# 1

# Introduction

> The top 1% of the population controls 35% of the wealth. On Twitter, the top 2% of users send 60% of the messages. In the health care system, the treatment for the most expensive fifth of patients create four-fifths of the overall cost. These figures are always reported as shocking, as if the normal order of things has been disrupted, as if [it] is a surprise of the highest order. It's not. Or rather, it shouldn't be.
> – Clay Shirky, in response to the question "What scientific concept would improve everybody's cognitive toolkit?" [194]

Introductory probability courses often leave the impression that the Gaussian distribution is what we should expect to see in the world around us. It is referred to as the "Normal" distribution after all! As a result, statistics like the ones in the quote above tend to be treated as aberrations, since they would never happen if the world were Gaussian. The Gaussian distribution has a "scale," a typical value (the mean) around which individual measurements are centered and do not deviate from by too much. For example, if we consider human heights, which are approximately Gaussian, the average height of an adult male in the US is 5 feet 9 inches and most people's heights do not differ by more than 10 inches from this. In contrast, there are order-of-magnitude differences between individuals in terms of wealth, Twitter followers, health care costs, and so on.

However, order-of-magnitude differences like those just mentioned are not new and should not be surprising. Over a century ago, Italian economist Vilfredo Pareto discovered that the richest 20 percent of the population controlled 80 percent of the property in Italy. This is now termed the "Pareto Principle," aka the "80-20" rule and variations of this principle have shown up repeatedly in widely disparate areas in the time since Pareto's discovery. For example, in 2002 Microsoft reported that 80 percent of the errors in Windows are caused by 20 percent of the bugs [188], and similar versions of the Pareto principle apply (though not always with 80/20) to many aspects of business, for example, most of the profit is made from a small percentage of the customers and most of the sales are made by a small percentage of the sales team.

Statistics related to the Pareto principle make for compelling headlines, but they are typically an indication of something deeper. When we see such figures, it is likely that there is not a Gaussian distribution underlying them, but rather a heavy-tailed distribution is the reason for the "surprising" statistics. The most celebrated such distribution again carries Vilfredo Pareto's name: *the Pareto distribution*. Heavy-tailed distributions such as the Pareto distribution are just as prominent as (if not more so than) the Gaussian distribution and have been observed in hundreds of applications in physics, biology, computer science, the social sciences, and beyond over the past century. Some examples include the sizes of cities [92, 163], the file sizes in computer systems and networks [52, 146], the size of avalanches and earthquakes [109, 144], the length of protein sequences in genomes [130, 145], the size of meteorites [13, 162], the degree distribution of the web graph [36, 116], the returns of stocks [49, 94], the number of copies of books sold [14, 110], the number of households affected during blackouts in power grids [114], the frequency of word use in natural language [77, 227], and many more.

Given the breadth of areas where heavy-tailed phenomena have been observed, one might guess that, by now, observations of heavy-tailed phenomena in new areas are expected – that heavy tails are treated as *more normal than the Normal*. After all, Pareto's work has been widely known for more than a century. However, despite a century of experience, statistics related to the Pareto Principle and, more broadly, heavy-tailed distributions are still typically presented as surprising curiosities – anomalies that could not have been anticipated. Even in scientific communities, observations of heavy-tailed phenomena are often presented as mysteries to be explained rather than something to be expected a priori. In many cases, there is even a significant amount of controversy and debate that follows the identification of heavy-tailed phenomena in data.

### *Surprising? Mysterious? Controversial?*

Given the century of mathematical and statistical work around heavy tails, it certainly should not be the case that heavy tails are surprising, mysterious, and controversial. In fact, there are many reasons why one should *expect* to see heavy-tailed distributions arise. Perhaps the main reason why they are still viewed as surprising is that the version of the central limit theorem taught in introductory probability courses gives the impression that the Gaussian will occur everywhere. However, this introductory version of the central limit theorem does not tell the whole story. There is a "generalized" version of the central limit theorem that states that either the Gaussian *or a heavy-tailed distribution* will emerge as the limit of sums of random variables. Unfortunately, the technical nature of this result means it rarely features in introductory courses, which leads to unnecessary surprises about the presence of heavy-tailed distributions. Going beyond sums of random variables, when random variables are combined in other natural ways (e.g., products or max/min) heavy tails are even more likely to emerge, whereas the Gaussian distribution is not.

So heavy-tailed phenomena should not be considered surprising. What about mysterious? The view of heavy tails as mysterious is, to some extent, a consequence of unfamiliarity. People are familiar with the Gaussian distribution because of its importance in introductory probability courses, and when something emerges that has qualitatively and quantitatively different properties it seems mysterious and counter-intuitive. The Pareto Principle is one illustration of the counterintuitive properties that make heavy-tailed distributions seem mysterious, but there are many others. For example, while the Gaussian distribution has a clear "scale" – most samples will be close to the mean – samples from heavy-tailed distributions frequently differ by orders of magnitude and may even be "scale free" (e.g., in the case of the Pareto distribution). Another example is that, while the moments (the mean, variance, etc.) of the Gaussian distribution are all finite, it is not uncommon to see data that fits a heavy-tailed distribution having an infinite variance, or even an infinite mean! For example, the degree distribution of many complex networks tends to have a tail that matches that of a Pareto with infinite variance (see, for example, [23]). This can potentially lead to mind-bending challenges when trying to apply statistical tools, which often depend on averages and variances.

The combination of surprise and mystery that surrounds heavy-tailed phenomena means that there is often considerable excitement that follows the discovery of data that fits a heavy-tailed distribution in a new field. Unfortunately, this excitement often sparks debate and controversy – often enough that an unfortunate pattern has emerged. A heavy-tailed phenomenon is discovered in a new field. The excitement over the discovery leads researchers to search for heavy tails in other parts of the field. Heavy tails are then discovered in many settings and are claimed to be a universal property. However, the initial excitement of discovery and lack of previous background in statistics related to heavy tails means that the first wave of research identifying heavy tails uses intuitive but flawed statistical tools. As a result, a controversy emerges – which settings where heavy tails have been observed really have heavy tails? Are they really universal? Over time, more careful statistical analyses are used, showing that some places really do exhibit heavy tails while others were false discoveries. By the end, a mature view of heavy tails emerges, but the whole process can take decades.

At this point, the pattern just described has been replicated in many areas, including computer science [68], biology [119], chemistry [160], ecology [10], and astronomy [216]. Maybe the most prominent example of this story is still ongoing in the area of *network science*. Near the turn of the century, the study of complex networks began to explode in popularity due to the growing importance of networks in our lives and the increasing ease of gathering data about large networks. Initial results in the area were widely celebrated and drove an enormous amount of research to look at the universality of scale-free networks. However, as the field matured and the statistical tools became more sophisticated, it became clear that many of the initial results were flawed. For example, claims that the internet graph [80] and the power network [24] are heavy-tailed were refuted [4, 222], among others. This led to a controversy in the area that continues to this day, 20 years later [37, 212].

### *Demystifying Heavy Tails*

The goal of this book is to demystify heavy-tailed phenomena. Heavy tails are not anomalies – and their emergence should not be surprising or controversial either! Heavy tails are an unavoidable part of our lives, and viewing statistics like the ones that started this chapter as anomalies prevents us from thinking clearly about the world around us. Further, while properties of heavy-tailed phenomena like the Pareto Principle may initially make heavy-tailed distributions seem counterintuitive, they need not be. This book strives to provide tools and techniques that can make heavy tails as easy and intuitive to reason about as the Gaussian, to highlight when one should expect the emergence of heavy-tailed phenomena, and to help avoid controversy when identifying heavy tails in data.

Because of the ubiquitousness and seductive nature of heavy-tailed phenomena, they are a topic that has permeated wide ranging fields, from astronomy and physics, to biology and physiology, to social science and economics. However, despite their ubiquity, they are also, perhaps, one of the most misused and misunderstood mathematical areas, shrouded in both excitement and controversy. It is easy to get excited about heavy-tailed phenomena as you start to realize the important role they play in the world around us and become exposed to the beautiful and counterintuitive properties they possess. However, as you start to dig into the topic, it quickly becomes difficult. The mathematics that underlie the analysis of heavy-tailed distributions are technical and advanced, often requiring prerequisites of graduate-level probability and statistics courses. This is the reason why introductory probability courses typically do not present much, if any, material related to heavy-tailed distributions. If they are mentioned, they are typically used as examples illustrating that "strange" things can happen (e.g., distributions can have an infinite mean). Thus, a scientist or researcher in a field outside of mathematics who is interested in learning more about heavy tails may find it difficult, if not impossible, to learn from the classical texts on the topic.

It is exactly this difficulty that led us to write this book. In this book we hope to introduce the fundamentals of heavy-tailed distributions using only tools that one learns in an introductory probability course. The book intentionally does not spend much time on describing the settings where heavy tails arise – there are simply too many different areas to do justice to even a small subset of them. Instead, we assume that if you have found your way to this book, then heavy tails are important to you. Given that, our goal is to provide an introduction to how to think about heavy tails both intuitively and mathematically.

The book is divided into three parts, which focus on three foundational guiding questions.

- **Part I: Properties.** *What leads to the counterintuitive properties of heavy-tailed phenomena?*
- **Part II: Emergence.** *Why do heavy-tailed phenomena occur so frequently in the world around us?*
- **Part III: Estimation.** *How can we identify and estimate heavy-tailed phenomena using data?*

In Part I of the book we provide insight into some of most mysterious and elegant properties of heavy-tailed distributions, connecting these properties to formal definitions of subclasses of heavy-tailed distributions. We focus on three foundational properties: "scale-invariance" (aka, scale-free), the "catastrophe principle," and "increasing residual life." We illustrate that these properties provide qualitatively different behaviors than what is seen under light-tailed distributions like the Gaussian, and provide intuition underlying the properties. The three chapters that make up Part I strive to demystify some of the particularly exotic properties of heavy-tailed distributions and to provide a clear view of how these properties interact with each other and with the broader class of heavy-tailed distributions.

In Part II of the book we explore simple laws that can "explain" the emergence of heavy-tailed distributions in the same way that the central limit theorem "explains" the prominence of the Gaussian distribution. We study three foundational stochastic processes in order to understand when one should expect the emergence of heavy-tailed distributions as opposed to light-tailed distributions. Our discussions in the three chapters that make up Part II highlight that heavy-tailed distributions should not be viewed as anomalies. In fact, heavy tails should not be surprising at all; in many cases they should be treated as something as natural as, if not more natural than, the emergence of the Gaussian distribution.

In Part III of this book we focus on the statistical tools used for the estimation of heavy-tailed phenomena. Unfortunately, there is no perfect recipe for "properly" detecting and estimating heavy-tailed distributions in data. Our treatment, therefore, seeks to highlight a handful of important approaches and to provide insight into when each approach is appropriate and when each may be misleading. Combined, the chapters that make up Part III highlight a crucial point: one must proceed carefully when estimating heavy-tailed phenomena in real-world data. It is naive to expect to estimate *exact* heavy-tailed distributions in data. Instead, a realistic goal is to estimate the *tail* of heavy-tailed phenomena. Even in doing this, one should not rely on a single method for estimation. Instead, it is a necessity to build confidence through the use of multiple, complementary estimation approaches.

## 1.1 Defining Heavy-Tailed Distributions

Before we tackle our guiding questions, we start with the basic question: *What is a heavy-tailed distribution?*

One of the reasons for the mystique that surrounds heavy-tailed distributions is that if you ask five people from different communities this question, you are likely to get five different answers. Depending on the community, the term heavy-tailed may be used interchangeably with terms like scale-free, power-law, fat-tailed, long-tailed, subexponential, self-similar, stable, and others. Further, the same names may mean different things to different communities!

Sometimes the term "heavy-tailed" is used to refer to a specific distribution such as the Pareto or the Zipf distribution. Other times, it is used to identify particular properties of a distribution, such as the fact that it is scale-free, has an infinite (or very large) variance, a decreasing failure rate, and so on. As a result, there is often a language barrier when discussing heavy-tailed distributions that stems from different associations with the same terms across communities.

Hopefully, reading this book will equip you to navigate the zoo of terminology related to heavy-tailed distributions. Each of the terms mentioned earlier does have a concrete, precise, established mathematical definition. It is just that these terms are often used carelessly, which leads to confusion. It will take us most of the book to get through the definitions of all the terms mentioned in the previous paragraph, but we start in this section by laying the foundation – defining the term "heavy-tailed" and discussing some of the most celebrated examples.

The term "heavy-tailed" is inherently relative – heavier than what? A Gaussian distribution has a heavier tail than a Uniform distribution, and an Exponential distribution has a heavier tail than a Gaussian distribution, but neither of these is considered "heavy-tailed." Thus, the key feature of the definition is the comparison point chosen.

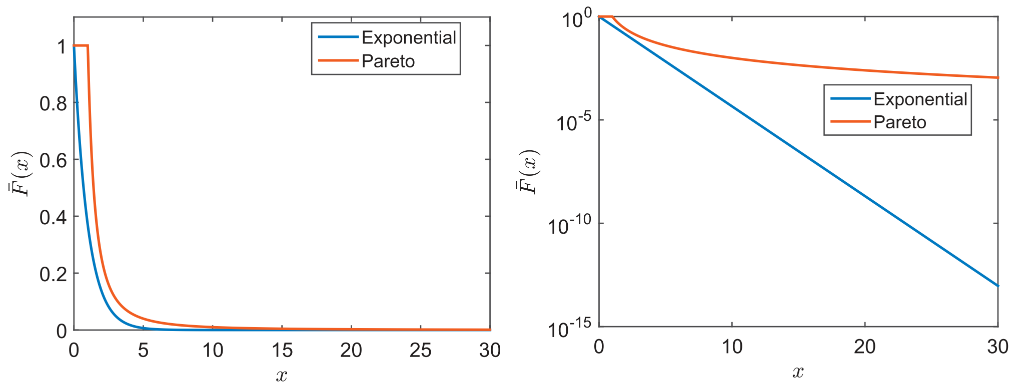
The comparison point that is used to define the class of heavy-tailed distributions is the Exponential distribution. That is, a distribution is considered to be heavy-tailed if it has a heavier tail than any Exponential distribution. Formally, this is stated in terms of the cumulative distribution function (c.d.f.) $F$ of a random variable $X$, that is, $F(x) = \Pr(X \leq x)$, and the complementary cumulative distribution function (c.c.d.f.) $\bar{F}$, that is, $\bar{F}(x) = 1 - F(x)$.

**Definition 1.1** A distribution function $F$ is said to be heavy-tailed if and only if, for all $\mu > 0$,

$$\limsup_{x \to \infty} \frac{1 - F(x)}{e^{-\mu x}} = \limsup_{x \to \infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \infty.$$

Cambridge University Press & Assessment
978-1-316-51173-2 — The Fundamentals of Heavy Tails
Jayakrishnan Nair , Adam Wierman , Bert Zwart
Excerpt
More Information






(a) Linear scale (b) Log scale

**Figure 1.1** Contrasting heavy-tailed and light-tailed distributions: The plots show the c.c.d.f. of the exponential distribution (with mean 1) and a heavy-tailed Pareto distribution (with minimal value $x_m = 1$, scale parameter $\alpha = 2$). While the contrast in tail behavior is difficult to discern on a linear scale (Fig. $(a)$), it is quite evident when the probabilities are plotted on a logarithmic scale (Fig. $(b)$).

Otherwise, $F$ is light-tailed. A random variable $X$ is said to be heavy-tailed (light-tailed) if its distribution function is heavy-tailed (light-tailed).

Note that the definition of heavy-tailed distributions given above applies to the *right* tail of the distribution, that is, it is concerned with the behavior of the probability of taking values larger than $x$ as $x \to \infty$. In some applications, one might also be interested in the *left* tail. In such cases, the definition of heavy-tailed can be applied to both the right tail (without change) and the left tail (by considering the right tail of $-X$).

The definition of heavy-tailed is, in some sense, natural. It looks explicitly at the "tail" of the distribution (i.e., the c.c.d.f. $\bar{F}(x)$), and it is easy to see from the definition that the tails of distributions that are heavy-tailed are "heavier" (i.e., decay more slowly) than the tails of distributions that are light-tailed; see Figure 1.1.

The particular choice of the Exponential distribution as the boundary between heavy-tailed and light-tailed may, at first, seem arbitrary. In fact, without detailed study of the class of heavy-tailed distributions, it is difficult to justify this particular choice. But, as we will see throughout this book, the Exponential distribution serves to separate two classes of distributions that have qualitatively different behavioral properties and require fundamentally different mathematical tools to work with.

To begin to examine the distinction between heavy-tailed and light-tailed distributions, it turns out to be useful to consider two alternative, but equivalent, definitions of "heavy-tailed."

**Lemma 1.2** *Consider a random variable $X$. The following statements are equivalent.*

*(i)* $X$ *is heavy-tailed.*
*(ii)* *The moment generating function* $M(s) := \mathbb{E}\left[e^{sX}\right] = \infty$ *for all* $s > 0$.
*(iii)* $\liminf_{x\to\infty} -\frac{\log \Pr(X>x)}{x} = 0$.

The proof of this lemma provides useful intuition about heavy-tailed distribution; however, before proving this result, let us interpret the two new, equivalent definitions of heavy-tailed that it provides.

First, consider *(ii)*, which states that a random variable is heavy-tailed if and only if its moment generating function $M(s) := \mathbb{E}\left[e^{sX}\right]$ is infinite for all $s > 0$. This definition highlights that heavy-tailed distributions require a different analytic approach than light-tailed distributions. For light-tailed distributions the moment generating function often provides an important tool for characterizing the distribution. It can be used to derive the moments of the distribution, but it also can be inverted to characterize the distribution itself. Further, it is a crucial tool for analysis because of the simplicity of handling convolutions via the moment generating function, for example, when deriving concentration inequalities such as Chernoff bounds. In contrast, the definition given by *(ii)* shows that such techniques are not applicable for heavy-tailed distributions.

Next, consider *(iii)*, which states that a random variable $X$ is heavy-tailed if and only if the log of its tail, $\log \Pr(X > x)$, decays sublinearly. This again highlights that heavy-tailed distributions require a different analytic approach than light-tailed distributions. In particular, when studying the tail of light-tailed distributions it is common to use concentration inequalities such as Chernoff bounds, which inherently have an exponential decay. As a result, such bounds focus on determining the optimal decay rate, which is characterized by deriving a maximal $\mu$ such that $\Pr(X > x) \leq Ce^{-\mu x}$. However, the definition given by *(iii)* highlights that the maximum possible $\mu$ for heavy-tailed distributions is zero, and so fundamentally different analytic approaches must be used.

To build more intuition on the relationship between these three equivalent definitions of "heavy-tailed," as well as to get practice working with the definitions, it is useful to consider the proof of Lemma 1.2.

*Proof of Lemma 1.2* To prove Lemma 1.2, we need to show the equivalence of each of the three definitions of heavy-tailed. We do this by showing that *(i)* implies *(ii)*, that *(ii)* implies *(iii)*, and finally that *(iii)* implies *(i)*.

**$(i) \Rightarrow (ii)$.** Suppose that $X$ is heavy-tailed, with distribution $F$. By definition, this implies that for any $s > 0$, there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ satisfying $\lim_{k\to\infty} x_k = \infty$, such that

$$\lim_{k\to\infty} e^{sx_k}\bar{F}(x_k) = \infty. \tag{1.1}$$

We can now bound $\mathbb{E}\left[e^{sX}\right]$ as follows.

$$\begin{aligned}\mathbb{E}\left[e^{sX}\right] &= \int_0^\infty e^{sx}dF(x)\\ &\geq \int_{x_k}^\infty e^{sx}dF(x)\\ &\geq e^{sx_k}\bar{F}(x_k).\end{aligned}$$

Since the above inequality holds for all $k$, it now follows from (1.1) that $\mathbb{E}\left[e^{sX}\right] = \infty$. Therefore, Condition *(i)* implies Condition *(ii)*.

**(ii) ⇒ (iii).** Suppose that $X$ satisfies Condition *(ii)*. For the purpose of obtaining a contradiction, let us assume that Condition *(iii)* does not hold. Since $-\frac{\log \Pr(X>x)}{x} \geq 0$, this means that

$$\liminf_{x\to\infty} -\frac{\log \Pr(X > x)}{x} > 0.$$

The above statement implies that there exist $\mu > 0$ and $x_0 > 0$ such that

$$-\frac{\log \Pr(X > x)}{x} \geq \mu \iff \Pr(X > x) \leq e^{-\mu x} \quad \forall\, x \geq x_0. \tag{1.2}$$

Now, pick $s$ such that $0 < s < \mu$. We may now bound the moment generating function of $X$ at $s$ as follows:

$$\begin{aligned} M(s) = \mathbb{E}\left[e^{sX}\right] &= \int_0^\infty \Pr\left(e^{sX} > x\right) dx \\ &= \int_0^{e^{sx_0}} \Pr\left(e^{sX} > x\right) dx + \int_{e^{sx_0}}^\infty \Pr\left(X > \frac{\log(x)}{s}\right) dx. \end{aligned}$$

Here, we have used the following representation for the expectation of a nonnegative random variable $Y$: $\mathbb{E}[Y] = \int_0^\infty \Pr(Y > y)\, dy$. While the first term above can be bounded from above by $e^{sx_0}$, we may bound the second using (1.2), since $x \geq e^{sx_0}$ is equivalent to $\log(x)/s \geq x_0$.

$$\begin{aligned} M(s) &\leq e^{sx_0} + \int_{e^{sx_0}}^\infty e^{-\mu \frac{\log(x)}{s}} dx \\ &= e^{sx_0} + \int_{e^{sx_0}}^\infty x^{-\mu/s} dx. \end{aligned}$$

Since $\mu/s > 1$, we have $\int_1^\infty x^{-\mu/s} dx < \infty$, which implies that $M(s) < \infty$, giving us a contradiction. Therefore, Condition *(ii)* implies Condition *(iii)*.

**(iii) ⇒ (i).** Suppose that the random variable $X$, having distribution $F$, satisfies Condition *(iii)*. Thus, there exists a strictly increasing sequence $(x_k)_{k\geq 1}$ satisfying $\lim_{k\to\infty} x_k = \infty$, such that

$$\lim_{k\to\infty} -\frac{\log \bar{F}(x_k)}{x_k} = 0.$$

Given $\mu > 0$, this in turn implies that there exists $k_0 \in \mathbb{N}$ such that

$$\begin{aligned} &-\tfrac{\log \bar{F}(x_k)}{x_k} < e^{-\frac{\mu}{2}} && \forall\, k > k_0 \\ \iff\ &\bar{F}(x_k) > e^{-\frac{\mu x_k}{2}} && \forall\, k > k_0 \\ \iff\ &\tfrac{\bar{F}(x_k)}{e^{-\mu x_k}} > e^{\frac{\mu x_k}{2}} && \forall\, k > k_0. \end{aligned}$$

The last assertion above implies that $\lim_{k\to\infty} \frac{\bar{F}(x_k)}{e^{-\mu x_k}} = \infty$, which implies $\limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \infty$. Since this is true for any $\mu > 0$, we conclude that Condition *(iii)* implies Condition *(i)*.

□

## 1.2 Examples of Heavy-Tailed Distributions

We now have three equivalent definitions of heavy-tailed distributions and, through the proof, we understand how these three definitions are related. But, even with these restatements, the definition of heavy-tailed is still opaque. It is difficult to get behavioral intuition about the properties of heavy-tailed distributions from any of the definitions. Further, it is very hard to see much about what makes heavy-tailed distributions have the mysterious properties that are associated with them using these definitions alone.

In part, this is due to the breadth of the definition of heavy-tailed. The important properties commonly associated with heavy-tailed distributions, such as scale invariance, infinite variance, the Pareto principle, etc., do not hold for all heavy-tailed distributions; they hold only for certain subclasses of heavy-tailed distributions.

As a result, it is important to build intuition for the class of heavy-tailed distributions by looking at specific examples. That is the goal of the remainder of this chapter. In particular, we focus in detail on the Pareto distribution, the Weibull distribution, and the LogNormal distribution with the goal of providing both the mathematical formalism for these distributions and some insight in their important properties and applications. Additionally, we briefly introduce some of the other important examples of heavy-tailed distributions that come up frequently in applications, including the Cauchy, Fréchet, Lévy, Burr, and Zipf distributions.

Perhaps the most important thing to keep in mind as you read these sections is the contrast between the properties of the heavy-tailed distributions that we discuss and the properties of light-tailed distributions, such as the Gaussian and Exponential distributions, with which you are likely more familiar. To set the stage, we summarize the important formulas for these two distributions next.

### *The Gaussian Distribution*

The Gaussian distribution, also called the Normal distribution or the bell curve, is perhaps the most widely recognized distribution and is extremely important in statistics and beyond. It is defined using two parameters, the mean $\mu$ and the variance $\sigma^2$, and is expressed most conveniently through its probability density function (p.d.f.), $f(x)$, or its moment generating function (m.g.f.), $M(s)$. Given a random variable $Z \sim \text{Gaussian}(\mu, \sigma)$, we have

$$f_Z(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$
$$M_Z(s) = E[e^{sZ}] = e^{\mu s + \frac{1}{2}\sigma^2 s^2}.$$

Since $M_Z(s) < \infty$ for all $s > 0$, it follows that the Gaussian distribution is light-tailed. The light-tailedness of the Gaussian distribution can also be deduced directly by bounding its c.c.d.f. (see Exercise 2).

The particular Gaussian distribution with zero mean and unit variance ($\mu = 0$, $\sigma = 1$) is commonly referred to as the *standard Gaussian*.

### *The Exponential Distribution*

The Exponential distribution is a widely known and broadly applicable distribution that serves as the light-tailed distribution on the boundary between light-tailed and heavy-tailed distributions. It is a nonnegative distribution defined in terms of one parameter: $\lambda$, which is referred to as the "rate" since the mean of the distribution is $1/\lambda$. Given a random variable $X \sim \text{Exponential}(\lambda)$, the p.d.f., c.c.d.f., and m.g.f., can be expressed as

$$f_X(x) = \lambda e^{-\lambda x} \qquad (x \geq 0),$$
$$\bar{F}(x) = e^{-\lambda x} \qquad (x \geq 0),$$
$$M_X(s) = \frac{1}{(1 - s/\lambda)} \qquad (s < \lambda).$$

Note that the tail of the Exponential distribution is heavier than that of the Gaussian because $e^{-x}$ goes to zero more slowly than $e^{-x^2}$. Additionally, unlike the Gaussian, the moment generating function is not finite everywhere.

### *1.2.1 The Pareto Distribution*

Vilfredo Pareto originally presented the Pareto distribution, and introduced the idea of the Pareto Principle, in the study of the allocation of wealth. But since then, it has been used as a model in numerous other settings, including the sizes of cities, the file sizes in computer systems and networks, the price returns of stocks, the size of meteorites, casualties and damages due to natural disasters, frequency of words, and many more. It is perhaps the most celebrated example of a heavy-tailed distribution, and as a result, the term Pareto is sometimes, unfortunately, used interchangeably with the term heavy-tailed.

Formally, a random variable $X$ follows a Pareto$(x_m, \alpha)$ distribution if

$$\Pr(X \geq x) = \bar{F}(x) = \left(\frac{x}{x_m}\right)^{-\alpha}, \text{ for } \alpha > 0, x \geq x_m > 0.$$

Here, $\alpha$ is the shape parameter of the distribution and is also commonly referred to as the *tail index*, while $x_m$ is the minimum value of the distribution, that is, $X \geq x_m$. Given the c.c.d.f. above, it is straightforward to differentiate and obtain the p.d.f.

$$f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}, \quad x \geq x_m.$$

It is easy to see from the c.c.d.f. that the Pareto is heavy-tailed. In particular, using Definition 1.1, we can compute

$$\limsup_{x\to\infty} \frac{\bar{F}(x)}{e^{-\mu x}} = \limsup_{x\to\infty} \left(\frac{x_m}{x}\right)^\alpha e^{\mu x} = \infty, \tag{1.3}$$

since the exponential $e^{\mu x}$ grows more quickly than the polynomial $x^\alpha$.

This highlights the key contrast between the Pareto distribution and common light-tailed distributions like the Gaussian and Exponential distributions: the Pareto tail decays *polynomially*, as $x^{-\alpha}$, instead of *exponentially* (as $e^{-\mu x}$) in the case of the Exponential, or *superexponentially* (as $e^{-x^2/2\sigma^2}$) in the case of the Gaussian. As a consequence, large values are much more likely to occur under a Pareto distribution than under a Gaussian or Exponential distribution. For example, you are much more likely to meet someone whose income is 10 times the average than someone whose height is 10 times the average.

This contrast is present visually too. Figure 1.2 shows that the tail of the Pareto is considerably heavier. The figure illustrates the p.d.f. and c.c.d.f. of the Pareto for different values of