STATISTICA, anno LXIX, n. 1, 2009

NONPARAMETRIC ESTIMATION IN RANDOM SUM MODELS

H.S. Bakouch, T.A. Severini

1. INTRODUCTION

The motivation for this paper arises from the fact that the random sum models are widely used in risk theory, queueing systems, reliability theory, economics, communications, and medicine. In insurance contexts, the compound distribution of random sum variables (see, e.g., Cai and Willmot (2005), Charalambides (2005) and Panjer (2006)) arises naturally as follows. Let X_i be the number of persons involved in the i th accident on a particular day in a certain city and let N be another random variable that represents the number of accidents occurring on that day. Then, the random variable given by $S_N = \sum_{i=1}^{N} X_i$, $S_0 = X_0 = 0$, denotes the total number of persons involved in accidents in a day, and is called compound random variable. In queueing model, at a bus station, assume that the number of passengers on the i th bus is X_i and the number of arriving buses is the random variable N; then the number of passengers arriving buses during a period of time is the random sum $S_N = \sum_{i=1}^{N} X_i$. In communications, let us consider the random variable N as the number of data packets transmitted over a communication link in one minute such that each packet is successfully decoded with probability p, independent of the decoding of any other packet. Hence, the number of successfully decoded packets in one minute span is the random sum $S_N = \sum_{i=1}^{N} X_i$, where X_i is 1 if the i th packet is decoded correctly and 0 otherwise. That is, the compound random variable S_N follows a binomial distribution with random and fixed parameters N and p, respectively. The distribution of the random variable S_N, is known as a compound distribution of the random variables X_i and N, where $i = 1, 2, ..., N$; see, e.g., Chatfield and Thoebald (1973), Lundberg (1964), Medhi (1972), Peköz and Ross (2004), and Sahinoglu (1992). Thus, the distribution of a random sum is based on two distributions: the distribution of N, which we will call the event distribution and the distribution of the random variables $X_1, X_2, \ldots, X_N$, which we will call the compounding distribution.

Many authors (e.g. Pitts (1994), Buchmann and Grübel (2003, 2004), and Cai and Willmot (2005)) have been investigated the distributional properties of the random sum S_N including its distribution function and the asymptotic behaviour of the tail. Pitts (1994) discussed the nonparametric estimation of the distribution function of S_N based on topological concepts and assumed that the event distribution is known and the compounding distribution is unknown. Buchmann and Grübel (2003, 2004) considered the estimation problem of a probability set of the compounding distribution of X_i, $i = 1, 2, ..., N$, and the compound distribution of X_i and N assuming that the event distribution follows Poisson distribution. Weba (2007) investigated prediction of the compound mixed Poisson process assuming that the event distribution is a mixed Poisson process and the compounding distribution is unknown. Meanwhile, we here deal with unknown event and compounding distributions. Also, our work is more general than Pitts (1994),and Buchmann and Grübel (2003, 2004), since it can be applied to any event and compounding distributions as long as they obey the assumptions that are in section 4.1. Recently, Bakouch and Ristić (2009), and Ristić *et al.* (2009) employed this random sum, S_N, to model integer-valued time series that are used in forecasting and regression analysis of count data. Also, they discussed some properties of S_N for a known event distribution and an unknown compounding distribution, and then obtained estimators of the parameters of such distributions after giving a known compounding distribution. In section 4.2 of this paper, we consider a nonparametric estimation of the compounding distribution in the compound Poisson model. The approach in this case is similar to the estimation method of Buchmann and Grübel (2003, 2004) but our approach is of interest to any problem concerning the compound Poisson model. On the other hand, Buchmann and Grübel (2003, 2004) are interested only in the compound Poisson model that is commonly used in queueing theory for modeling the arrival distribution of the number of customers.

Let $Y_1, \ldots, Y_n$ denote independent random variables, each with the distribution of S_N. Based on the information provided by the random variables $Y_1, \ldots, Y_n$, we wish to estimate features of the distributions of X_i and N. A common used approach to this problem is to use parametric models for the event and compounding distributions. This leads to a parametric model for the distribution of $Y_1, \ldots, Y_n$. Estimating the parameters of this model leads to estimators for the parameters of the distributions of X_i and N. The drawback of this approach is that it requires strong assumptions about the distributions of X_i and N. However, since we do not observe a random sample from both distributions, these assumptions are difficult to verify empirically.

The purpose of this paper is to consider two methods of nonparametric estimation in random sum models; that is, we consider methods that do not require full parametric models for the compounding distribution and the event distribution. The first considers estimation of the means of the distributions of X_i and N based on the second-moment assumptions regarding compounding and event distributions. The second problem is the nonparametric estimation of the compounding distribution based on a parametric model for the event distribution.

The outline of the paper is as follows. In section 2 some basic results regarding random sum distributions are presented; although these results are well-known, they are central to the proposed estimation procedures and, hence, they are presented for completeness. The approach based on second-moment assumptions is presented in section 3. In section 4, nonparametric estimation of the compounding distribution is developed.

2. SOME PROPERTIES OF RANDOM SUMS

Let $S_N = \sum_{i=1}^{N} X_i$, where X_i is independent, identically distributed, non-negative, integer-valued random variables and let N be a non-negative, integer-valued random variable independent of $X_1, X_2, \ldots, X_N$. In this section, we briefly review some statistical properties of S_N (see, e.g., Klugman et al., 2004).

Let P_N denote the probability generating function of N and let P_{X_i} denote the probability generating function of X_i. Therefore, the probability generating function of S_N is given by

$$P_{S_N}(t) = P_N[P_{X_i}(t)]. \tag{2.1}$$

The function $P_{S_N}(t)$ is well defined for $|t| \le 1$. Similarly, let M_N and M_{X_i} denote the moment generating functions of N and X_i, respectively. Hence, the moment generating function of S_N is

$$M_{S_N}(t) = P_N[M_{X_i}(t)] = M_N[\ln M_{X_i}(t)]. \tag{2.2}$$

The function $M_{S_N}(t)$ is well defined for $|t| \le a$ for some $a > 0$. Hence, it follows that

$$E[S_N] = E[N]E[X_i], \tag{2.3}$$

$$E[S_N^2] = E[N]E[X_i^2] + E[N(N-1)]E[X_i]E[X_j], \tag{2.4}$$

$$Var[S_N] = E[N]Var[X_i] + Var[N](E[X_i])^2. \tag{2.5}$$

3. ESTIMATION OF THE DISTRIBUTION MEANS UNDER SECOND-MOMENT ASSUMPTIONS

Let $\theta = E(X_i;\theta)$ and let $\lambda = E(N;\lambda)$. Suppose we observe $Y_1,\ldots,Y_n$, independent, identically distributed random variables such that Y_1 has the distribution of $S_N = X_1 + \cdots + X_N$. Note that neither N nor $X_1,\ldots,X_N$ are directly observed; the data only consists of n realizations of the random variable S_N and are denoted by $Y_1,\ldots,Y_n$.

Our goal is to estimate θ and λ, the means of the distributions of X_i and N, respectively, based on $(Y_1,\ldots,Y_n)$. Clearly, this will be impossible without further assumptions regarding the distributions of X_i and N.

Define the functions $V:\Lambda \to [0,\infty)$ and $W:\Theta \to [0,\infty)$ by

$$V(\lambda) = Var(N;\lambda) \quad \textit{and} \quad W(\theta) = Var(X_i;\theta)$$

and assume that V and W are known. Using equations (2.3) and (2.5) we can show that Y_1 has mean $\theta\lambda$ and variance

$$\theta^2 V(\lambda) + \lambda W(\theta).$$

We propose to estimate (θ,λ) using the first two moments of the distribution of Y_1. Expressing the parameters θ and λ in terms of the first two moments of the distribution of Y_1 proceeds this approach. The assumption that V and W are known is meant to be a weaker version of the assumption of parametric models for the distributions. This method might be used when one is willing to assume that $V(\lambda) = \lambda$, as in the Poisson distribution, but is not willing to assume that the Poisson distribution actually holds. Thus, our approach in this section is similar to the quasi-likelihood method described, e.g., in McCullagh and Nelder (1989, chapter 9).

Since the first moment of Y_1 is $\theta\lambda$, it is necessary that the second moment of Y_1 is not a function of $\theta\lambda$; hence, we require the following assumption:

Assumption 1

For $(\theta,\lambda) \in \Theta \times \Lambda$, $\theta^2 V(\lambda) + \lambda W(\theta)$ is not a function of $\theta\lambda$.

Let

$$\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

and

$$S^2 = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \bar{Y})^2.$$

Define the estimator $(\bar{\theta}, \bar{\lambda})$ as the solution to

$$\bar{\theta}\,\bar{\lambda} = \bar{Y}, \quad \bar{\theta}^2 V(\bar{\lambda}) + \bar{\lambda} W(\bar{\theta}) = S^2, \tag{3.1}$$

provided that such solution exists.

Determination of the asymptotic distribution of $(\bar{\theta}, \bar{\lambda})$ requires the following regularity conditions:

Assumption 2

The parameter space $\Theta \times \Lambda$ is an open subset of $\Re^2$.

Under assumption 2, the true parameter value (θ, λ) is an interior point of the parameter space.

Assumption 3

Define a function $H : \Theta \times \Lambda \to A$ by

$$H(\theta, \lambda) = (\theta\lambda, \theta^2 V(\lambda) + \lambda W(\theta)).$$

Then A is an open subset of $\Re^2$ and $H(\theta, \lambda)$ is a one-to-one function with a continuously differentiable inverse.

Under assumption 3, the estimators $\bar{\theta}$ and $\bar{\lambda}$ can be written as functions of $\bar{Y}$ and S^2. This fact, together with the asymptotic distribution of $(\bar{Y}, S^2)$ can be used to determine the asymptotic distribution of $(\bar{\theta}, \bar{\lambda})$.

Assumption 4

For all $(\theta, \lambda) \in \Theta \times \Lambda$, $E(Y_1^4; \theta, \lambda) < \infty$.

Assumption 4 is used to apply the strong law of large numbers and the central limit theorem to sums of the form $\sum_{j=1}^{n} Y_j$ and $\sum_{j=1}^{n} Y_j^2$.

For a non-negative definite 2×2 matrix L, let $N_2(\mathbf{0}, L)$ denote a random variable with a bivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix L; let $\mu_3(\theta, \lambda)$ and $\mu_4(\theta, \lambda)$ denote the third and fourth central moments, respectively, of Y_1.

Theorem 3.1

Assume that assumptions 2-4 hold. Then, with probability approaching 1 as $n \to \infty$, the estimator $(\overline{\theta}, \overline{\lambda})$ exists and

$$\sqrt{n}\begin{pmatrix} \overline{\theta} - \theta \\ \overline{\lambda} - \lambda \end{pmatrix} \to N_2(\mathbf{0}, \Sigma(\theta,\lambda)) \quad as \quad n \to \infty$$

where

$$\Sigma(\theta,\lambda) = M(\theta,\lambda)\begin{pmatrix} Var(Y_1;\theta,\lambda) & \mu_3(\theta,\lambda) \\ \mu_3(\theta,\lambda) & \mu_4(\theta,\lambda) - Var(Y_1;\theta,\lambda)^2 \end{pmatrix}^{-1} M(\theta,\lambda)^T$$

and

$$M(\theta,\lambda) = \begin{pmatrix} \lambda & \theta \\ 2\theta V(\lambda) + \lambda W'(\theta) & V'(\lambda)\theta^2 + W(\theta) \end{pmatrix}.$$

Proof:

Since here we are concerned with the limiting behaviour of the estimators $\overline{\theta}, \overline{\lambda}$, we write $\overline{Y}_n$ and S_n^2 for $\overline{Y}$ and S^2, respectively, to emphasize their dependence on n.

First consider consistency of $(\overline{\theta}, \overline{\lambda})$. Under assumption 4,

$$\overline{Y}_n \to E(Y_1;\theta,\lambda) = \theta\lambda \ w.p. \quad 1 \quad as \quad n \to \infty,$$

where "w.p." means "with probability", and

$$S_n^2 \to Var(Y_1;\theta,\lambda) = \theta^2 V(\lambda) + \lambda W(\theta) \ \ w.p. \quad 1 \quad as \quad n \to \infty$$

so that $(\overline{Y}_n, S_n^2)$ converges to $H(\theta,\lambda)$ with probability 1. It follows that, with probability 1, $(\overline{Y}_n, S_n^2)$ lies in A for sufficiently large n.

By assumption 3 there exists a continuous function $h = H^{-1}(.,.)$, the inverse of $H(.,.)$, such that, when $(\overline{Y}_n, S_n^2) \in A$, then $(\overline{\theta}, \overline{\lambda}) = h(\overline{Y}_n, S_n^2)$. Thus, with probability 1, $(\overline{\theta}, \overline{\lambda}) = h(\overline{Y}_n, S_n^2)$ for sufficiently large n and $(\overline{Y}_n, S_n^2)$ converges to $H(\theta,\lambda)$. It follows that $(\overline{\theta}, \overline{\lambda})$ is a consistent estimator of (θ,λ).

Now consider the asymptotic distribution of $(\overline{\theta}, \overline{\lambda})$. Under assumption 4,

$$\sqrt{n}\left(\begin{pmatrix} \overline{Y}_n \\ S_n^2 \end{pmatrix} - H(\theta,\lambda)\right) \to N_2(\mathbf{0}, B(\theta,\lambda)) \quad as \quad n \to \infty$$

where

$$B(\theta,\lambda)=\begin{pmatrix} Var(Y_1;\theta,\lambda) & \mu_3(\theta,\lambda) \\ \mu_3(\theta,\lambda) & \mu_4(\theta,\lambda)-Var(Y_1;\theta,\lambda)^2 \end{pmatrix};$$

see, e.g., example 13.3 of Severini (2005).

Since, for sufficiently large n, $(\overline{\theta},\overline{\lambda})=h(\overline{Y}_n,S_n^2)$, it follows from the δ-method (see, e.g., Severini, 2005, section 13.2), that

$$\sqrt{n}\begin{pmatrix} \overline{\theta}-\theta \\ \overline{\lambda}-\lambda \end{pmatrix} \to N_2(\mathbf{0},D(\theta,\lambda))$$

where

$$D(\theta,\lambda)=h'(H(\theta,\lambda))B(\theta,\lambda)h'(H(\theta,\lambda))^T$$

and h' denotes the matrix of partial derivatives of h. Note that, since $h=H^{-1}(.,.)$, $h(H(\theta,\lambda))=(\theta,\lambda)$; it follows that

$$h'(H(\theta,\lambda))=H'(\theta,\lambda)^{-1},$$

the matrix inverse of $H'(.,.)$, where H' denotes the matrix of derivatives of $H(\theta,\lambda)$ with respect to (θ,λ). It follows immediately from the expression for $H(\theta,\lambda)$ that

$$H'(\theta,\lambda)=\begin{pmatrix} \lambda & \theta \\ 2\theta V(\lambda)+\lambda W'(\theta) & V'(\lambda)\theta^2+W(\theta) \end{pmatrix}.$$

The result follows.

Let

$$\hat{\mu}_3=\frac{1}{n}\sum_{j=1}^{n}(Y_j-\overline{Y})^3,$$

$$\hat{\mu}_4=\frac{1}{n}\sum_{j=1}^{n}(Y_j-\overline{Y})^4$$

and

$$B_n=\begin{pmatrix} S^2 & \hat{\mu}_3 \\ \hat{\mu}_3 & \hat{\mu}_4-S^4 \end{pmatrix}.$$

Define

$$\hat{\Sigma}_n = M(\bar{\theta},\bar{\lambda})^{-1} B_n M(\bar{\theta},\bar{\lambda})^T.$$

Theorem 3.2

Assume that assumptions 2-4 hold. Then

$$\hat{\Sigma}_n \to \Sigma(\theta,\lambda) \quad as \quad n \to \infty.$$

Proof:

Under assumption 4, as $n \to \infty$,

$$S^2 \to Var(Y_1;\theta,\lambda), \quad \hat{\mu}_3 \to \mu_3(\theta,\lambda), \quad and \quad \hat{\mu}_4 \to \mu_4(\theta,\lambda).$$

It follows that $B_n \to B(\theta,\lambda)$ as $n \to \infty$. Since $M(\theta,\lambda)$ is a continuous function of (θ,λ), and $(\bar{\theta},\bar{\lambda})$ is a consistent estimator of (θ,λ), it follows that $M(\bar{\theta},\bar{\lambda}) \to M(\theta,\lambda)$ as $n \to \infty$. The result follows.

4. NONPARAMETRIC ESTIMATION OF THE COMPOUNDING DISTRIBUTION

4.1 *General approach*

Let P denote the probability generating function of Y_i, let R denote the probability generating function of the event distribution, and let Q denote the probability generating function of the compounding distribution. Assume that we have a parametric model with parameter λ for the event distribution; hence, we write $R(\cdot;\lambda)$ for R as

$$R(Q(t);\lambda) = P(t), \quad |t| \le 1.$$

Let

$$\hat{P}(t) = \frac{1}{n}\sum_{j=1}^{n} t^{Y_j}$$

denote the empirical probability generating function based on $Y_1,\ldots,Y_n$. Then $\hat{P}(t)$ is a consistent estimator of $P(t)$ for $|t| \le 1$. Hence, an estimator of the compounding distribution can be obtained by setting $R(Q(t);\hat{\lambda})$ equal to $\hat{P}(t)$, where $\hat{\lambda}$ is an appropriate estimator of λ.

In order to carry out this approach, there are several issues which must be considered. First, we must ensure that λ and Q are identified in this model. Second, an estimator of λ must be constructed. Finally, we must be able to use the identity

$$R(Q(t);\hat{\lambda}) = \hat{P}(t) \tag{4.1}$$

in order to obtain an estimator of the compounding distribution.

First consider identification. The following assumption is sufficient for Q and λ to be identified.

Assumption 5

1. $R(t_1;\lambda) = R(t_2;\lambda)$ if and only if $t_1 = t_2$.
2. $R(0;\lambda_1) = R(0;\lambda_2)$ if and only if $\lambda_1 = \lambda_2$.

Lemma 4.1

Assume that assumption 5 is satisfied. Then $R(Q_1(t);\lambda_1) = R(Q_2(t);\lambda_2)$ for $|t| \le 1$ if and only if $Q_1 = Q_2$ and $\lambda_1 = \lambda_2$.

Proof:

Clearly, if $Q_1 = Q_2$ and $\lambda_1 = \lambda_2$, then $R(Q_1(t);\lambda_1) = R(Q_2(t);\lambda_2)$ for $|t| \le 1$. Hence, assume that $R(Q_1(t);\lambda_1) = R(Q_2(t);\lambda_2)$ for $|t| \le 1$. Note that $Q_1(0) = Q_2(0) = 0$. Hence, setting $t = 0$, it follows that $R(0;\lambda_1) = R(0;\lambda_2)$; by part (1) of assumption 5, it follows that $\lambda_1 = \lambda_2$.

Fix t. By part (2) of assumption 5, $R(Q_1(t);\lambda) = R(Q_2(t);\lambda)$ implies that $Q_1(t) = Q_2(t)$. Since t is arbitrary, $Q_1 = Q_2$, proving the result.

Now consider estimation of λ. It follows from the identification result that an estimator of λ can be obtained by solving

$$R(0;\hat{\lambda}) = \hat{P}(0) \tag{4.2}$$

for $\hat{\lambda}$. However, the estimator given in (4.2) will only be useful for those distributions for which $P(0)$ is not close to 0. An alternative approach is to base an estimator of a second-moment assumption, as discussed in section 3.

Finally, we must use the identity in (4.1) to obtain an estimator of the compounding distribution. The details of this will depend on the parametric model used for the event distribution. In the remainder of this section we consider two choices for this distribution. In section 4.2, the event distribution is taken to be a

Poisson distribution; this case was also considered by Buchmann and Grübel (2003), who developed a similar method of estimation. In section 4.3, the event distribution is taken to be a geometric distribution.

4.2 *Estimation under a compound Poisson model*

The compound Poisson distribution has a major importance in the class of compound distributions. This is because of using it in the probability theory and its applications in biology, risk theory, meteorology, health science, etc.

Under the assumption that N has a Poisson distribution with mean λ, it follows from the discussion in section 4.1 that

$$P(t) = \exp\{\lambda[Q(t) - 1]\}, \quad |t| \leq 1.$$

Suppose that the parameter λ is known. Hence, we can estimate Q by $\hat{Q}$ where

$$\exp\{\lambda[\hat{Q}(t) - 1]\} = \hat{P}(t) \tag{4.3}$$

and let

$$\hat{p}_k = \frac{1}{n}\sum_{j=1}^{n} I_{\{Y_j = k\}}$$

so that $\hat{P}$ can also be written as

$$\hat{P}(t) = \sum_{k=0}^{\infty} t^k \hat{p}_k.$$

Here $I_{\{Y_j=k\}} = 1$ if $Y_j = k$ and 0 otherwise.

For $j = 1, 2, \ldots$, let $q_j = \Pr(X_i = j)$. Our goal is to estimate $q_1, q_2, \ldots$ based on $\hat{p}_0, \hat{p}_1, \ldots$ or, equivalently, based on $\hat{P}(t)$. To do this, we make the following assumption:

Assumption 6

There exists a positive integer M such that $\Pr(X_i > M) = 0$ and, hence,

$$q_{M+1} = q_{M+2} = \cdots = 0.$$

Under assumption 6, the number of unknown parameters is finite, which simplifies certain technical arguments. Note that M can be taken to be very large so that assumption 6 is nearly always satisfied in practice.

We can estimate $q_1, q_2, \ldots$ by finding $\hat{Q}$ to solve (4.3) and then differentiating in order to obtain the estimators $\hat{q}_1, \hat{q}_2, \ldots$:

$$\hat{q}_1 = \hat{Q}'(0),\ \hat{q}_2 = \hat{Q}''(0)/2!,\ \ \hat{q}_3 = \hat{Q}'''(0)/3!$$

and so on.

Define $\bar{Q}$ to be the solution to

$$\exp\{\bar{Q}(t) - 1\} = \hat{P}(t),\ |t| \le 1. \tag{4.4}$$

Then

$$\hat{Q}(t) = \frac{1}{\lambda}[\bar{Q}(t) - 1] + 1$$

and $\bar{q}_1, \bar{q}_2, \ldots$, the probabilities corresponding to $\bar{Q}$, are related to $\hat{q}_1, \hat{q}_2, \ldots$ through

$$\hat{q}_k = \frac{1}{\lambda}\bar{q}_k,\ k = 1, 2, \ldots.$$

In the usual case in which λ is unknown, it must be estimated. Let $\hat{\lambda}$ denote an estimator of λ. Then $q_1, q_2, \ldots$ can be estimated by

$$\hat{q}_k = \frac{1}{\hat{\lambda}}\bar{q}_k,\ k = 1, 2, \ldots.$$

One approach to estimate λ is to use an estimator based on equation (4.2). This yields the estimator

$$\hat{\lambda} = -\log \hat{p}_0. \tag{4.5}$$

An alternative approach is to use the method described in section 3. In this case, we assume the existence of a known function W such that $W(\theta) = Var(X_i)$ where $\theta = E(X_i;\theta)$. Then an estimator $\hat{\lambda}$ is given by the solution to equation (3.1), which may be written as

$$\frac{1}{\hat{\lambda}}\sum_{j=1}^{M} j^2 \bar{q}_j = \frac{1}{\hat{\lambda}^2}\sum_{j=1}^{M}\left(\sum_{j=1}^{M} j\,\bar{q}_j\right)^2 + W\left(\frac{1}{\hat{\lambda}}\sum_{j=1}^{M} j\,\bar{q}_j\right). \tag{4.6}$$

Other estimators of λ can be used, provided that the following condition is satisfied.

Assumption 7

The estimator $\hat{\lambda}$ is a continuously differentiable function of $(\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_M)$ and is a consistent estimator of λ.

Let $\hat{q} = (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$ denote the vector of probabilities corresponding to $\hat{Q}$; similarly, define $\overline{q} = (\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_M)$ and $\hat{p} = (\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_M)$. A method of calculating $\hat{q}_1, \hat{q}_2, \ldots$ is given in the following theorem.

Theorem 4.1

Assume that assumptions 5-7 hold and that N has a Poisson distribution with mean λ.

Define an $M \times M$ matrix H as follows:

$$H_{ij} = \begin{cases} 0, & \textit{if} \quad i < j, \\ \dfrac{j}{i} \hat{p}_{i-j} & \textit{if} \quad i = j, \ldots, M, \end{cases} \quad j = 1, 2, \ldots, M.$$

Hence

$$\overline{q} = H^{-1} \hat{p}$$

and

$$\hat{q} = \frac{1}{\hat{\lambda}} \overline{q}.$$

Proof:

Let $\alpha(t) = \hat{P}(t)$ and let $\beta(t) = \overline{Q}(t) - 1$. Therefore

$$\beta(t) = \sum_{j=1}^{M} \overline{q}_j t^j - 1,$$

$$\alpha(t) = \sum_{j=0}^{\infty} \hat{p}_j t^j,$$

and $\beta(t) = \log \alpha(t)$, using equation (4.4). It now follows from Severini (2005, lemma 4.1) that

$$\hat{p}_{r+1} = \sum_{j=0}^{r} \frac{j+1}{r+1} \overline{q}_{j+1} \hat{p}_{r-j}, \quad r = 0, 1, 2, \ldots. \tag{4.7}$$

Expressing in terms of the matrix H and the vectors $\hat{q}$ and $\hat{p}$, equation (4.7) can be written as

$$H\overline{q} = \hat{p}.$$

The result for $\hat{q}$ now follows from the relationship between $\hat{q}$ and $\overline{q}$.

The following result shows that $(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$ is asymptotically normally distributed.

Theorem 4.2

Assume that assumptions 5-7 hold and that N has a Poisson distribution with mean λ. Therefore, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \hat{q}_1 - q_1 \\ \hat{q}_2 - q_2 \\ \vdots \\ \hat{q}_M - q_M \end{pmatrix}$$

converges in distribution to an M dimensional multivariate normal distribution with mean vector $\mathbf{0}$.

Proof:

It follows from equation (4.7) that $(\overline{q}_1, \overline{q}_2, \ldots, \overline{q}_M)$ is a continuously differentiable function of $(\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_M)$. Under assumption 7, it follows that $(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$ is a continuously differentiable function of $(\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_M)$.

The result now follows from the δ-method (e.g., Severini, 2005, section 13.2) together with the fact that $(\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_M)$ is asymptotically normally distributed (e.g., Severini, 2005, example 12.7).

The asymptotic covariance matrix of $(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$, which is not specified in theorem 4.2, can be obtained using the δ-method (see, e.g., Severini, 2005, theorem 13.1). Since $(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$ is a complicated function of $(\hat{p}_0, \ldots, \hat{p}_M)$, we do not give its expression here. The computation of such expression can be obtained via the formula in Severini (2005, theorem 13.1). Note that the estimators $\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M$ are not guaranteed to be non-negative nor are they guaranteed to sum to 1. Thus, in practice, any negative estimate should be replaced by 0. It may also be helpful to normalize the estimates by dividing them by their sum.

4.3 *Estimation under a compound geometric model*

In this subsection, we consider the compound geometric distribution which plays an important role in reliability, queueing, regenerative processes, and insurance applications.

We now consider the case in which N has a geometric distribution with the parameter λ and probability function

$$P(N=n)=(1-\lambda)\lambda^{n},\ \ n=0,1,2,\ldots \tag{4.8}$$

where $0<\lambda<1$. The analysis in this case is very similar to the analysis given in section 4.2 for the compound Poisson model and, hence, we keep the treatment here in brief.

For the compound geometric model,

$$P(t)=\frac{1-\lambda}{1-\lambda Q(t)},\ \ |t|\leq 1.$$

Let $\hat{\lambda}$ denote an estimator of λ satisfying assumption 7 and define $\hat{Q}$ by

$$\hat{Q}(t)=\frac{1}{\hat{\lambda}}+\frac{1-1/\hat{\lambda}}{\hat{P}(t)}. \tag{4.9}$$

A method of calculating $\hat{q}_1,\hat{q}_2,\ldots$ is given in the following theorem.

Theorem 4.3

Assume that assumptions 5-7 hold and that the distribution of N is given by (4.8). Hence

$$\hat{q}_k=\frac{\hat{p}_k}{\hat{\lambda}\hat{p}_0}-\sum_{j=1}^{k-1}\binom{k}{j}\frac{\hat{p}_{k-j}}{\hat{p}_0}\hat{q}_j,\quad k=1,2,\ldots. \tag{4.10}$$

Proof:

Rewriting (4.9) as

$$\hat{\lambda}\hat{Q}(t)\hat{P}(t)=\hat{P}(t)+\hat{\lambda}-1$$

for $|t|\leq 1$, and differentiating this expression implies

$$\hat{\lambda}\sum_{j=0}^{k}\binom{k}{j}\hat{Q}^{(j)}(t)\hat{P}^{(k-j)}(t)=\hat{P}^{(k)}(t),\ \ k=1,2,\ldots.$$

Evaluating this expression at $t=0$, and setting $\hat{q}_0=0$, lead to (4.10).

The following result shows that $(\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_M)$ is asymptotically normally distributed. The proof is essentially the same as the proof of theorem 4.2 and, hence, is omitted.

Theorem 4.4

Assume that assumptions 5-7 hold and that the distribution of N is given by (4.8). Therefore, as $n \to \infty$,

$$\sqrt{n}\begin{pmatrix} \hat{q}_1 - q_1 \\ \hat{q}_2 - q_2 \\ \vdots \\ \hat{q}_M - q_M \end{pmatrix}$$

converges in distribution to an M dimensional multivariate normal distribution with mean vector $\mathbf{0}$.

The comments following theorem 4.2 can be applied here as well.

Remark 4.1.

Our approach in section 4.1 works well with other event distributions (e.g. binomial and negative binomial) as long as they obey the assumptions in this section. Also, the results that can be obtained to such distributions are very similar to those already given in subsections 4.2 and 4.3, so it is not necessary to include such results here.

Mathematics Department, Faculty of Science, HASSAN S. BAKOUCH
Tanta University, Tanta, Egypt

Department of Statistics, THOMAS A. SEVERINI
Northwestern University, USA

ACKNOWLEDGMENTS

The authors thank the referees for their suggestions and comments which have significantly improved the paper. The second author gratefully acknowledges the support of the U.S. National Science Foundation.

REFERENCES

H. S. BAKOUCH, M. M. RISTIĆ (2009), *Zero truncated Poisson integer-valued AR(1) model*, "Metrika", doi 10.1007/s00184-009-0252-5.

G. BUCHMANN, R. GRÜBEL (2003), *Decompounding: an estimation problem for Poisson random sums*, "Annals of Statistics", 31, pp. 1054-1074.

G. BUCHMANN, R. GRÜBEL (2004), *Decompounding Poisson random sums: Recursively truncated estimates in the discrete case*, "Annals of the Institute of Statistical Mathematics", 56, pp. 743-756.

J. CAI, G. E. WILLMOT (2005), *Monotonicity and aging properties of random sums*, "Statistics & Probability Letters", 73, pp. 381-392.

C. A. CHARALAMBIDES (2005), *Combinatorial methods in discrete distributions*, New York: Wiley.

C. CHATFIELD, C. M. THEOBALD (1973), *Mixtures and random sums*, "The Statistician", 22, pp. 281-287.

S. A. KLUGMAN, H. H. PANJER, G. E. WILLMOT (2004), *Loss models: from data to decisions*, 2nd ed., Wiley-Interscience.

O. LUNDBERG (1964), *On random processes and their application to sickness and accident statistics*, 2nd ed., Almqvist and Wiksell (Uppsala).

P. MCCULLAGH, J. A. NELDER (1989), *Generalized linear models*, 2nd ed., Chapman & Hall/CRC.

J. N. MEDHI (1972), *Stochastic processes*, 3rd ed., John Wiley: New York.

H. H. PANJER (2006), *Operational risk: modeling analytics*, New York: Wiley.

E. PEKÖZ, S. M. ROSS (2004), *Compound random variables*, "Probability in the Engineering and Informational Sciences", 18, pp. 473-484.

S. M. PITTS (1994), *Nonparametric estimation of compound distributions with applications in insurance*, "Annals of the Institute of Statistical Mathematics", 46, pp. 537-555.

M. M. RISTIĆ, H. S. BAKOUCH, A. S. NASTIĆ (2009), *A new geometric first-order integer-valued autoregressive (NGINAR(1)) process*, "Journal of Statistical Planning and Inference", 139, pp. 2218-2226.

M. SAHINOGLU (1992), *Compound Poisson software reliability model*, "IEEE Transactions on Software Engineering", 18, pp. 624-630.

T. A. SEVERINI (2005), *Elements of distribution theory*, Cambridge University Press: Cambridge.

M. WEBA (2007), *Optimal prediction of compound mixed Poisson processes*, "Journal of Statistical Planning and Inference", 137, pp. 1332-1342.

SUMMARY

Nonparametric estimation in random sum models

Let $X_1, X_2, \ldots, X_N$ be independent, identically distributed, non-negative, integer-valued random variables and let N be a non-negative, integer-valued random variable independent of $X_1, X_2, \ldots, X_N$. In this paper, we consider two nonparametric estimation problems for the random sum variable $S_N = \sum_{i=1}^{N} X_i$, $S_0 = X_0 = 0$. The first is the estimation of the means of X_i and N based on the second-moment assumptions on distributions of X_i and N. The second is the nonparametric estimation of the distribution of X_i given a parametric model for the distribution of N. Some asymptotic properties of the proposed estimators are discussed.