COMMUN. STATIST.-THEOR. METH., 12(7), 841-848 (1983)

A GENERATING MODEL INVOLVING PASCAL AND LOGARITHMIC SERIES DISTRIBUTIONS

John Panaretos

Department of Statistics and Actuarial Science
The University of Iowa, Iowa City, Iowa 52242

1. INTRODUCTION

Let X be a positive, integer-valued random variable (r.v.) with $P(X=n) = g_n$, $n = 1,2,\ldots$. Let $\gamma(r|n)$ denote the probability that the value n of the r.v. X is increased by a generating process to the value r. If Y denotes the resulting r.v., then

$$P(Y=r) = \sum_{n=1}^{r} g_n \gamma(r|n), \quad r = 1,2,\ldots$$

where $\gamma(r|n) = P(Y=r|X=n)$. In the sequel, we will refer to the triplet $\{X,Y,\gamma(r\ n)\}$ as the generating model.

0361-0926/83/1207-0841$3.50/0

In its mathematical form the generating model can be considered as a mixture of the distribution $\gamma(r|n)$ over n where the mixing distribution is g_n with $r \geq n$. Models of similar type have been studied before. Thomas (1949), for example, studied the number of plants in a quadrant under the assumption that plants appear in clusters within the quadrant. Her "generating model" however was a generalization (random sum) of distributions rather than a mixture.

Situations where the generating model is applicable arise very often in practice. For example, in stochastic abundance models and especially in partitioning a random sample of individuals from a certain population into species. Suppose that we first classify the individuals according to the size of the group each species forms. Then X is the size of the group ($X > 0$) and Y is the total number of individuals in our sample. Alternatively, X may represent the number of groups observed ($X > 0$) and Y the number of individuals in the sample. (For various models and details see Engen (1978, 1979)).

In operational research problems, particularly in the study of a certain class of inventory control problems the generating model can also be considered. X can be defined as the number of demands placed per unit time ($X > 0$) and Y as the total number of items requested per unit time (Williamson and Bretherton (1964)).

In the context of accident theory X may represent the number of fatal accidents and Y the number of fatalities. Alternatively, X may be the number of injury accidents in a given locality within a certain period of time and Y the number of resulting injuries. (For a similar problem see Panaretos (1981)).

In the next section we study a special case of this model by assuming that the generating mechanism $\gamma(r|n)$ is known. In particular, we consider the "generation" to be effected through a Pascal distribution (not an unreasonable choice for the generating process, as this ensures that $Y \geq X$). We also consider X to follow the logarithmic series distribution (LSD), a distribution with a wide range of applications (see Johnson and Kotz (1969)). We then

derive the form of the unconditional distribution of Y and also that of the distribution of Y given that X=Y (i.e., given that we sample only observations when we know that no generation has taken place). We also establish a relationship between these two distributions.

In section 3 we point out the similarities and differences of the generating model to the well-known in the literature damage model (Rao (1965)). Finally, in section 4 an example is given where the results are illustrated and their implications are discussed.

2. THE LSD-PASCAL GENERATING MODEL.

Before giving the results we introduce some notation.

We say that X follows the LSD (a) if

$$P(X=n) = d\,\frac{a^n}{n}\,,\quad n = 1,2,\ldots;\ a < 1 \tag{2.1}$$

and $d = -1/\log(1-a)$.

We will also say that the conditional distribution of $Y|(X=n)$ is Pascal (n,p) if

$$P(Y=r|X=n) = \binom{r-1}{n-1} p^n q^{r-n} \tag{2.2}$$

$$r = n,n+1,\ldots;n = 1,2,\ldots;0 < p < 1;q = 1-p$$

Theorem 1. For the generating model $\{X,Y,\gamma(r|n)\}$ considered in the introduction, suppose that $X \sim$ LSD (a) as in (2.1) and $\gamma(r|n)$ is Pascal (n,p) as in (2.2). Then

(i) $Y \sim$ weighted LSD (q)

(ii) $Y|(X=Y) \sim$ LSD (ap).

Moreover, Y and $Y|(X=Y)$ are related in the following way.

$$k_1 P(Y=r) w^r = k_2 P(Y=r|X=Y)\{(w+1)^r - 1\} \tag{2.3}$$

$$r = 1,2,\ldots$$

where w is a weight function ($w = \frac{ap}{q}$) and k_1, k_2 are normalizing constants ($k_1 = -1/\log(1-ap)$, $k_2 = d = -1/\log(1-a)$).

Note. Condition (2.3) implies that the distribution of Y weighted by w^r is the same as the distribution of $Y|(X=Y)$ weighted by $(w+1)^r - 1$. (For a comprehensive account of weighted distributions and their applications the reader is referred to the work of Patil and Rao (1977).

Proof of theorem 1. Because of the nature of the generating model we have

$$P(Y=r) = \sum_{n=1}^{r} \gamma(r|n) g_n = \sum_{n=1}^{r} d\,\frac{a^n}{n}\,\frac{(r-1)!}{(n-1)!(r-n)!}\,p^n q^{r-n}$$

$$= \frac{d}{r}\{(ap+q)^r - q^r\} = d\,\frac{q^r}{r}\{(\frac{ap}{q}+1)^r - 1\},\quad r=1,2,\ldots . \quad (2.4)$$

Hence, Y ~ weighted LSD (q) with weight $\left\{\left(\frac{ap}{q}+1\right)^r - 1\right\}$.

On the other hand,

$$P(Y=r|X=Y) = \frac{g_r \gamma(r|r)}{\sum_{i=1}^{\infty} g_i \gamma(i|i)} = \frac{\frac{(ap)^r}{r}}{\sum_{i=1}^{\infty} \frac{(ap)^i}{i}}$$

Therefore $Y|(X=Y)$ follows the LSD (ap).

It can be seen now that (2.3) is an immediate consequence of (2.4) and (2.5).

We will now show that a relationship of the type of (2.3) leads to the LSD as the distribution of X.

Theorem 2. Suppose that in the generating model $\{X,Y,\gamma(r|n)\}$ we have that $\gamma(r|n)$ is Pascal (n,p), p fixed and independent of n, as in (2.2) with $p > 1/2$. Let the distribution of Y be the same as the distribution of $Y|(X=Y)$ weighted by $(2^r - 1)$ i.e.,

$$P(Y=r) = cP(Y=r|X=Y)(2^r - 1) \quad r=1,2,\ldots \qquad (2.6)$$

where c is the normalizing constant. Then X ~ LSD (a), $a = q/p$.

Proof. It can be observed that (2.6) with $P(Y=r|X=n)$ as given by (2.2) uniquely determines the distribution of X. Since (2.6) is

a special case of (2.3) (for w = 1) and because the LSD satisfies (2.3), as seen in theorem (1), the result follows.

3. GENERATING MODEL AND DAMAGE MODEL.

In this section we discuss the analogy of the generating model introduced in this paper to the celebrated damage model of Rao (1965). In the damage model the non-negative integer-valued r.v. X is subjected to a destructive process. If $s(r|n)$ denotes the probability that the value n of the r.v. X is reduced to the value r and if Y denotes the resulting r.v., then

$$P(Y=r) = \sum_{n=r}^{\infty} P(X=n)\, s(r|n), \quad r = 0,1,2,\ldots$$

Rao pointed out that when $X \sim$ Poisson (λ) and $s(r|n)$ binomial (n,p) with p fixed and independent of n, Y and $Y|(X=Y)$ are both Poisson (λp) i.e.,

$$P(Y=r) = P(Y=r|X=Y), \quad r = 0,1,2,\ldots \tag{3.1}$$

Rao and Rubin (1964) showed that (3.1) is a property enjoyed only by the Poisson distribution for X and thus derived a characterization of the Poisson distribution. Many variants and extensions of this result have since appeared in the literature, the most recent and general of which is due to Panaretos (1979).

The main difference between the two models lies in the fact that our condition (2.3) refers to the distribution of the larger of the two variables while the Rao-Rubin condition (3.1) refers to the distribution of the smaller of the two variables.

We should perhaps, point out that we have used the notion of weighted distributions in a somewhat different context than Patil and Rao. Their observed distribution is the actual distribution weighted by some function of a weight. Here, the observed distribution is weighted to yield the actual distribution.

4. AN APPLICATION

In this section we suggest a possible interpretation and application of the previous results in terms of the generating model.

The LSD has been extensively used in the literature, because of its simplicity, in a wide variety of cases. Boswell and Patil (1971) mentioned some of its applications. More recently, Wani (1978) made an interesting study of the LSD in connection with species abundance models. In fact he used the LSD to describe the distribution of species abundance in the population and showed that, under certain assumptions, the LSD is also the distribution of species abundance in the sample. Our earlier results enable us to start with the distribution of species abundance in the sample and make inference about the species abundance in the population.

For our purposes let X denote the number of individuals contributed to the sample by a species and Y denote the number of individuals contributed to the population by the same species. Then $\gamma(r|n)$ will be the probability that there are r individuals of a particular species in the population given that we found n of them in the sample. It is well known that X can reasonably be assumed to follow the LSD (see, for example, Engen (1979)). If the individuals enter the sample independently of each other, the Pascal distribution (n,p) is not an unreasonable choice for $\gamma(r|n)$ (since r has to be greater than or equal to n, a restriction satisfied by the Pascal distribution as given by (2.2)). Our results suggest that if the above assumptions are valid we can determine the distribution of the number of individuals Y contributed to the population by a species. This will be weighted logarithmic (1-p). Further, suppose that we are only interested in species for which all of the individuals in the population are included in the sample. (This may happen in the case of rare species.) Then we would be talking about the distribution of $Y|(X=Y)$. According to the results of section 2 this distribution can also

be thought of as a logarithmic distribution. Further, relation (2.3) shows that one can specify the distribution of Y from information on the distribution of $Y|(X=Y)$ and vice versa.

ACKNOWLEDGEMENT

I would like to thank a referee for a suggestion that made the proof of theorem 2 shorter.

BIBLIOGRAPHY

Boswell, M.T. and Patil, G.P. (1971). Chance Mechanisms Generating the Logarithmic Series Distribution Used in the Analysis of Number of Species and Individuals. In Statistical Ecology, Patil, G.P., Pielou, E.C. and Waters, W.E. (eds.), 1, 99-130. Penn. State University Press.

Engen, S. (1978). Stochastic Abundance Models. Chapman and Hall, London.

Engen, S. (1979). Stochastic Abundance Models in Ecology. Biometrics 35, 331-338.

Johnson, N.L. and Kotz, S. (1969). Discrete Distributions. Houghton Mifflin, Boston.

Panaretos, John (1979). On Characterizing Some Discrete Distributions Using an Extension of the Rao-Rubin Theorem. Sankhyā A, 44 (2), 1983 (to appear).

Panaretos, John (1981). Unique Properties of Some Distributions and Their Applications. Statistica, XLI, n.4, 567-574.

Patil, G.P. and Rao, C.R. (1977). The Weighted Distributions and Their Applications. In Applications in Statistics. Krishnaiah, P.R. (ed.), North-Holland, 383-405.

Rao, C.R. (1965). On Discrete Distributions Arising out of Methods of Ascertainment. In Classical and Contagious Discrete Distributions. Patil, G.P. (ed). Pergamon Press and Statistical Publishing Society, Calcutta, 320-332.

Rao, C.R. and Rubin, H. (1964). On a Characterization of the Poisson Distribution. Sankhyā A, 25, 295-298.

Thomas, M. (1949). A Generalization of Poisson's Binomial Limit for Use in Ecology. Biometrika, 36, 18-25.

Wani, J.K. (1978). Measuring Diversity in Biological Populations with Logarithmic Abundance Distributions. Canad. J. Stat. 6, (2), 219-228.

Williamson, E. and Bretherton, M.H. (1964). Tables of the Logarithmic Series Distribution. Ann. Math. Statist. 35, 284-297.

Received by Editorial Board member, July, 1982; Revised December, 1982.

Recommended by Samuel Kotz, University of Maryland at College Park, MD