On the Relationship between the Conditional and Unconditional Distribution of a Random Variable

Panaretos, John

1981

ON THE RELATIONSHIP BETWEEN THE CONDITIONAL AND UNCONDITIONAL DISTRIBUTION OF A RANDOM VARIABLE

J. PANARETOS*

Trinity College, Dublin

1. INTRODUCTION

An interesting problem in distribution theory with many potential applications is the study of the relation between the distribution of a random variable (r.v.) Y and its conditional distribution on another r.v. X $(Y|X)$ when the form of the distribution of X is known. In many cases, some form of complete or partial independence between Y and X-Y is involved. This problem sometimes is referred to as the problem of mixtures of distributions and their identifiability in the sense studied by Teicher (1961) and Blischke (1963). Skibinsky (1970) used the term reproducibility

**Present Address:* Department of Statistics, University of Missouri, Columbia, Missouri 65211 USA.

C. Taillie et al. (eds.), Statistical Distributions in Scientific Work, Vol. 4, 379–387.

to describe the fact that when X is binomial (n,p) $(X \sim b(n,p))$ then Y is $b(N,\theta)$ if and only if (iff) the distribution $s(r|n)$ of $Y|(X=n)$ is hypergeometric. Janardan (1973) and Nevil and Kemp (1975) showed that a similar characterization may be obtained by replacing the binomial by a hypergeometric distribution. They also generalized the result to the multivariate case. Grzegorska (1977) considered X to be Poisson and proved that Y has an inflated Poisson distribution iff $Y|X$ is inflated binomial. Problems referring to the case where the distribution of X is given and the relationship between Y and $X|Y$ is of interest, have been investigated by Seshadri and Patil (1964).

An interesting situation arises when either the parameter of the distribution of X, or the parameter of the distribution of Y, is assumed to be a random variable with a given distribution function (d.f.). An attempt in this direction was made recently by Xekalaki and Panaretos (1979). They assumed that $Y|X$ is $b(n,p)$ with p following a distribution with d.f. $F(p)$ $(Y|X \sim b(n,p) \wedge F(p))$. Then, they showed that X is Poisson (λ) iff Y is Poisson$(\lambda p) \wedge F(p)$.

In the next section we consider the case where X is Poisson(λ). Then we establish a characteristic relation between $Y|X$ and Y when either of these have a mixed form of distribution. In Section 3 we examine the implications of changing our basic assumption by allowing X to have a distribution of a mixed Poisson form. We discuss the problem arising and we impose some further restrictions on the distribution of X in order to establish a one to one correspondence between the distributions of Y and $Y|X$. Finally, in Section 4 some potential applications of the results are discussed.

2. THE SIMPLE POISSON MODEL

Theorem 1. Let X,Y be two non-negative, integer-valued r.v.'s Suppose that the distribution of X is power-series Poisson with parameter λ and that the distribution $s(r|n)$ of $Y|(X=n)$ is independent of λ. Then $Y \sim$ Poisson$(\lambda p) \wedge F(p)$ iff $Y|X \sim b(n,p) \wedge F(p)$.

Proof. Necessity follows easily. (It can also be found in the results of Xekalaki and Panaretos, 1979, and Krishnaji, 1974.) For sufficiency we observe that

$$P(Y=r) = \int_0^1 e^{-\lambda p} \frac{(\lambda p)^r}{r!} \, dF(p). \tag{1}$$

On the other hand

$$P(Y=r) = \sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} P(Y=r|X=n). \tag{2}$$

Hence

$$\sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} s(r|n) = \int_0^1 e^{-\lambda p} \frac{(\lambda p)^r}{r!} dF(p). \tag{3}$$

This is a functional equation in $s(r|n)$. The $b(n,p) \wedge F(p)$ is a solution. This is so, because when $s(r|n)$ is $b(n,p) \wedge F(p)$ the left hand-side of (3) can be written as

$$\sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \int_0^1 \binom{n}{r} p^r q^{n-r} dF(p) = \int_0^1 e^{-\lambda} \frac{(\lambda p)^r}{r!} \sum_{n=r}^{\infty} \frac{(\lambda q)^{n-r}}{(n-r)!} dF(p)$$

$$= \int_0^1 e^{-\lambda p} \frac{(\lambda p)^r}{r!} dF(p)$$

which in fact is the right hand-side of (3). We will now show that this is the only solution. Suppose that there exists another solution say $s^*(r|n)$. Then we would have

$$P(Y=r) = \sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} s^*(r|n). \tag{4}$$

Hence (2) and (4) would imply that

$$\sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \{s^*(r|n) - s(r|n)\} = 0. \tag{5}$$

But we are given that $s(r|n)$ is independent of λ. Thus, equating the coefficients of $\frac{\lambda^n}{n!}$ in both sides of (5) we see that $s^*(r|n) = s(r|n)$ and the theorem follows.

As a result of Theorem 1 we can see that if $F(p)$ is degenerate then Y is Poisson iff $Y|X$ is binomial. Other interesting cases can also be derived for different forms of $F(p)$. For example, if $F(p)$ is beta(a,b) we can easily see that $Y|X$ is $b(n,p) \wedge \text{beta}(a,b)$ (negative hypergeometric) iff Y follows the Gurland distribution with p.g.f. ${}_1F_1\{a;\ a+b;\ \lambda(t-1)\}$. Here

${}_1F_1(a;\ b;\ t)$ denotes the confluent hypergeometric function given by

$${}_1F_1(a;\ b;\ t) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \int_0^1 e^{tu}\, u^{a-1}(1-u)^{b-a-1}du.$$

This distribution was first studied by Gurland (1958) in connection with larvae survivors in a field. Another corollary of Theorem 1 results if $F(p) \sim \text{gamma}(a,b)$ truncated to the right at the point 1. In this case $Y|X$ is $b(n,p) \wedge \text{gamma}(a,b)$ truncated to the right at the point 1 iff Y follows a distribution with a p.g.f. proportional to ${}_1F_1(a;\ a+1;\ \lambda(t-1) - \frac{1}{b})$. This distribution was suggested by Kemp (1968) as a model in collective risk theory. Kemp argued that for insurance applications it is more realistic to assume limited risk having the form of a tail-truncated gamma distribution.

Note. The sufficient parts of the above corollaries have already been mentioned in Panaretos (1979).

3. THE MIXED POISSON MODEL

Let us now turn to the situation where X is $\text{Poisson}(\lambda) \wedge F_1(\lambda)$. It can be seen easily that if $Y|X \sim b(n,p) \wedge F_2(p)$ then $Y \sim \text{Poisson}(\lambda p) \wedge F_1(\lambda) \wedge F_2(p)$. However, in this case there does not exist a one-to-one correspondence between the distributions of Y and $Y|X$. In other words, the fact that $Y \sim \text{Poisson}(\lambda p) \wedge F_1(\lambda) \wedge F_2(p)$ does not imply that the only form of $Y|X$ is $b(n,p) \wedge F_2(p)$. In fact, if we assume that $Y \sim \text{Poisson}(\lambda p) \wedge F_1(\lambda) \wedge F_2(p)$ we can see, using the argument employed in Theorem 1, that the distribution of $Y|X$ must satisfy the equation

$$P(Y=r) = \sum_{n=r}^{\infty} \{\int_0^{\infty} e^{-\lambda} \frac{\lambda^n}{n!}\, dF_1(\lambda)\} s(r|n). \tag{6}$$

Clearly the $b(n,p) \wedge F_2(p)$ for $s(r|n)$ is a solution. It can be proved, however, that it is not unique. Suppose that $s^*(r|n)$ is a second solution. Then we would have

$$\int_0^{\infty} \sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \{s(r|n) - s^*(r|n)\} dF_1(\lambda) = 0 \tag{7}$$

But this does not necessarily imply that

$$\sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \{s(r|n) - s^*(r|n)\} = 0$$

i.e., that $s(r|n) = s^*(r|n)$. Consider for example the case where $s^*(r|n)$ is defined as

$$s^*(r|n) = \begin{cases} s(r|n) & n = k+1,\ k+2, \cdots \\ s(r|n) - c_{r,n} & n = 0,1,\cdots,k. \end{cases}$$

Then (7) is equivalent to

$$\int_0^{\infty} \sum_{n=r}^{k} e^{-\lambda} \frac{\lambda^n}{n!} c_{r,n}\, dF_1(\lambda) = 0$$

i.e., to $$\sum_{n=r}^{k} c_{r,n}\, \xi_n = 0 \tag{8}$$

where $$\xi_n = \int_0^{\infty} e^{-\lambda} \frac{\lambda^n}{n!}\, dF_1(\lambda).$$

Since we can find $c_{r,n} \neq 0$ for which (8) holds we come to the conclusion that there exist solutions of (6) other than the $b(n,p) \wedge F_2(p)$. Hence, to obtain a one-to-one correspondence between Y and $Y|X$ (when $X \sim \text{Poisson}(\lambda) \wedge F_1(\lambda)$ and $Y|X \sim b(n,p) \wedge F_2(p)$) we have to impose some restrictions on the structure of $F_1(\lambda)$. A possible way of doing this is stated and proved in the following theorem.

Theorem 2. Suppose that the distribution of the r.v. X is $\text{Poisson}(\lambda) \wedge F_1(\lambda)$ where $F_1(\lambda)$ is absolutely continuous with density

$$f_1(\lambda|\theta) = e^{-\theta\lambda} \frac{\phi(\lambda)}{\psi(\theta)}$$

where $\lambda \in (a,b)$ and $\phi(\lambda)$, $\psi(\theta)$, $\theta > 0$. Assume that the distribution $s(r|n)$ of $Y|X$ is independent of θ. Then the distribution of Y is mixed Poisson of the form Poisson$(\lambda p) \wedge F_1(\lambda) \wedge F_2(p)$ iff $s(r|n)$ is $b(n,p) \wedge F_2(p)$.

Proof. The 'if' part is straightforward. To show that the $b(n,p) \wedge F_2(p)$ is the only possible form for $s(r|n)$ satisfying the above property, let us assume that there exists another such form say, $s^*(r|n)$. Then we would have

$$\int_0^{\infty} e^{-\theta\lambda} \sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \frac{\phi(\lambda)}{\psi(\theta)} \{s(r|n) - s^*(r|n)\} d\lambda = 0,$$

which implies that

$$\sum_{n=r}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} \phi(\lambda) \{s(r|n) - s^*(r|n)\} = 0 \quad \text{for} \quad \lambda \in (a,b)$$

i.e. that $s^*(r|n) = s(r|n)$. This establishes the result.

For different forms of $f_1(\lambda)$ and $F_2(p)$ characterizations can be obtained for various distributions. For example, assume that $f_1(\lambda)$ is gamma(θ,ρ) i.e., that

$$f_1(\lambda) = \frac{1}{\theta^{\rho}\Gamma(\rho)} e^{-\frac{\lambda}{\theta}} \lambda^{\rho-1} \qquad \rho,\theta > 0$$

and $F_2(p)$ is degenerate. Then on the assumption that $X \sim$ negative binomial$(\frac{1}{1+\theta}, \rho)$ we have that $Y|X \sim b(n,p)$ iff $Y \sim$ negative binomial$(\frac{1}{1+p\theta}, \rho)$. This indicates that in Skibinsky's terminology, the negative binomial distribution is reproducible with respect to sampling with replacement.

4. SOME POSSIBLE APPLICATIONS

It is known that characteristic properties of distributions in general, apart from their mathematical interest, can be useful in applied statistics. The main reason lies in the fact that these properties are unique for the characterized distribution. This fact can guide the choice of assumptions that one has to impose in a given problem or enable him to reduce a complicated problem to an equivalent but possibly simpler one.

The characterizations of the two previous sections in particular can be useful when the r.v. X describing the phenomenon under investigation is of either a Poisson or a mixed Poisson form. They can also offer help when a mixed binomial distribution is appropriate for the conditional r.v. $Y|X$ that may be involved. Both situations arise very often in practice especially when sampling takes place over an extended area or period of time; data derived in this way do not always conform to the simple Poisson or simple binomial type. This implies that the parameter of the assumed distribution varies according to some probability law. In most of the cases concerned with the binomial parameter p this law is reasonably assumed to be beta(a,b). On this assumption a potential application of Theorem 1 may arise in the following situation.

Assume that the distribution of the number X of cars passing through a junction with traffic lights in a given period of time is Poisson(λ) distributed. Let the number of cars out of n which pass while the red light is on be binomially distributed with parameter p. Assume that p is not constant. Instead, take it as a random variable associated with the drivers tendency to commit an offense. If the beta(a,b) model is suitable for the distribution of p, we have that the number of cars out of n passing when the red light is on will have the binomial-beta distribution. Consequently, according to Theorem 1 the distribution of the number Y of cars passing against a red light is Poisson(λp) $\underset{p}{\wedge}$ beta(a,b). In fact, Theorem 1 provides more information. It tells us that if we have reasons to believe that Y is Poisson(λp) $\wedge F(p)$, then the only possible form for the distribution of cars out of n passing against the red light (i.e., for the distribution of $Y|X$) is $b(n,p) \wedge F(p)$. Moreover, if Y is simple Poisson with parameter $\mu < \lambda$ we have that $Y|X \sim$ binomial(n,p) where $p = \mu/\lambda$.

A number of other cases in which the mixed binomial model was shown to be appropriate can be viewed in a similar way in the light of the results of Sections 2 and 3 (see for example Chatfield and Goodhart's (1970) work where they adopted the beta-binomial distribution for the description of consumer purchasing behavior).

Applications of mixed Poisson distributions, on the other hand, go as far back as 1920 when Greenwood and Yule used the Poisson $\wedge$ gamma distribution to describe accident data. More recently, other workers (e.g. Froggatt *et al.*, 1969; Ashford, 1972) used the same model to examine problems concerning patient contacts with the doctor (GP). In the first place, the patient contacts were regarded as events in a Poisson process with parameter λ characteristic of the individual patient. Then it is argued that λ represents the "proneness" of the patient to illness and a skew distribution of a gamma form is suggested for λ in a population of patients. Thus, for a given individual the number of contacts X with the GP in a given time interval has a Poisson $\wedge$ gamma distribution. If we now further assume that each contacting patient is referred to a consultant with probability p (p fixed) our corollary of Theorem 2 indicates a one-to-one correspondence between the distribution of the number Y of visits to the consultant (negative binomial) and the distribution of the number of visits to the consultant given the total number X of contacts with the GP (binomial). If however, either the negative binomial does not seem appropriate for Y or the binomial does not explain well the distribution of $Y|X$ this might signal the need for considering a variable p. A reasonable explanation for this is that p may vary from patient to patient depending on the seriousness of his illness.

The above examples are only a collection of possible cases where the results of Sections 2 and 3 can be of some use. Clearly one can find similar situations for other forms of $F_1(\lambda)$ and $F_2(p)$.

Finally, it should be pointed out that another major area that the results fit, is that of the damage model introduced by Rao (1963). Here X is the original observation produced by some natural process, $Y|X$ is the destructive process and Y is the observed (undamaged) part of X. Clearly, this model can be looked into in the light of our results.

ACKNOWLEDGEMENT

I would like to thank Dr. D. N. Shanbhag for his helpful comments and suggestions.

REFERENCES

Ashford, J. (1972). Patient contacts in general practice in the National Health Service. *The Statistician*, 21, 265-289.
Blischke, W. R. (1963). Mixtures of discrete distributions. In *Classical and Contagious Discrete Distributions*, G. P. Patil, ed. Statistical Publishing Society, Calcutta. Pages 351-372.
Chatfield, C. and Goodhart, G. J. (1970). The beta-binomial model for consumer purchasing behavior. *Applied Statistics*, 19, 240-250.
Froggatt, P., Dugeon, M. Y. and Merrett, J. D. (1969). Consultations in general practice, analysis of individual frequencies. *British Journal of Preventive and Social Medicine*, 23.

Greenwood, M. and Yule, G. U. (1920). An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attack of disease or repeated accidents. *Journal of the Royal Statistical Society*, 83, 255-279.

Grzegórska, L. (1977). A characterization of the Poisson distribution in discrete models with perturbation (Polish). *Matematyka Stosowana*, 10, 55-64.

Gurland, J. (1958). A generalized class of contagious distributions. *Biometrics*, 14, 229-249.

Janardan, K. G. (1973). A characterization of multivariate hypergeometric and inverse hypergeometric models. Technical report, Math. Systems Program, Sangamon State University.

Kemp, A. W. (1968). A limited risk CPp. *Skandinavisk Aktuarietidskrift*, 51, 198-203.

Krishnaji, N. (1974). Characterization of some discrete distributions based on a damage model. *Sankhyā, Series A*, 36, 204-213.

Nevil, A. M. and Kemp, C. D. (1975). On characterizing the hypergeometric and multivariate hypergeometric distributions. In *Statistical Distributions in Scientific Work, Vol. 3*, G. P. Patil, S. Kotz and J. K. Ord, eds. Reidel, Dordrecht-Holland, Pages 353-358.

Panaretos, J. (1979). An extension of the damage model. *Metrika* (to appear).

Rao, C. R. (1963). On discrete distributions arising out of methods of ascertainment. *Sankhyā, Series A*, 25, 311-324.

Seshadri, V. and Patil, G. P. (1964). A characterization of a bivariate distribution by the marginal and the conditional distribution of the same component. *Annals of the Institute of Statistical Mathematics*, 15, 215-221.

Skibinsky, M. (1970). A characterization of hypergeometric distributions. *Journal of the American Statistical Association*, 65, 926-929.

Teicher, H. (1961). Identifiability of mixtures. *Annals of Mathematical Statistics*, 32, 244-248.

Xekalaki, E. and Panaretos, J. (1979). Characterization of the compound Poisson distribution. *Bulletin of the International Statistical Institute*, 48, 577-580.

[*Received May* 1980]