193

Reprinted from INFORMATION AND CONTROL, Volume 4, Nos. 2-3, September 1961

INFORMATION AND CONTROL 4, 305–308 (1961)

Reply to Dr. Mandelbrot's Post Scriptum

HERBERT A. SIMON

Carnegie Institute of Technology, Pittsburgh, Pennsylvania

Dr. Mandelbrot has proposed a new set of objections to my 1955 models of the Yule distribution. Like his earlier objections, these are invalid.

If Dr. Mandelbrot's *Post Scriptum* were limited to points previously discussed in our interchange on the Yule distribution, I would not bore readers with a reply denying that I "implicitly concede" all his points. The reader would have discerned for himself that no such concession has been made. However, as in his previous replies, Dr. Mandelbrot in his *Post Scriptum* abandons his earlier arguments, which I refuted, and brings forth a whole spate of new ones. These should not go unchallenged—especially since they are as faulty as those they replace.

Again, I shall follow Dr. Mandelbrot's *Post Scriptum*, section by section. In I, having admitted that his criticism was based on an assumption I did not use, he now claims to derive the same conclusion from the correct assumption. In II, he retreats from his criticism of the case of slowly decreasing $n'(k)$, by now calling it "practically equivalent" to the case of $n'(k)$ constant, and hence "asymptotically analytically circular," whatever that means. In III, he repeats an earlier argument based on an erroneous assumption about the initial conditions of the process. In Section IV he states a preference for one approximation over another, then restates a conclusion that cannot validly be derived from either approximation. In V, he repeats his earlier incorrect assertions on several minor matters.

I. REMARKS ON THE FIRST ASSUMPTION

Assumption I′, which I did not use, is an assumption about the probability that any particular word will be the next one chosen; the much weaker Assumption I, which I did use, is an assumption about the probability that the next word chosen will come from a particular stratum of words. Dr. Mandelbrot tries to equate these two assumptions by pointing out that at the upper end of the distribution each word (or city) belongs to a distinct stratum. What he fails to mention is that I raised this same objection in my 1955 paper, and showed how to modify the model to answer it (1955, p. 437):

> "This assumption [Assumption I] is certainly satisfied at least roughly. Moreover, it need not hold for each city, but only for the aggregate of cities in each population band. Finally, the equation would still be satisfied if there were net migration to and from cities of particular regions, provided the net addition or loss of population of individual cities within any region was proportional to city size."

Dr. Mandelbrot is correct in pointing out that this generalization leads to models of the type proposed by Champernowne, a fact I earlier pointed out on page 438 of my 1955 paper. In fact, this illustrates one of the central points of my 1955 paper—that there is a whole group of closely related stochastic models that yield the Yule distribution and variants thereof. As I said in my conclusion (p. 440):

> "The probability assumptions we need for the derivations are relatively weak, and of the same order of generality as those commonly employed in deriving other distribution functions. . . . Hence, the frequency with which the Yule distribution occurs in nature . . . should occasion no great surprise."

What all these models have in common are assumptions corresponding to my Assumption I (constant returns to scale), and corresponding to my Assumption II (relatively steady rate of new entry). The particular virtue of my model and its variants is that it explains why ρ is very often close to 1, and it gives good predictions of $f(i, k)$ for small as well as large i in cases where this is appropriate (e.g., word frequencies). Thus my "implicit concession" regarding city sizes and income distributions amounts to a reiteration of the conclusions of my 1955 paper.

As far as words are concerned, the predicted frequencies of the most frequent words do not, contrary to Dr. Mandelbrot's assertions, become "ridiculously low" for the actually observed values of $n'(k)$,which give a ρ much closer to 1 than to 1.1. I have now confirmed this by Monte Carlo runs of my model with samples up to 2.2 x 10^5, using $n'(k)$ taken from actual word counts. In the largest sample tested thus far, the observed frequency of the most frequent word was 9868, the frequency predicted by my model, 9241.

II. REMARKS ON SLOWLY DECREASING $n'(k)$

I find nothing in my reply corresponding to the alleged "narrowing of claims." The words "very slow decrease" were, again, not initially mine, but quoted from Dr. Mandelbrot, and were defined in my comments not to mean "such that $\sum n'(k)/k$ diverges," but to mean "rates . . . like those encountered in the data." The approximate model also fits the data well in cases where this sum appears to converge; its divergence is certainly a sufficient, but not a necessary condition for applying the model.

My models, contrary to Dr. Mandelbrot's unsupported assertion, are not based on assumptions "relative to first occurrences of words in a child's speech," nor do I recall claiming to have data on a child's first speech; but there exist several pieces of data that give $n'(k)$ for continuous pieces of prose, and these are quite consistent with my "absurd rationalization."

All the discussion of asymptotic behavior is as irrelevant as before [V − 2 of my previous reply (1961)]. "Analytic circularity" now seems to mean merely that the premises are strong enough to support the conclusions; it has been observed before this that all (correct) mathematics is a tautology. That my approximate model also handles the case where $\rho < 1$ I have now shown several times, beginning with pages 430–431 of my 1955 paper.

III. REMARKS ON THE DERIVATION OF $f(i, k)$

Far from failing to understand the integral transform, I showed in my reply that the correct transform is given by the Chapman-Kolmogoroff equations, which do not lead to Dr. Mandelbrot's conclusions, in particular the conclusion that $f(i)$ is "only a locally smoothed-out form" of $n'(k)$. For this conclusion, it is not sufficient for all the words "to have sometime passed through the value i^0." The conclusion would only hold if they had all passed through that value at the same time, which the Chapman-Kolmogoroff equations show they do not. Hence, I do not "implicitly accept" Dr. Mandelbrot's point about small i. In those cases (e.g., incomes, city sizes) where the model only holds for i above some minimum, the members of the population do not reach this minimum all at a single time, but at a relatively constant rate over time (Assumption II).

IV. REMARKS CONCERNING THE MATHEMATICS

I am sorry that Dr. Mandelbrot does not like my approximation; I do not like his. As I have just shown, neither approximation supports his conclusions, which depend on the mythical minimum i^0. The remainder of Section IV of the *Post Scriptum* is simply a series of assertions unsupported by evidence.

V. MISCELLANEOUS

On "improper" distributions, I have already had my say in Section VI-2 of my reply. Finally, Dr. Mandelbrot now defends his incorrect estimate of ρ on Zipf's Figure 9.7 by claiming the graph is incorrectly plotted! With this, I am left speechless, and I too hope that the question is now settled.

References

Mandelbrot, B. (1961). Final note on a class of skew distribution functions: Analysis and critique of a model due to H. A. Simon. *Information and Control* **4,** 198–216.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika* **42,** 425–440.

Simon, H. A. (1961). Reply to "Final Note" by Benoit Manbelbrot. *Information and Control* **4,** 217–223.