NEGATIVE BINOMIAL RATIONALE

THOMAS O. CARLSON

Volume XLIX, Page 177

DISCUSSION BY JOHN W. CARLETON

Mr. Carlson sets forth one of the reasons for writing his paper in these words: ". . . We are all interested in finding tools that work. But we should not be satisfied as actuaries without probing into any unfamiliar mathematical model until we know why it works, because thus only do we learn whether it is the best model for the purpose or whether is can be improved upon, and also what extensions of its utility may be available. . . ."

For some of us the utility of a model increases to the extent that it makes possible a visual image of something physical: Gears turning other gears where there is causal linkage, or colored balls being drawn out of an urn where the problem is that of defining the particular degree of absence of causal linkage. Models that make possible visual imagery may be a handicap to the investigator while he is pursuing his investigation, but they will help him communicate his findings to a larger audience after he has found something.

Thus, the concept of a Bernoulli distribution has a comforting tangible aspect when it is built upon a coin-tossing activity that anyone can easily picture, even if he has no intention of actually trying it out. The concepts of "likelihood" and even "equal likelihood," which are difficult to define without some circuity, are communicated painlessly by pointing at a coin. Each item of the distribution is understood to be determined quantitatively as the sum of a fixed number of contributions, additive or essentially additive, all small with respect to the total, and the variation of each contribution being independent of the variations of all others. The physical model gives clues as to what kinds of empirical distributions might be expected to follow the Bernoulli pattern, and perhaps some clues as to why others do not.

If the coins are thought of as being similar, then the information required to describe any Bernoulli distribution is very small and it should not be surprising that the formulas, even in their limiting forms, can be expressed by a very small number of parameters.

The next best thing to models that permit visualizing something physical are those that can be pictured on coordinate paper with one dimension of complexity partialed out. I think the recent papers on the negative binomial, at least in some respects, lend themselves to this treatment.

Picture a distribution of events occurring in a large number of exposure items as being the sum of some subdistributions, each generated by a subgroup of the exposure items. Spread the exposure groupings vertically along the Y axis of a piece of graph paper so that each can generate its subdistribution from left to right at some distance up from the bottom of the page. If the exposures are grouped and distributed by inherent hazards, and if each inherent hazard is assumed to generate a Poisson distribution of events or accidents, then you will be looking at a frequency contour analogous to the one prepared by Mr. A. L. Bailey and shown on Page 71 of the 1942 *PCAS*. I have trouble staying in touch with imaginary three-dimensional contours, so I'd prefer to think of the exposure items for each inherent hazard generating its Poisson distribution of events separately, after which all of the subdistributions can be added up and recorded as a total distribution across the bottom of the page.

If the distribution of inherent hazards, running up and down the Y axis, follows a Pearson Type III curve, then according to the authors of a number of recent papers, the distribution of accidents across the bottom of the page will follow a negative binomial or depart from it only by chance.

In Mr. Carlson's paper there is developed a distribution of the number of claims which I think can be set up and looked at in a similar way. Instead of using the Y axis to sort out the different inherent hazards into a frequency distribution of its own, it can be used to show on separate lines the separate distributions of accidents producing different numbers of claims per accident. The total line will be a claim count distribution. If the parameters of the Poisson formula for each of the subdistributions are connected in a particular way, then the claim count distribution will also follow the negative binomial pattern.

To the extent that I understand what Mr. Carlson has done, the Y axis would represent only the formula with which he connects the variables generating the distributions for each of the different numbers of claims per accident. It would not represent anything tangible that can be pictured in the imagination, like different numbers of exposure items (insured cars) grouped and arrayed by inherent hazards (the bad drivers at one end of the street, the good drivers at the other). I feel more comfortable with the latter and want to go back to it.

In the real world that brings forth empirical data on accidents, the inherent hazards that are arrayed up and down the Y axis will have certain quantitative characteristics that, whatever they are, can be described in a manner intelligible to statisticians by specifying the moments of their distribution. The more moments that can be measured, the more nearly the general characteristics of the frequency for curve can be bounded. Thus, I assume all frequency curves having the same first five moments look pretty much alike on graph paper, although I know of no reason why there should not be a very large number of curves, including freehand curves, that would satisfy the same five values.

It is believed the moments of the inherent hazard distribution can be determined from empirical data by comparing the empirical distribution of numbers of accidents with those that would be predicted by a Poisson distribution for the same average hazard. The greater the number of differences that can be taken with confidence, the greater the number of inherent hazard moments that can be estimated with a little algebra. These are the moments of an inherent hazard distribution that one infers must exist, if one is satisfied that the accidents generated by any single magnitude of inherent hazard should follow a Poisson distribution, and if one finds, as people have, that the empirical data don't quite do that.

(Parenthetically, I don't believe any of the recent contributors to the *PCAS* have commented on the correspondence between the model that underlies the Poisson distribution and the actual behavior of what Mr. Simon would call "isohazardous" exposure groups. One writer suggested, perhaps for a special development, that the hazard of each member of such a group must be assumed to be constant for the period of time over which the exposure unit is being accumulated. If so, the model is contra-indicated by the obvious changes in hazard as an insured car moves from a freeway to a garage. I don't believe the requirement is necessary. It is thought sufficient if (a) the members of the isohazardous group each have the same average hazard, and (b) fluctuations in the hazard of an individual member from hour to hour and day to day are unrelated to the accidents that fortuity occasionally brings forth. However, even these easy requirements suggest a possible difficulty: Would cyclical fluctuations in hazard intensity impair the criterion (b)? There is a feeling that they might.)

Is there any reason to believe that these moments of the inherent hazard distribution should lend themselves to being reproduced by a formula that has only a few parameters? I know of none. Aside from a few platitudes about continuity in natural phenomena, I know of no reasons why the inherent hazard distribution should not be multimodal, or at best the sum of a few subdistributions each of which has its own pattern.

The Pearson Type III is found to fit the inherent hazard distribution in the sense that when it (implicit in the negative binomial) is used along the Y axis, the total line fits the empirical data better than a Poisson distribution (zero variance along the Y axis) would. Since common sense suggests that some exposures have more inherent hazard than others, it seems possible that any inherent hazard distribution that can contribute a suitable amount of variance would be apt to permit a better fit than a single value distribution, which can contribute none. Is it known if the negative binomial (with its implied Pearson Type III distribution of inherent hazards) permits a better fit than could be accomplished if the Pearson Type III were replaced along the Y axis by some other distributions having the same mean and the same second moment, particularly by some freehand distributions?

"Freehand distribution" suggests a fuction that is obtained that way. I an using it to mean one that requires a very large number of parameters for its sufficient expression. Investigators, trying to find useful and meaningful descriptions of nature, usually grope for formulas with small numbers of parameters. In spite of this tendency a good deal of the world's work is done with smoothed tables of empirical data (mortality tables, seasonal corrections, magnetic compass adjustments, even Table M). Empirical data may have been smoothed by one device or another, but the smoothing devices seldom have any derivation from the structure of the multiple parameter formula that might have been there if there had only been enough data or enough insight to permit its discovery. Also, much of the world's work is done with tables prepared from simple functions like that of the normal curve. Thus, it's difficult to say that practical applications prefer formulas and accept tables only when formulas can't be found. What then is the fascination of the search for simple formulas to fit empirical data?

One motive might be to find or test an explanation of why the empirical data are as they are. The distinction between "to explain" and "to describe" may have become blurred at some levels of epistemology, but for immediate purposes I want to use the word "explanation" to cover something that helps me visualize a model within which I can see what produces the result.

Does the Type III Pearson curve purport to be the frequency distribution that can be expected when some definable factors are working on the individual items? In other words, is there a model that underlies it? I do not know whether there is or is not such a model. Has an analysis of the sources of hazard differences among exposure items suggested that they should be subject to analogous factors? In other words, does the Type III model, if it exists, look promising? With affirmative answers to both questions, a good fit would tend to support the inferences drawn from the analysis. Absent affirmative answers to either or both questions, the fit would seem to be coincidental. Moreover, searches for such fits, prior to dealing with such questions, would seem to be searches for such coincidences.

Such searches may be well worthwhile and yield many useful results, including those turned up through serendipity. However, some questions suggest themselves to which answers would be interesting: Do the conventional tests of Goodness of Fit apply to an undirected or trial and error search for a formula to fit some empirical data? Does testing a single hypothesis against some data call for different testing mathematics than starting with the data and then drawing at random from an infinite (or very large) available supply of formulas until one is found that seems suitable? Was the chi-square test built on the latter model? There is the intuitive notion that the random search should be shorter if the data are too thin to carry much information about the higher moments. Probably the notion is unfounded.

I hope these comments have some bearing on Mr. Carlson's concern with the rationale and the utility of models. Certainly his paper will stimulate others on claim count distributions.

DISCUSSION BY KENNETH L. McINTOSH

In this paper, deceptively simple in concept though perhaps not simple in mathematical detail, Mr. Carlson has accomplished three things, one of which possibly exceeds the limits of his own original objectives. First, the paper constitutes an excellent historical summary of various approaches to the negative binomial distribution in general, including presentation of one such approach in some detail. Secondly, the use of the factorial moment generating function is demonstrated. This extremely powerful mathematical tool is ignored by