Informetric Distributions, Part II: Resilience to Ambiguity

A. Bookstein
University of Chicago, Center for Information and Language Studies, Chicago, IL 60631

This article continues the discussion of the informetric distributions begun in a companion paper. In the earlier paper, the informetric distributions were introduced and found to be variants of a single distribution. It was suggested that this might be explained in terms of that distribution being unusually resilient to ambiguity. In this paper the notion of resilience to ambiguity is made precise. By way of introduction, a number of simple examples of resilience, taken from the social sciences, are discussed. This approach is then applied to the informetric distributions themselves. It is argued that the form taken by the informetric regularities does indeed make them insensitive to the wide range of ambiguities that occur when measuring the output of social activity, and that this ubiquitous form is unusual in having this property.

Introduction

In "Informetric Distributions, Part I," the companion to this article, I described a variety of regularities that have been discovered in the social sciences (Bookstein, 1990). In that article, I first argued that, if we abstract from the subject content, many of these regularities are in fact describing, in different ways, the same type of data: yields as distributed over a population of items. I then tried, beginning with one mode of description, to find the exact form taken if the description were expressed in one of the other modes; on doing this, I was able to show that the variety of regularities commonly appearing in the literature are, to good approximations, simply reexpressions of the same distribution. That any regularity at all appears in the phenomena we are studying is surprising. That these regularities are actually identical is even more so. We are, in fact, studying a broad range of different phenomena. Thus, if we can understand better the properties and occasions of occurrence of this one recurring form, we at the same time gain insight into a wide range of social science regularities.

One approach to understanding the regularities is to seek a substantive explanation for each. While such an approach might be successful for one or another regularity, it seems unlikely that it can explain why a vast range of phenomena have descriptions that reduce to a single form. Rather, to be fully satisfying, it would seem that an explanation is called for that is insensitive to the detailed underlying causes. Further, given the persistence of ambiguities, which seems to be inherent in the social sciences, we must also demand that the form taken by these regularities be resistant to the ambiguities in which the phenomena these regularities describe are enmeshed.

Actually, an examination of some of the mathematical functions most commonly and successfully used in the social sciences reveals the existence of a resiliency of this kind, though this property has not generally been made explicit in discussions of these functions. Perhaps examining a number of prominent, simple examples is the most natural way to introduce the more extended discussion for the informetric laws.

Examples of Robustness in the Social Sciences

In this section, I shall describe a few of the most prominent mathematical functions occurring in the social sciences, emphasizing their resilience against ambiguity.

Poisson Distribution

The Poisson distribution often appears in the social sciences to describe the number of events that occur in a given time interval. Examples include automobile accidents, typing errors, occurrences of noncontent bearing words in text, and fatalities in the Prussian cavalry due to horse kicks (Feller, 1957). In all of these, the time span is arbitrary. Upon consideration, we would expect that if, in some sense, an underlying regularity exists, the form that ultimately emerges should not be strongly dependent on the time interval that is actually selected for collecting the data. This is true of the Poisson process, though this property is not usually noted as an example of this distribution's robustness characteristics. For example, if the random variables $\tilde{x}_1, \ldots, \tilde{x}_n$, are Poisson distributed, so is their sum $\tilde{x} = \tilde{x}_1 + \cdots + \tilde{x}_n$, where the expected value of the sum is the sum of the expected values. Thus, if the above

Received August 21, 1989; revised November 20, 1989; accepted November 27, 1989.

counts are Poisson distributed for a one-month span, they will also be Poisson distributed for a one-year span, as the number of events in a year is equal to the sum of the number of events in each of the months. This argument can be generalized: The Poisson distribution is often derived as the solution of a Markov birth process, where the probability, $P_n(t)$, that the number of events occurring over a period of time, t, is n is given by

$$P_n(t) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}; \tag{1}$$

in equation (1), λ is a constant denoting the rate at which events are being generated. Here the independence of the Poisson *form* on the time variable is explicit; that is, if this process describes the phenomenon of interest, a Poisson distribution will describe the number of events that occur, no matter what time interval we choose. Were this not the case, that is, if it had turned out that a Poisson distribution would be appropriate, but only for a single, specified time interval, then it is highly unlikely that this regularity would have been discovered, or that the Poisson distribution would be as popular as it is. This property is parallel to the well-known stability of the normal distribution to sums and means.

Exponential Decay

The exponential distribution has been used to describe phenomena as varied as the decline in the number of citations to an article over time to the service time in a queue. As is the case for equation (1), the unit of time is implicit in the detailed form taken by the exponential function; however, it is easily seen that changing this unit merely changes the distribution's parameter, leaving the exponential form resilient to changes in time unit.

To educe other resiliency properties of the exponential distribution, consider the phenomenon of exponential growth. For example, suppose the number of scientists who have ever lived up to time t, $N(t)$, grows exponentially; thus

$$N(t) = N(0)e^{\alpha t} \tag{2a}$$

for some constants $N(0)$ and α. It has been noted (deSolla Price, 1963) that this implies that the number of "new" scientists also grows exponentially. For, if a new scientist is one who came into existence in the last t_0 years, then there are

$$\begin{aligned} n(t) &= N(0)\,[e^{\alpha t} - e^{\alpha(t-t_0)}] = N(0)\,(1 - e^{-\alpha t_0})e^{\alpha t} \\ &= n(0)e^{\alpha t} \end{aligned} \tag{2b}$$

new scientists at time t. This is also an exponentially growing function. It is interesting to note in the context of this discussion that the exponential form does not depend on the time interval, t_0, chosen to define a "new" scientist. One reason the exponential function is so popular is that if it describes a phenomenon, it appears for a wide range of ways in which one might collect the data in search for a regularity: whether one collects cumulative data, or whether one collects data in increments, the form taken is the same, and further, the interval of time over which the data are collected is irrelevant.

Learning Curves

It has often been noted that the cost of producing an item, whether it be an airplane or a microprocessor, decreases by a constant factor every time the cumulative quantity produced doubles (Johnson, 1976). The functional form describing this is

$$C(Y) = C_0 Y^{\alpha} \tag{3}$$

where $C(Y)$ is the cost per item at the point in time at which Y is the cumulative amount produced, α is a constant less than zero, and C_0 is the cost per item of producing the initial item. Suppose we are at the point in time when the yield is Y and progress to the point when Y becomes $2Y$. We then find that $C(Y)$ has become $2^{\alpha}C(Y)$, revealing 2^{α} to be the factor, independent of Y, describing the reduction in cost due to learning.

The description of the learning process by means of equation (3) is (approximately) true whether we are describing the incremental cost per item, as is the case here, or the average cost of the total number of the items produced. For suppose that in equation (3) $C(Y)$ described the average cost per item. Then the total cost would be $YC = C_0 Y^{\alpha+1}$ and the incremental cost for the next item would be given by

$$\begin{aligned} &C_0(Y + 1)^{\alpha+1} - C_0 Y^{\alpha+1} \\ &\qquad = C_0 Y^{\alpha+1} + (\alpha + 1)C_0 Y^{\alpha} + \cdots - C_0 Y^{\alpha+1} \\ &\qquad = (\alpha + 1)C_0 Y^{\alpha}\left[1 + O\left(\frac{1}{Y}\right)\right] \end{aligned}$$

Thus, for Y well above one, the incremental cost also has the form of equation (3). In particular, plotting C vs Y on log-log paper would show an approximately straight line *regardless of which way the cost was assessed*. This resiliency has increased the likelihood of that functional form being discovered by increasing the modes of data collection for which it is valid.

Also important is the independence of the form of the learning curve on the initial unit of output used—for example, whether units of output are measured in pounds or tons. For suppose equation (3) does indeed hold for some initial unit of output, but that I had decided to search for a regularity using some other unit—say, k times the original unit. Then, where I measure Y units now, I would have measured Y/k units originally. Since, for the original units, the law holds, I can conclude that now the relation between C and yield will be given by

$$C = C_0\left(\frac{Y}{k}\right)^{\alpha} = \left(\frac{C_0}{k^{\alpha}}\right) Y^{\alpha} = C_0' Y^{\alpha}.$$

That is, if the learning curve takes the form of equation (3) in the original units, it will do so in the new units as well;

it will even have the same percent decline in cost per doubling period,with only the initial cost changing. This perhaps explains why so many different industries report figures around a 20% decline in cost at each doubling of output. Since the unit of yield may indirectly reflect the choice of a unit of time, for example a month or a year, the above argument also indicates a resilience against a change of time unit.

Morse's Markov Model

It is interesting and useful to study patterns of library book circulation from the vantage point introduced above. The most influential model here was introduced by Morse (1968). Morse's model relates circulation magnitudes in one year to that of following years. It is made up of two parts:

(1) A book that circulates C_{t-1} times in year $t - 1$ will circulate, *on the average,*

$$\lambda_t = \alpha + \beta C_{t-1} \tag{4a}$$

times in the following year. That is, the *expected* circulation this year is linearly related to the *actual* circulation last year and is not influenced by earlier years' experience. I am considering the expected circulation of a book to be a parameter intrinsic to that book for the year considered.

(2) Given the expected circulation of a book for a given year (say, year t), the actual circulation in that year, C_t, is governed by a Poisson distribution:

$$Pr\{C_t\} = e^{-\lambda_t}\frac{\lambda_t^{C_t}}{C_t!}. \tag{4b}$$

Thus the probability of a book circulating C times in any given year is influenced through the parameter λ by the previous year's circulation; however, given the previous year's circulation experience, earlier years have no impact. That is, book circulation exhibits the Markov property.

In Morse's development, the states of the Markov model are the actual number of times a book circulates in a given year. The state transition probabilities are governed by the Poisson distribution, through the intermediation of the parameter λ. Specifically, $\tilde{C}_t = \lambda_t + \tilde{E}_t$, where $\tilde{E}_t$ is a random error, with a mean of zero, described in terms of the Poisson distribution.

Morse used an academic year as the natural time interval over which to measure circulations. To analyze the resilience of this model against a change of time interval, it is illuminating first to modify the model slightly. I shall assume that the states of the model are the λ's (not the actual circulation) and that it is the λ's that satisfy the linear transition equation,

$$\lambda_{t+1} = \alpha + \beta\lambda_t \tag{5}$$

In any given year, each book has a value of λ associated with it, and this parameter is a measure of the book's potential for circulation. The λ's themselves are unobservable, but actual circulation will be an approximation of this quantity. In this model, the λ_t's either monotonically increase without bound as t increases ($\beta \geq 1$) or monotonically converge to a limit ($\beta < 1$). As such it is not realistic as a model of a single book, though perhaps it may describe a *class* of books.

To test this model for resilience against a change of time interval, consider what happens when we consolidate successive pairs of years into single intervals. Then the new states will be $\Lambda_1 = \lambda_1 + \lambda_2$, $\Lambda_2 = \lambda_3 + \lambda_4, \cdots, \Lambda_n = \lambda_{2n-1} + \lambda_{2n}, \ldots$. The transition equation implies

$$\begin{aligned}\Lambda_n &= \lambda_{2n-1} + \lambda_{2n} = (\alpha + \beta\lambda_{2n-2}) + (\alpha + \beta\lambda_{2n-1}) \\ &= 2\alpha + \beta(\lambda_{2n-2} + \lambda_{2n-1}),\end{aligned}$$

and $\lambda_{2n-2} + \lambda_{2n-1}$ can similarly be expanded, finally yielding

$$\Lambda_n = 2\alpha(1 + \beta) + \beta^2(\lambda_{2n-3} + \lambda_{2n-2}) = \alpha' + \beta'\Lambda_{n-1},$$

where I am setting $\alpha' = 2\alpha(1 + \beta)$ and $\beta' = \beta^2$.

We thus find that, for the simplified model, the robustness property *does* hold for the transition equation for two-year merging; one can generalize this argument for arbitrary n-period merging.

Another interesting robustness criterion, especially given that the λ's are not observable, is robustness under class ambiguity. That is, at time t a class of N books is made up of n_i books from class λ_{it}, for $i = 1, 2, \cdots$. Given the interpretation of λ as an expected circulation, λ_t for the combined class will equal

$$\frac{(n_1\lambda_{1t} + n_2\lambda_{2t} + \cdots)}{N}.$$

But $\lambda_{1,t+1} = \alpha + \beta\lambda_{1t}$, etc., so

$$\lambda_{t+1} = \frac{\sum n_i}{N}\alpha + \beta\frac{\sum n_i\lambda_{1t}}{N} = \alpha + \beta\lambda_t.$$

Thus if it is meaningful to speak of a "pure" λ book, an arbitrary collection of such books from different classes still obey the transition rule: this rule is robust under class ambiguity. We can also speak in a meaningful way of the λ of a class of books, with the state, λ, developing linearly from year to year.

How might such a model be tested empirically? λ itself can't be observed, but since $\tilde{C} = \lambda + \tilde{\varepsilon}$ for a single book, the average circulation over a class of books will equal $\bar{\lambda} + \bar{\varepsilon}$, where the bars denote averages over the class. However, the standard deviation of $\bar{\varepsilon}$ for a class of any significant size will be small, and $\bar{C}$ becomes a good estimator of $\bar{\lambda}$. Thus, for a class of books, it is difficult to distinguish C and λ. It is interesting to note that this is precisely the test performed by Morse, who, in table 5.3 and Figure 5.1 of his book, shows excellent fit to the model for a one-year transition: the linear transition component of the modified model has both theoretical and empirical strength.

Although the simplified model may describe a set of books, for a single book the state must have a stochastic influence. Let us, then, consider the somewhat more complex (though still simpler than Morse's) model

$$\lambda_{t+1} = \alpha + \beta\lambda_t + \varepsilon_t, \tag{6}$$

that is, the *expected* circulations are still linearly related from year to year, but there is now an additional random error, ε_t. I shall assume this error has zero mean and is independent of other error terms. If we take an average of a large number of such books, the impact of the error becomes negligible and the preceding model returns. Let us apply the same argument used above to prove resilience under time-interval merging to the new model. If we do so, we now find

$$(\lambda_{2n-1} + \lambda_{2n}) = \alpha' + \beta'(\lambda_{2n-3} + \lambda_{2n-2}) + \varepsilon_{2n-1} + (1 + \beta)\varepsilon_{2n-2} + \beta\varepsilon_{2n-3},$$

with constants α' and β' simply related to α and β. This has the superficial form

$$\Lambda_n = \alpha' + \beta'\Lambda_{n-1} + E_{n-1},$$

which is similar to the model we are studying. However, because of the component ε_{2n-1}, the error E_n in successive intervals of the new model *will not be independent.* Since the assumption of independent error terms is critical, we must conclude that the modified model is not resilient to time-interval merging. (The argument for robustness under class-ambiguity is still valid, however.)

Since $C = \lambda + \varepsilon$, the model as proposed by Morse is similar to my second model, but introduces, in effect, yet another error term. Thus, regarding the linear transition assumption, we conclude:

(1) The Morse model is not robust: it depends critically on the choice of time interval being correct; however,
(2) Taken as a group property, the mischievous error term becomes small and the model approximates a simpler model for which the robustness property holds; finally,
(3) It is the group version which in effect Morse has tested and verified in his book.

I can deal much more briefly with the remaining assumptions in the Morse model. If $\tilde{C}_1$ and $\tilde{C}_2$ are governed by Poisson distributions with means λ_1 and λ_2 respectively, then $\tilde{C}_1 + \tilde{C}_2$ is indeed governed by a Poisson distribution, and with mean $\lambda_1 + \lambda_2$. Thus, if the rule for the time development of the means were robust under time-interval merging, the Poisson assumption would combine with it to form a coherent model that was also robust. Given that the rule, as defined by Morse, is *not* robust, this consideration is not relevant. But do the data support the Poisson assumption? An examination of Morse's tables 5.5 and 5.9 shows that the Poisson distribution consistently has too small a variance, although the distribution does roughly describe the data. For example, for the 240 books circulating once, the model predicts none will circulate more than four times. In fact, three of the books circulated four times and one nine times. Similarly, for books that have a large circulation this year, the model systematically underestimates the number of items that will have low circulation values next year.

Interestingly, a possible explanation for this discrepancy that could save the attractive Poisson assumption may be the Morse model's adoption of circulation rather than expected circulation as the Markov state. That is, interpreted in terms of the modified model, the class of books circulating, say, once this year would be made up of books from *several* states (that is, having several values for the parameter, λ). Thus, the circulation rates next year would be governed by a mixture of Poisson distributions, which would increase its variance.

Clearly, given the discussion above, we similarly conclude that the Markov assumption is not robust. Of course, while robustness is highly desirable for social data, it is not required: there may be processes for which a designated interval is natural. Book circulation in a university library may well be such a process. The ultimate test is whether the data support the model. Do Morse's data satisfy the Markov property? Curiously, although Morse's book circulation model has universally come to be called "Morse's *Markov* model," and Morse himself referred to it as a Markov model, the Markov property was not tested in Morse's book—nor, so far as we are aware, anywhere else. To test the Markov assumption, at least three years of data are needed. An example of such a test would be as follows: One might, for example, look at books that circulated once in year $t - 2$ and five times in year $t - 1$; these books might be compared to the class of books that circulated nine times in year $t - 2$ and also five times in year $t - 1$. Thus both sets of books circulated five times in $t - 1$, although they differ in their circulation yield the preceding year. If the Markov property is met, both sets should have an identical pattern of circulation for year t—for example, the means of the two sets should be approximately equal. Morse himself does not report on such a test.

Morse's model is interesting as an example of the analysis we are describing here. It is a very important and frequently cited model, so its properties are interesting in their own right. But theoretically, it offers us an example of a model that is *not* robust, but which, especially when group properties are the focus of study, approximates a model for which robustness criteria are met.

ARIMA Models

Morse's Markov model is one example of a prediction model. Others exist and are widely applied. Very prominent, and potentially competitive with Morse's model, is the class of ARIMA models (Box & Jenkins, 1976).

Suppose we select a time interval and denote the measured yield in that interval by y_t. Simple special cases of the ARIMA model would include the following:

(1) AR(1) model: $y_t = \alpha + \beta y_{t-1} + \varepsilon_t$ (auto regressive model) and,
(2) MA(1) model: $y_t = \alpha + \varepsilon_t + \beta\varepsilon_{t-1}$ (moving average model).

In these models, the ε's are independent, identically distributed random disturbances.

Such models have been applied in many fields and have been intensively studied. The AR(1) model is very similar to the modified Morse model discussed above (where expected values rather than yield were modeled). Here I only wish to note that, although these models are very attractive and probably underutilized for informetric research, they are not robust under time-interval merging. The analysis showing this is similar to that of the preceding section: the occurrence of the error terms implies that the error terms in the merged models will no longer be independent. For example, given the MA(1) model: the merged yield, $Y_n = y_{2n-1} + y_{2n}$ satisfies $Y_n = 2\alpha + (\varepsilon_{2n-1} + \varepsilon_{2n}) + \beta(\varepsilon_{2n-2} + \varepsilon_{2n-1})$. The error term is $\varepsilon_{2n-1} + \varepsilon_{2n}$. Note that the term reflecting past disturbances, $(\varepsilon_{2n-2} + \varepsilon_{2n-1})$, is no longer independent of the current disturbance, and furthermore, does not equal the preceding disturbance of the merged model. Thus, for both models, careful attention to time period is crucial when these models are used.

Robustness of Informetric Laws

In this section I shall explore the consequences of demanding that the informetric laws be resilient to ambiguity.We already saw that Bradford's law exhibits such a property inasmuch as it does not depend on identifying a specific core of items (Bookstein, 1990). I showed that, at minimum, if Bradford's law is valid for any given core, it is also valid for any multiple of that core; but if we begin with Leimkuhler's variant of Bradford's law, and assume we can extrapolate to non-integral values, Bradford's law is found to be valid for any core at all.

Up to now, we have not considered the dynamics of the processes yielding the informetric laws nor the consequences of assuming such resilience to ambiguity. I shall now turn to this task. The model I shall use is a refinement of the one first developed in Bookstein (1977). In this model, I shall consider a set of objects producing events over time (or some dimension resembling time) and ask, if we concentrate on a given time interval, how many objects will have a yield of y events.

One of the difficulties in discussing the informetric laws is the variety of ways in which one can interpret what exactly it is that the law is claiming. The most direct interpretation, that the laws are to be understood literally (that, for example, *exactly* $f(y)$ scientists publish y papers), must immediately be rejected. This interpretation makes no allowances for inevitable statistical fluctuation, and imposes the inconvenience of having to interpret predictions of fractional items, as when $3\frac{1}{2}$ scientists are predicted to have published a given number of papers. More satisfying is to consider the laws as statements about probability distributions, or, alternatively, as statements about expected values. Pareto's law, for example, is often expressed in probabilistic terms: the *probability* that a person earn more than y dollars is proportional to $y^{-\alpha}$. I shall find an interpretation in terms of expected values more convenient. Suppose, then, that the regularity we are studying can be expressed as follows: Each object in our population has an integer valued yield that can be measured. But also associated with each object is a parameter that cannot be measured and that may possibly take nonintegral values: its expected yield. To emphasize the distinction between yield and expected yield, I will use the letter x to denote the latter. The expected value is probabilistically related to the actual yield, y. I shall be interpreting the informetric distributions as statements about the expected yield: the number or density of objects, $f(x)$, having an *expected* yield of x is proportional to some function, h, of x:

$$f(x) = Ah(x). \tag{7}$$

A, which is a positive constant, can be fixed by setting $h(1) = 1$. Since we are interpreting x as an expected value, it can take any value, even though the quantity we ultimately measure is restricted to take integer values. This is very important for our considerations below, since it permits h to be a continuous, indeed, differentiable, function on the real numbers.

The relation between an expected value and the value that is actually measured is given in terms of a probability model. Suppose the probability that an item with an expected yield of x will in fact have a yield of n (an integer) is given by $P(n|x)$, with $P(n|x)$, as a function of n, sharply peaked around x. Let $\tilde{d}_i(n)$ be an indicator random variable for the ith item: it takes the value one if the ith item has a yield of n; otherwise it takes the value zero. Then the number of items with a yield of n is given by $\sum_i \tilde{d}_i(n)$; the expected value of the number of items with a yield of n is given by

$$E\left(\sum_i \tilde{d}_i(n)\right) = \sum_i E(\tilde{d}_i(n)) = \sum_i P(n|x_i). \tag{8}$$

Each term in the sum is conditional on x, which varies from item to item. If, as assumed in equation (7), the distribution density of x is proportional to $h(x)$, this sum is proportional to $\int P(n|x)h(x)\,dx$, which will approximate $h(n)$, given the assumption that $P(n|x)$ is sharply peaked at n near x. Thus, we expect the number of items with a yield of n, for n an integer, to be approximated by $Ah(n)$, if the density of the expectations is proportional to $h(x)$, and that this will be true even if x can take on non-integer values. For example, if $h = 1/x^2$ and $P(n|x)$ is a Poisson distribution, then the expected number of items generating n events is given exactly by $A/n(n - 1)$ (Bookstein, 1976), which is, indeed, approximately A/n^2 *for* $n \gg 1$. Taking this approach allows us explicitly to include variability in the model, and to recognize that even when we measure integer numbers of events, the potential for production, interpreted as an expected value, is most realistically measured along a continuum. Below, I shall indeed assume a continuous yield variable, consistent with our expected value interpretation. I am interested in uncovering the ambiguities inherent in this process, and learning how to characterize those functions, h, that are insensitive to these ambiguities. Please keep in mind, though, that these models are of the abstract variable, expected yield, and not of the quantity actually measured. The above argument, however, suggests that these are closely related.

Change of Time Segment

In Bookstein (1977), I was concerned with the arbitrariness of the time segment chosen when yield data are collected. For example, suppose some regularity is found to hold for some fixed time segment. Will it hold for a different time segment? To answer this question analytically, we must have some mechanism that relates the generation of events in one period to that in another. The approach I shall be taking here and below is to create a conceptual model population of entities that observe an informetric regularity under a given set of conditions. I shall further design this model population so that its members behave in a simple and predictable way over time, which allows me to consult them to check how their yield patterns change as conditions change. Thus, if I cannot perform real experiments with live scientists, I shall be able to perform thought experiments with the scientists dwelling in the cooperating model population. If the regularities we are studying are robust within this model population, and this robustness persists over a wide variety of stresses, we have some reason to believe that in a real population, the regularity would be at least approximately stable over the types of changes we are considering. I thus shall be carrying out experiments on an available, artificial population to draw inferences about a more realistic, but inaccessible and analytically more complex population.

An especially simple example of this methodology is its application to the question of stability under time interval change. As a workable first approximation, I shall have the members of the model population produce events at a constant rate. With this assumption, it is possible to change the time segment into one s times as long and ask what form the density function will take when I count events in the new interval.

On the one hand, I am assuming within my model that items are producing events at a constant rate, so any item that produced an expected value of x events in the first interval will produce an expected value of $x' = sx$ events in the second interval, and the number of items having expected values between x' and $x' + \Delta$ in the new interval will have expected values between x'/s and $(x'/s) + (\Delta/s)$ in the old interval—that is, between x and $x + (\Delta/s)$ events. But we know how many items produced between x and $x + (\Delta/s)$ events in the original interval—our model reveals this to be $(Ah(x)/s)\Delta$. Thus, the value of the density function for the new interval is $Ah(x)/s$: so much follows from the assumptions about our model population. (The factor $1/s$ is in effect the Jacobean of the transformation between the old and new intervals.)

But I am also arguing that the law ought to take the same form in all intervals. If so, the value that we just computed for the density function must be proportional to $h(x')$; that is, in terms of the values taken in the earlier interval, the probability density at x' should equal $A'h(sx)$. Thus, we can conclude

$$A'h(sx) = \frac{Ah(x)}{s}.$$

Setting $x = 1$ and recalling that $h(1) = 1$, we find

$$\frac{A}{s} = A'h(s),$$

so $h(s)$ must satisfy

$$h(sx) = h(s)h(x). \tag{9}$$

Thus, if we insist on the regularity being valid for *any* time span, we are rather severely constrained in the choice of function, h. Indeed, suppose $h(x)$ is defined in a neighborhood of $x = 1$, is differentiable at $x = 1$, and satisfies the above constraint everywhere that all terms are defined. Then, for small Δ,

$$h(x + \Delta) = h\left(x\left(1 + \frac{\Delta}{x}\right)\right) = h(x)h\left(1 + \frac{\Delta}{x}\right),$$

and,

$$\frac{h(x + \Delta) - h(x)}{\Delta} = \frac{h(x)}{x} \cdot \frac{h\left(1 + \frac{\Delta}{x}\right) - h(1)}{\frac{\Delta}{x}}.$$

Since we are assuming that $h(x)$ has a derivative at $x = 1$, the following limit is defined:

$$\frac{d}{dx}h(x) = \lim_{\Delta \to 0}\left[\frac{h(x + \Delta) - h(x)}{\Delta}\right] = \frac{h(x)}{x}\frac{d}{dx}h(1).$$

The constraint of equation (9) is seen to be a very strong one: to satisfy it, if $h(x)$ has a derivative at one value of x, it must also have a derivative for all x at which it is defined. (If $h(x)$ is defined at $x \neq 0$, the definition can be extended, in a manner consistent with equation (9), to a neighborhood of x—the proof of this claim is implicit in the above derivation of the derivative of $h(x)$.)

Since $(d/dx)h(1)$, the derivative of $h(x)$ evaluated at $x = 1$, is just a constant, the derivative at a general x takes the form

$$\frac{d}{dx}h(x) = A\frac{h(x)}{x}.$$

This first order differential equation has as its general solution

$$h(x) = Ax^{\alpha} \tag{10}$$

for α an arbitrary constant. Thus we find not only that the function Ax^{α} satisfies the constraint, but that it does so uniquely, and we expect A/n^{α} to estimate the number of items with an actual integral yield of n.

Entries and Exits from Population

The arguments leading to the functional form, $h(x) = x^{\alpha}$, were based on the assumption that we were describing a stable population of researchers. Although the finding that the function x^{α} is the only "nice" one that can describe such a population is in itself quite remarkable, it is only natural to ask what happens if we relax the restriction that the population be stable. Specifically, I am asking what happens if we allow researchers to enter and leave the population over the period of time, T, that we are examining. In Bookstein (1977), I showed that the form $h(x) = x^\alpha$ is maintained if items enter and leave the population at arbitrary rates, provided that the distribution in yield production of the items entering and leaving the population is the same as that of those initially in the population. Further, once again, x^α is the only form for which this is true. This extension makes the model more realistic as a description of, for example, scientists or journals, which come into existence, flourish, and then depart. It also permits the model to describe Wyllis' law, in which genera disappear and new ones come into existence.

Varying Rates of Production

But in a realistic population, items don't generate events at a fixed rate. For example, there are periods in which science is more productive than in other periods, and there are geological periods in which new species appear at a greater or lesser rate than usual. To the extent that this is explained by, for example, an increase in the number of productive scientists, we have seen that the model already includes such growth. But, to some extent, this is also a result of the individual productivity of each individual changing. I shall now examine this possibility.

I shall first define a reference time interval, T. Suppose, then, that over this reference interval time, $Ah(x)$ is the density of items expected to produce x events each. Consider a different time period, T'. Given our model, there are two ways we can determine how many items are expected to produce x events in the new time period. First, recalling the robustness property being assumed, this quantity must be given by $A'h(x)$, for some constant A'.

On the other hand, if we know how the rate of productivity is changing, and if we know how many items can be expected to produce x events in the reference period, we can calculate how many items should produce x events in the new period. For suppose that the changing rate of productivity can be expressed in terms of some function $r(t)$: that is, for every event that an item can be expected to generate over the reference time span, it can be expected to produce $r(t)\,dt$ events in the period from t to $t + dt$ within the new time span, T'. For example, if we had assumed exponential growth with a rate equal to α, then $r(t) = e^{\alpha t}$. However, our results will be more general if we do not restrict the pattern of growth to be of any specific kind.

Denoting by R the integral $\int_T r(t)\,dt$, we may conclude that an item generating a single event through interval T will generate R events in T'. Thus, an item producing x events over the reference period will produce approximately xR events in the new time period, and the density function for the new time period is given by $(Ah(x/R))/R$; as in the earlier derivation, the factor $1/R$, corresponding to the Jacobean of our transformation, is required because we are dealing with a continuous x variable and must take into account the change in size of the differential dx over the two time periods. Equating the two expressions for the density of items expected to produce x events in the reference time span, we conclude

$$A'h(x) = \frac{Ah\left(\frac{x}{R}\right)}{R},$$

or, letting $s \equiv 1/R$,

$$A'h(x) = Ash(sx).$$

Since $h(1)$ is assumed to be 1, A' must equal $Ash(s)$, and h must satisfy the constraint of equation (9).

$$h(s)h(x) = h(sx).$$

As above, I shall assume that $h(x)$ has a derivative at $x = 1$, and that this constraint holds over all values of x. If so, we once again obtain equation (10).

We find that the functional form of equation (9) is preserved even if a population is changing its rate of event production in an arbitrary way, provided the new rates tend to influence all items in the same way. These results give us new insight into the meaning of laws such as Lotka's. The existence of Lotka's law can be considered to be a statement regarding the inertia of productivity patterns. We now see that such an interpretation cannot be taken to mean that the productivity of researchers in a discipline is fixed over time. Indeed, the stability of Lotka's law is completely consistent with the recognition that a discipline will experience slow periods, when, for whatever reason, research levels off, and periods of accelerated growth—for example, following a major discovery. So long as these variations over time tend to influence all the members of the discipline in the same way, Lotka's law will be preserved.

Multiple Authorship

Another difficulty that must be resolved by investigators trying to verify Lotka's law of scientific productivity is how to treat multiple authorship. The problem of multiple authorship and its effect on the law was of concern even for Lotka, at a time when multiple authorship was less common than it is today. Lotka refers to this problem in explaining deviations from the law he was proposing:

> Fortunately, however, there are somewhat more persons of very great productivity than would be expected under this simple law. The very high figures (e.g., Abderhalden, 346 contributions in 10 years) should perhaps be considered separately, since they are not the product of one person unassisted. Joint contributions have in all cases been credited to the senior author only (Lotka, 1926).

Today, with multiple authorship much more the norm, the problem of how to treat it has become aggravated. Consideration of this problem, however, suggests that it is much more complex than may immediately meet the eye. For once we recognize the ambiguity in deciding for a paper with, say, three authors, whether each author is to be credited with a single paper or one third of a paper—that is, once we accept the possibility that a person associated with a paper may be given credit for a fraction of a paper—it is natural to ask whether every paper necessarily deserves a unit of credit even if it is written by a single author. For example, should a minor note be given the same weight as a report of a major piece of research? And, in the case of multiple authorship, should each author be given the same degree of credit—Lotka, it will be noted, resolved the problem by giving full credit to the senior author and to him alone.

Once we recognize the degree of ambiguity in the determination of authorship, we must again marvel that a regularity as striking as that discovered by Lotka be found. For supposing that Lotka's law did indeed hold for some distinguished system of counting authorship, it would be very unlikely that the system would in fact be the one used by anyone trying to verify, or discover, regularities in scientific productivity. Yet many researchers have searched for such regularities in a number of different fields, each with their own means of resolving the problem of multiple authorship, and the pattern first formulated by Lotka seems always to recur. I shall now show that if Lotka's law is valid for one accounting system, it will be valid for any other, provided certain regularities obtain; furthermore, Lotka's $1/x^\alpha$ law is unique in being invariant under changes of counting, for x the expected yield in published articles. Thus, if we would find a $1/x^\alpha$ relation to describe productivity when we give a full publication to every author whose name appears on a paper, this will also be the case if we had assigned fractional authorship instead.

Let the number of people whom we expect to publish between x and $x + dx$ articles over a certain time span be given by $Ah(x)\,dx$ for some function $h(x)$ and constant A; once again, since we are dealing with expected values, x need not be an integer. The constant A is again defined by the constraint that $h(1) = 1$. Given that $h(x)$ describes the population for *some* basis of accounting, we ask how this function is changed if we change the accounting system, for example, to the one we are actually using.

Suppose then, that an individual has his name on N papers: n_1 with his name alone, n_2 with one other person, etc. Then we might say, for example, that:

(1) he published N papers, or, perhaps, that
(2) he published

$$n_1 + \frac{n_2}{2} + \frac{n_3}{3} + \cdots = \left[\frac{n_1}{N} + \frac{n_2}{2N} + \cdots\right]N$$
$$= rN \text{ papers}.$$

More generally, if we adopt an accounting system that gives the author an amount of credit v_i for paper i, we will attribute to him $x = \sum_{i=1}^{N} v_i = rN$ publications, where the last equation defines r. The value r is a measure of the average fractional credit per paper given to a particular author. It is on the basis of this system of accounting that we find there are $Ah(x)\,dx$ people who are expected to publish between x and $x + dx$ papers.

If we had adopted a different accounting system, the author would have been assigned instead $x' = r'N$ papers; thus, for that author, $x' = (r'/r)x$. If we let $\Theta \equiv r'/r$, we have $x' = \Theta x$ for that author; in general, Θ, which will take different values for different authors, translates the amount of credit given, per paper, to an author by one system of accounting to that given him by another such system.

We can now ask approximately how many authors will be expected to publish x' papers in our new accounting system. We can see that those authors who will now be credited with x' papers include those who previously were credited with x' papers and had $\Theta = 1$, as well as those, for example, who were previously credited with $4/5x'$ papers, but for whom $\Theta = 5/4$, etc.; in general, we now credit with x' papers all those authors with translation factor Θ, who, before, we credited with x'/Θ papers. To get $A'h'(x')$, the density function for the total number of scientists expected to publish x' papers, we must "add up" all contributions of this type. More precisely, if $f(\Theta_0)\,d\Theta$ denotes the fraction of scientists for whom Θ is between Θ_0 and $\Theta_0 + d\Theta$, we can conclude

$$A'h'(x') = A\int \frac{1}{\Theta} h\left(\frac{x'}{\Theta}\right) f(\Theta)\,d\Theta;$$

the extra factor $1/\Theta$ compensates, as before, for the change in the size of the interval dx before and after the transformation. I shall assume below that $f(\Theta)$, here defined as describing the change of credit as we change accounting methods, does not depend strongly on x'.

In terms of this model, the original problem can be posed by the question: is it possible that $h'(x) = h(x)$? That is, can

$$A'h(x) = A\int \frac{1}{\Theta} h\left(\frac{x}{\Theta}\right) f(\Theta)\,d\Theta.$$

If so, note that, since $h(1) = 1$, we must have

$$A' = A\int \frac{1}{\Theta} h\left(\frac{1}{\Theta}\right) f(\Theta)\,d\Theta,$$

so,

$$h(x)\int h\left(\frac{1}{\Theta}\right)\frac{f(\Theta)}{\Theta}\,d\Theta = \int h\left(\frac{x}{\Theta}\right)\frac{f(\Theta)}{\Theta}\,d\Theta. \qquad (11)$$

We can immediately observe that this is true for Lotka's function $h(x) = 1/x^2$, or more generally, $h(x) = 1/x^\alpha$, as $h(x)h(1/\Theta)$ in this case equals $h(x/\Theta)$; that is,

$$\frac{1}{x^\alpha}\frac{1}{\left(\frac{1}{\Theta}\right)^\alpha} = \frac{1}{\left(\frac{x}{\Theta}\right)^\alpha}.$$

Thus, we can conclude that if Lotka's law holds for one accounting method, it will hold for any other one in which the change in the typical amount of credit given to authors per paper may vary from author to author, but does not depend strongly on how much the author published. If this is true, the investigator is free to adopt any reasonable system of assigning credit, and can be confident that if Lotka's system isn't observed, it is not because he chose the "wrong" means of attributing articles to authors. This is most fortunate because, as I noted before, it would be very unlikely that a law not enjoying this property could be observed.

But it is possible to make an even stronger and more startling assertion. If we insist that $h(x)$ take a form such that equation (11) would be true no matter what form $f(\Theta)$, and thus $f(\Theta)/\Theta$, happened to take, then $h(x)$ *must* satisfy the now familiar functional equation $h(sx) = h(s)h(x)$, for s and x arbitrary. This result follows from equation (11), after substituting $s = 1/\Theta$, and recognizing that $\int f(x)g(x) = \int h(x)g(x)$ for all possible integrable functions g only if $f(x) = h(x)$ almost everywhere (in the Lebesgue sense). If the functions f and h can be assumed to be continuous, then $f(x) = h(x)$ everywhere. As we saw, the only continuous functions satisfying this constraint take the form $h(x) = 1/x^{\alpha}$. Thus, not only is the generalized Lotka law independent of how we count publications, it is the only law of that type that has this property!

I began this section by asking the questions: What would happen to Lotka's law if we tried to verify it by using the "wrong" method of counting articles? and, implicitly, how could a law of the kind Lotka discovered be found at all? The surprising results we found were that Lotka's law is not sensitive to how we count articles, so that two people testing the law for a single population, but using different counting methods, will very likely come up with the same law. Furthermore, it seems unlikely that any other law has this property, or, for that reason, could have been discovered. We focused on the problem for multiple authorship, but any counting ambiguity would be covered as well.

The problem I am formulating here, with regard to Lotka's law, has direct parallels in other areas of information science. I showed, for example, that Lotka's law and Bradford's law of scattering are formally equivalent. Parallel questions, regarding Bradford's law, are: How do we count articles when attributing them to journals? How should we treat peripheral articles, touching on the subject of interest but not centrally about it? Can deviations from Bradford's law be explained in terms of our not having counted articles properly? We now see that Bradford's law is not sensitive to how we count articles: if one way of deciding how much weight to assign an article will do, any other way of counting will do as well. Recognition of this resiliency property of the Bradford distribution helps us to understand how a law of this type is capable of being discovered and verified in a wide variety of different fields.

Combining Classes

The next example of ambiguity I shall consider is that of classification. For example, when Lotka studied the publication productivity of chemists, he had to define a chemist as someone cited in *Chemical Abstracts*. Although this definition has the merit of convenience, it is easy to imagine someone else arguing that the authors listed in *Chemical Abstracts* in fact represent a range of intellectually separate disciplines; indeed, if Lotka had not found the pattern he did find, he very reasonably might have explained this as a consequence of how his source collected its authors. A similar difficulty arises in the other regularities. For example, the confusion as to what constitutes a word (are "boy" and "boys" one or two words?) has always haunted researchers interested in Zipf's law. Yet, despite these difficulties, seeking after examples of both Lotka's law and Zipf's law tends to meet with success. How is this to be explained?

The Lotka formulation is the easiest to work with, and, since these laws are very similar, explaining the phenomena in terms of Lotka will explain it for all the forms. Suppose, then, that unknown to us, several disciplines exist that independently observe a Lotka type of law: in each, the number of items generating an expected value of x events is given by A/x^{α}. Suppose, however, that because of our ignorance of what properly constitutes a category, we make our measurements on the union of the underlying classes. Then, the number of items, f, producing x events will be given by

$$f(x) = \sum_i \frac{A_i}{x^{\alpha_i}}.$$

First note that if α_i is constant, then $f(x) = A/x^{\alpha}$, for $A = \sum A_i$ and α the common value of the α_i's; thus, in this simplest case, the Lotka form is retained. Suppose that the α_i's are *not* the same, but, consistent with experience, vary but slightly about a common value, α. Specifically, let $\alpha_i = \alpha + d_i$ for d_i small compared to α, and α chosen so $\sum A_i d_i = 0$ (that is, α is a weighted mean of the α_i's). Then, taking a Taylor series around α,

$$\begin{aligned} f(x) &\approx \sum A_i x^{-\alpha} \\ &\quad - \left(\sum A_i d_i\right) x^{-\alpha} \ln(x) + \left(\sum \frac{A_i d_i^2}{2}\right) x^{-\alpha} \ln(x)^2 \\ &= \sum \frac{A_i\left(1 + \dfrac{d_i^2 \ln^2(x)}{2}\right)}{x^{\alpha}}. \end{aligned}$$

Thus, instead of $1/x^{\alpha}$ being multiplied by a constant factor A, it is multiplied by A plus a correction term: $(\sum A_i d_i^2/2)\ln^2(x)$, which introduces an additional x dependency. This term can be rewritten as

$$\sum \frac{A_i}{2}\left(\frac{d_i}{\alpha}\right)^2 \ln^2(x^{\alpha}).$$

Since the d_i are small compared to α, and since $\ln(x^{\alpha})$ varies slowly compared to x^{α}, the correction term will be small and approximately constant for large ranges of x. Thus, $f \approx A'/x^{\alpha}$, for stretches of x, and the Lotka format is approximated again, though a plot of $\log(f)$ vs. $\log(x)$ should show slight curvature over a wide range of x. In fact, on log-log paper, at large values of x the graph would be a straight line corrected by the addition of a term in $log(log(x))$.

Another concern we might have is that we may not be combining complete classes of items, but rather taking a portion of each class. However, if the selection mechanism within each class is not strongly dependent on the productivity of the items, then it is easily seen that the remnant of each class that appears in our sample will still obey Lotka's law, and thus so will the combined population. Once again, the distribution is resistant to ambiguity. However, unlike the preceding cases, this distribution is not unique in enjoying this property; also, the immunity is partial, depending upon the exponent, α, varying only slightly about its (weighted) mean.

Grouping

Often classes are formed by grouping ranges of values; for example, income may be indicated as \$15,000–\$15,999, \$16,000–\$16,999, etc. If A/x^{α} is the density of people with expected yield x, then, if our classes are of the form $0 - s, s - 2s, \ldots, (n-1)s - ns$, etc., we expect $\int_{(n-1)s}^{ns} Ax^{-\alpha}\,dx$ people in the nth class. At large n, this quantity approximates $(A/s^{\alpha-1})/n(n-1)^{\alpha-1}$, which itself quickly converges to A'/n^{α}; here the frequency is expressed in terms of class interval position. Alternatively, this quantity is equal to As/x^{α}, with x, or sn, denoting the number of events defining a class boundary.

It has been observed that people comprehend magnitudes logarithmically, so that natural classes would be $0 - s$, $s - s^2, \ldots, s^{n-1} - s^n$, etc. Using such classes, the limits of the integral are s^{n-1} and s^n. The integral evaluates to $A'/s^{\alpha(n-1)}$, or A'/x^{α}, with s^{n-1} designating the nth class. Thus, for this case, grouping returns Lotka's law exactly.

Conclusions

In the earlier paper (Bookstein, 1990) and the preceding sections of this article, I examined a variety of regularities taken from different fields and exhibiting a variety of forms. I first showed that these can be conceptualized as versions of a single regularity, so that we can properly speak of *the* informetric law and its manifestations. I next indicated a number of properties we would expect of a regularity in the social sciences if that regularity is to be discoverable, and argued that the informetric law did satisfy these conditions, often doing so uniquely. I have found that making these observations reduces the value of these laws for many people, who see this as a "debunking" of the laws. Perhaps this response is a consequence of our having learned to associate scientific laws with causal rather than technical explanations. But this has not been my intention. Rather, I am arguing that the place in statistics of the informetric law should be very much the same as that of the normal distribution, which also describes a wide variety of phenomena, and for a similar reason. Although we have to be suspicious of causal explanations within informetrics, it is nonetheless reasonable to try to understand the circumstances under which a distribution of this form should be expected. Much of this paper explores this issue.

We also would like to learn how to take advantage of the law when it does appear. It is reasonable to ask, then, how might one use these regularities? Do they make a difference? Ultimately, I believe the answer to this question involves searching for situations in which decisions must be made based upon distributional assumptions. To defend the use of the informetric distributions, it must be shown that decisions made on the basis of these distributions will be better than, or comparable to, those made without them (Bookstein, 1988).

Many decision making problems in information science depend on a frequency distribution describing activity. An example from coding theory is the creation of a Huffman code, which depends on the frequency of occurrence of characters making up the source alphabet (Heaps, 1978). Planning the allocation of space for an index of an information retrieval system is another. In Bookstein (1988), the problem of dividing a library book budget over segments of the collection is examined, showing in detail how distributional assumptions make a difference.

A number of alternative distributions have been suggested as being better fits of the data we are studying (Burrell, 1988; Sichel, 1985). An advantage of the informetric distributions, besides their familiarity, is their simplicity. But whether this simplicity justifies their use will depend on how well or poorly they assist us in making decisions as compared to their competition.

Acknowledgment

This article and its companion have been brewing for some time and reflect many debts. I would like in particular to acknowledge the insights and encouragement of Dr. Leo Egghe, who has seen several early drafts and who has discussed with me some of the more difficult problems inherent in this topic. This article began with research supported by NSF Grant IST-8026564 and is based on the report submitted in 1982. The author wishes to acknowledge the NSF support.

References

Bookstein, A. (1976). The bibliometric distributions. *Library Quarterly, 46*, 416–423.

Bookstein, A. (1977). Patterns of scientific productivity and social change: A discussion of Lotka's law and bibliometric symmetry. *Journal of the American Society for Information Science, 28*, 206–210.

Bookstein, A. (1990). Informetric distributions, part I: A unified overview. *JASIS, 41*.

Bookstein, A. (1988). Applications of the bibliometric distributions. In Egghe, L., and Rousseau, R., (eds), Informetrics. 87/88. Amsterdam: Elsevier Science Publishers, 1988.

Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco: Holden Day.

Burrell, Q. L. (1988). Predictive aspects of some bibliometric processes. In Egghe, L., and Rousseau, R., (eds), Informetrics. 87/88. Amsterdam: Elsevier Science Publishers, 1988.

deSolla Price, D. J. (1963). *Little science, big science*. New York: Columbia University Press.

Feller, W. (1957). *An introduction to probability theory and its applications* (2d ed), vol. 1. New York: Wiley.

Heaps, H. S. (1978). *Information retrieval—computational and theoretical aspects*. New York: Academic Press.

Johnson, R. D. (1976). *Quantitative techniques for business decisions*. Englewood Cliffs, New Jersey: Prentice-Hall.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16, 317–23.

Morse, P. (1968). *Library effectiveness: A systems approach*. Cambridge, MA: MIT Press.

Sichel, H. S. (1985). A bibliometric distribution which really works. *JASIS*, *36*, 314–321.