# E7

# Rank-size plots, Zipf's law, and scaling

✦ **Abstract.** Rank-size plots, also called Zipf plots, have a role to play in representing statistical data. The method is somewhat peculiar, but throws light on one aspect of the notions of concentration. This chapter's first goals are to define those plots and show that they are of two kinds. Some are simply an analytic restatement of standard tail distributions but other cases stand by themselves. For example, in the context of word frequencies in natural discourse, rank-size plots provide the most natural and most direct way of expressing scaling.

Of greatest interest are the rank-size plots that are rectilinear in log-log coordinates. In most cases, this rectilinearity is shown to simply rephrase an underlying scaling distribution, by exchanging its coordinate axes. This rephrasing would hardly seem to deserve attention, but continually proves its attractiveness. Unfortunately, it is all too often misinterpreted and viewed as significant beyond the scaling distribution drawn in the usual axes. These are negative but strong reasons why rank-size plots deserve to be discussed in some detail. They throw fresh light on the meaning and the pitfalls of infinite expectation, and occasionally help understand upper and lower cutoffs to scaling. ✦

THIS LARGELY SELF-CONTAINED CHAPTER covers a topic that goes well beyond finance and economics and splits into two distinct parts. Hence, the points to be made are best expressed in terms of two definite and concrete contexts. The bulk is written in terms of "firm sizes," as measured by sales or number of employees, but would be unchanged if firm sizes were replaced by such quantities as city populations. The second context to be invoked, word frequencies, warrants a digression from this book's thrust , if only because the straightness of a log-log rank-size plot is explained most readily and simply in that context.

*Restatement of the probabilists' notation*. A capital letter, say $U$, denotes a quantity whose value is random, for example the height of man or the size of an oil reservoir selected at random on the listing of the data. The corresponding lower case letter, say $u$, denotes the sample value, as measured in numbers of inches or in millions of barrels.

## 1. INTRODUCTION

### 1.1 Rank-size plots for concrete quantities

A concrete random variable is a quantity that is measured on an "extrinsic" or "physical" scale. Humans are measured by *height*, firms by sales or numbers of employees, and cities by *numbers* of inhabitants. More generally, statistical quantities such as "height" and "number of inhabitants" are originally defined in a non-stochastic context. Their physical scale serves to rank those random variables by increasing or decreasing value, through either $F(u) = \Pr\{U \leq u\}$ or the tail distribution $P(u) = 1 - F(u)$.

The concrete reality that underlies the notions of $F(u)$ and $P(u)$ can be also represented in the following alternative fashion. The first step is to *rank* the elements under investigation by decreasing height, size, and number. The largest item will be indexed as being of rank $r = 1$; the largest of the remaining items will be of rank $r = 2$, and so on. The second step is to specify size, or any other suitable quantity $Q$, as a function of rank. One way to specify the distribution of a random quantity is to specify the corresponding function $Q(r)$.

By definition, $Q(r)$ varies inversely with $r$ : it decreases as $r$ increases. Granted the possibility of more than one item of equal size, $Q(r)$ must be non-increasing. This is the counterpart of the fact that $F(u)$ and $P(u)$ are non-decreasing and non-increasing, respectively.

Special interest attaches to the positive scaling case, when the assertion that $Q$ varies *inversely* with $r$ can be strengthened to the assertion that $Q$ is *proportional to the inverse* of $r$, or perhaps that $\log Q$ varies *linearly* with $\log r$. Unfortunately, some careless rank-size studies confuse different meanings of "inverse variation."

### 1.2 "Static" rank-frequency plots in the absence of an extrinsic scale

The occurrence of a word in a long text is not accompanied by anything like "a human's height" or "a city's number of inhabitants". But there is a simple and beautiful way out. Even when extrinsic "physical" quantities are not present, every random event involves at least one intrinsic quantity: it is the event's own probability.

Thus, in the case of word frequencies, rank-size does not involve the usual functions $F(u)$ and $P(u)$, but begins with a function $Q(r)$ that gives the probability of the word whose rank is $r$ in the order of decreasing probabilities. To some authors, this looks like a snake biting it's tail, but the paradox is only apparent and the procedure is quite proper. In the scaling case, $\log Q$ varies linearly with $\log r$.

Furthermore, this ranking happens to be justified a posteriori in the theory of word frequencies introduced in M 1951, sketched in Section 1.2.4 of Chapter E8, and developed in M 1961b. That theory introduces a quantity that is always defined and often has desirable additivity properties similar to those of "numbers of inhabitants;" it is the function $-\log p$, where $p$ is a word's probability. By introducing $-\log p$, the ranking based on frequency is reinterpreted as conventional ranking based on $-\log p$ viewed as an intrinsic random variable. In practice, of course, one does not know the probability itself, only an estimate based upon a sample frequency.

There are strong reasons to draw attention to a wide generalization of my derivation of the law of word frequencies. One reason is that it may bear on the problem of city population via a reinterpretation of the central place theory. A second reason is that this generalization involves a phenomenon described in the next section, namely a built-in crossover for low ranks, that is, frequent words and large city population. The reader interested in the derivation is referred to M 1995f, and the reader prepared to face an even more general but old presentation is referred to M 1955b.

### 1.3 Distinction between the terms, Zipf distribution and Zipf law

The term "Zipf law" is used indiscriminately, but the concepts behind this terms *distribution* and *law* are best kept apart. The fairest terminology seems to be the following one.

*Zipf distribution* will denote all instances of rank-size relation $Q(r)$ such that, with a suitable "prefactor" $\Phi$, the expression

$$Q(r) \sim \Phi r^{-1/\alpha}$$

is valid over an intermediate range of values of $r$, to be called *scaling range.* This range may be bounded by one or two crossovers, $r_{\min} \geq 1$ to $r_{\max} \leq \infty$, to which we return in Section 1.7. Allowing crossovers automatically allows all values of $\alpha > 0$. When $\alpha < 1$, the scaling range need not, but can, extend to $r \to \infty$ with no crossovers; when $\alpha \geq 1$, the scaling range is necessarily bounded from above.

Zipf emphasized the special case $\alpha = 1$. If so, $Q(r)$ does not only vary inversely with $r$ but varies in inverse proportion to $r$. In the special case of word frequencies, Zipf asserted $\alpha = 1$ and $\Phi = 1/10$, which are very peculiar values that demand $r_{\max} < \infty$.

*Zipf law* will denote all empirical cases when the Zipf distribution is found to hold.

### 1.4 Zeta and truncated zeta distributions

"Zeta" and "truncated zeta" distributions are the terms to be used to denote exact statements valid for all values of $r$.

*The zeta distribution.* When $\alpha < 1$, hence $1/\alpha > 1$, the function

$$\zeta(1/\alpha) = \sum_{s=1}^{\infty} s^{-1/\alpha}.$$

is the mathematicians' Riemann zeta function. This suggests "zeta distribution" to denote the one-parameter discrete probability distribution

$$p(r) = \frac{r^{-1/\alpha}}{\sum_{s=1} \text{to} \infty s^{-1/\alpha}} = \frac{r^{-1/\alpha}}{\zeta(1/\alpha)} = \Phi r^{-1/\alpha}.$$

In the coordinates $\log r$ and $\log p(r)$, the zeta distribution plots as an exact straight line of slope $-1/\alpha$. Clearly,

$$\int_1^{\infty} s^{-1/\alpha} ds = \frac{\alpha}{1-\alpha} < \sum_{s=1}^{\infty} s^{-1/\alpha} < 1 + \int_1^{\infty} s^{-1/\alpha} ds = \frac{1}{1-\alpha} .$$

When $\alpha$ is near 1, the two bounds are close to each other.

Under the zeta distribution, the relative size of the largest firm is $\zeta^{-1}(1/\alpha)$. The joint share of the $r$ largest firms is

$$\zeta^{-1}(1/\alpha)\sum_{s=1}^{r} s^{-1/\alpha}.$$

The ratio: sum of sizes of firms of rank strictly greater than $r$, divided by the size of the $r$-th firm, is

$$\left\{\sum_{s=r+1}^{\infty} s^{-1/\alpha}\right\} r^{1/\alpha}.$$

As $r$ increases, the sum in braces becomes increasingly closer to the integral $\int_r^\infty x^{-1/\alpha}dx$, and the preceding ratio becomes

$$\frac{r^{(1-1/\alpha)}r^{1/\alpha}}{(1/\alpha-1)} = \frac{r\alpha}{(1-\alpha)}.$$

*Truncated zeta distribution*. When $\alpha < 1$ and $V > -1$, define

$$\zeta(1/\alpha, V) = \sum_{s=V+1}^{\infty} s^{-1/\alpha} = \sum_{1}^{\infty}(r+V)^{-1/\alpha}.$$

I use the term "truncated zeta distribution" to denote the two-parameter discrete probability distribution

$$p(r) = \frac{(r+V)^{-1/\alpha}}{\zeta(1/\alpha, V)} = \Phi(r+V)^{-1/\alpha}.$$

Plotted in the coordinates $\log r$ and $\log p(r)$, the tail is straight, of slope $-1/\alpha$, as in the truncated zeta distribution, but this tail is preceded, for small ranks, by an appreciable flattening that extends to values of $r$ equal to a few times $V$.

### 1.5 Dynamic evolution of a rank-size plot as the sample-size increases

The considerations in Sections 1.1 and 1.2 are called static because they concern a fixed sample. Some sort of dynamics enters if the rank-size plot is continually updated as data are drawn from this sample. Let us show that the examples of Section 1.1 and 1.2 behave very differently from that viewpoint. In other words, the commonality of structure that seems to be implied by the term Zipf law is misleading.

*Firms*. Create an increasing sample of firms from an industry by picking them at random. One approximation is to follow a list ordered lexicographically. As the sample develops, the largest firm will repeatedly change, and a given firm's rank will increase as new firms flow in. The rank-size plot will grow by its *low rank* end. Furthermore, however long a list of prices may be, it is certainly finite. Therefore, as the sample size increases, the straightness of the rank-size plot must eventually break down at the high-rank end. Additional reasons for breakdown will be examined in the next sub-section.

*Words*. By way of contrast, increase a sample of words by reading a scrambled text, or perhaps a book by James Joyce. The most probable word will soon establish and maintain itself and other words' rank will gradually settle down to those words' probabilities. Experience suggests that in most cases the number of distinct words is so extremely high, that fresh words keep being added as the sample increases. Therefore, the rank-size plot will grow at its *high rank* end.

### 1.6 Large estimated values of $\alpha$ are not reliable, hence not significant

The scaling range from $(r_{\min}, Q_{\max})$ to $(r_{\max}, Q_{\min})$. might be reported in the form of a "number of decades," defined as the decimal logarithm of either of two ratios, namely $\log_{10}(Q_{\max}/Q_{\min})$ or $\log_{10}(r_{\max}/r_{\min})$. When $\alpha \sim 1$, the two ratios are close to each other. When $\alpha$ is large and $1/\alpha$ is small, the two ratios differ significantly. One is tempted to report the larger of the two values, $\log_{10}(r_{\max}/r_{\min})$, but the proper value is the smaller. The reason is that the intrinsic quantity is not $r$ but $Q$. The issue is discussed in Chapter E3.

For example, consider the reports of phenomena for which $1/\alpha = 1/4$ holds over a seemingly convincing range of 2 decades in terms of $r$. Restated in terms of $Q$, this range reduces to an unconvincing one-half decade.

### 1.7 The many forms taken by the crossovers

The difference Section 1.5 draws between the cases of firms and words is essential from the viewpoint of the width of the scaling ranges from $(r_{\min}, Q_{\max})$ to $(r_{\max}, Q_{\min})$. Let us run through a few examples.

*Personal income*. Scaling was observed by Pareto and is discussed in several chapters of this book. But scaling breaks down for large values of the rank, because small incomes *do not* follow a scaling distribution. There is also an operational reason for breakdown: small incomes are neither defined nor reported with accuracy. As a result, the log-log rank-size plot is expected to cross-over for high values of $r$ into a near-vertical portion. Once again, however, and this is important to the discussion of incomes in this book, the evidence suggests that scaling holds for unboundedly large incomes, implying a straight log-log plot for small ranks. An exception is that the straightness is not expected to hold for $r = 1$, because, as Section 3.2 will show, the largest value $U(1, N)$ is expected to have extraordinarily high sample scatter.

*Firms*. This notion breaks down into artificiality and irrelevance for very small sizes, because of legal reasons to register or not to register.

*City sizes*. Both ends of the graph are affected by artificiality, for example by political boundaries that represent nothing worth studying quantitatively.

*Word frequencies*. As already mentioned, Section 1.2.4 of Chapter E8 describes my reasons why one should expect word frequencies to follow Zipf's law in the form $Q(r) \sim \Phi r^{-1/\alpha}$. But those reasons rely on limit theorem of probability and say nothing about small values of $r$. In general, the model yields unrelated values of $\alpha$ and $\Phi$, which fail to satisfy the equality $\Phi^{-1} = \zeta(1/\alpha)$ that is characteristic of the zeta distribution. When such is the case, a crossover is inevitable. One can *define* a correction factor $V$ by the relation

$$\Phi^{-1} = \zeta(1/\alpha, V),$$

and use as approximation the truncated zeta expression

$$Q(r) = \Phi(r + V)^{-1/\alpha}.$$

In the context of word frequencies, this relation is often referred to as the *Zipf-Mandelbrot law.*

*Summary*. All told, the expectation that one or both ends of the curve will cross over implies that the estimation of $\alpha$ must often neglect the values of very low or very high rank.

*Analytic expressions for the behavior of a non-scaling distribution beyond the scaling interval: limitations to their usefulness*. Many specialists in curve-fitting insist that one can account for crossovers by replacing a linear log, log plot, by the plot of a second-order polynomial. When the second order is not enough, one moves to a polynomial of higher order.

A different approach is suggested by a different tradition that is very strong both in physics and in economics (where it goes back to Pareto; see (Chapter E2, Section 3.3). In that tradition, one multiplies $r^{-1/\alpha}$ by $r^{-1/\alpha}\exp(-\beta r)$, or perhaps by a factor that varies more slowly than an exponential, such as $r^{-1/\alpha}/\log r$.

Those "all-purpose" traditional corrective terms may improve the fit, or broaden the range in which a single formula prevails. But they are not useful, in my judgement, and draw attention away from the impact of approximate straightness. The only corrective terms I find valuable are not those imitated from physics, but those suggested by theory.

### 1.8. The power of words and pictures

Words are powerful. Probabilists who now speak of *distribution* used to speak of *law*, which sounds or "feels" more impressive. "Scaling" distribution and "power-law distribution" are neutral terms that do not seek mystery and do not promise much in common between the various occurrences of scaling. By contrast, experience shows that "Zipf's law" is a repulsive magnet to professional students of randomness, but an attractive magnet for non-professional dabblers of all kind. The same is true of "$1/f$ noise," a term that necessity often forces me to both use and fight. Its near-synonym "self-affine function" makes no ringing statement, but experience proves that "$1/f$ noise" suggests a single underlying phenomenon, which happens to be very far off the mark.

*Zipf's law as attractor*. Zipf 1949 put forward the bold claim that scaling is the "norm" for all social phenomena, while for physical phenomena the "norm" is the Gaussian. His claims created quite a stir when I was a post-doc at MIT, in search for unusual facts to investigate.

In 1953, I gained durable praise from linguists for having shown that a straight rank-size plot for word frequencies is devoid of meaning for linguistics; there is nothing in it for syntax or semantics. However, Zipf's law proved interesting in probabilistic terms and (as told in Chapter 42 of M 1982F{*FGN*}) somehow started me on a path that led, first, to finance and economics, and eventually to fractals.

*Zipf's law as repeller*. Very different is the conventional conclusion, already mentioned in Chapter E4, that is recorded in Aitchison & Brown 1957. On pp. 101-2, we read that "A number of distributions are given by Zipf, who uses a mathematical description of his own manufacture on which he erects some extensive sociological theory; in fact, however, it is likely that many of these distributions can be regarded as lognormal, or truncated lognormal, with more prosaic foundations in normal probability theory." This statement proves two things: a) Aitchinson and Brown did not feel it necessary to check; b) they did not know what they were talking about. Few other technically competent persons knew.

As I write in 1997, the "bad vibes" that overselling had created in the nineteen fifties are forgotten, and Zipf's law is again oversold as a fresh and mysterious key to complexity or to a "linguistic" analysis of DNA structure. Those old dreams should crawl back in some hole.

## 2. FAST TRACK FROM A SCALING DISTRIBUTION TO A STRAIGHT RANK-SIZE PLOT

The themes of this section will be discussed again rigorously in Section 3.

### 2.1 From scaling to straight rank-size plots

The quantity $U$ is called *scaling* when one has the relation

$$\Pr\{U \geq u\} = \text{ probability that } U \geq u = P(u) \sim Fu^{-\alpha}.$$

$\alpha$ is called *scaling exponent*, and $F$ is a *numerical prefactor* that includes a *scale factor*. The sign ~ expresses that the relation is valid only for large values of $u$. Scaling does not exclude negative values of $u$, but this chapter does not dwell on them.

Assimilating the relative number of cases to a probability, a sample made of $N$ independent drawings from a scaling distribution yields

$$\mathrm{Nr}\,\{U \geq u\} \sim NFu^{-\alpha}.$$

The quantity Nr $\{U \geq u\}$ becomes the rank $r$ of an item in the ordering by decreasing frequency, population or income. Once again, the biggest firm has rank $r = 1$ and size $U(1, N)$, the second biggest has rank $r = 2$ and size $U(2, N)$, etc..

Plotting this expression on transparent paper and turning the sheet around the main diagonal of the axes will yield $u$ as function of $r$,

$$u(r, N) = F^{-1/\alpha} r^{-1/\alpha} N^{1/\alpha}.$$

Diagrams are not neutral, and different presentations of the same set of data emphasize one thing or another. The eye tends to be drawn to the top of a figure. In the plot of $\Pr\{U > u\} \sim u^{-\alpha}$, this position contains the many cases where $u$ is small, while the other cases hide in the tail. In the plot of $u \sim r^{-1/\alpha}$, the opposite is true. When the values of $u$ that matter most are the few largest ones, they are seen best in rank-size plots.

### 2.2 Relative size, the prefactor and criticality of the exponent $\alpha = 1$

Careful discussions of the rank-size relation consider the relative size

$$u_R(r, N) = \frac{\overline{u}(r, N)}{\sum_{s=1}^{N} \overline{u}(s, N)} .$$

We shall write

$$U_R(r, N) = \Phi r^{-1/\alpha}.$$

This formula involves a new prefactor $\Phi$ for which a numerical value is often reported with no comment. This implies the belief that $\Phi$ is independent of $N$. This strong statement is not obvious at all, in fact, it expresses a specific and unusual property. An essential role of this chapter is to tackle the case where the scaling range continues to $u_{\max} = \infty$, and to give necessary conditions for the prefactor $\Phi$ to be independent of $N$. One condition is that the exponent should satisfy $\alpha < 1$. Another condition is $1 < \alpha < 2$ combines with $EU = 0$. In all other scaling cases, we shall see that $\Phi$ is a decreasing function of $N$.

## 3. CAREFUL DERIVATION: FROM A SCALING DISTRIBUTION TO A STRAIGHT RANK-SIZE PLOT

This Section begins informally and becomes rigorous thereafter.

### 3.1 Typical absolute size as function of rank

A) Select the unit of "firm size" so that the tail distribution is

$$\Pr\{U \geq u\} = P(u) \sim u^{-\alpha},$$

and take a random sample of $N$ firms. A "typical value" of the number of firms larger than $u$ is the expectation

$$r(u, N) = Nu^{-\alpha}.$$

B) Exchange the role of variable and function, rank the firms in decreasing order of size, and define $U(r, N)$ as the size of the $r$-th largest firm in this ranking. Inverting the preceding function for a group of $N$ firms, the number of those of size $U(r, N)$ or larger, will "typically" be

$$r \sim NU(r, N)^{-\alpha}.$$

C) Draw $N$ firms independently from the same scaling distribution and rank them as in B). For given $r$, a "typical value" of $U(r, N)$ will be

$$\bar{u}(r, N) = (N/r)^{1/\alpha} = N^{1/\alpha} r^{-1/\alpha}.$$

### 3.2 Rigorous results replacing the "typical" values in Section 3.1

The standard statistical theory of extreme values confirms that, as the number of firms in an industry increases, the size of the largest increases proportionately to $N^{1/\alpha}$. The precise results are as follows.

*Theorem concerning weighting by* $N^{1/\alpha}$. (Many references, including Arov & Bobrov 1960, Formula 19). As $N \to \infty$, the sampling distribution of the ratio $U(r, N)N^{-1/\alpha}$ converges to the truncated gamma distribution

$$\lim_{N \to \infty} \Pr\left\{ \frac{U(r, N)}{N^{1/\alpha}} < x \right\} = \frac{1}{\Gamma(k)} \int_{x^{-\alpha}}^{\infty} z^{r-1} e^{-z} dz.$$

The most probable value of $U^{-\alpha}(r, N)/N$ is $r - 1$, giving some legitimacy to $U \sim (r-1)^{-1/\alpha}$. More importantly,

$$\lim_{N \to \infty} E\left\{ \frac{[U(r, N)]^q}{N^{q/\alpha}} \right\} = \frac{\Gamma(r - q/\alpha)}{\Gamma(r)} \text{ when } r > \frac{q}{\alpha}, \text{ and } = \infty \text{ otherwise }.$$

*Double asymptotics*. As $N \to \infty$ and $r \to \infty$, the Stirling formula yields

$$E\{[U(r, N)]^q N^{-q/\alpha}\} \sim r^{-q/\alpha}.$$

For $q=1$, this Stirling approximation for $r\to\infty$ agrees with the "typical value" $\overline{u}(r,N)$. Moreover, as $N\to\infty$ and $r\to\infty$, the variability factor $EU^2/(EU)^2-1$ tends to 0; more generally, $U(r,N)N^{-1/\alpha}$ becomes for all practical purposes non-random. This was implicitly taken for granted in the heuristic argument of Section 2, and has now been justified.

*Preasymptotic behavior*. The Stirling formula is not a good approximation until very large values of $r$ are reached.

When $\alpha$ is estimated from the low range portion of the plot, there is a clear statistical bias. It is due to averaging of $U(r,N)$ for fixed $r$, therefore represents a self-inflicted complication due to the use of rank-size plots.

The fact that $E\{U(r,N)\}=\infty$ for $r<1/\alpha$ is unfortunate; it is avoided using the following result.

*Theorem concerning weighting by* $U(1,N)$. (Arov & Bobrov 1960, formula 21) As $N\to\infty$, the sampling distribution of the ratio $U(r,N)/U(1,N)$ converges to

$$\lim_{N\to\infty}\Pr\left\{\frac{U(r,N)}{U(1,N)}<x\right\}=1-(1-x^{\alpha})^{r-1}.$$

It follows that

$$\lim_{N\to\infty}E\left\{\left[\frac{U(r,N)}{U(1,N)}\right]^q\right\}=\frac{\Gamma(1+q/\alpha)\Gamma(r)}{\Gamma(r+q/\alpha)}$$

As $r\to\infty$,

$$\lim_{N\to\infty}E\left\{\left[\frac{U(r,N)}{U(1,N)}\right]^q\right\}\sim\Gamma\left(1+\frac{q}{\alpha}\right)r^{-q/\alpha}.$$

For $q=1$, this formula agrees with the ratio of "typical values" $\overline{u}(r,N)/\overline{u}(1,N)$, except for the prefactor $\Gamma(1+1/\alpha)$ which is greater than 1, and implies that $U(r,N)/U(1,N)$ remains scattered even when it is large. Since the variability of $U(r,N)$ tends to 0 as $r\to\infty$, the variability of $U(r,N)/U(1,N)$ solely reflects the scatter of $N^{1/\alpha}/U(1,N)$. Of course, the moments of $N^{1/\alpha}/U(1,N)$ follow from the fact that $N^{-1/\alpha}U(1,N)$ follows the Fréchet distribution $\Pr\{X<x\}=\exp(-x^{\alpha})$.

Clearly,

$$\lim_{N\to\infty}\Pr\left\{\left[\frac{U(r,N)}{U(1,N)}\right]^{\alpha} > y\right\} = (1-y)^{r-1},$$

hence,

$$\lim_{N\to\infty}E\left\{\frac{U(r,N)}{U(1,N)}\right\}^{\alpha} = \frac{1}{r}\ .$$

This means that $E\{U(1,N)^{-\alpha}\}$ is near 1. The sizes of firms of low rank are very sample dependent, hence are not necessarily close to their typical values. To avoid this variability, it is best to take a different point of comparison.

*Weighting $U(r,N)$ by the cumulative size of the firms of rank higher than $r$.* From the rank-size argument, the ratio of the sizes of the $r'$ largest firms and the $r''$ largest firms is approximately equal to

$$\frac{1+\cdots s^{-1/\alpha}+\cdots r'^{-1/\alpha}}{1+\cdots s^{-1/\alpha}+\cdots r''^{-1/\alpha}}\ .$$

This expression is the same as for the zeta distribution. It varies continuously with $\alpha$; for $\alpha$ near one, and large $r'$ and $r''$, its order of magnitude is $\log r'/\log r''$.

### 3.3 Additional considerations

*Logarithmic plots.* Log-log plots involve the expectation of $\log[U(r,N)]/\log[U(1,N)]$ rather than of $[U(r,N)/U(1,N)]^{\alpha}$. This change brings no difficulty as long as $r$ is not too small: $U(r,N)$ clusters tightly around its own expectation, which validates the approximation

$$E\left\{\frac{\log U(1,N)}{\log U(r,N)}\right\} \sim \frac{E[\log U(1,N)]}{E[\log U(r,N)]} = \frac{EV(1,N)}{EV(r,N)}\ ,$$

where $V(r,N)$ is $r$-th largest among $N$ exponential variables $V=\log_e U$.

*Visual estimation of $\alpha$ from the rank-size plot on doubly logarithmic paper.* Despite the encouraging values of the various expected values reported in this section, the small $r$ sampling distributions of $U(r,N)/U(1,N)$ are usually too scattered for complete statistical comfort. As the term of comparison, it is better *not* to use the size of the largest firm but rather a firm of rank as large as practical; the larger, the safer. This feature bears on the problem of the estimation of $\alpha$. The usual procedure is to fit a straight line to the tail of $\log U(r, N)$ considered as a function of $\log r$, and to measure the slope of that line. When this is done, the points of rank 1, 2 or 3 are too sample-dependent, and should be given little weight. The resulting informal procedure can be approximated in several stages.

The *first* approximation would be to choose two values of $r$ (say $r'' = 5$ and $r' = 20$ ), and draw a line through the corresponding points on a doubly logarithmic graph; the sampling distribution of this estimator of alpha could be derived from the second theorem of this section.

A *second* approximation is to choose two couples $(r', r'')$ and fit a straight line to 4 points. The sampling distribution would no longer be known exactly because the $U(r, N)$ are so defined that they do not provide independent information about $\alpha$, but the precision of estimation naturally increases with the number of sampling points. The commonly practiced visual fitting amounts to weighting the estimates corresponding to various couples $(r', r'')$, thus eliminating automatically the outlying estimates and averaging the others. It would be desirable to formalize this procedure and informal visual fitting should be studied more carefully, but it does not deserve its shady reputation.

### 3.4 Total industry size when $U > 0$ : contrast between the cases $\alpha > 1$ (hence $EU < \infty$) and $\alpha < 1$ (hence $EU = \infty$)

The size of the industry is the sum of the sizes of the $N$ firms it contains, $\sum_{s=1}^{N} U_s$. While the arguments in Sections 3.1 and 3.2 hold for all $\alpha$, it is now necessary to distinguish between $\alpha > 1$ and $\alpha < 1$.

D1) *The case when* $\alpha > 1$, *hence* $EU < \infty$. Firm size being positive, $EU > 0$, and "common sense" and the tradition of practical statistics take it for granted that the law of large numbers hold, so that the *total industry size is approximately N times the expected size of a randomly selected firm.*

D2) *The case when* $\alpha < 1$, *hence* $EU = \infty$. The inequality $EU < \infty$ cannot and must not be taken for granted: it fails when the random variable $U$ is scaling with $\alpha < 1$. Many authors describe this feature as being "improper," and failed to face it. But *it is not* improper, and *must* be faced.

Applied blindly to the case $EU = \infty$, the law of large numbers claims that the total industry size is approximately infinite. This ridiculous result shows that one can no longer rely on common sense that is based on expectations.

Heuristically, if expectation is replaced by a different "typical" value, the total industry size is the sum of the above-written typical values $\overline{u}(r, N)$

$$\sum_{s=1}^{N} \overline{u}(s, N) = N^{1/\alpha} \sum_{s=1}^{N} s^{-1/\alpha}.$$

The most important feature is that the customary proportionality to $N$ has disappeared. For very large $N$, it must be replaced by proportionality to $N^{1/\alpha}$. For moderately large $N$,

$$\tilde{u} N^{1/\alpha} \sum_{s=1}^{N} s^{-1/\alpha} \sim \tilde{u} N^{1/\alpha} [\zeta(1/\alpha) - \frac{N^{1-1/\alpha}}{1/\alpha - 1}] = \tilde{u}[\zeta(1/\alpha) N^{1/\alpha} - \alpha(1-\alpha)^{-1} N].$$

Because of $\alpha < 1$, the factor in $N^{1/\alpha}$ grows faster than the factor in $N$.

### 3.5 Relative shares when $U > 0$: contrast between $\alpha > 1$ and $\alpha < 1$; when $\alpha < 1$ and $N \to \infty$, $\Phi$ has a limit and the largest addend does not become relatively negligible

The two paths started in Section 3.3 continue in profoundly different fashions.

E1) *The case* $\alpha > 1$. As $N \to \infty$, (due to point C), the $r$ -th largest firm increases proportionately to the power $N^{1/\alpha}$, and (due to point D1)) the sum of all firm sizes increases proportionately to $N$.

$$\overline{U}_R(r, N) \sim \frac{N^{1/\alpha} r^{-1/\alpha}}{NEU} = N^{-1+1/\alpha} r^{-1/\alpha}.$$

As $N \to \infty$, this ratio tends to zero. This is a familiar and widely used property; for example, the relative size of the largest of $N$ Gaussian, exponential, Poisson, Gamma, or lognormal variables becomes negligible.

E2) *The case* $\alpha < 1$. *Heuristics*. As $N \to \infty$, both the $r$-th largest firm (due to point C)) and the sum of all firm sizes (due to point D2)) increase proportionately to the power $N^{1/\alpha}$. It follows that the relative share of the $r$-th largest firm behaves roughly like

$$\bar{u}_R(r, N) = \frac{r^{-1/\alpha}}{\sum_{s=1}^{N} s^{-1/\alpha}}.$$

When size is measured by the work force, the preceding relation gives an estimate of the probability that a worker chosen at random is an employee of the $r$-th firm.

E3) *The case* $\alpha < 1$, *continued. Rigorous results*. Given its significance, the argument yielding $\bar{u}_R(r, N)$ must be scrutinized carefully. This assumption that the numerator and denominator are statistically independent as $N \to \infty$ is false, but the conclusion is correct. Darling 1952 shows that $U_R(1, N)$ indeed has a distribution that is asymptotically independent of $N$. The formulas look forbidding and are not needed here, therefore, were put in the Appendix.

### 3.6 Comments

Chapter E9 will study the lognormal distribution, and show that this chapter's uncomfortable conclusion can be "papered over" by asserting that the observed facts concern an ill-defined "transient", but it is better to face it squarely. Against the background of the usual practical statistics, the fact that it is possible for $\Phi$ to be independent of $N$ is astounding. The usual inference, once again, is that when an expression is the sum of many contributions, each due to a different cause, then the relative contribution of each cause is negligible. Here, we find, not only that the predominant cause is not negligible, but that it is independent of $N$.

The reader may be reminded of the distinction that Chapter E5 makes between *mild, slow,* and *wild* fluctuations. Most scientists' intuition having been nourished by examples of *mild* randomness, the preceding conclusion is *wild* and "counter-intuitive," but it will not go away.

## APPENDIX A: THEOREMS CONCERNING LONG-RUN CONCENTRATION FOR THE WILD SCALING DISTRIBUTIONS

*Theorem (Darling 1952).* There exists a family of distribution functions, $G(1, \alpha, y)$, a special case of the distributions $G(r, \alpha, y)$ which will be examined later, such that

(A) if $0 < \alpha < 1$, $\lim_{N \to \infty} \Pr\left\{ \dfrac{\sum_{n=1}^{N} U_n - U(1, N)}{U(1, N)} \leq y \right\} = G(1, \alpha, y)$.

(B) if $1 < \alpha < 2$, $\lim_{N \to \infty} \Pr\left\{ \dfrac{\sum_{n=1}^{N} U_n - NE(U) - U(1, N)}{U(1, N)} \leq y \right\} = G(1, \alpha, y)$.

(C) if $1 < \alpha < 2$ and $EU \neq 0$, one has, in addition

$$\lim_{N \to \infty} \Pr\left\{ \frac{\sum_{n=1}^{N} U_n}{U(1, N)} \leq y \tilde{u} N^{-1-1/\alpha} \right\} = \exp\{ -[y/E(U)]^{\alpha} \}.$$

The distribution $G(1, \alpha, y)$ cannot be written as a simple analytic expression but its characteristic function $\hat{G}(\alpha, z)$ is known. It is as follows:

$$\text{If } 0 < \alpha < 1, \hat{G}(\alpha, z) = \frac{1}{1 - \alpha \int_0^1 (e^{izs} - 1) s^{-(\alpha+1)} ds} = \frac{1}{e^{iz} \int_0^1 e^{isz} s^{-\alpha} ds}$$

$$\text{If } 1 < \alpha < 2, \hat{G}(\alpha, z) = \frac{1}{-1 + \dfrac{iz\alpha}{(\alpha - 1)} - \alpha \int_0^1 (e^{isz} - 1 - isz) s^{-(\alpha+1)} ds}.$$

The essential thing about $G$ is that it does not reduce to the degenerate value 0 as is the case in the distributions cited in Section 5, but has finite and non-vanishing moments of all orders. It is important to note the following: when $1 < \alpha < 2$, then $NE(U)$ must be subtracted from $\sum U_n$ in order to make its expectation even to zero. If $0 < \alpha < 1$, one finds

$$E\left\{ \frac{\sum_{n=1}^{N} U_n - U(1, N)}{U(1, N)} \right\} = \frac{\alpha}{1 - \alpha}.$$

*Theorems (Arov & Bobrov 1960).* These theorems generalize the results in Darling 1952 to firms of ranks 2, 3, etc.. We have the following:

$$\text{If } 0 < \alpha < 1, \lim_{N\to\infty}\Pr\left\{\frac{\sum_{n=1}^{N} U_n - U(1,N) - \cdots - U(r,N)}{U(r,N)} \le y\right\} = G(r, \alpha, y),$$

where the distribution $G(r, \alpha, y)$ relates to the sum of $r$ independent random variables, each following the law of distribution $G(1, \alpha, y)$; in particular, its expected value is $r\alpha/(1-\alpha)$,

$$\text{If } 1 < \alpha < 2, \lim_{N\to\infty}\Pr\left\{\frac{\sum_{n=1}^{N} U_n - NE(U) - U(1,N) \ldots U(r,N)}{U(r,N)} \le y\right\} = G(r, \alpha, y).$$

## APPENDIX B: TWO MEASURES OF CONCENTRATION AND THEIR DEPENDENCE ON THE FINITENESS OF *EU* AND $EU^2$

To establish the usefulness of the rank-size rule and of the preceding heuristics, it is good to examine their bearing on existing techniques of statistical economics.

### B.1 An index that measures inequality by a second moment

Herfindahl proposed the following statistical index of inequality

$$H = \sum_{s=1}^{N}\left\{\frac{U(s,N)}{\sum U_n}\right\}^2 \le 1.$$

This index has no independent motivation, and we shall see that its behavior is very peculiar. It is odd that it should ever be mentioned in the literature, even solely to be criticized because it is an example of inconsiderate injection of a sample second moment in a context where even the existence of expectation is controversial. Three cases must be distinguished.

*The case where* $EU^2 < \infty$. For large N, the law of large numbers applies to both $U$ and $U^2$ and yields

$$H \sim \frac{NEU^2}{N^2(EU)^2} = \frac{1}{N}\frac{EU^2}{(EU)^2}.$$

The ratio $EU^2/(EU)^2$ is a normalized second moment, and $H$ is expected to depend inordinately on the sample size $N$, in a way that is inextricably intertwined with its dependence on the distribution.

*The case where* $EU = \infty$*, in particular where* $U$ *is scaling with* $0 < \alpha < 1$. For large $N$, the law of large numbers applies to neither $U$ nor $U^2$. To obtain a first order of magnitude of $H$, one can take the heuristic step that uses the rank-size argument. This yields

$$H \sim \tilde{H} = (\text{ a constant }) \sum_1^N s^{-2/\alpha} \left\{ \sum_{s=1}^N s^{-1/\alpha} \right\}^{-2}.$$

As $N \to \infty$, this ratio tends to the positive and finite limit

$$\tilde{H}(\alpha) = (\text{ a constant }) \zeta(2/\alpha) \zeta^{-2}(1/\alpha).$$

When $\alpha$ is close to 1, which is the value claimed for firm sizes,

$$\tilde{H}_\infty(\alpha) \sim (\text{a constant }) \zeta(2/\alpha)(1-\alpha)^2.$$

The values of $H$ do not depend much on $N$, but it amplifies the statistical fluctuations around the rank-size typical value.

*The case when* $EU < \infty$ *but* $EU^2 = \infty$*, in particular when* $U$ *is scaling and* $1 < \alpha < 2$. According to the rank-size argument, Herfindahl's index is of the order of $N^{-2+2/\alpha}$ and tends to 0 as $N \to \infty$.

According to reports, Herfindahl's index is taken seriously in some publications. This is hard to believe.

### B.2 Lorenz curves

As a measure of concentration, Lorenz proposed the function

$$L(x) = \frac{\sum_{s=1}^{xN} U(s, N)}{\sum_{s=1}^{N} U(s, N)}.$$

This function yields the proportion $L(x)$ of the total size as function of the proportion $x$ of the number of firms, starting from the largest. It is taken for granted that the function $L(x)$ is obtained by a simple transformation from the size distribution $F(u) = \Pr\{U \le u\}$, and that the graph of $L(x)$, to be denoted by $\mathcal{L}$, is visually "more telling" than either the graph of $F(u)$ or the corresponding rank frequency graph $Q(r)$.

Skeptics respond that Lorenz curves emphasize a concept of inequality that involves the whole distribution and may be very misleading because the data in the bell of the distribution are frequently very incomplete. However, Lorenz curves also encounter a more serious theoretical objection. Indeed, it seems to be implicitly assumed that $\mathcal{L}$ only depends as the degree of concentration within a sample, not on the sample size $N$. Let us show that this implicit assumption is correct when $EU < \infty$, but *not* when $EU = \infty$. For scaling distribution, the implicit assumption is correct for $\alpha > 1$ but not for $\alpha < 1$.

A) For distributions with moments of all orders, Lorenz curves are theoretically unobjectionable. But their reputation for being visually telling is undeserved.

Indeed, in terms of $P(u) = \Pr\{U > u\}$, the number of firms of size greater than $u$ is $NP(u)$ and their share of the sum of all firm sizes is $-N\int_u^\infty s dP(s)$. One can therefore write relative numbers and relative shares as functions of $u$ as follows:

$$x = P(u) \text{ and } L(x) = -\int_u^\infty s dP(s).$$

This means that both $x$ and $L(x)$ are independent of $N$, and define a curve $\mathcal{L}$. When $N$ is small, the sample Lorenz curve will be made up of segments of straight line; but it will tend towards the well-defined limit $\mathcal{L}$ as $N \to \infty$.

For example, if $U$ is scaling with $\alpha > 1, P(u) \sim u^{-\alpha}$ and the curve behaves as follows near the point $L = 0; x = 0$:

$$x(L) \sim u^{-\alpha} \sim [(\alpha-1)\alpha^{-1}L]^{\alpha/(\alpha-1)}; \text{ or } L(x) = \alpha(\alpha-1)^{-1}x^{(\alpha-1)/\alpha}.$$

This behavior has an unfortunate by-product: $\mathscr{L}$ may be well-defined independently of $L$, yet fails to deliver on its promise of being "visually telling." Indeed, if $\alpha - 1$ is small, a large number of derivatives of the function $x(L)$ vanish at the point $x = 0$, meaning that the curve $\mathscr{L}$ has a contact of very high order with the axis $x = 0$. On computer-drawn Lorenz curves, the exact order of this contact is not at all clear to the eye. As to old-time draftsmen, they near-invariably destroyed any evidence of contact by drawing Lorenz curves with a slope that is neither zero nor infinite near $x = L = 0$.

B) When $U$ is scaling with $\alpha < 1$, sample Lorenz curves are unjustified and misleading, because they are greatly dependent on sample size. Indeed, we know that the relative share of the $r$ largest firms is independent of $N$. Therefore, any prescribed ordinate $L(x)$ will be achieved for an abscissa $x$ that tends towards zero as $N \to \infty$. This means that for $0 < \alpha < 1$, the sample Lorenz curve will tend towards the "degenerate" limit made up of the lower edge and the right edge of the unit square. Hence, the sample curves for finite $N$ will not be representative of anything at all. In particular, sample Lorenz curves will depend even more critically upon the thoroughness with which small firms have been tabulated.

When $U$ is scaling near the borderline value $\alpha \sim 1$, the convergence of $\mathscr{L}$ to its degenerate limit is very slow, which makes $\mathscr{L}$ especially misleading.