STATISTICA, anno XLVII, n. 1, 1987

CONVOLUTION AND MIXING PROPERTIES OF THE DAMAGE MODEL

John Panaretos

1. INTRODUCTION

The damage model was introduced by Rao (1965). According to this model, let X denote an original observation produced by some natural process and let $P(X=n), n = 1, 2, \ldots$ denote the probability distribution of X. This original observation may be partially destroyed, or may be only partially ascertained. Let Y denote the resulting random variable (what is actually observed). Then $Z=X-Y$ represents the damaged part. The distribution of Y given X is called the survival distribution. In this paper we study the effect of the convolution in the damage model (section 2). In particular we examine the damage model where the original observation X is the sum of independent observations. In section 3 we present a property of the damage model when the parameter of the distribution of the original random variable follows a discrete distribution. Finally in section 4 we derive a result that extends a property of the Poisson distribution given by Moran (1952).

2. THE EFFECT OF CONVOLUTION IN THE BINOMIAL DAMAGE MODEL

Suppose that the distribution of Y given $X=n$ in the damage model is binomial with parameter p independent of n. Then the following properties hold.

Property 2.1. Let $X = \sum_{i=1}^{s} X_i$ be the original observation, where $X_1, X_2, \ldots, X_s$ are independent and identically distributed random variables with probability generating function (p.g.f.) $G(t)$. Let

$$P(Y_i = r_i \mid X_i = n_i) = \binom{n_i}{r_i} p^{r_i} q^{n_i - r_i} \qquad i = 1, 2, \ldots, s, \tag{2.1}$$

$0 < p < 1$, $q = 1-p$, $r_i = 0 = 1, 2, \ldots, n_i$; $n_i = 1, 2, \ldots$ and p independent of n_i for all $i = 1, 2, \ldots, s$.

Let $Y = \sum_{i=1}^{n} Y_i$ denote the resulting random variable. Then the p.g.f.'s of the distribution of Y and the distribution of $Y|(X=Y)$ (the distribution of the resulting random variable when no damage has occurred) are, respectively,

$$G_Y(t) = G(q+pt)^s \tag{2.2}$$

and

$$G_{Y|X=Y}(t) = \frac{G(pt)^{\ s}}{G(p)} \ . \tag{2.3}$$

Proof: The proof is straightforward.

Property 2.2. Let X_i, $i = 1, 2, \ldots, s$ be a sequence of independent random variables with $X_i \sim$ Poisson (λ_i), $i = 1, 2, \ldots, s$, $0 < \lambda_i < \infty$ and let λ_i be random variables with absolutely continuous density functions $f_i(\lambda_i)$, $i = 1, 2, \ldots, s$. Further, let $\lambda = \sum_{i=1}^{s} \lambda_i$ and consider a random variable $X \sim$ Poisson (λ) and let $f(\lambda)$ be the density function of λ. Suppose that $Y_i | (X_i = n_i)$, $i = 1, 2, \ldots, s$ and $Y|(X=n)$ follow binomial distributions as in (2.1.). Then

$$G_Y(t) = \prod_{i=1}^{s} G_{Y_i}(t) \tag{2.4}$$

and

$$G_{Y|X=Y}(t) = \prod_{i=1}^{s} G_{Y_i|X_i=Y_i}(t). \tag{2.5}$$

Proof: We have that

$$G_{Y_i}(t) = \int_0^\infty e^{\lambda_i p(t-1)} f_i(\lambda_i)\, d\lambda_i\ , \qquad i = 1, 2, \ldots, s \tag{2.6}$$

and

$$G_{Y_i|X_i=Y_i}(t) = \frac{\int_0^\infty e^{\lambda_i(pt-1)} f_i(\lambda_i)\, d\lambda_i}{\int_0^\infty e^{\lambda_i(p-1)} f_i(\lambda_i)\, d\lambda_i}\ , \qquad i = 1, 2, \ldots, s. \tag{2.7}$$

Also,

$$f(\lambda) = \int_0^\infty f_s(\lambda_s) \int_0^\infty f_{s-1}(\lambda_{s-1}) \ldots \int_0^\infty f_2(\lambda_2) f_1(\lambda - \sum_{k=2}^{s} \lambda_k) d\lambda_2\, d\lambda_3 \ldots d\lambda_s$$

Thus

$$G_Y(t) = \int_0^\infty e^{\lambda p(t-1)} \int_0^\infty f_s(\lambda_s) \int_0^\infty f_{s-1}(\lambda_{s-1}) \dots$$

$$\int_0^\infty f_2(\lambda_2) f_1\left(\lambda - \sum_{k=2}^{s} \lambda_k\right) d\lambda_2\, d\lambda_3 \dots d\lambda_s \qquad (2.8)$$

If we denote by

$$\mathscr{L}\{F(\lambda), t\} = \int_0^\infty e^{-\lambda t} F(\lambda)\, d\lambda \qquad (2.9)$$

the Laplace transform of $F(\lambda)$, then (2.6), (2.7) and (2.8) become, respectively,

$$G_{Y_i}(t) = \mathscr{L}\{f_i(\lambda_i), p(1-t)\} \qquad i = 1, 2, \dots, s \qquad (2.10)$$

$$G_{Y_i \mid X_i = Y_i}(t) = \frac{\mathscr{L}\{f_i(\lambda_i), 1-pt\}}{\mathscr{L}\{f_i(\lambda_i), 1-p\}} \qquad i = 1, 2, \dots, s \qquad (2.11)$$

$$G_Y(t) = \mathscr{L}\left\{ \int_0^\infty f_s(\lambda_s) \int_0^\infty f_{s-1}(\lambda_{s-1}) \dots \int^\infty f_2(\lambda_2) \times \right.$$

$$\left. f_1\left(\lambda - \sum_{k=2}^{\infty} \lambda_k\right) d\lambda_2\, d\lambda_3 \dots d\lambda_s, p(1-t) \right\}$$

$$= \prod_{i=1}^{s} \mathscr{L}\{f_i(\lambda_i), p(1-p)\}.$$

Hence, using (2.10),

$$G_Y(t) = \prod_{i=1}^{s} G_{Y_i}(t).$$

In a similar manner we arrive at (2.5).

Property 2.3. Let us now consider the general case where $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_s)$. Define $\underset{\sim}{X} = (X_1, X_2, \dots, X_s)$ to be a vector of s-independent Poisson variables such that $E(X_j) = \lambda_j$, $j = 1, 2, \dots, s$. Let $\underset{\sim}{Y} = (Y_1, Y_2, \dots, Y_s)$ be a vector of independent and nonnegative integer valued r.v.'s such that

$$P(Y_j = r_j \mid X_j = n_j) = \binom{n_j}{r_j} p_j^{r_j} q_j^{n_j - r_j} \qquad j = 1, \dots, s.$$

Further assume that $\underset{\sim}{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_s)$ is a random vector of positive components with distribution function $F(\lambda_1, \lambda_2, \dots, \lambda_s)$. Then the p.g.f. of the distribution of $(\underset{\sim}{X}, \underset{\sim}{Y})$ will be given by

$$G_{\underset{\sim}{X},\underset{\sim}{Y}}(\underset{\sim}{t},\underset{\sim}{z}) = \sum_{n_1,\ldots,n_s} G_{\underset{\sim}{Y}\mid\underset{\sim}{X}}(\underset{\sim}{z}) P_{n_1,\ldots,n_s} t_1^{n_1}\ldots t_s^{n_s}$$

$$= \sum_{n_1,\ldots,n_s} \prod_{j=1}^{s} (q_j + p_j z_j)^{n_j} P_{\underset{\sim}{n}} t_1^{n_1}\ldots t_s^{n_s}$$

$$= G_{\underset{\sim}{X}}(\underset{\sim}{t}(\underset{\sim}{q} + \underset{\sim}{p}\underset{\sim}{z})) = \int G_{\underset{\sim}{X}\mid\underset{\sim}{\lambda}}\, dF(\lambda)$$

(where $\underset{\sim}{u}\underset{\sim}{v} = (u_1 v_1, \ldots, u_s v_s)$)

$$= E(G_{\underset{\sim}{X}\mid\underset{\sim}{\lambda}}) = E\left(\prod_{i=1}^{s} e^{-\lambda_i(1 - t_i(q_i + p_i z_i))}\right)$$

i.e., by

$$G_{\underset{\sim}{X},\underset{\sim}{Y}}(\underset{\sim}{t},\underset{\sim}{z}) = E\left[\prod_{i=1}^{s} e^{-\lambda_i\{1 - t_i(q_i + p_i z_i)\}}\right].$$

For the p.g.f. of $\underset{\sim}{Y}$ we have

$$G_{\underset{\sim}{Y}}(\underset{\sim}{z}) = E\left[\prod_{i=1}^{s} e^{-\lambda_i p_i(1 - z_i)}\right].$$

We also have (for $\sum_{i=1}^{s} Y_i = Y,\ \lambda = \sum_{i=1}^{s} \lambda_i$)

$$G_Y(z) = E\left[\prod_{i=1}^{s} e^{-\lambda_i p_i(1 - z)}\right]$$

which, if all $p_i = p$, simplifies to

$$G_Y(z) = E\left[e^{-\lambda p(1 - z)}\right].$$

3. A MIXING PROPERTY OF THE DAMAGE MODEL

In a recent paper, Panaretos (1982) studied some aspects of an extension of Rao's damage model, where either the original distribution or the survival distribution were mixtures of distributions. In fact it was observed that when $X \sim$ Poisson (λ) and λ is a random variable with distribution function $F(\lambda)$, then, under the assumption of a binomial (n, p) survival distribution, $G_Y(t) = M_\lambda(p(t-1))$ where $M_\lambda(t)$ denotes the moment generating function of λ. An interesting special case of this result occurs when λ follows a discrete distribution with p.g.f. of the form $\{g(t)\}^k$. In this case we have the following property of the damage model.

Property 3.1. Suppose that $X \sim$ Poisson (λ) and λ is a discrete random variable with p.g.f. of the form $\{g(t)\}^k$. Suppose that $Y \mid (X = n) \sim b(n, p)$. Then the resulting random variable Y will have the distribution of λ, generalized with a Poisson distribution with parameter p, i.e.,

$$G_Y(t)=\{g(e^{p(t-1)})\}^k.$$

Proof. The proof follows immediately from the above-mentioned result and the well-known fact that $M_\lambda(t)=G_\lambda(e^t)$.

As an example, we may observe that if λ takes the values $k, k+1, \ldots$, $k=1, 2, \ldots$ with probabilities

$$g_\lambda=\binom{\lambda-1}{k-1}a^k(1-a)^{\lambda-k} \qquad \lambda=k, k+1, \ldots; \quad k=1,2, \quad 0<a<1,$$

i.e., when λ follows a Pascal distribution with parameters a and k, then Y is Pascal (a, k) generalized with a Poisson (p), i.e.,

$$G_Y(t)=\left\{\frac{ae^{p(t-1)}}{1-(1-a)e^{p(t-1)}}\right\}^k.$$

4. A VARIANT OF MORAN'S PROPERTY OF THE POISSON DISTRIBUTION

Moran (1952) has shown that if $Y|(X=n)$ is binomial (n, p) and Y and $Z=X-Y$ are independent random variables such that there exists at least one integer i so that $P(Y=i)>0$ and $P(Z=i)>0$ then Y and Z are Poisson distributed. Panaretos (1983) has shown that Moran's result cannot be extended to a situation where $Y|(X=n)$ follows a truncated binomial distribution. Here we present a variant of Moran's result.

Suppose that X, Y, Z are three nonnegative, integer-valued random variables with $X=Y+Z$. Assume that the conditional distribution of $Y|X$ is independent of the parameter of the distribution of X. Then,

(i) If either Y or Z follow Poisson distributions, and $P(Y=r|X=n)= \binom{n}{r}p^rq^{n-r}$, $0<p<1$, $p+q=1$, then Y, Z are independent Poisson r.v's.

(ii) If $P(Y=r)=P(Y=r|X=Y)$ and $P(Y=r|X=n)=\binom{n}{r}p^rq^{n-r}$, then Y, Z are independent Poisson r.v.'s.

(iii) If (Y, Z) follows a bivariate Poisson with p.g.f.

$$G(t_1, t_2)=e^{\lambda_1(t_1-1)+\lambda_2(t_2-1)+\lambda_{12}(t_1t_2-1)}$$

and $P(Y=r|X=Y)=P(Y=r)$, then

$$P(Y=r|X=n)=\binom{n}{r}p^rq^{n-r}.$$

Proof. (i) Because $Y|(X=n)\sim$ binomial we have the joint p.g.f. of Y and Z

as $G_{Y,Z}(t_1, t_2) = G_X(pt_1 + qt_2)$. Without loss of generality we may assume that $Y \sim$ Poisson. Then $G_Y(t_1) = e^{-\lambda + \lambda t_1}$. We also have $G_Y(t_1) = G_X(pt_1 + q)$, i.e.,

$$G_X(pt_1 + q) = e^{-\lambda + \lambda t_1} \; .$$

Consequently,

$$G_{Y,Z}(t_1, t_2) = G_X(pt_1 + qt_2) = G_X\{(pt_1 + qt_2 - q) + q\}$$
$$= e^{-\lambda + \frac{\lambda}{p}(pt_1 + qt_2 - q)} = e^{-\frac{\lambda}{p} + \frac{\lambda}{p}(pt_1 + qt_2)} ,$$

i.e.,

$$G_{Y,Z}(t_1, t_2) = G_Y(t_1)\, G_Z(t_2) ,$$

which implies that Y and Z are independent Poisson r.v.'s.

(ii) It has been shown by Rao and Rubin (1964) that when $Y|(X = n) \sim$ $\sim$ binomial (n, p) and $P(Y=r) = P(Y=r\,|\,X=Y)$ then X follows a Poisson distribution. It is easy to see that in this case Y and Z are independent and hence they have Poisson distributions.

(iii) Since (Y, Z) has a bivariate Poisson distribution we find that

$$P(Y=y, Z=z) = e^{-(\lambda_1 + \lambda_2 + \lambda_{12})} \sum_{i=0}^{\min(y,z)} \frac{\lambda_1^{y-i}\,\lambda_2^{z-i}\,\lambda_{12}^{i}}{(y-i)!\,(z-i)!\,i!} \; .$$

Hence

$$P(Y=r\,|\,X=Y) = P(Y=r\,|\,Z=0) = \frac{P(Y=r, Z=0)}{P(Z=0)} = \frac{e^{-\lambda_1 + \lambda_2 + \lambda_{12}}\,\frac{\lambda_1^r}{r!}}{e^{-(\lambda_r + \lambda_{12})}} ,$$

i.e., $P(Y=r\,|\,X=Y) = e^{-\lambda_1}\,\dfrac{\lambda_1^r}{r!}$.

However,

$$P(Y=r) = e^{-(\lambda_1 + \lambda_{12})}\,\frac{(\lambda_1 + \lambda_{12})^r}{r!}$$

and $P(Y=r) = P(Y=r\,|\,X=Y)$. Hence $e^{\lambda_1 (t-1)} = e^{(\lambda_1 + \lambda_{12})(t-1)}$ which implies that $\lambda_1 = \lambda_1 + \lambda_{12}$, i.e., $\lambda_{12} = 0$. As a result of this it is clear that Y, Z are independent Poisson r.v.'s and so $Y|(X=n) \sim$ binomial.

5. SOME APPLICATIONS

The results of the previous sections may be applicable to a number of statistical problems. Properties 2.1 and 2.2 for example, could be used in a cumulative damage situation, when damage to each of the X_i's occurs according to the binomial law.

Let X_i be the number of missprints on page i of a book. Let p be the probability that a missprint is detected by the proofreader. Let Y_i be the number of detected missprints on page i. If p is constant, we can reasonably assume that $Y_i \mid (X = n_i) \sim b\ (n_i, p)$. Then property 2.1 provides the connection between the distributions of the total number of missprints X in the book and the total number of detected missprints Y. Property 2.2 could be used in problems like the above when more elaborate analysis is needed because the X_i's are assumed to follow Poisson distributions with means λ_i, which are not constant for each i. X_i, for example, can be the number of telephone calls that arrive at a switchboard during a time period i of the day. It is not unreasonable to assume that X_i follows a Poisson distribution with mean λ_i which is not constant (it varies from day to day). Let Y_i be the number of telephone calls placed during the period i that the switchboard was able to handle. Further let p be the probability that a call placed will be handled within the period i. If X is the total number of calls placed during the day and if X is assumed to follow a Poisson distribution with parameter $\lambda = \Sigma\lambda_i$ then property 2.2 provides the connection between the distribution of the number of calls Y_i handled during the time period i and the distribution of the total number of calls Y handled in a given day.

Property 2.3 is useful when we are dealing with random vectors. In a traffic offence study for example, suppose that we split the total period of observation into s subperiods. Let X_i $i = 1, 2, \ldots, s$ be the number of offences in a given locality during the period i and let Y_i be the number of offences that traffic wardens were able to record. P_i is the probability that an offence is recorded during that period i. Assuming that the X_i's follow Poisson distributions with parameters λ_i, $i = 1, 2, \ldots, s$ which are random variables for each i property 2.3 gives the joint distribution of the vectors $\underset{\sim}{X} = (X_1, X_2, X_3, \ldots, X_s)$ and $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_s)$. It also provides the distribution of the total number $Y = \Sigma Y_i$ of offences recorded during the total period of observation.

Quality control is an area where property 3.1 can find an application. Let X be the number of deffective items produced by a factory in a given day. Suppose that X follows a Poisson distribution with mean λ which varies from day to day. Suppose that Y is the number of defective items detected by the quality control process in a given day. Then if λ follows a discrete distribution with p.g.f. of the form $\{g\ (t)\}^k$, property 3.1 suggests that the distribution of the number of detected defective items Y will follow the distributio of λ generalized by a Poisson distribution with parameter p (p being the constant probability that a defective item will be detected).

Finally the results of section 4 could be used in testing of hypothesis problems where the Poisson and the binomial distributions are involved. Properties (i) and (ii) of this section can be used for testing the assumption of independent Poisson r.v.'s Y and Z. On the other hand property (iii) is very important in damage model testing of hypothesis problems. Usually in such problems testing the hypothesis of binomial damage is not easy because the parameter n of the binomial is not fixed. As property (iii) suggests testing this hypothesis is equivalent to testing the hypothesis that Y, Z follow a bivariate Poisson distribution and that they satisfy the condition $Y \overset{d}{=} Y \mid (Z=0)$.

University of Patras JOHN PANARETOS

REFERENCES

[1] P.A.P. MORAN (1952), *A Characteristic Property of the Poisson Distribution.* "Proc. Camb. Phil. Soc." 48, 206-207.

[2] J. PANARETOS (1983), *On Moran's Property of the Poisson Distribution.* "Biom. J." 25 (1), 69-76. (Also in: Proceedings of the 6th Conference in Probability Theory Brasov, Romania. B. Beréanu et al. (Eds.). Ed. Ac. Rep. Soc. Rom. Bucuresti, 1981, 319-325).

[3] J. PANARETOS (1982), *An Extension of the Damage Model.* "Metrika" 29, 189-194.

[4] C.R. RAO (1963), *On Discrete Distributions Arising Out of Methods of Ascertainment.* "Classical and Contagious Discrete Distributions, Statistical Publishing Society", Calcutta, 320-332. (Also reprinted in Sankhyā, A, 1963, 25, 311-344).

[5] C.R. RAO and H. RUBIN (1964), *On a Characterization of the Poisson Distribution.* "Sankhyā A", 26, 295-298.

RIASSUNTO

Convoluzione e proprietà di mistura del modello di danno

Il modello di danno è applicabile nei casi in cui un'osservazione originale sia sottoposta ad un processo distruttivo. In questo articolo sono trattati i modelli di danno con una distribuzione di sopravvivenza binomiale. La forma della distribuzione risultante viene trovata nel caso in cui la variabile originale casuale sia la convoluzione di "S" variabili casuali indipendenti. E' discusso inoltre il caso in cui la distribuzione originale sia una mistura discreta. E' stabilita infine una relazione fra le distribuzioni delle variabili casuali originali, di sopravvivenza e risultante.

RÉSUMÉ

Convolution et propriété de mixture du modèle de dommage

Le modèle de dommage est applicable dans les cas où une observation originale est soumise à un procès destructif. Dans cet article on considère les modèles de dommage avec une distribution binomiale de survie. La forme de la distribution résultante est trouvée quand la variable originale aléatoires est la convolution de "S" variables aléatoires indépendantes. Est traité, en autre, le cas où la distribution originale est une mixture discrète. En fin on établi une relation entre les distributions des variables aléatoires originale, de survie et resultante.