The Rank-Frequency Form of Zipf's Law

Bruce M. Hill

Journal of the American Statistical Association, Volume 69, Issue 348 (Dec., 1974), 1017-1026.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28197412%2969%3A348%3C1017%3ATRFOZL%3E2.0.CO%3B2-M

The Rank-Frequency Form of Zipf's Law

BRUCE M. HILL*

Suppose that there are K regions in a country, with N_i people and M_i cities in the ith region. Let N_i be large, M_i random, given N_i, and such that the distribution of $M_i N_i^{-1}$, given N_i, converges to a limiting distribution F, with $F(x) \sim Cx^\gamma$ as $x \to 0$, $\gamma > 0$. Let $L^{(r)}$ be the size of the rth largest city in the country. If, given M_i and N_i, there is a Bose-Einstein allocation of the N_i people to the M_i cities in region i, independently for the various regions, then a plot of $L^{(r)}$ against r will be approximately proportional to $r^{-(1+\alpha)}$, for $1 + \alpha = \gamma^{-1}$.

1. INTRODUCTION AND SUMMARY

This article presents a theoretical derivation of the rank-frequency form of Zipf's Law based on a Bose-Einstein form of the classical occupancy model, with the additional feature that the number of cells is itself random. By the rank-frequency form of Zipf's Law we mean a relationship in which the rth largest of a set of quantities is in some specified sense approximately proportional to $r^{-(1+\alpha)}$ for some $\alpha > 0$, with $\alpha \leq 1$ being the case of greatest interest. In many different areas it is common that a plot of frequency against rank yields a surprisingly close fit to such a law. For example, a plot against r of the population of the rth largest city [1], or of the income of the rth richest unit (the Pareto Law [6]), or of the number of articles on a given subject in the journal having the rth largest number of such articles [8], all seem to fit the Zipf form. In his book [11] Zipf presents a massive array of graphs drawn from a great many areas including, besides those already mentioned, linguistics, music, warfare and several others, which also seem to fit such a law surprisingly well. Sometimes, of course, there are substantial discrepancies, and what is surprising to one is not to another.[1] Nevertheless it seems appropriate to inquire whether a law of such apparent universality might not arise from a simple probabilistic mechanism. Earlier work of interest in this connection was done by Yule [10], Simon [8] and Mandelbrot [5, 6].

The present work is an outgrowth of my earlier work on the generic-specific form of Zipf's Law [3], where by the generic-specific form we mean that the proportion of genera with exactly s species is in some specified sense approximately proportional to $s^{-(1+\alpha)}$, for some $\alpha > 0$. The two forms of Zipf's Law are closely related, and it is shown later that a modification of my original model for the generic-specific form yields the rank-frequency form as well. For completeness and unity, the basic models and their implications will be described in some detail in the remainder of this section. This presentation should also serve those readers primarily interested in the nature of the underlying model and its applications. These readers may wish to skip the mathematical derivations in Sections 2 and 3, and resume with Section 4 where further discussion ensues.

We begin with the model of [3]. Suppose there are N units which are to be allocated to M nonempty cells with, say, L_i units in the ith cell so that $L_i \geq 1$, $i = 1, 2, \cdots, M$, and $\sum_1^M L_i = N$. For vividness and suggestiveness as to applications we shall sometimes refer to the units as either species or people, and to the cells as either genera or cities, respectively. The method of allocation will be either of the Bose-Einstein form

$$\Pr\{\underline{L} \mid M, N\} = \binom{N-1}{M-1}^{-1}, \qquad (1.1)$$

or of the Maxwell-Boltzmann form

$$\Pr\{\underline{L} \mid M, N\} = (N-M)!\{\textstyle\prod_{i=1}^{M}(L_i - 1)!\}^{-1} M^{-(N-M)}, \qquad (1.2)$$

where $\underline{L} = (L_1, \cdots, L_M)$, and with the usual abuse of notation we use the same symbol to refer to both the random variable and to its value. Although the Maxwell-Boltzmann allocation is of interest in its own right, and results for this case will be derived below, it is the Bose-Einstein allocation that yields Zipf's Law, and for simplicity the discussion here will be confined to this case.

The novel aspect of the model consists in allowing M, the number of cells, to be random, given N, say with distribution $F_N(x) = \Pr\{MN^{-1} \leq x \mid N\}$, while the allocation of units to cells is conditional upon both M and N. We shall suppose throughout this discussion that $F_N(\cdot)$ converges to a distribution $F(\cdot)$, which is absolutely continuous with density $f(\cdot)$ (in fact many of the results hold more generally), and that N is large. Thus we may visualize the model in terms of a large number N of units (species, people), a random number M of cells (genera, cities), with $1 \leq M \leq N$, and MN^{-1} having approximately the limiting distribution $F(\cdot)$, and finally with a Bose-Einstein allocation of units to cells, conditional on M and N. Now let $G(s)$ be the number of cells

* Bruce M. Hill is professor, Department of Statistics, University of Michigan, Ann Arbor, Mich. 48104. The substance of this article was an invited paper at the annual meeting of the American Statistical Association at Colorado State University, August 1971. The author is grateful to Richard Olshen for helpful comments.

[1] Any formal statistical test of goodness of fit seems out of the question, given the quantity and nature of the data, even if the philosophical validity of such tests were agreed upon.

© Journal of the American Statistical Association
December 1974, Volume 69, Number 348
Theory and Methods Section

with s units, and let Θ denote a random variable having the distribution $F(\cdot)$. It was shown in [3] that the proportion of genera with s species, $G(s)/M$, then converges in distribution to that of $\Theta(1-\Theta)^{s-1}$. If, for example, Θ has a uniform distribution on the unit interval, then as $N \to \infty$,

$$E\{G(s)/M|N\} \to \int_0^1 y(1-y)^{s-1}dy = [s(s+1)]^{-1} ,$$

which is a special and important form of the Zipf relationship. By appropriate choice of the distribution of Θ approximate behavior of the form $s^{-(1+\alpha)}$ can similarly be achieved for the limiting expected proportion of genera with s species, and so this provides a weak sense in which the generic-specific form of Zipf's Law can be expected to hold. Stronger forms in which $G(s)/M$ converges in probability to such a function of s are derived in [4].

Now consider the ordered values of the L_i, say, $L^{(1)} \geq L^{(2)} \geq \cdots \geq L^{(M)}$, e.g., the populations of the cities in nonincreasing order. The relationship between $L^{(R)}$ and $G(s)$ is that the event $L^{(R)} \leq a$ occurs if and only if $G(1) + \cdots + G(a) \geq M - R + 1$. For fixed a, as N grows large, $\Pr\{L^{(R)} \leq a|N\}$ goes to 0, but as is shown below, a limiting distribution will exist if $a \equiv a_N$ goes to infinity appropriately. Thus while [3] concerns the probability distribution of $G(s)$, for fixed s, the present article concerns the probability distribution of $\sum_1^{a_N} G(s)$, i.e., the cumulative sum. The first result obtained is that for any fixed R, and with MN^{-1} converging to a constant $\Theta_o \neq 0$, that $L^{(R)}/\ln N$ converges in probability to $-[\ln(1-\Theta_o)]^{-1}$. This result parallels that of [3], where under the same conditions $G(s)/M$ converges in probability to $\Theta_o(1-\Theta_o)^{s-1}$. Similarly, if M is random, given N, with $F_N(x) = \Pr\{MN^{-1} \leq x|N\}$ converging to an absolutely continuous distribution $F(x)$, then it will be shown that

$$\Pr\{L^{(R)}/\ln N \leq a|N\}$$

converges to

$$\Pr\{\Theta \geq 1 - e^{-1/a}\} = H(a), \text{ say,}$$

where Θ is a random variable having distribution F. This result also parallels that of [3] in which under the same conditions $G(s)/M$ is in the limit distributed like $\Theta(1-\Theta)^{s-1}$.

However, since under the preceding model the limiting distribution of $L^{(R)}/\ln N$ does not depend on R, we would expect that a plot of $L^{(r)}/\ln N$ against r would be flatter than that implied by Zipf's Law, where $L^{(r)}$ is roughly proportional to $r^{-(1+\alpha)}$. We now show how our original model can be modified so as to yield precisely such behavior. We use the people-city terminology for vividness but the argument is of course general.

Thus suppose that a country consists of K regions, the ith such region having N_i people and M_i cities, and that within each of these regions a Bose-Einstein allocation of people to cities takes place, and that moreover, these allocations are mutually independent, given the N_i. This new model is then simply an extended version of the original model, and $K = 1$ is precisely the original model. Now let r_i, $i = 1, \cdots, K$, be fixed, and let $L_i^{(r_i)}$ be the population of the r_ith largest city in the ith region. According to the result mentioned above, the limiting distribution of each $Z_i = L_i^{(r_i)}/\ln N_i$ is $H(\cdot)$,[2] and given the N_i, these Z_i are independent. Supposing that the N_i are sufficiently large so that the Z_i can be regarded as a sample from the limiting distribution, $H(\cdot)$, we can now apply standard limit theorems to discuss the behavior of the ordered values $Z^{(1)} \geq Z^{(2)} \geq \cdots \geq Z^{(K)}$. In particular, by the Renyi representation theorem [7], we can express $Z^{(r)}$ as

$$Z^{(r)} = H^{-1}\left\{\exp\left[-\left(\frac{\delta_1}{K} + \frac{\delta_2}{K-1} + \cdots + \frac{\delta_r}{K-r+1}\right)\right]\right\} , \quad r = 1, \cdots, K ,$$

where the δ_i are independent random variables each having the exponential distribution with expectation 1. Using this result it is now easily seen that by appropriate choice of $F(\cdot)$, Zipf's Law will result. For example, if $\Theta \sim U[0, 1]$, i.e., $F(x) = x$, then

$$H(a) = \Pr\{\Theta \geq 1 - e^{-1/a}\} = e^{-1/a} ,$$

and so $H^{-1}(x) = -[\ln x]^{-1}$, $0 \leq x \leq 1$. Hence,

$$Z^{(r)} = \left[\frac{\delta_1}{K} + \frac{\delta_2}{K-1} + \cdots + \frac{\delta_r}{K-r+1}\right]^{-1} .$$

Now let $X_{(r)}$ denote a random variable having the gamma distribution with shape parameter r, and scale parameter 1, and thus density

$$[\Gamma(r)]^{-1}x^{r-1}e^{-x}, \quad \text{for} \quad x \geq 0, r > 0 ,$$

where $\Gamma(r)$ is the standard gamma function. Since $K\{(\delta_1/K) + [\delta_2/(K-1)] + \cdots + [\delta_r/(K-r+1)]\}$ is, in the limit as $K \to \infty$, distributed like $\sum_{i=1}^r \delta_i$, i.e., like $X_{(r)}$, it follows that $K^{-1}Z^{(r)}$ is in the limit distributed like $X_{(r)}^{-1}$. For sufficiently large K we can then regard $K^{-1}Z^{(r)}$ as being distributed approximately like $X_{(r)}^{-1}$, so that

$$E[K^{-1}Z^{(r)}] \approx E[X_{(r)}^{-1}] = (r-1)^{-1}, \quad \text{for} \quad r > 1 ,$$

$$\text{Var}[K^{-1}Z^{(r)}] \approx \text{Var}[X_{(r)}^{-1}] = [(r-1)^2(r-2)]^{-1} \quad \text{for} \quad r > 2 .$$

To this extent a plot of the $Z^{(r)}$ against r should approximate a Zipf Law with $\alpha = 0$. Recalling that $Z_i = L_i^{(r_i)}/\ln N_i$, we see that if the $\ln N_i$ are nearly equal, and if the r_i are all small enough so that the limiting results are applicable, then a plot of the ordered values of the $L_i^{(r_i)}$, i.e., of the ordered city sizes, should also approximate a Zipf Law with $\alpha = 0$. A related result holds, as shown in Section 4, if instead of selecting the

[2] H is assumed to have an inverse.

r_ith largest city in region i, we merely select a city at random from that region and use its size in place of $L_i^{(r_i)}$.

In a similar way we can show that a Zipf Law with parameter α will arise if the density of Θ is

$$f(x) \sim (1 + \alpha)^{-1} x^{-(\alpha/1+\alpha)}$$

as x goes to 0, where $\alpha > 0$, and the symbol "$\sim$" here and throughout means that the ratio goes to 1. It then follows easily from the Renyi representation theorem that $Z^{(r)}/K^{1+\alpha}$ is, in the limit as $K \to \infty$, distributed like $X_{(r)}^{-(1+\alpha)}$. Hence for sufficiently large K we can regard $Z^{(r)}/K^{1+\alpha}$ as distributed approximately like $X_{(r)}^{-(1+\alpha)}$, so that

$$E[Z^{(r)}/K^{1+\alpha}] \approx EX_{(r)}^{-(1+\alpha)} = \Gamma(r - 1 - \alpha)/\Gamma(r) \ ,$$
$$\mathrm{Var}\,[Z^{(r)}/K^{1+\alpha}] \approx \mathrm{Var}\, X_{(r)}^{-(1+\alpha)}$$
$$= [\Gamma(r)\Gamma(r - 2 - 2\alpha) - \Gamma^2(r - 1 - \alpha)]/\Gamma^2(r) \ ,$$

where we require $r > 1 + \alpha$ for the expectation to exist, and $r > 2 + 2\alpha$ for the variance to be finite. Since $\Gamma(r - 1 - \alpha)/\Gamma(r)$ is approximately proportional to $r^{-(1+\alpha)}$ for large r, it follows that a plot of $Z^{(r)}$ against r should now approximate a Zipf Law with parameter α. Just as for the case $\alpha = 0$ this then implies that a plot of ordered city sizes should also approximate such a Zipf Law.

So far the results quoted state that an approximation to Zipf's Law is obtained if either the R_ith largest city in region i, or alternatively a randomly selected city from region i, is used in forming the ranked city sizes from the regions. A final result in Section 4 shows that Zipf's Law still holds if we plot the size of the rth largest city in the *country* against r. This result seems to be the most relevant to the graphs displayed by Zipf [11]. In Section 4 it is also shown that a Maxwell-Boltzmann allocation leads to a very different result.

This concludes the description of the model yielding the rank frequency form of Zipf's Law. In Sections 2 and 3 the preceding results are derived, while in Section 4 further discussion ensues. In Section 5 motivation for the Bose-Einstein distribution and other assumptions of the model is proposed.

2. EXACT DISTRIBUTION OF $L^{(R)}$

Let $\underline{L} = (L_1, \cdots, L_M)$ have any exchangeable distribution, and let $L^{(1)} \geq L^{(2)} \geq \cdots \geq L^{(M)}$ be the ordered values of the L_i. If $G(s)$ is the number of L_i having the value s, then the event $L^{(R)} \leq a$ occurs if and only if $G(1) + \cdots + G(a) \geq M - R + 1$. For if $L^{(R)} \leq a$, then $L^{(R+1)}, \cdots, L^{(M)}$, must also be $\leq a$, and so the number of L_i with value $\leq a$, must be at least $M - R + 1$. The converse follows similarly. It is worth noting also that

$$RL^{(R)} \leq \sum_{i=1}^{R} L^{(i)} \leq N - M + R \ ,$$

since each $L_i \geq 1$, so that

$$L^{(R)} \leq R^{-1}(N - M + R) \ .$$

Now let V_j denote the event $L_j > a$, $j = 1, \cdots, M$. Then

$$\Pr\{G(1) + \cdots + G(a) = r\}$$
$$= \Pr\{\text{exactly } M - r \text{ of the } V_j \text{ occur}\}.$$

But from the familiar result in Feller [2, p. 96], for any m with $0 \leq m \leq M$,

$$\Pr\{\text{exactly } m \text{ of the } V_j \text{ occur}\}$$
$$= \sum_{j=0}^{M-m} (-1)^j \binom{m+j}{m} S_{m+j} \ ,$$

where by exchangeability

$$S_m = \binom{M}{m} \Pr\{V_1 \cap V_2 \cap \cdots \cap V_m\} \ , \quad \text{for} \quad m \geq 1 \ ,$$

and by definition, $S_0 = 1$. Hence,

$$\Pr\{\text{exactly } m \text{ of the } V_j \text{ occur}\}$$
$$= \sum_{j=0}^{M-m} (-1)^j \binom{m+j}{j} \binom{M}{m+j} \Pr\{V_1 \cap \cdots \cap V_{m+j}\}$$
$$= \binom{M}{m} \sum_{j=0}^{M-m} (-1)^j \binom{M-m}{j} \Pr\{V_1 \cap \cdots \cap V_{m+j}\} \ .$$

Putting $m = M - r$, we have

$$\Pr\{G(1) + \cdots + G(a) = r\}$$
$$= \binom{M}{r} \sum_{j=0}^{r} (-1)^j \binom{r}{j} \Pr\{V_1 \cap \cdots \cap V_{M-r+j}\} \ .$$

This yields our main result for the exact distribution of $L^{(R)}$,

$$\Pr\{L^{(R)} \leq a\}$$
$$= \Pr\{G(1) + \cdots + G(a) \geq M - R + 1\}$$
$$= \sum_{r=M-R+1}^{M} \binom{M}{r} \sum_{j=0}^{r} (-1)^j \binom{r}{j} \cdot \Pr\{V_1 \cap \cdots \cap V_{M-r+j}\}$$
$$= \sum_{t=0}^{R-1} \binom{M}{M-R+t+1} \sum_{j=0}^{M-R+t+1} (-1)^j \cdot \binom{M-R+t+1}{j} \Pr\{V_1 \cap \cdots \cap V_{R-t+j-1}\}. \quad (2.1)$$

In particular, for $R = 1$,

$$\Pr\{L^{(1)} \leq a\}$$
$$= \sum_{j=0}^{M} (-1)^j \binom{M}{j} \Pr\{V_1 \cap \cdots \cap V_j\}. \quad (2.2)$$

Equation (2.1) holds for any exchangeable distribution of the finite sequence $(L_1, \cdots, L_M)$, where the L_i are positive integers.

Now suppose, in addition, that the Bose-Einstein model of Hill [3] holds, so that given M, N, the vector

$\underline{L}$ is uniformly distributed over the set of its possible values, with $\sum_1^M L_i = N$. Then, conditional on M and N,

$$\begin{aligned}\Pr\{V_1 \cap \cdots \cap V_m\} &= \Pr\{L_1 > a, \cdots, L_m > a\} \\ &= \sum_{i_1>a,\cdots,i_m>a} \Pr\{L_1 = i_1, \cdots, L_m = i_m\} \\ &= \binom{N-1}{M-1}^{-1} \sum_{i_1>a,\cdots,i_m>a} \binom{N - i_1 - \cdots - i_m - 1}{M - m - 1} \\ &= \binom{N-1}{M-1}^{-1} \binom{N - ma - 1}{M-1}.\end{aligned}$$

Noting that if a is replaced by Na, the right side tends to $(1 - ma)^{M-1}$ as N goes to infinity, and we have

$$\begin{aligned}\lim_{N\to\infty} \Pr\{N^{-1}L^{(R)} \le a\} &= \sum_{t=0}^{R-1} \binom{M}{M - R + t + 1} \sum_{j=0}^{M-R+t+1} (-1)^j \\ &\cdot \binom{M - R + t + 1}{j} [1 - (R - t + j - 1)a]^{M-1}\end{aligned}$$

For $R = 1$ this limit is $\sum_{j=0}^M (-1)^j \binom{M}{j} (1 - ja)^{M-1}$, agreeing with the result in Wilks [9], since $L^{(1)}/N$ is asymptotically distributed like the largest observation from a Dirichlet distribution with all parameters equal to unity. We note that to obtain this limiting distribution we have let M/N tend to 0. This situation is of little interest in regard to Zipf's Law. Rather, we are concerned with the corresponding limiting distributions when M/N tends to some $\Theta_o \neq 0$, or when M/N has a nondegenerate limiting distribution. I am not aware of any literature concerning limiting distributions of order statistics from such triangular arrays, although there are, of course, well known results for sequences of independent identically distributed random variables.

3. LIMITING DISTRIBUTION OF $L^{(R)}/\ln N$ AS $M/N \to \Theta$

We now prove two theorems concerning the limiting distribution of $L^{(R)}/\ln N$. The Bose-Einstein model is assumed throughout this section. The first theorem deals with the case where M/N converges to a constant Θ_o, $0 < \Theta_o < 1$, with all probabilities conditional on M and N.

Theorem 1: If M and N go to infinity in such a way that $M/N - \Theta_o = o(1/\ln N)$, where $0 < \Theta_o < 1$, then for fixed R, $L^{(R)}/\ln N$ converges in probability to $-[\ln(1 - \Theta_o)]^{-1}$.

Proof: From the equation

$$\Pr\{L^{(R)} \le a\} = \Pr\{G(1) + \cdots + G(a) \ge M - R + 1\} \ ,$$

it suffices to prove the theorem for the case $R = 1$. Now let $k > 0$ and define

$$a_N \equiv a_N(\Theta_o, k) = \ln[k/N\Theta_o]/\ln[1 - \Theta_o] \ ,$$

so that

$$N\Theta_o(1 - \Theta_o)^{a_N} = k \ .$$

Using Stirling's formula it is easy to show that for fixed j,

$$\binom{N-1}{M-1}^{-1} \binom{N - ja_N - 1}{M-1} \sim (1 - \Theta_o)^{ja_N} , \quad (3.1)$$

where the symbol "$\sim$" means that the ratio of the two sides tends to unity as M and N tend to infinity. Hence,

$$\binom{M}{j} \binom{N-1}{M-1}^{-1} \binom{N - ja_N - 1}{M-1} \sim (1 - \Theta_o)^{ja_N} \frac{M^j}{j!} \sim k^j/j! \ .$$

Since the terms of the series are uniformly bounded in absolute value by those of a summable function, it therefore follows that

$$\begin{aligned}\lim \Pr\{L^{(1)} \le a_N\} &= \lim \sum_{j=0}^M (-1)^j \binom{M}{j} \binom{N-1}{M-1}^{-1} \binom{N - ja_N - 1}{M-1} \\ &= e^{-k} \text{ as } M \text{ and } N \text{ go to } \infty \ .\end{aligned}$$

Now define

$$a_o \equiv a(\Theta_o) = -[\ln(1 - \Theta_o)]^{-1} ,$$

so $a_N \sim a_o \ln N$. If $\epsilon > 0$ and if M and N are sufficiently large, then $(a_o - \epsilon)\ln N \le a_N$, so $\Pr\{L^{(1)}/\ln N \le a_o - \epsilon\} \le \Pr\{L^{(1)} \le a_N\}$. Since the right side tends to e^{-k}, and since this holds for all $k > 0$, it follows that $\lim \Pr\{L^{(1)}/\ln N \le a_o - \epsilon\}$ exists and is 0. Similarly, for sufficiently large M and N,

$$\Pr\{L^{(1)}/\ln N \le a_o + \epsilon\} \ge \Pr\{L^{(1)} \le a_N\} \ ,$$

and again the right side tends to e^{-k}, with the relationship holding for all $k > 0$. Thus $\lim \Pr\{L^{(1)}/\ln N \le a_o + \epsilon\}$ exists and is 1. It follows that $L^{(1)}/\ln N$ converges in probability to $a_o = -[\ln(1 - \Theta_o)]^{-1}$, as was claimed, which completes the proof.

In our next theorem we allow M to be random, having a conditional distribution, given N. Given both M and N, the Bose-Einstein model is assumed to hold, as before.

Theorem 2: Let $F_N(x) = \Pr\{M/N \le x | N\}$. If, as N goes to infinity, $F_N(x)$ converges weakly to an absolutely continuous distribution function $F(x)$, then for any fixed R,

$$\lim_{N\to\infty} \Pr\{L^{(R)}/\ln N \le a | N\}$$

exists and is equal to

$$\int_{\Theta(a)}^1 dF(x) \ ,$$

where $\Theta(a) = 1 - e^{-1/a}$.

Proof: Just as with Theorem 1, it suffices to prove the result for $R = 1$. We have

$$\Pr\{L^{(1)}/\ln N \le a|N\}$$
$$= \textstyle\sum_{j=0}^{M} \Pr\{L^{(1)}/\ln N \le a|M = j, N\} \Pr\{M = j|N\}$$
$$= \int_0^1 \nu_N(x; a) dF_N(x) \;,$$

where $\nu_N(x; a) = \Pr\{L^{(1)}/\ln N \le a | M = [Nx], N\}$, and $[y]$ is the largest integer less than or equal to y. Our proof consists in showing that $\nu_N(x; a)$ converges to the indicator function $I_a(x)$ which is 1 when $x \ge \Theta(a)$, and 0 otherwise; and that this convergence is uniform in closed intervals not including 0, 1 or $\Theta(a)$. It will then follow from the absolute continuity of $F(x)$ and the Helly-Bray Lemma, just as in [3, p. 1225], that

$$\int_0^1 \nu_N(x; a) dF_N(x)$$

converges to

$$\int_0^1 I_a(x) dF(x) = \int_{\Theta(a)}^1 dF(x) \;,$$

as claimed.

Let x replace Θ_o in Theorem 1, and let

$$a(x) = -[\ln(1 - x)]^{-1} \;,$$
$$a_N(x; k) = \ln[k/Nx]/\ln(1 - x) \;,$$

for $x \ne 0$, 1. Then, according to Theorem 1, $\nu_N(x; a)$ converges to 0 if $a < a(x)$, and to 1 if $a \ge a(x)$. Since $a \ge a(x)$ if and only if $x \ge \Theta(a)$, the pointwise convergence to $I_a(x)$ follows. Next, in the proof of Theorem 1 it was shown that $\Pr\{L^{(1)} \le a_N(x; k) | M = [Nx], N\}$ converges to e^{-k}. An easy extension of that proof shows that this convergence is in fact uniform in any closed interval not containing 0 or 1. Now let $x < \Theta(a) - \delta$, where $\delta > 0$, so that $a(x) > a + \epsilon$, say, where $\epsilon > 0$. Then, for N sufficiently large,

$$\nu_N(x; a) \le \nu_N[x; a(x) - \epsilon]$$
$$\le \Pr\{L^{(1)} \le a_N(x; k) | M = [Nx], N\} \;.$$

Since the right side goes to e^{-k} uniformly for x in any closed interval not containing 0 or 1, and since this holds for all $k > 0$, it follows that $\nu_N(x; a)$ converges uniformly to 0 for $\delta \le x \le \Theta(a) - \delta$. Similarly, it can be shown that $\nu_N(x; a)$ converges uniformly to 1 for $\Theta(a) + \delta \le x \le 1 - \delta$. Hence, $\nu_N(x; a)$ converges uniformly to $I_a(x)$ in any closed interval not containing 0, 1, or $\Theta(a)$, and Theorem 2 follows.

It should be noted that Theorems 1 and 2 actually prove more than was stated. Thus in the proof of Theorem 1, it was shown that

$$\lim \Pr\{L^{(1)} \le \ln[k/N\Theta_o]/\ln[1 - \Theta_o]\} = e^{-k} \;,$$

so that $N\Theta_o(1 - \Theta_o)^{L^{(1)}}$ has a limiting distribution, namely the exponential distribution. Similarly it can be shown that $N\Theta_o(1 - \Theta_o)^{L^{(R)}}$ is in the limit distributed like $X_{(R)}$, i.e., has a gamma distribution with shape parameter R. Thus $L^{(R)}/\ln N$ can be viewed as distributed like $\ln[X_{(R)}/N\Theta_o]/\ln N \ln[1 - \Theta_o]$ under the conditions of Theorem 1. Under conditions of Theorem 2 this holds with Θ random.

4. EXAMPLES AND DISCUSSION

According to Theorem 2, $\Pr\{L^{(R)}/\ln N \le a|N\}$ tends to $H(a) = \int_{\Theta(a)}^1 F'(x)dx$, where $\Theta(a) = 1 - e^{-1/a}$. Let us now examine the behavior of the limit for those F of interest in regard to Zipf's Law. Suppose that $F'(x) \sim \gamma x^{\gamma-1}$ as $x \to 0$, where $\gamma > 0$. Then

$$1 - H(a) \sim \int_0^{\Theta(a)} \gamma x^{\gamma-1} dx = [\Theta(a)]^\gamma \sim a^{-\gamma} \text{ as } a \to \infty \,.$$

If $Z_1, \cdots, Z_K$, are independent identically distributed random variables, each with distribution H, and if $Z^{(1)} \ge Z^{(2)} \ge \cdots \ge Z^{(K)}$ are their ordered values, then by the Renyi representation theorem referred to in the introduction, it follows that $K[Z^{(R)}]^{-\gamma}$ is, in the limit as $K \to \infty$, distributed like $X_{(R)}$, i.e., the gamma distribution with shape parameter R. This result, which is easily proved, depends upon the fact that $1 - H(a) \sim a^{-\gamma}$. A heuristic derivation, not using the representation theorem, is instructive. For $R = 1$, we have from (2.2),

$$\Pr\{Z^{(1)} \le a_K\}$$
$$= \textstyle\sum_{j=0}^{K} (-1)^j \binom{K}{j} \Pr\{Z_1 > a_K, \cdots, Z_j > a_K\}$$
$$\sim \textstyle\sum_{j=0}^{K} (-1)^j \binom{K}{j} a_K^{-j\gamma} = (1 - a_K^{-\gamma})^K \text{ as } a_K \to \infty \,.$$

Hence, if $a_K^{-\gamma} = \lambda/K$, for some $\lambda > 0$, then

$$\Pr\{K(Z^{(1)})^{-\gamma} \ge \lambda\} = \Pr\{Z^{(1)} \le (K/\lambda)^{1/\gamma}\} \to e^{-\lambda} \,,$$

so that the limiting distribution of $K(Z^{(1)})^{-\gamma}$ is that of $X_{(1)}$. Similarly, it is easily seen that

$$\Pr\{K(Z^{(R)})^{-\gamma} \ge \lambda\}$$
$$= \Pr\{Z^{(R)} \le (K/\lambda)^{1/\gamma}\} \to e^{-\lambda} \textstyle\sum_{t=0}^{R-1} \lambda^t/t! \;,$$

so that the limiting distribution of $K(Z^{(R)})^{-\gamma}$ is that of $X_{(R)}$. If $\gamma = (1 + \alpha)^{-1}$, then for sufficiently large K, the distribution of $Z^{(R)}/K^{1+\alpha}$ will be approximately that of $X_{(R)}^{-(1+\alpha)}$. Noting that $EX_{(R)}^{-(1+\alpha)}$ is approximately proportional to $R^{-(1+\alpha)}$, and that the coefficient of variation of $X_{(R)}^{-(1+\alpha)}$ goes to 0 as $R \to \infty$, we anticipate that a plot of $Z^{(r)}$ against r, for $r = 1, \cdots, R$, should be approximately a Zipf Law with parameter α. When R is so small that the expectation or variance of $[X_{(R)}]^{-(1+\alpha)}$ does not exist, then the Zipf interpretation can be made in terms of the median or some other characteristics of the distribution. For example, the median for $[X_{(1)}]^{-1}$ is 1.4, while the median for $[X_{(2)}]^{-1}$ is .6. It should be noted that the dependencies between the $Z^{(r)}$ are explicitly revealed by the Renyi representation theorem. For example, when $\alpha = 0$,

$$Z^{(r)}/K = K^{-1}\left[\frac{\delta_1}{K} + \frac{\delta_2}{K-1} + \cdots + \frac{\delta_r}{K-r+1}\right]^{-1} ,$$

and since the δ_i are mutually independent, the dependence between $Z^{(r)}$ and $Z^{(s)}$ arises from the fact that they have min $[r, s]$ common δ_i. Realizations of such a $Z^{(r)}$ process, $r = 1, 2, \cdots, R$, tend to be even closer to the Zipf form (here with $\alpha = 0$) than would be suggested by consideration of the marginal distributions of the $Z^{(r)}$ alone.

Now return to our extended model of the introduction, i.e., with K regions, M_i cities and N_i people in the ith such region, and with the basic Bose-Einstein allocation scheme holding within the various regions independently of one another. Suppose that the M_i/N_i are mutually independent, and that each has limiting distribution F as $N_i \to \infty$, where F is such that

$$1 - H(a) = \int_0^{\Theta(a)} F'(x)dx \sim a^{-\gamma} \text{ as } a \to \infty ,$$

and $\gamma = (1 + \alpha)^{-1}$. Let R_i, $i = 1, \cdots, K$, be any fixed integers, and let $L_i^{(R_i)}$ be the size of the R_ith largest city in region i. If $Z_i = L_i^{(R_i)}/\ln N_i$, it follows from Theorem 2 that each Z_i has limiting distribution H. Since the Z_i are independent, it follows from the preceding discussion that a plot of the ordered values $Z^{(r)}$ against r should be approximately of the Zipf form with parameter α, provided that the N_i are sufficiently large and the R_i sufficiently small so that the limit obtained in Theorem 2 yields a good approximation. Finally, if the $\ln N_i$ are nearly equal, then a plot of the ordered values of the $L_i^{(R_i)}$ against their rank order should also be approximately of the Zipf form. For example, if the country were divided into K regions with nearly equal values of $\ln N_i$, and if the largest city was taken from each region, then the ordered values of the $L_i^{(1)}$ should yield a curve of the Zipf form. This constitutes the first sense in which our model yields Zipf's Law.

We now show that a similar results holds if, instead of choosing the R_ith largest city in region i, we merely select a city at random from each region, independently for the various regions. Let L_i^* be the size of the randomly selected city from region i, $i = 1, 2, \cdots, K$. Since the Bose-Einstein model holds within each region, from (3.1) we have

$$\Pr \{L_i^* > a | M_i, N_i\} = \binom{N_i - 1}{M_i - 1}^{-1} \binom{N_i - a - 1}{M_i - 1} \sim (1 - \Theta_o)^a ,$$

if, as in Theorem 1, $M_i/N_i \to \Theta_o$. If, as in Theorem 2, F is the limiting distribution of M_i/N_i, then as $N_i \to \infty$,

$$\Pr \{L_i^* > a | N_i\} \to \int_0^1 (1 - x)^a F'(x)dx .$$

For $F'(x) \sim \gamma x^{\gamma - 1}$ as $x \to 0$,

$$\int_0^1 (1 - x)^a F'(x)dx \sim \Gamma(\gamma + 1)\Gamma(a + 1)/\Gamma(a + 1 + \gamma) \sim Ca^{-\gamma}$$

as $a \to \infty$, where C is constant. Just as with the $L_i^{(R_i)}$, it follows from the Renyi representation theorem that a plot of the ordered values of the L_i^* against their rank order will approximate a Zipf Law. This constitutes the second sense, or situation, in which our model leads to Zipf's Law.

For the final sense in which our model leads to Zipf's Law, let L_{ij} be the population of the jth city in the ith region, $j = 1, \cdots, M_i$, $i = 1, \cdots, K$, and let $L^{(r)}$ be the rth largest of the L_{ij}, i.e., $L^{(r)}$ is the size of the rth largest city in the country. We wish to show that a plot of $L^{(r)}$ against r, $r = 1, 2, \cdots, R$, will be approximately of the Zipf form. This would be immediate if each $L^{(r)}$ was in fact an $L_i^{(1)}$ for some i, in other words was the size of the largest city in some region, and if the $\ln N_i$ were nearly equal. For in this case a plot of $L^{(r)}$ against r would in fact be a plot of the ordered values of the $L_i^{(1)}$ against their rank order, and it has already been shown that the latter follows a Zipf Law. However, it can occur, for example, that the second largest city in one region is larger than the largest in other regions, so that two or more cities from the same region may be among the R largest L_{ij}. It is clear that such an event could lead to a violation of Zipf's Law. To take an extreme case, if the R largest cities all come from the same region, then we are back in the single region model, and a plot of $L^{(r)}$ against r will be too flat. However, in the Appendix it is shown that in fact the probability of two or more of the R largest cities coming from the same region goes to 0 if K and the N_i go to infinity in an appropriate manner. This can be made intuitively clear by the following line of argument. Consider the case $R = 2$. We would like to show that the probability that the two largest cities come from the same region goes to 0. Supposing for simplicity that the N_i are equal, this probability is $K \Pr \{L^{(1)} = L_1^{(1)}, L^{(2)} = L_1^{(2)}\}$, where we have suppressed the conditioning on K and the N_i. Now the event $L^{(2)} = L_1^{(2)}$ implies that $L_1^{(2)} \geq L_j^{(1)}$, $j = 2, \cdots, K$. But $L_1^{(2)}, L_2^{(1)}, \cdots, L_K^{(1)}$ form an independent sequence of random variables, each $L_j^{(1)}$ having the same distribution. Since, for each x,

$$\Pr \{L_j^{(1)} \leq x\} = \Pr \{L_1^{(1)} \leq x\} \leq \Pr \{L_1^{(2)} \leq x\} ,$$

$L_1^{(2)}$ is stochastically smaller than the $L_j^{(1)}$. It is now plausible that the probability that $L_1^{(2)}$ is the maximum of the above sequence goes to 0 faster than K^{-1}, since K^{-1} would be the exact probability if $L_1^{(2)}$ had the same distribution as the others. It then would follow that $K \Pr \{L^{(1)} = L_1^{(1)}, L^{(2)} = L_1^{(2)}\}$ goes to 0 as $K \to \infty$, as desired. Now consider general R. Suppose that two or more of the R largest cities come from the same region. Then it follows that some $L_i^{(2)}$ must be among the R largest city sizes. This can only occur if that $L_i^{(2)}$ is larger than at least $K - R + 1$ of the $L_j^{(1)}$. Again, just as in the case $R = 2$, the fact that $L_i^{(2)}$ is stochastically smaller than the $L_j^{(1)}$ leads to the desired result, i.e., that the probability that two or more of the R largest cities come from the same region goes to 0. This result is proved in the appendix under a modest condition. We remark that for our application to Zipf's Law we only require that this probability be small enough so that most of the $L^{(r)}$, $r = 1, \cdots, R$, are likely to come from different regions. In this case a plot of $L^{(r)}$ against r will be equivalent to a plot of the rth largest of the $L_i^{(1)}$ against r, and thus follow Zipf's Law.

The three different situations, in which we have shown that the Zipf Law arises, suggest that the phenomenon should occur, at least approximately, under fairly general conditions. Violations of some of the conditions under which the limit theorems were obtained should not have too much effect—for example the N_i could be random and some dependencies between the various (M_i, N_i) could be allowed. The really critical assumptions are the assumptions of a Bose-Einstein allocation within regions, and that the M_i/N_i have a limiting distribution of the required form. We now show that a Maxwell-Boltzmann allocation within regions leads to a very different result. The proof follows closely the proof of Theorem 1, with the Bose-Einstein assumption of that theorem replaced by a Maxwell-Boltzmann assumption.

Suppose that, given M and N, the vector $\underline{L}$ has the Maxwell-Boltzmann distribution as defined by (1.2), i.e., the $L_i - 1$ have a multinomial distribution with M cells, $N - M$ units, and probability M^{-1} that a given unit falls in a given cell. By (2.2),

$$\Pr\{L^{(1)} \le a | M, N\} = \sum_{j=0}^{M} (-1)^j \binom{M}{j} \Pr\{V_1 \cap \cdots \cap V_j\} ,$$

where V_j is the event $L_j > a$. Let $M/N \to \Theta_o$, where $\Theta_o \neq 0, 1$. Since each $L_i - 1$ has a binomial distribution with $N - M$ trials and probability M^{-1} of success on a given trial, and since such a limit of binomial distributions converges to a Poisson distribution, it follows that

$$\Pr\{V_j\} = \Pr\{L_j - 1 \ge a\} \to p(a) = \sum_{i=a}^{\infty} e^{-\lambda_o}\lambda_o^i/i! ,$$

where $\lambda_o = \Theta_o^{-1} - 1$. Similarly, $\Pr\{V_1 \cap \cdots \cap V_j\} \to [p(a)]^j$ for each j. Just as in the proof of Theorem 1, we now define $a_N = a_N(\Theta_o, k)$ to be such that $Mp(a_N) \to k$, so that $\Pr\{L^{(1)} \le a_N | M, N\} \to e^{-k}$, where $k > 0$. But $p(a) \sim e^{-\lambda_o}\lambda_o^a/a!$ as $a \to \infty$, and using Stirling's formula it is easily verified that for $C_N = C \ln N/\ln \ln N$, we have $Mp(C_N)$ goes to 0 if $C > 1$, and goes to ∞ if $C < 1$. It follows that $\Pr\{L^{(1)} \le C_N | M, N\} \to 1$ if $C > 1$, and $\to 0$ if $C < 1$, so that $L^{(1)}[\ln N/\ln \ln N]^{-1}$ converges in probability to 1. This behavior is very different from that in the Bose-Einstein case, where $L^{(1)}/\ln N$ converges in probability to $-[\ln(1 - \Theta_o)]^{-1}$. Since the same result holds for $L^{(R)}$, we see that Zipf's Law cannot arise for a Maxwell-Boltzmann allocation scheme.

5. MOTIVATION AND DISCUSSION OF ASSUMPTIONS

I have presented a simple mathematical model, of which both the rank-frequency and generic-specific forms of Zipf's Law are a consequence. It is therefore a model which yields results that are in substantial accord with great quantities of data drawn from a variety of scientific areas. For this reason alone it would be of interest to inquire as to the plausibility of the assumptions of that model in each of the areas to which Zipf's Law seems to apply. In each such area, however, such a study would be a major undertaking of itself. Use of the geographic and taxonomic examples in this article was not meant to imply that such assumptions have been verified in these situations, but merely that they are worthy of consideration, and of being tentatively entertained, precisely because the examples are known to provide remarkably good fits to Zipf's Law. I prefer to think of the model as a skeletal model having a simple and general structure, which can be fleshed out in different ways in different scientific areas. The basic structure, ingredients and assumptions of the model will now be described, and so far as possible, motivated. Although, for vividness, the city-size and generic-specific terminology will again be used, it is hoped that the reader will view the model in a more general sense.

The first assumption is that of regions (families), in which the Θ_i for different regions (families) are approximately independent, and similarly the allocations of people to cities within regions (species to genera within families) are also approximately independent. It is also assumed that despite such approximate independence (or autonomy) between regions, similar underlying forces or constraints operate within the various regions, thus tending to produce approximately the same form of distribution for the Θ_i, and the same method of allocation of people to cities, within the various regions. In any application there will of course be some flexibility as to the precise definition of a region. For example, on a global scale, such regions might in fact each be countries of a certain type, while on a national scale, they might be certain natural geographic-political-economic entities. In the context of personal income distribution, the role of a region might be played by a corporation or even an entire industry, while in the context of word frequency usage, it might be played by a paragraph or perhaps a chapter. The model only requires that there be some way of decomposing the totality of units into subsets for which the assumptions are approximately valid. So far, the assumptions are of a largely qualitative nature, and do not seem wholly unreasonable.

The next assumption concerns the settling down, for large N_i, of the distribution of Θ_i, to some limiting distribution F, common to the various regions. Such settling down seems plausible on the basis of continuity considerations. Thus one would not anticipate that slight changes in N_i, for large N_i, would greatly alter the distribution of Θ_i, given N_i. The use of a continuous distribution F as the limiting distribution provides, in the usual way, an approximation to the true discrete distribution, for large N_i. The assumed similarity of the underlying forces and constraints within the various regions then suggests that the same F should be used for each region.

The next assumption is that $F(x) \sim Cx^\gamma$ as $x \to 0$, for some $\gamma > 0$. The family of such distributions is a large one, so this assumption is not particularly restrictive. Moreover, the values $0 \leq \alpha \leq 1$, for the rank-frequency form of Zipf's Law, proportional to $r^{-(1+\alpha)}$, correspond to particularly simple distributions F, e.g., $\alpha = 0$ corresponds to $\gamma = 1$ (which includes the uniform distribution), while $\alpha = 1$ corresponds to $\gamma = \frac{1}{2}$ (which includes the arc sine distribution). Since there is substantial evidence that $0 \leq \alpha \leq 2$ in a variety of scientific areas, this seems to me to lend considerable support to the entire model. By contrast, if in order to obtain Zipf's Law it had been necessary to choose F of some unusual and exotic form, then this would have made the model somewhat less plausible.

In my opinion, the really critical assumption of the model is that of a Bose-Einstein allocation of people to cities within regions (species to genera within families). Although the Bose-Einstein distribution need not hold literally to obtain Zipf's Law, gross violations will lead to a very different result. For example, if the Bose-Einstein allocation were replaced by a Maxwell-Boltzmann allocation, then, as shown previously $L^{(1)}[\ln N/\ln \ln N]^{-1}$ converges in probability to 1 as N grows large, and a plot of $L^{(r)}$ against r will be much flatter than under Zipf's Law. The essential difference between the two allocation schemes seems to be that under the Bose-Einstein model there will be much greater variability in city sizes within a region, including typically a few extremely large cities, while under the Maxwell-Boltzmann model such city sizes tend to be more nearly equal. I conjecture that other allocation schemes implying substantial variability of city sizes within regions will also yield an approximation to Zipf's Law. In fact, I would now like to give an interpretation of the Bose-Einstein model which may serve both to make it more palatable as an assumption, and also to shed light on the mathematics and robustness of the Zipf Laws. Suppose that N units are to be allocated to M nonempty cells, and define U_j to be the index of the cell to which the jth unit goes, $j = 1, 2, \cdots, N$. The cells are indexed say by the integers $1, 2, \cdots, M$. Let $U_1 = 1$, $U_2 = 2$, $\cdots$, $U_M = M$, to insure that all cells are occupied, so that the remaining $n = N - M$ units can be allocated without any constraints. Letting L_i be the total number of units in the ith cell, $i = 1, \cdots, M$, it is well known that a Bose-Einstein distribution for $\underline{L} = (L_1, \cdots, L_M)$, $L_i \geq 1$, $\sum_{i=1}^M L_i = N$, is obtained by introducing a probability vector $\underline{p} = (p_1, \cdots, p_M)$, $\sum_{i=1}^M p_i = 1$, such that given $\underline{p}$, the n remaining units generate a multinomial distribution, with the p_i as cell probabilities, while $\underline{p}$ itself has a uniform distribution on the simplex $\sum_{i=1}^M p_i = 1$, i.e., $\underline{p}$ has the Dirichlet distribution with all parameters equal to unity. This follows as a special case of the more general result for $\underline{p}$ having a Dirichlet distribution with density

$$\phi(\underline{p}) = \Gamma(\textstyle\sum_{i=1}^M \alpha_i) \prod_{i=1}^M p_i^{\alpha_i - 1} \{\prod_{i=1}^M \Gamma(\alpha_i)\}^{-1}, \quad \alpha_i > 0 \ .$$

Here, if $\underline{l} = (l_1, \cdots, l_M)$, $l_i \geq 0$, $\sum_{i=1}^M l_i = n$, then it is easily verified that

$$\Pr\{\underline{l}\} = \frac{n!\Gamma(\sum_{i=1}^M \alpha_i) \prod_{i=1}^M \Gamma(l_i + \alpha_i)}{\Gamma(n + \sum_{i=1}^M \alpha_i) \prod_{i=1}^M \Gamma(\alpha_i)\Gamma(l_i + 1)}, \quad (5.1)$$

which reduces to $\binom{N-1}{M-1}^{-1}$ when $\alpha_i \equiv 1$. The Maxwell-Boltzmann model can be interpreted as the limiting case $\alpha_1 = \alpha_2 = \cdots = \alpha_M \to \infty$ of (5.1). For the general Dirichlet-multinomial model, the conditional distribution of $\underline{p}$, given $U_1, \cdots, U_{M+r}$, has a Dirichlet density proportional to

$$\prod_{i=1}^M p_i^{l_i(r)+\alpha_i-1} ,$$

where $l + l_i(r)$ is the number of these $M + r$ units in the ith cell, and so

$$\begin{aligned} \Pr\{U_{M+r+1} = i \mid U_1, \cdots, U_{M+r}\} &= E[p_i \mid U_1, \cdots, U_{M+r}] \\ &= [l_i(r) + \alpha_i]/(r + \textstyle\sum_{i=1}^M \alpha_i) \ . \end{aligned}$$

In particular, under the Bose-Einstein model, where $\alpha_i \equiv 1$, the probability that a next unit goes to cell i, given the allocation of the previous units, is proportional to the number of units already in that cell. Similarly, if n^* additional units are to be allocated to the cells, then the expected number to be added to cell i is $n^*\{l_i(r) + 1\}/(M + r)$ under this model. For city sizes, the Bose-Einstein model thus implies that expected increases in city size are proportional to existing size, which seems plausible as an approximation. Furthermore, I would conjecture that replacing the Bose-Einstein model by a general Dirichlet-multinomial model, where $\underline{p}$ has a non-degenerate symmetric distribution, should also yield Zipf Laws in much the same way that the Bose-Einstein model does, which would perhaps partially explain the wide applicability of these laws. The degeneracy of the distribution of $\underline{p}$ under the Maxwell-Boltzmann model, which tends to make city sizes too nearly equal, would also perhaps account for the very different behavior under this model.

From one point of view, we have merely replaced the mysterious phenomenon of Zipf's Law by an equally mysterious set of assumptions. However, such assumptions, which are of a simple and general character, can be studied separately in each of the areas where Zipf's Law seems to apply, and by noting discrepancies, one can hope to modify them appropriately. Furthermore, it seems to me that it is the surprising quantity and variety of data fitting Zipf's Law that strongly suggests the need for a general skeletal model, such as the one presented here, which can then be fleshed out in a corresponding variety of ways. Support for this model derives from the fact that it is both simple and general, not wholly unreasonable as an approximation, and most importantly, that it is consistent with the vast collections of data obtained by Zipf and others.

I anticipate, however, that many scientists will reject the Bose-Einstein model, even as an approximation, in regard to their particular area, on the grounds of its being too simple to describe the complex phenomena in question. I sympathize with such a view, but at the same time I wonder whether this is not possibly a matter of not seeing the forest because of the trees. For example, in regard to city size, to someone who is familiar with the enormous complexity involved in birth rates, migration, etc., the Bose-Einstein Law will surely seem far too simple. Similarly, to a taxonomist, familiar with all the tedious and complex analysis that goes into the classification of organisms, not to mention the complexity of the underlying evolutionary process, the Bose-Einstein distribution must seem unreasonably simple. But perhaps such intimate knowledge of the details only obscures what, from the view of an outsider, might in some respects be a relatively simple process. It does not seem too outlandish to imagine that the net result of a large number of complex interacting forces might be such a simple process. It would then be of interest to try to determine how the Bose-Einstein allocation might arise out of such complexity. In this connection, it should be understood that ultimately one is dealing with a dynamic process where changes occur as a function of time. For example, Zipf's Law for city sizes has held until very recently, but the development of suburbia seems to have altered matters to a certain extent. A more sophisticated model than that presented here would deal with the dynamics of the situation, and not merely the one-dimensional view obtained at a given point in time.

Of course, other models have been formulated, and other approaches taken, in regard to justifying the Zipf (or Pareto) Law, most notably those of Mandelbrot [6] and Simon [8]. There are some interesting relationships between the various models for Zipf's Law, and comparisons between these models will be considered in a subsequent article.

APPENDIX

We now prove a lemma relative to the probability that two or more of the R largest cities come from the same region.

Lemma: Let $Y_1, \cdots, Y_K$ be K independent nonnegative random variables, Y_1 having absolutely continuous distribution $W_1(\cdot)$, and $Y_2, \cdots, Y_K$ all having the common absolutely continuous distribution $W_2(\cdot)$, where $W_2'(y) > 0$ for $y > 0$. Suppose $W_1'(y)/W_2'(y) \to 0$ as $y \to \infty$. Then for any fixed R, $K \Pr\{Y_1$ has rank R in the sequence$\} \to 0$ as $K \to \infty$.

Proof:

$$K \Pr\{Y_1 \text{ has rank } R\} = K\binom{K-1}{R-1}\int_0^\infty [1 - W_2(y)]^{R-1}[W_2(y)]^{K-R}W_1'(y)dy$$

$$= K(K-R+1)^{-1}\binom{K-1}{R-1}\int_0^\infty [W_1'(y)/W_2'(y)] \cdot [1 - W_2(y)]^{R-1}dP(y) \ ,$$

where $P(y) = [W_2(y)]^{K-R+1}$. Thus,

$$K \Pr\{Y_1 \text{ has rank } R\} = K(K-R+1)^{-1}\binom{K-1}{R-1} \cdot E\{[W_1'(X)/W_2'X][1 - W_2(X)]^{R-1}\} \ ,$$

where X is a random variable having distribution $P(\cdot)$, i.e., X is distributed like the maximum of $K - R + 1$ independent random variables from the distribution $W_2(\cdot)$. But

$$K(K-R+1)^{-1}\binom{K-1}{R-1}E\{[1 - W_2(X)]^{R-1}\} = 1 \ ,$$

since each of K independent random variables with distribution $W_2(\cdot)$ is equally likely to have rank R. Next, since X converges in probability to ∞ as $K \to \infty$, and since $W_1'(y)/W_2'(y) \to 0$ as $y \to \infty$, it follows that $W_1'(X)/W_2'(X) \xrightarrow{p} 0$ as $K \to \infty$. In fact, breaking the expectation up into the integral from 0 to C_K, and the integral from C_K to ∞, where $C_K \to \infty$, we have

$$E\left\{K(K-R+1)^{-1}\binom{K-1}{R-1}\left[\frac{W_1'(X)}{W_2'(X)}\right][1 - W_2(X)]^{R-1}\right\} \to 0 \, .$$

Here it is required only that $C_K \to \infty$ and be such that

$$K^R[W_2(C_K)]^{K-R} \to 0 \quad \text{as} \quad K \to \infty \, .$$

This completes the proof of the lemma.

An important illustration of this lemma occurs when $W_1'(y) = e^{-y}$, $W_2'(y) = ye^{-y}$.

The application of the lemma is as follows. As discussed in Section 4, if two or more of the R largest cities are from the same region, then some $L_i^{(2)}$ must be among the R largest city sizes, and so some $L_i^{(2)}$ must be larger than at least $K - R + 1$ of the $L_j^{(1)}$. Thus

$$\Pr\{\text{two or more of the } R \text{ largest are from same region}\}$$
$$\leq K(R-1)\Pr\{L_1^{(2)} \text{ has rank } R - 1 \text{ in the sequence } L_1^{(2)}, L_2^{(1)}, \cdots, L_K^{(1)}\} \, .$$

Assuming for simplicity that the N_i are equal, it follows that the $L_j^{(1)}$ are identically distributed. Since the $M_iN_i^{-1}$ are mutually independent, and the Bose-Einstein allocations within the various regions are also independent, $L_1^{(2)}, L_2^{(1)}, \cdots, L_K^{(1)}$ form a sequence of independent random variables. If we could apply the lemma with $Y_1 = L_1^{(2)}$, $Y_j = L_j^{(1)}$, $j = 2, \cdots, K$, we would obtain $K \Pr\{L_1^{(2)}$ has rank $R - 1$ in sequence$\} \to 0$, as $K \to \infty$, yielding the desired result. However, the $L_j^{(r)}$ have a discrete distribution. But although, for simplicity, the lemma was proved in the absolutely continuous case, where the possibility of ties could be ignored, a similar result holds in the discrete case too. Here the condition $W_1'(y)/W_2'(y) \to 0$ would be replaced by $\Pr\{L_1^{(2)} = y\}/\Pr\{L_2^{(1)} = y\} \to 0$ as $y \to \infty$. Since $L_1^{(2)}$ is stochastically smaller than $L_2^{(1)}$, it is at least plausible that such a condition might hold.

It should be remarked that to obtain our result we must let the $\ln N_i$ go to ∞ relatively slowly as compared to K. For, by the remarks following the proof of Theorem 2, $L_i^{(r)}/\ln N_i$ is distributed approximately like $\ln[X_{(r)}/N_i\Theta_i]/\ln N_i \ln(1 - \Theta_i)$ as $N_i \to \infty$, where $\Theta_i = M_iN_i^{-1}$; and so if the N_i went to ∞ with K fixed, then the $L_i^{(r)}[\ln N_i]^{-1} + [\ln(1 - \Theta_i)]^{-1}$ would all converge in probability to 0, $r = 1, 2, \cdots, R$. In this case, the region with the smallest Θ would tend to have the R largest cities, which is contrary to the desired result. On the other hand, if $K \to \infty$ sufficiently fast relative to the $\ln N_i$, we obtain the desired result. For moderate K and $\ln N_i$, it seems plausible that the probability that several of the R largest cities come from the same region will be small, but the question is delicate.

[*Received September 1973. Revised March 1974.*]

REFERENCES

[1] Berry, B. and Garrison, W., "Alternate Explanations of Urban Rank Size Relations," *Annals of the Association of American Geographers*, 48 (March 1958), 83–91.

[2] Feller, W., *An Introduction to Probability Theory and its Applications, Vol. 1*, 2nd ed., New York: John Wiley and Sons, Inc., 1957.

[3] Hill, B.M., "Zipf's Law and Prior Distributions for the Composition of a Population," *Journal of the American Statistical Association*, 65, No. 331 (September 1970), 1220–32.

[4] ——— and Woodroofe, M., "Stronger Forms of Zipf's Law," University of Michigan Technical Report No. 31, 1974.

[5] Mandelbrot, B., "A Class of Long-Tailed Probability Distributions and the Empirical Distribution of City Sizes," in *Mathematical Explorations in Behavorial Science*, F. Massarik and P. Ratoosh, eds., Homewood, Ill.: Richard D. Irwin Inc. and the Dorsey Press, 1965, 322–32.

[6] ———, "The Pareto-Lévy Law and the Distribution of Income," *International Economic Review*, 1 (May 1960), 79–106.

[7] Rényi, A., "On the Theory of Order Statistics," *Acta Mathematica Academiae Scientiarum Hungaricae*, 4 (December 1953), 191–232.

[8] Simon, H.A., "On a Class of Skew Distribution Functions," *Biometrika*, 42 (December 1955), 425–40.

[9] Wilks, S.S., *Mathematical Statistics*, New York: John Wiley and Sons, Inc., 1962.

[10] Yule, G.U., "A Mathematical Theory of Evolution Based on the Conclusions of Dr. J.C. Willis, F.R.S.," *Philosophical Transactions B*, 213 (May 1924), 21–87.

[11] Zipf, G.K., *Human Behavior and the Principle of Least Effort*, Reading, Mass.: Addison-Wesley Publishing Co., 1949.