated above are only relevant for distributions of nonnegative random variables. For other distributions, one needs Tauberian theorems for the characteristic function. An example of such a Tauberian theorem is the following, which is due to Pitman [176] (see also page 336 of [31]). Note that this Tauberian theorem uses only the real component of the characteristic function, $U_X(t)$, that is,

$$U_X(t) := \mathrm{Re}(\phi_X(t)) = \int_{-\infty}^{\infty} \cos(tx) dF(x).$$

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

Theorem 2.17 (Pitman's Tauberian Theorem) *For slowly varying $L(x)$, and $\alpha \in (0, 2)$, the following are equivalent:*

$$\Pr(|X| > x) \sim x^{-\alpha} L(x) \text{ as } x \to \infty,$$
$$1 - U_X(t) \sim \frac{\pi}{2\Gamma(\alpha)\sin(\pi\alpha/2)} t^\alpha L(1/t) \text{ as } t \downarrow 0.$$

While there are many versions of Abelian and Tauberian theorems for the characteristic function, we choose to highlight this one because we make use of it later in the book in Chapter 5 when introducing and proving the generalized central limit theorem. Like Karamata's Tauberian theorem, this result connects the tail of the distribution to the behavior of a "transform" around zero, only in this case the "transform" considered is the characteristic function. Note that, because this Tauberian theorem applies to the tail of $|X|$ rather than X, it cannot be used to distinguish the behavior of the right and left tails of the distribution. Rather, it provides information only about the sum of the two tails. However, because of this fact, it deals only with the real part of the characteristic function, which makes it much simpler to work with analytically. The interested reader can find more general Abelian and Tauberian theorems in [31].

We have not focused on examples in this section; however, there are a number of illustrative examples of how to apply the theorems in this section scattered throughout the book. Two particularly important examples are in Chapter 5: we apply the Abelian part of Theorem 2.17 to prove the generalized central limit theorem, and the Tauberian part of Theorem 2.16 to study the return time of a one-dimensional random walk.

2.4 An Example: Closure Properties of Regularly Varying Distributions

Regularly varying distributions play a central role in this book, showing up in nearly every chapter. So, as you work through the book you will encounter a variety of applications of the properties and theorems discussed in the previous sections of this chapter. For example, regularly varying distributions play a foundational role in the generalized central limit theorem discussed in Chapter 5, the analysis of a multiplicative processes in Chapter 6, and the discussion of the extremal central limit theorem in Chapter 7.

Here, as a "warm-up" to those applications we provide some simple illustrations of the properties we have studied so far in order to prove some important *closure* properties about the set of regularly varying distributions. These closure properties, while intuitive, should not be taken for granted. In fact, these closure properties do not always hold for the more general classes of heavy-tailed distributions we study in the next two chapters.

Lemma 2.18 *Suppose that the random variables X and Y are independent, and regularly varying of index $-\alpha_X$ and $-\alpha_Y$ respectively.*

(i) $\min(X, Y)$ is regularly varying with index $-(\alpha_X + \alpha_Y)$.
(ii) $\max(X, Y)$ is regularly varying with index $-\min\{\alpha_X, \alpha_Y\}$.
(iii) $X+Y$ is regularly varying with index $-\min\{\alpha_X, \alpha_Y\}$. Moreover, $\Pr(X + Y > t) \sim \Pr(\max(X, Y) > t)$.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

Lemma 2.18 shows that the class of regularly varying distributions is closed with respect to min, max, and convolution. These properties should be exactly what you should expect, given the intuition that regularly varying distributions are generalizations of Pareto distributions. For example, the convolution of two Pareto distributions does not yield another Pareto distribution, of course, but when one considers only the tail, the resulting convolution will certainly continue to have a tail that decays like a polynomial, and thus be regularly varying. A similar intuition holds for both the $\min$ and $\max$ of two Pareto random variables. Though the resulting distributions are certainly not Pareto distributions, they still have a tail that decays like a polynomial, and thus are regularly varying.

While simple and intuitive, these closure properties often turn out to be powerful. For example, the third property in Lemma 2.18 can be extended to the case of n i.i.d. regularly varying random variables $Y_i, i \geq 1$ easily, that is, $\Pr(Y_1 + Y_2 + \cdots + Y_n > t) \sim n\Pr(Y_1 > t)$ (see Exercise 7). This fact is used crucially in our analysis of random walks in Chapter 7, specifically in the proof of Theorem 7.6.

Proof of Lemma 2.18 We begin by using the representation of regularly varying distributions given by Theorem 2.8. Since X and Y are regularly varying, there exist slowly varying functions L_X and L_Y such that $\Pr(X > t) = t^{-\alpha_X} L_X(t)$ and $\Pr(Y > t) = t^{-\alpha_Y} L_Y(t)$. Now, using these representations, we can prove each closure property in turn.

(i) Note that $\Pr(\min(X,Y) > t) = \Pr(X > t)\Pr(Y > t) = t^{-(\alpha_X+\alpha_Y)} L_X(t) L_Y(t)$. Since the product of slowly varying functions is also slowly varying, Claim *(i)* of the lemma follows.

(ii) Since $\{\max(X,Y) > t\} = \{X > t\} \cup \{Y > t\}$, we have

$$\Pr(\max(X,Y) > t) = \Pr(X > t) + \Pr(Y > t) - \Pr(X > t)\Pr(Y > t). \quad (2.12)$$

Without loss of generality, we can consider the following cases separately: $\alpha_X < \alpha_Y$, and $\alpha_X = \alpha_Y$.

If $\alpha_X < \alpha_Y$, it follows from (2.12) that $\Pr(\max(X,Y) > t) \sim \Pr(X > t)$, which then implies that $\max(X,Y)$ is regularly varying with index $-\alpha_X$.

If $\alpha_X = \alpha_Y$, then it follows from (2.12) that $\Pr(\max(X,Y) > t) \sim \Pr(X > t) + \Pr(Y > t)$, that is, $\Pr(\max(X,Y) > t) \sim t^{-\alpha_X}(L_X(t) + L_Y(t))$. Since the sum of slowly varying functions is also slowly varying, it follows that $\max(X,Y)$ is regularly varying with index $-\alpha_X$.

This completes the proof of Claim *(ii)*.

(iii) This final claim is the most involved. The first step in our proof is to establish an upper bound and a lower bound on the probability of the event $\{X + Y > t\}$. Then we analyze those bounds to obtain the result.

To begin, note that the event $\{X > t\} \cup \{Y > t\}$ implies $\{X + Y > t\}$. This gives us the following lower bound.

$$\Pr(X + Y > t) \geq \Pr(X > t) + \Pr(Y > t) - \Pr(X > t)\Pr(Y > t). \quad (2.13)$$

Next, let us fix $\delta \in (0, 1/2)$. It is easy to see that the event $\{X + Y > t\}$ implies the event $\{X > (1-\delta)t\} \cup \{Y > (1-\delta)t\} \cup \{X > \delta t, Y > \delta t\}$. This implication, along with the union bound, leads to the following upper bound.

$$\Pr(X+Y>t) \leq \Pr(X>(1-\delta)t)+\Pr(Y>(1-\delta)t)+\Pr(X>\delta t)\Pr(Y>\delta t). \tag{2.14}$$

Now, to complete the proof we consider the following two cases separately: $\alpha_X < \alpha_Y$, and $\alpha_X = \alpha_Y$.

Let us first consider the case $\alpha_X < \alpha_Y$. It follows from (2.13) that

$$\liminf_{t\to\infty} \frac{\Pr(X+Y>t)}{\Pr(X>t)} \geq 1.$$

Similarly, it follows from (2.14) that

$$\limsup_{t\to\infty} \frac{\Pr(X+Y>t)}{\Pr(X>t)} \leq \lim_{t\to\infty} \frac{\Pr(X>(1-\delta)t)}{\Pr(X>t)} = (1-\delta)^{-\alpha_X}.$$

Letting δ approach zero, we conclude that $\Pr(X+Y>t) \sim \Pr(X>t)$. This implies that $X+Y$ is regularly varying with index $-\alpha_X$, and also that $\Pr(X+Y>t) \sim \Pr(\max(X,Y)>t)$ (since we have established in the proof of Claim *(ii)* that $\Pr(\max(X,Y)>t) \sim \Pr(X>t)$).

Finally, we consider the case $\alpha_X = \alpha_Y$. In this case, using the same steps as above, it can be shown that

$$\Pr(X+Y>t) \sim \Pr(X>t) + \Pr(Y>t).$$

This, of course, implies that $X+Y$ is regularly varying with index $-\alpha_X$, and also that $\Pr(X+Y>t) \sim \Pr(\max(X,Y)>t)$ (we have established in the proof of Claim *(ii)* that $\Pr(\max(X,Y)>t) \sim \Pr(X>t)+\Pr(Y>t)$).

This completes the proof of Claim *(iii)*.

□

2.5 An Example: Branching Processes

Branching processes are a fundamental and widely applicable area of stochastic processes. While they were born from the study of surnames in genealogy, at this point they have found applications broadly in the study of reproduction, epidemiology, queueing theory, statistics, and many other areas. Here we use one of the first, and most famous branching process models – the Galton–Watson process – as an illustrative example of the power of the properties of regular variation that we have explored in this chapter.

Not only is the Galton–Watson model one of the most prominent examples of a branching process, it has an interesting story behind it. As the story goes, Victorian aristocrats were concerned about keeping their surnames from going extinct and wanted to understand how many children they needed to have to ensure the survival of their name. This prompted Sir Francis Galton to pose the following question in the *Educational Times* in 1873 [95]:

How many children (on average) must each generation of a family have in order for the family name to continue in perpetuity?

Just a year later, Reverend Henry William Watson came up with the answer, and the two wrote a paper [215]. By now, the model named after them has become the canonical model of branching processes and has been used in wide-reaching areas from biology (see [11]), to the analysis of algorithms (see [65]), to the spread of epidemics (see, for example, [35, 164]).

The modern version of this model is defined formally as follows. In particular, a Galton–Watson process $\{X_n\}_{n\geq 0}$ is defined by

$$X_0 = 1,$$
$$X_{n+1} = \sum_{j=1}^{X_n} N_j^{(n+1)} \quad (n \geq 0),$$

where $N_j^{(n+1)}$ are i.i.d. random variables taking nonnegative integer values, distributed as N. In the Victorian context, N was interpreted as the number of male children (since the woman took the man's surname at marriage) in a family, and X_n as the number of men in the $n+1$st generation. Given this model, the question asked by Victorian aristocrats can be studied by asking, given the distribution of N, will the process go on forever (i.e., $X_n > 0$ for all n) or will it go extinct (i.e., for some n_o, $X_n = 0$ for all $n \geq n_0$)? And, if it goes extinct, how many total distinct descendants (across all generations) would exist?

It turns out that the answers to these questions depend on the expected number of male children each person has, that is, $\mu := \mathbb{E}[N]$. It is not hard to see that the probability of extinction, η, is given by $\eta = \lim_{n\to\infty} \Pr(X_n = 0)$. The foundational theorem for Galton–Watson processes illustrates that there are three cases, depending on whether μ is greater than, less than, or equal to 1. Basically, to have a positive probability of avoiding extinction, the expected number of children of each generation needs to be strictly greater than one.

Theorem 2.19 *The probability of extinction, η, in a Galton–Watson branching process satisfies the following:*

(i) Subcritical case: If $\mu < 1$ then $\eta = 1$.
(ii) Critical case: If $\mu = 1$ and N has positive variance, then $\eta = 1$.
(iii) Supercritical case: If $\mu > 1$, then $\eta \in (0, 1)$.

Note that in the subcritical case and the critical case, extinction is guaranteed (except in the trivial case where N equals 1 with probability 1). On the other hand, in the supercritical case, the lineage has a positive probability of surviving in perpetuity. Theorem 2.19 is classically proven using an approach based on probability generation functions; and we refer the interested reader to [104, Section 5.4] for the proof.

While the subcritical and critical cases are identical from the standpoint of extinction probability, they differ in terms of the distribution of the total number of distinct male descendants $Z := \sum_{n\geq 0} X_n$ as well as the time to extinction $\tau := \min\{n : \ X_n = 0\}$ (see [20, 103]). Here, our goal in studying branching processes is to illustrate the power of the Tauberian theorems we have introduced in this chapter; thus, we study the tail of Z in the critical case. Specifically, we prove the following result.

Theorem 2.20 *Suppose that $\mu = 1$ and that N has finite, positive variance. Then, the total number of distinct male descendants, Z, is regularly varying with*

$$\Pr(Z > t) \sim \frac{1}{\sqrt{\pi(\mathbb{E}[N^2] - \mathbb{E}[N])}} t^{-1/2}.$$

Before moving to the proof, it is important to notice that the total number of distinct male descendants has the following recursive structure:

$$Z \stackrel{d}{=} 1 + \sum_{i=1}^{N} Z_i, \tag{2.15}$$

where $\{Z_i\}$ are i.i.d. random variables with the same distribution as Z and independent of N. To see this, think of N as the number of descendants of the "first" individual, and Z_i as the number of descendants of his ith child. That each Z_i has the same distribution as Z is referred to as the *branching property*. In the proof that follows, we exploit this branching property to characterize the tail of Z.

Proof of Theorem 2.20 Since we are going to apply a Tauberian theorem, the transform of Z is important. Here, we use the LST of Z, denoted by ψ_Z. Let $G_N(t) := \mathbb{E}\left[t^N\right] = \sum_{i=0}^{\infty} t^i \text{Pr}\left(N = i\right)$ denote the *probability generating function* of N. It is now not hard to show, using (2.15), that ψ_Z satisfies the following functional equation:

$$\psi_Z(s) = e^{-s} G_N(\psi_Z(s)); \tag{2.16}$$

see Exercise 13.

Analogously to the LST, the probability generating function admits the following Taylor expansion around $t = 1$:

$$G_N(t) = 1 + m_1(t-1) + m_2(t-1)^2(1 + o(1)) \quad \text{as } t \uparrow 1,$$

where $m_1 = \mathbb{E}\left[N\right], m_2 = \mathbb{E}\left[N^2\right] - \mathbb{E}\left[N\right]$. Given that N has finite, positive variance, $m_2 > 0$ (see Exercise 14) and using the fact that $m_1 = \mu = 1$ in the critical case that we are studying, the above expansion simplifies to

$$G_N(t) = t + m_2(t-1)^2(1 + o(1)).$$

Now, we combine this with the functional equation for $\psi_Z(s)$ to obtain

$$\psi_Z(s) = e^{-s}[\psi_Z(s) + m_2(1 - \psi_Z(s))^2(1 + o(1)) \quad \text{as } s \downarrow 0. \tag{2.17}$$

Our goal is to apply Theorem 2.16, but we must first simplify the above expression. To do so, we use a few Taylor expansions. First, note that $e^{-s} = 1 - s(1 + o(1))$. Similarly, we also know that $\psi_Z(s) = 1 - o(1)$, since the total size of Z is finite with probability 1 by Theorem 2.19. Now, if we first move the term $e^{-s}\psi_Z(s)$ to the left-hand side of (2.16) and then use the two expansions, we get

$$s = m_2(1 - \psi_Z(s))^2(1 + o(1)) \quad \text{as } s \downarrow 0.$$

It follows that

$$1 - \psi_Z(s) \sim \sqrt{s/m_2} \quad \text{as } s \downarrow 0.$$

Finally, we are ready to apply Theorem 2.16 with $n = 0$, $\alpha = 1/2$, and $L(x) = 1/\sqrt{m_2}$. Using the identity $\Gamma(1/2) = \sqrt{\pi}$, we get

$$\text{Pr}\left(Z > x\right) \sim \frac{1}{\sqrt{\pi m_2}} x^{-1/2} \quad \text{as } x \to \infty.$$

□

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

2.6 Additional Notes

In the chapter we gave an overview of several properties of regularly varying *distributions*. While we did not focus much on regularly varying *functions*, the properties we described also apply more broadly. However, the interested reader can find much more on regularly varying *functions* in [31]. That book also contains results for sums that are discrete analogues of Section 2.3.1. That section is the only section in this chapter that assumes continuity.

While we described many analytic and closure properties of regularly varying distributions within this chapter, there are many other useful properties we did not have space to cover. For example, with respect to closure properties, an additional important property is the closure of products of random variables, for which we refer to one of the exercises at the end of the chapter; see also [63]. In addition, it can be shown that certain generalized inverses of regularly varying distributions are still regularly varying. On the analytical side, an important property that should be mentioned is the *uniform convergence theorem*, stating that the convergence of $L(at)/L(t)$ in the definition of slowly varying functions is necessarily uniform on any interval $a \in [g, d]$ for $0 < g < d < \infty$. For an overview of these and many other properties, we refer to the landmark monograph on regular variation [31].

We have focused on the most classical version of regularly varying distributions in this chapter, but it is important to be aware that there are several important extensions of regular variation, some of which will appear in later chapters. Two particularly useful extensions are (i) *intermediate regular variation* and (ii) *dominated variation*. A function f is of intermediate regular variation if

$$\lim_{\epsilon \downarrow 0} \limsup_{x\to\infty} \frac{f(x(1+\epsilon))}{f(x)} = \lim_{\epsilon \downarrow 0} \liminf_{x\to\infty} \frac{f(x(1+\epsilon))}{f(x)} = 1; \tag{2.18}$$

and a function f is of dominated variation if

$$\limsup_{x\to\infty} \frac{f(xa)}{f(x)} < \infty, \liminf_{x\to\infty} \frac{f(xa)}{f(x)} > 0, \tag{2.19}$$

for every $a > 0$.

It is straightforward to see that regularly varying functions of index $\neq 0$ satisfy both intermediate regular variation and dominated variation. In some particular situations, for example in queueing theory, the assumption of intermediate regular variation is the most general possible assumption for particular approximations of the tail to hold, or for particular proof methods to work; see [33, 165] for examples.

In statistical applications such as establishing asymptotic normality of estimators, it is convenient to consider a subclass of slowly varying functions, which is the class of second-order slowly varying functions. In particular, a function L is second-order slowly varying of index γ if there exists a function g that is regularly varying with index γ such that

$$\lim_{t\to\infty} \frac{\frac{L(tx)}{L(t)} - 1}{g(t)} = K \frac{x^\gamma - 1}{\gamma}. \tag{2.20}$$

See, for example, Chapter 2 in [31].

As we have mentioned already, regular variation plays an important in queuing theory. In addition, the concept of regular variation is also paramount in financial and insurance

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

mathematics [75], [17]. Another application area where the concept of scale-freeness and regular variation is important is that of complex networks, where the definition of power laws versus the more flexible class of regularly varying distributions sometimes seems to cause some confusion; see the discussion in [212]. For an introduction to the field of complex networks, see [23]. In Chapter 6 of this book, we come back to this application, when we discuss the mechanism of preferential attachment as a classical example of the emergence of heavy-tailed phenomena.

Regular variation will reappear at many other places in this book. A non-exhaustive list of examples is

- Chapter 3, where we use regularly varying distributions to investigate the behavior of random sums.
- Chapter 4, where we apply regularly varying distributions to study residual life and illustrate a connection between slowly varying functions and long-tailed distributions.
- Chapter 5, where we use Tauberian theorems to derive the generalized Central Limit Theorem.
- Chapter 6, where we use regular variation to understand variations of the multiplicative Central Limit Theorem.
- Chapter 7, where characterizing the classes of distributions that admit a limit law for their maxima rely on analytic tools of regular variation, as does the analysis of the all-time maximum of a random walk with negative drift.
- Chapters 8 and 9, where regular variation plays a key role in the development of statistic tools for estimating heavy-tailed distributions from data.

2.7 Exercises

1. Show that the following distributions are asymptotically scale invariant (i.e., regularly varying).
   (a) The Cauchy distribution.
   (b) The Burr distribution.
   (c) The Lévy distribution.
   The definitions of these distributions can be found in Section 1.2.4.
2. Consider a distribution F over $\mathbb{R}_+$ with finite mean μ. Recall that the *excess* distribution corresponding to F, denoted by F_e, is defined as
$$\bar{F}_e(x) = \frac{1}{\mu}\int_x^\infty \bar{F}(y)dy.$$
   If $F \in \mathcal{RV}(-\alpha)$ for $\alpha > 1$, show that $F_e \in \mathcal{RV}(-(\alpha-1))$. Specifically, show that
$$\bar{F}_e(x) \sim \frac{x}{\alpha-1}\bar{F}(x).$$
3. Prove that the LogNormal distribution is *not* regularly varying.
4. Prove Lemma 2.7.
5. Prove Theorem 2.14.
6. Prove that if the function f satisfies $xf'(x) \sim \rho f(x)$, then $f \in \mathcal{RV}(\rho)$.
   Hint: Use the Karamata representation theorem (Theorem 2.12).

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press

7. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. regularly varying random variables with index $-\alpha$, where $n \geq 2$. Prove that
$$\Pr(X_1 + X_2 + \cdots + X_n > x) \sim n\Pr(X_1 > x).$$
8. Suppose that the random variables X and Y are independent, with $X \in \mathcal{RV}(-\alpha_X)$ and
$$\Pr(Y > t) = o(\Pr(X > t)) \quad (t \to \infty).$$
Prove that $X + Y \in \mathcal{RV}(-\alpha_X)$.
9. Suppose that the random variable $X \in \mathcal{RV}(-\alpha)$. Show that the same property holds for its integer part $\lfloor X \rfloor$.
10. Prove Theorem 2.16 for the case $n = 1$. Specifically, for a nonnegative random variable with finite mean, prove that for $\alpha \in (1, 2)$,
$$\bar{F}(x) \sim \frac{-1}{\Gamma(1-\alpha)} x^{-\alpha} L(x) \iff \psi_X(s) - 1 + \mathbb{E}[X]\, s \sim s^{\alpha} L(1/s).$$
Note: The above exercise should give the reader an idea of how to prove Theorem 2.16 for general n.
11. Let X be a nonnegative random variable of which the distribution function F is regularly varying with index $-\alpha$, and let Y be a random variable independent of X for which $E[Y^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$. Show that
$$\frac{\Pr(XY > t)}{\Pr(X > t)} \to \mathbb{E}[Y^{\alpha}] \tag{2.21}$$
as $t \to \infty$. [Hint: condition on the value of Y and use the condition $E[Y^{\alpha+\epsilon}] < \infty$ and a useful property of slowly varying functions to justify the interchange of limit and integral.]
12. Suppose that $f(t)$ is regularly varying of index $\alpha > 0$. Define $f^{\leftarrow}(x) = \inf\{t : f(t) = x\}$. Prove that $f^{\leftarrow}(x)$ is regularly varying of index $1/\alpha$.
13. Suppose that $\{X_i\}_{i \geq 1}$ is a sequence of i.i.d. random variables. The random variable N takes nonnegative integer values, and is independent of $\{X_i\}_{i \geq 1}$. Let $G_N(\cdot)$ denote the probability generating function corresponding to N. Define $S_N = \sum_{i=1}^{N} X_i$. Prove that
$$\psi_{S_N}(s) = G_N(\psi_X(s)).$$
Here, $\psi_Y(\cdot)$ denotes the LST corresponding to random variable Y.
14. Suppose that the random variable N takes nonnegative integer values. If the variance of N is positive and finite, show that $\mathbb{E}[N^2] > \mathbb{E}[N]$.

https://doi.org/10.1017/9781009053730.003 Published online by Cambridge University Press