CIRJE-F-361

On a Limiting quasi-multinomial distribution

Nobuaki Hoshino
Kanazawa University

August 2005

On a limiting quasi-multinomial distribution

Nobuaki Hoshino*
Faculty of Economics, Kanazawa University

August 2005

Abstract

A random clustering distribution is useful for modeling count data. The present article derives a new distribution of this type from the Lagrangian Poisson distribution, based on the result that any infinitely divisible distribution over nonnegative integers produces a random clustering distribution through conditioning and a limiting argument that is equivalent to the law of small numbers. The resulting distribution is shown to be tractable. Its application is also presented.

Keywords: Borel distribution, Compound Poisson, Consul's generalized Poisson, Random partitioning, Species abundance

1 Introduction

Hoshino (2005) considered a general method to produce a random partitioning distribution of a positive integer from an infinitely divisible distribution over nonnegative integers; see Steutel and van Harn (2004) for the concept of infinite divisibility. As an instance of a random clustering distribution produced in this way, the present article proposes a new distribution useful for modeling count data. The adopted method reads as follows.

Let N_0 and N be the sets of nonnegative integers and positive integers respectively. Let us denote the set of all unordered partitions of a positive integer n by

$$\mathcal{S}_n := \{\mathsf{s}_n := (s_1, s_2, \ldots, s_n) : s_i \in \mathsf{N}_0, i = 1, 2, \ldots, n, \sum_{i=1}^{n} i s_i = n\}.$$

We will mainly consider a random vector

$$\mathsf{S}_n := (S_1, S_2, \ldots, S_n),$$

where $\mathrm{P}(\mathsf{S}_n = \mathsf{s}_n)$ is defined for $\mathsf{s}_n \in \mathcal{S}_n$.

Suppose that random variables $F_1, F_2, \ldots, F_J$ are independently and identically distributed (i.i.d.) over N_0. Let

$$S_i := \sum_{j=1}^{J} 1(F_j = i), \quad i \in \mathsf{N}_0,$$

*Address for correspondence : Nobuaki Hoshino, Faculty of Economics, Kanazawa University, Kakuma-machi, Kanazawa-shi, Ishikawa, Japan. E-mail: hoshino@kenroku.kanazawa-u.ac.jp

$$\begin{tikzcd}
\boxed{\mathsf{S}_\mathsf{n} \text{ over } \mathcal{S}_\mathsf{n}(J)} \arrow[d, "\text{limiting}"'] & \boxed{\mathsf{S} \text{ over } \mathcal{S}_\infty(J)} \arrow[l, "\text{conditioning}"', "N"] \arrow[d, "\text{limiting}"] \\
\boxed{\mathsf{S}_\mathsf{n} \text{ over } \mathcal{S}_\mathsf{n}} & \boxed{\mathsf{S} \text{ over } \mathsf{N_0}^\infty} \arrow[l, "\text{conditioning}"', "N"]
\end{tikzcd}$$

Figure 1: Hoshino's (2005) construction of random clustering distributions

and write

$$\mathsf{S} := (S_1, S_2, \ldots).$$

Then $\mathrm{P}(\mathsf{S} = \mathsf{s})$ is defined over

$$\mathcal{S}_\infty(J) := \{\mathsf{s} := (s_1, s_2, \ldots) : s_\mathsf{i} \in \mathsf{N}_0, i = 1, 2, \ldots, \sum_{\mathsf{i}=1}^{\infty} s_\mathsf{i} \leq J\},$$

where S is multinomially distributed. An infinite dimensional distribution is formally defined by the sequence of distributions; see Hoshino (2005, Appendix A) for the formal derivation of this multinomial distribution.

Conditioning random variables on their total is a common idea to produce random partitioning distributions; see Arratia et al. (2003) for a review. The total sum $F_1 + F_2 + \cdots + F_\mathsf{J}$ is denoted by

$$N := \sum_{\mathsf{i}=1}^{\infty} i \cdot S_\mathsf{i}. \tag{1}$$

The conditional distribution $\mathrm{P}(\mathsf{S} = \mathsf{s}|N = n) = \mathrm{P}(\mathsf{S}_\mathsf{n} = \mathsf{s}_\mathsf{n}|N = n)$ is then defined over

$$\mathcal{S}_\mathsf{n}(J) := \{\mathsf{s}_\mathsf{n} : s_\mathsf{i} \in \mathsf{N}_0, i = 1, 2, \ldots, n, \sum_{\mathsf{i}=1}^{\mathsf{n}} i s_\mathsf{i} = n, \sum_{\mathsf{i}=1}^{\mathsf{n}} s_\mathsf{i} \leq J\}.$$

Suppose that F_j is subject to an infinitely divisible distribution over N_0. Then we can take $J \to \infty$, while the distribution of N remains unchanged. The limiting distribution of S_n constitutes a random partitioning distribution of n over $\mathcal{S}_\mathsf{n}$ actually.

In this construction depending on the infinite divisibility, the order of the conditioning on N and the limiting of $J \to \infty$ is exchangeable. That is, the multinomially distributed S over $\mathcal{S}_\infty(J)$ converges to a proper distribution defined over $\mathsf{N_0}^\infty$ by first taking $J \to \infty$, where the distribution of N remains unchanged. The limiting distribution of S is the joint distribution of independent Poisson variables because the law of small numbers holds; see Hoshino (2005, Theorem 2.1). The conditional distribution of the limiting distribution over $\mathsf{N_0}^\infty$ given $N = n$ then coincides with the random partitioning distribution of n over $\mathcal{S}_\mathsf{n}$. Figure 1 summarizes these arguments.

The quintessence of a random partitioning distribution of this type is the Ewens distribution (Ewens (1972)). This instance is produced from the negative binomial distribution, which is infinitely divisible over N_0; see Hoshino and Takemura (1998). Another example derived from the inverse Gaussian-Poisson mixture (Holla (1966)) is investigated in Hoshino (2002).

The present article derives another from the Lagrangian Poisson distribution; see Consul (1989) on this infinitely divisible distribution. The resulting random partitioning distribution seems new and is applicable.

In the statistical literature, S_i's are called size indices (Sibuya (1993)) or frequencies of frequencies (Good (1953)). They are used, for example, to summarize the data of numbers of many species. In this case, F_j corresponds to the number of individuals of the j-th species, and S_1 is important because it represents the number of endangered species. Moreover, in contingency table analysis, F_j expresses the number of individuals in the j-th cell, and S_0 is then the number of empty cells. We denote the number of nonempty cells by

$$U = \sum_{i=1}^{\infty} S_i = J - S_0, \tag{2}$$

which may correspond to the total number of species. The estimation of U is an interesting subject, and its vast context was surveyed by Bunge and Fitzpatrick (1993). When the total number of individuals N is fixed at n, we write in particular

$$U_n = \sum_{i=1}^{n} S_i.$$

The organization of the present article is as follows. Section 2 explains the Lagrangian Poisson distribution and its derivatives. Section 3 introduces a new distribution and elucidates its properties useful for application. Section 4 consists of the parameter estimation and an application result, with a conclusion.

2 Relating distributions

This section briefly describes a few distributions used in the main argument.

Consul and Jain (1973) proposed a generalized Poisson distribution or the Lagrangian Poisson distribution defined by the probability function:

$$\mathrm{P}(X = x; \theta, \lambda) = \frac{\theta(\theta + x\lambda)^{x-1}}{x!} \exp(-\theta - x\lambda), \quad x \in \mathbb{N}_0, \tag{3}$$

where $\theta > 0$, $0 < \lambda < 1$. This distribution (3) is referred to by $LP(\theta, \lambda)$ in the following. The parameter θ is proportional to the mean. When $\lambda = 0$, $LP(\theta, \lambda)$ degenerates into the Poisson distribution with mean θ; λ is an indicator of overdispersion. Negative λ, which produces an improper distribution, is not allowed in the present article.

The Lagrangian Poisson distribution is infinitely divisible because its probability generating function (pgf) is expressed as a compound Poisson form:

$$G(z) = \exp(\theta(g(z) - 1)),$$

where $g(z)$ is the pgf of the Borel distribution:

$$\mathrm{P}(X = i; \lambda) = \frac{(\lambda i)^{i-1}}{i!} \exp(-\lambda i), \quad i \in \mathbb{N}. \tag{4}$$

See Johnson et al. (1993, p.394) for more on the compound Poisson representation of the Lagrangian Poisson distribution.

The quasi-multinomial (QM) distribution was proposed by Janardan (1975). Its construction is given below. Let $F_j, j = 1, 2, \ldots, J$, be independently distributed as $LP(\theta_j, \lambda)$; then N is distributed as $LP(\sum \theta_j, \lambda)$, and the conditional distribution of F_j's given N becomes

$$\begin{aligned} &\mathrm{P}(F_1 = g_1, F_2 = g_2, \ldots, F_J = g_J | N = n; \theta_1, \theta_2, \ldots, \theta_J, \lambda) \\ &= \binom{n}{g_1 g_2 \cdots g_J} \frac{1}{\sum \theta_j (\sum \theta_j + n\lambda)^{n-1}} \prod_{j=1}^{J} \theta_j (\theta_j + g_j \lambda)^{g_j - 1}, \end{aligned} \tag{5}$$

where $g_j \in \mathsf{N}_0, \sum g_j = n$. This distribution (5) is called the QM distribution because it reduces to the quasi-binomial distribution (type 2) proposed by Consul and Mittal (1975) when $J = 2$. If $\lambda = 0$, (5) becomes the multinomial distribution. Reparameterizing as $p_j = \theta_j / \sum \theta_j$ makes this fact clearer.

When $\theta_1 = \theta_2 = \cdots = \theta_J = \theta$, i.e. cells are exchangeable, the QM distribution can be expressed simply in terms of size indices. Furthermore, the conditional distribution does not depend on θ, because N is complete and sufficient for θ in the exchangeable case; see Consul (1989, p.91). Let us adopt the reparameterization that $\lambda = \alpha\theta$, as "restricted generalized Poisson" (Consul (1989, p.5)). Then the exchangeable QM distribution is expressed for $0 < \alpha$ as

$$\begin{aligned} &\mathrm{P}(\mathsf{S}_n = \mathsf{s}_n | N = n; \alpha) \\ &= J!n! \frac{1}{J(J + n\alpha)^{n-1}} \prod_{i=0}^{n} \left(\frac{(1 + i\alpha)^{i-1}}{i!} \right)^{s_i} \frac{1}{s_i!}, \quad \mathsf{s}_n \in \mathcal{S}_n(J), \end{aligned} \tag{6}$$

where $0 \leq s_0 = J - \sum_{i=1}^{n} s_i$. Its unconditional distribution is multinomial:

$$\mathrm{P}(\mathsf{S} = \mathsf{s}; \theta, \lambda) = J!\theta^J \exp(-J\theta - n\lambda) \prod_{i=0}^{\infty} (\frac{(\theta + i\lambda)^{i-1}}{i!})^{s_i} \frac{1}{s_i!}, \quad \mathsf{s} \in \mathcal{S}_\infty(J), \tag{7}$$

where $n = \sum i s_i$; it should be noted that N is subject to $LP(J\theta, \lambda)$.

3 Main results

This section substantiates Hoshino's (2005) construction of a random partitioning distribution. As stated before, a new distribution is derived from the Lagrangian Poisson distribution. Some important properties of the derived distribution are investigated for application. All the proofs of theorems in this section are given in Appendix.

Our main theorem below derives a random partitioning distribution from (7) by first conditioning on N and second a limiting argument. The resulting distribution (8) is referred to by the Limiting Quasi-Multinomial (LQM) distribution.

Theorem 1 *If $J/\alpha \to \rho\, (> 0)$ as $J \to \infty$, the limiting distribution of* (6) *is*

$$\mathrm{P}(\mathsf{S}_n = \mathsf{s}_n | N = n; \rho) = n!\, \rho^{u-1} (\rho + n)^{1-n} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!} \right)^{s_i} \frac{1}{s_i!}, \quad \mathsf{s}_n \in \mathcal{S}_n, \tag{8}$$

where $u = \sum_{i=1}^{n} s_i$.

The limiting argument used in Theorem 1 implies that the number of cells goes to infinity while the distribution of N is unchanged: $J\theta$ is fixed at a positive constant $(\rho\lambda)$. In practice, the number of cells J tends to be very large. Thus this limiting has a sound basis.

To clarify the difference of the random partitioning distribution from similar one of Pitman (2003), let us mention the concept of partition structure proposed by Kingman (1978). A distribution that has the partition structure satisfies, for all $n \in \mathsf{N}$,

$$\begin{aligned} &\mathrm{P}(S_1 = s_1, S_2 = s_2, \ldots | N = n) \\ &= \mathrm{P}(S_1 = s_1 + 1, S_2 = s_2, \ldots | N = n+1) \frac{s_1 + 1}{n+1} \\ &\quad + \sum_{r=2}^{n+1} \mathrm{P}(S_1 = s_1, \ldots, S_{r-1} = s_{r-1} - 1, S_r = s_r + 1, \ldots | N = n+1) \frac{r(s_r + 1)}{n+1}, \end{aligned}$$

which implies that a given partition of n elements results from the deletion of one element uniformly at random from a partition of $n+1$ elements. The Ewens distribution has this partition structure, and Pitman (2003) discusses its generalized distributions that have the partition structure. The following fact, which is easily verified, shows that our construction is another generalization.

Remark 1 *The LQM distribution does not have Kingman's partition structure.*

Next we exchange the order of the conditioning and the limiting argument in the derivation of the LQM distribution (8). We apply first the limiting argument (Theorem 2) and second conditioning (Theorem 3).

Theorem 2 *Let $J\theta$ be fixed at finite and positive μ. If size indices are distributed as* (7)*, then $S_i, i \in \mathsf{N}$, becomes independently Poisson distributed with mean*

$$\mathrm{E}(S_i) = \frac{\mu(\lambda i)^{i-1}}{i!} \exp(-\lambda i), \tag{9}$$

as $J \to \infty$. Namely, the limiting distribution is

$$\mathrm{P}(\mathsf{S} = \mathsf{s}; \mu, \lambda) = \mu^u \exp(-\mu - n\lambda) \lambda^{n-u} \prod_{i=1}^{\infty} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!}, \quad \mathsf{s} \in \mathsf{N}_0{}^{\infty}, \tag{10}$$

where $u = \sum_{i=1}^{\infty} s_i, n = \sum_{i=1}^{\infty} i s_i$.

Theorem 3 *The conditional distribution of* (10) *given $N = n$ is* (8) *when $\mu = \rho\lambda$.*

In view of Theorem 2 and 3, the LQM distribution is the result of introducing dependence into independent Poisson variables by conditioning. Hence the dependence naturally diminishes when $n \to \infty$. The following theorem is an analogue of Sibuya (1993, Proposition 2.2), who dealt with the Ewens distribution. It is noteworthy that (11) equals (9) when $\lambda = 1$.

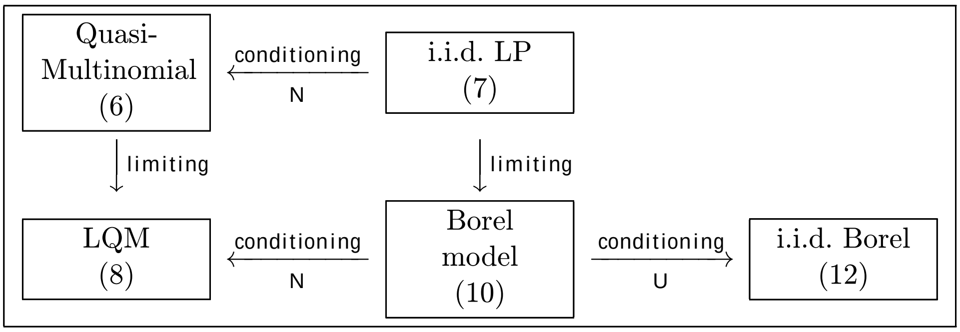

Figure 2: Relationships among 5 types of models

Theorem 4 *Let m be a finite fixed positive integer. As $n \to \infty$, the joint distribution of the first m components $(S_1, S_2, \ldots, S_m)$ of the argument of* (8) *converges to the joint distribution of independent Poisson variables with means*

$$\frac{\rho i^{i-1}}{i!} \exp(-i), \quad i = 1, 2, \ldots, m. \tag{11}$$

The expectation (9) is proportional to the Borel distribution's probability function (4), as it has to be; see Hoshino (2005, Theorem 2.1). This use of the Borel distribution is different from the orthodox use of independently and identically Borel distributed variables. The relationship between these uses, stated below, is evident when we note that U is Poisson distributed with mean μ under (10).

Theorem 5 *The conditional distribution of* (10) *given $U = u$ is multinomially distributed as*

$$\mathrm{P}(\mathsf{S} = \mathsf{s} | U = u; \lambda) = u! \prod_{i=1}^{\infty} \left(\frac{\exp(-\lambda i)(\lambda i)^{i-1}}{i!} \right)^{s_i} \frac{1}{s_i!}, \quad \mathsf{s} \in N_0{}^{\infty}(\sum s_i = u), \tag{12}$$

where $F_j, j = 1, 2, \ldots, u$, is independently and identically subject to the Borel distribution (4).

The conditional distribution (12) does not depend on μ, and thus U is sufficient for μ in (10). We realize that the uncertainty of U is conveniently modeled in (10), compared with the standard way (12) where u is fixed. This difference is especially significant when n varies subject to model's interpretation. For example, suppose that we estimate the total number of species in a population consisting of $n'(> n)$ individuals from n samples. Then the total number of species in the population should be larger than observed u, but the standard model (12) can not describe this situation. On the contrary in (10), the parameter μ' corresponding to the population may be set as $\mu \times n'/n$, which should result in a reasonable estimate of the total number of species. This ability is a common advantage of models such as (8), where U_n is random.

Figure 2 illustrates relationships stated in Theorem 1 to 5. It is comparable with Figure 1, which is the basic structure we exploited. The "i.i.d. Borel" part (12) was not mentioned in Section 1, but conditioning S on U to derive this part is possible for any proper infinitely divisible distribution of F_j. Hoshino (2004) introduced those 5 types of count data modeling.

ρ	$\mathrm{E}(S_1)$	$\mathrm{E}(S_2)$	$\mathrm{E}(S_3)$	$\mathrm{E}(S_4)$	$\mathrm{E}(S_5)$
1	0.368	0.136	0.075	0.049	0.035
10	3.684	1.357	0.750	0.491	0.353
100	36.678	13.453	7.401	4.826	3.457
300	107.064	38.207	20.451	12.973	9.041
500	171.329	58.701	30.165	18.370	12.288
700	228.885	74.828	36.688	21.316	13.604
900	280.098	87.153	40.667	22.485	13.655
10000	830.239	68.873	8.563	1.261	0.204

Table 1: The expectation of a size index (LQM, $n = 1000$)

The following investigation focuses upon some properties of the LQM distribution (8) useful for applications. The moments of size indices are of the first importance; see Section 4.2 for an application.

Theorem 6 *Suppose that size indices are subject to* (8). *For* $r_i \in \mathrm{N}_0$, *let us denote* $r := \sum r_i, l := \sum i r_i$. *Then, for* $l \leq n$,

$$\mathrm{E}(\prod_{i=1}^{n} S_i{}^{(r_i)}) = \frac{n!\rho^r(\rho+n)^{1-n}}{(n-l)!(\rho+(n-l))^{1-n+l}} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{r_i}, \tag{13}$$

where $n^{(r)} = n(n-1)\cdots(n-r+1), n^{(0)} = 1$.

In particular, for $i = 1, 2, \ldots, n$,

$$\mathrm{E}(S_i) = \frac{n!}{(n-i)!} \frac{\rho(\rho+n)^{1-n}}{(\rho+n-i)^{1-n+i}} \frac{i^{i-1}}{i!}, \tag{14}$$

and

$$\mathrm{E}(U_n) = \sum_{i=1}^{n} \mathrm{E}(S_i) = 1 + \frac{(n-1)\rho}{\rho+n}. \tag{15}$$

This expectation (15) is an easy consequence of Theorem 7 given later. Table 1 summarizes the numerical values of $\mathrm{E}(S_1)$ to $\mathrm{E}(S_5)$ for various ρ when $n = 1000$. We observe that individuals tend to be unique as ρ increases.

Next we rewrite (8) as

$$\mathrm{P}(\mathsf{S}_n = \mathsf{s}_n | N = n; \rho) = \exp((u-1)\log\rho + (1-n)\log(\rho+n))n! \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!},$$

which implies the following fact. See Lehmann (1991, p.46) for the completeness of a sufficient statistic of an exponential family.

Remark 2 *The LQM distribution* (8) *belongs to an exponential family, and* U_n *is complete and sufficient for* ρ.

It is thus important to elucidate the behavior of U_n. Also we should remember that an applied research often finds practical meaning in U_n, whose distribution is given below.

Theorem 7 *Suppose that size indices are distributed as* (8)*. Then U_n is shifted binomial distributed as*

$$\mathrm{P}(U_n = u|N = n; \rho) = \binom{n-1}{u-1}(\frac{\rho}{\rho+n})^{u-1}(\frac{n}{\rho+n})^{n-u}, \quad u = 1, 2, \ldots, n. \tag{16}$$

Theorem 8 *Suppose that U_n is subject to* (16)*. Then, as $n \to \infty$, U_n converges in distribution to $X + 1$ where X is Poisson distributed with mean ρ.*

Another law of small numbers holds in Theorem 8; U_n becomes shifted Poisson distributed. Observing these results, we realize that the LQM distribution behaves very simply.

4 Application

4.1 Parameter estimation

Since the distribution of the sufficient statistic U_n is simple, it may seem that there is not much to discuss. However, it is important to point out here that the increment of n does not necessarily improve the parameter estimation of the LQM distribution. To see this, we first obtain the Maximum Likelihood Estimator (MLE) of the LQM distribution. The Fisher information is also given in this section.

We denote the log likelihood of (8) by

$$L = (u-1)\log\rho + (1-n)\log(\rho+n) + \mathrm{Const.}$$

Then

$$\frac{dL}{d\rho} = (u-1)\frac{1}{\rho} + (1-n)\frac{1}{\rho+n}.$$

The MLE ($\hat{\rho}$) is the solution of $dL/d\rho = 0$:

$$\hat{\rho} = \frac{u-1}{1-u/n},$$

which equals the moment estimator based on (15).

Moreover

$$\frac{d^2L}{d\rho^2} = (u-1)\frac{-1}{\rho^2} + (1-n)\frac{-1}{(\rho+n)^2}.$$

Then, using (15), the Fisher information is

$$I(\rho) = \mathrm{E}\left(-\frac{d^2L}{d\rho^2}\right) = \frac{(n-1)n}{\rho(\rho+n)^2}.$$

$$\lim_{n\to\infty} I(\rho)^{-1} = \rho. \tag{17}$$

Hence due to the information inequality, the variance of an unbiased estimator of ρ remains to be strictly positive when $n \to \infty$. Consequently, any unbiased estimator does not converge in probability to ρ. There appears the following remark.

Remark 3 *Concerning the LQM distribution* (8), *no estimator of* ρ *is consistent as* $n \to \infty$.

Let us view this result in a different way. As $n \to \infty$, the MLE $\hat{\rho}$ is asymptotically equivalent to $U_n - 1$, whose variance converges to ρ according to Theorem 8. This limit equals the lower bound (17).

We have to repeat taking n samples to enjoy the reduction of an estimation error by the increase of observations.

4.2 An example of application

Very many zero counts (i.e. s_0) tend to be observed in practice. To describe this type of data set, a common model mixes a distribution with a point mass at frequency zero. Namely, when X is originally distributed over N_0, the adjusted variable Y is distributed as

$$\mathrm{P}(Y = 0) = \Delta + (1 - \Delta)\mathrm{P}(X = 0),$$

$$\mathrm{P}(Y = j) = (1 - \Delta)\mathrm{P}(X = j),$$

where $\Delta \leq 1$ and negative Δ deflates zeros. For example, Aitchison and Brown (1957) called this adjustment a Δ-distribution. See Johnson et al. (1993, p.312) for other literatures. The same idea is also called a Zero Inflated distribution; see Lambert (1992) for the Zero Inflated Poisson (ZIP) distribution. If X is subject to the Lagrangian Poisson distribution (3), Y should be called ZI Lagrangian Poisson (ZILP) distributed.

The LQM distribution adjusts the proportion of zero counts by tacitly assuming infinitely many zeros; the general method used to derive this distribution can be regarded as an alternative approach to many (or infrequent) zeros. Hence this section compares the fit of the LQM distribution with that of the ZILP distribution.

Leroux and Puterman (1992) recorded the number of movements by a fetal lamb in 240 consecutive 5-s intervals. These data can be briefly described with size indices; s_i expresses the number of intervals where i movements were observed. The number of movements n was 86, and the number of intervals u in which at least one movement was observed was 58. The total number of intervals J was 240. These size indices and the fits of the ZILP distribution and the LQM distribution are tabulated in Table 2.

Gupta et al. (1992, Table 1) fitted the ZILP distribution to the data set and gave the ML estimates as

$$\hat{\Delta} = -0.3143,\ \hat{\alpha} = 1.1254,\ \hat{\theta} = 0.2032,$$

where $\lambda = \alpha\theta$. The author calculated the fit for s_i by $J \times \mathrm{P}(Y = i)$, where Y is ZILP distributed under these estimates. The fits are slightly different from those given in Gupta et al. (1992, Table 2), but the author does not know the reason other than rounding errors. The ML estimate of the LQM distribution for the data set was $\hat{\rho} = 175$, under which the fit was the expectation of s_i given in (14).

Although the ZILP distribution has more parameters than the LQM distribution, the fits are similar in Table 2. It is so because $\hat{\theta}$ is close to zero; we took $\theta \to 0$ in the derivation of the LQM distribution. Hence, seemingly, the LQM distribution can be used for the approximation to the (ZI)LP distribution when θ nearly equals zero. The Borel distribution too appears from the (zero-truncated) LP distribution when $\theta \to 0$, and thus the LQM distribution should be rather compared with the Borel distribution.

i	s_i	ZILP	LQM
0	182	182.00	-
1	41	41.62	41.77
2	12	10.94	9.92
3	2	3.47	3.51
4	2	1.22	1.47
5	0	0.46	0.67
6	0	0.18	0.32
7+	1	0.12	0.33

Table 2: Fetal movements (Leroux and Puterman (1992))

A characteristic difference between these distributions lies in their upper tails. The support of the Borel distribution is unbounded, but under the LQM distribution, $s_i = 0$ if $i > n$. The Borel distribution has a very heavy tail, and sometimes the upper tail is truncated to obtain a better fit on the surface. This expedient of truncation seemingly lacks a reasonable basis. Conditioning is a less arbitrary treatment than the truncation of a tail.

In summary, the LQM distribution seems to be an advantageous substitute for the Borel distribution (or the (ZI)LP distribution at $\theta \doteq 0$) especially because (a) U_n is random as discussed in Section 3 and (b) it is free from an arbitrary truncation, despite of its tractable behavior.

Acknowledgements

This manuscript was written during the author's visit to Center for International Research on the Japanese Economy, Faculty of Economics, The University of Tokyo. The research was also supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology. The author would like to thank these organizations and Prof. Masaaki Sibuya and Prof. Akimichi Takemura for valuable suggestions.

Appendix

Proof of Theorem 1 By putting $\alpha = J/\rho$, the right hand side of (6) is rewritten as

$$\begin{aligned} &\frac{J!n!\rho^{n-1}}{J^n(\rho+n)^{n-1}} \prod_{i=0}^{n} \left(\frac{J^{i-1}(1/J+i/\rho)^{i-1}}{i!} \right)^{s_i} \frac{1}{s_i!} \\ &= \frac{J!n!\rho^{n-1}}{J^u(J-u)!(\rho+n)^{n-1}} \prod_{i=1}^{n} \left(\frac{(1/J+i/\rho)^{i-1}}{i!} \right)^{s_i} \frac{1}{s_i!}. \end{aligned}$$

Since $J!/((J-u)!J^u) \to 1$ and

$$\prod_{i=1}^{n} \left(\frac{(1/J+i/\rho)^{i-1}}{i!} \right)^{s_i} \to \rho^{u-n} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!} \right)^{s_i}$$

as $J \to \infty$, the probability function converges to (8). Q.E.D.

Proof of Theorem 2 When $J\theta = \mu$, we can rewrite the right hand side of (7) as

$$\frac{J!}{(J-u)!J^u}\frac{J^u\theta^J}{\theta^{J-u}}\exp(-\mu - n\lambda)\prod_{i=1}^{\infty}(\frac{(\theta+i\lambda)^{i-1}}{i!})^{s_i}\frac{1}{s_i!}$$
$$\to \quad \mu^u \exp(-\mu - n\lambda)\prod_{i=1}^{\infty}(\frac{(i\lambda)^{i-1}}{i!})^{s_i}\frac{1}{s_i!}$$

by taking $J \to \infty, \theta \to 0$. The last expression is tantamount to (10). Q.E.D.

Proof of Theorem 3 Under (7), N is subject to $LP(J\theta, \lambda)$. This distribution is unchanged by the limiting argument taken, since $J\theta$ is fixed at μ. Hence N is subject to $LP(\mu, \lambda)$ under (10). Then a simple division results in (8). Q.E.D.

Proof of Theorem 4 Following Sibuya (1993), we adopt the method of moments to show the convergence in distribution. This proof depends on the joint factorial moments (13), which will be shown later. Assuming that (13) is correct, the components' joint factorial moments is expressed for nonnegative integer r_i as

$$\mathrm{E}(\prod_{i=1}^{m} S_i{}^{(r_i)}) = \frac{n!\rho^r(\rho+n)^{1-n}}{(n-l)!(\rho+(n-l))^{1-n+l}}\prod_{i=1}^{m}\left(\frac{i^{i-1}}{i!}\right)^{r_i},$$

where $r = \sum r_i, l = \sum i r_i$. Hence

$$\lim_{n\to\infty}\mathrm{E}(\prod_{i=1}^{m} S_i{}^{(r_i)}) = \rho^r \exp(-l)\prod_{i=1}^{m}\left(\frac{i^{i-1}}{i!}\right)^{r_i} = \prod_{i=1}^{m}\left(\frac{\exp(-i)\rho i^{i-1}}{i!}\right)^{r_i}$$

The right hand side equals the joint factorial moments of Poisson variables with means (11). Because the convergence holds for all combinations of $(r_1, r_2, \ldots, r_m)$, the theorem holds. Q.E.D.

Proof of Theorem 5 To deal with the distribution of $U = \sum_{i=1}^{\infty} S_i$, first let us consider $U_m = \sum_{i=1}^{m} S_i$. Since S_i is independently Poisson distributed under (10), the pgf of U_m is

$$\prod_{i=1}^{m}\exp\left(\frac{\mu(\lambda i)^{i-1}}{i!}\exp(-\lambda i)(z-1)\right) = \exp\left(\sum_{i=1}^{m}\frac{\mu(\lambda i)^{i-1}}{i!}\exp(-\lambda i)(z-1)\right).$$

The logarithm of the right hand side is rewritten as

$$\sum_{i=1}^{m}\frac{\mu(\lambda i)^{i-1}}{i!}\exp(-\lambda i) = \mu\sum_{i=1}^{m}\mathrm{P}(X=i),$$

where X is subject to the Borel distribution (4). Therefore, since μ is assumed to be finite and $\sum_{i=1}^{m}\mathrm{P}(X=i)$ converges monotonically to unity as $m \to \infty$,

$$\lim_{m\to\infty}\mathrm{E}(z^{U_m}) = \exp(\mu(z-1)) = \mathrm{E}(z^U),$$

for an appropriate interval of z. The last equation implies that U is Poisson distributed with mean μ.

Thus the conditional distribution (12) is the result of dividing (10) by the probability function of the Poisson distribution with mean μ. The infinite dimensional multinomial distribution (12) is that of the size indices where $F_j, j = 1, 2, \ldots, u$, are independently and identically Borel distributed, by the same reason that S is multinomially distributed in (7). Q.E.D.

Proof of Theorem 6 The moments are calculated using the fact that $\sum_{\mathsf{s}_n} \mathrm{P}(\mathsf{S}_n = \mathsf{s}_n | N = n) = 1$ for all n.

$$\begin{aligned}
\mathrm{E}(\prod_{i=1}^{n} S_i^{(r_i)}) &= \sum_{\mathsf{s}_n \in \mathcal{S}_n} n! \rho^{u-1} (\rho+n)^{1-n} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{(s_i - r_i)!} \\
&= \sum_{\mathsf{s}_n \in \mathcal{S}_n} n! \rho^{r} (\rho+n)^{1-n} \frac{\rho^{u-r-1}(\rho+n-l)^{1-n+l}(n-l)!}{(\rho+n-l)^{1-n+l}(n-l)!} \\
&\quad \times \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i - r_i} \left(\frac{i^{i-1}}{i!}\right)^{r_i} \frac{1}{(s_i - r_i)!} \\
&= \frac{n! \rho^{r} (\rho+n)^{1-n}}{(n-l)!(\rho+(n-l))^{1-n+l}} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{r_i}.
\end{aligned}$$

Q.E.D.

Proof of Theorem 7 We would like to simplify

$$\mathrm{P}(U_n = u | N = n) = n!\, \rho^{u-1} (\rho+n)^{1-n} \sum_{\mathsf{s}_n \in \mathcal{S}_n} 1(\sum s_i = u) \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!}.$$

To evaluate

$$\sum_{\mathsf{s}_n \in \mathcal{S}_n} 1(\sum s_i = u) \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!}, \tag{18}$$

we use the fact that the sum of probability is unity:

$$1 = \sum_{\mathsf{s}_n \in \mathcal{S}_n} n!\, \rho^{u-1} (\rho+n)^{1-n} \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!},$$

which is equivalent to

$$(\rho+n)^{n-1} = n! \sum_{u=1}^{n} \rho^{u-1} \sum_{\mathsf{s}_n \in \mathcal{S}_n} 1(\sum s_i = u) \prod_{i=1}^{n} \left(\frac{i^{i-1}}{i!}\right)^{s_i} \frac{1}{s_i!}. \tag{19}$$

Using a binomial expansion, we also obtain

$$(\rho+n)^{n-1} = \sum_{l=0}^{n-1} \binom{n-1}{l} \rho^{l} n^{n-1-l}. \tag{20}$$

By comparing the coefficients of ρ between (19) and (20), we see that (18) has to equal

$$\binom{n-1}{u-1} n^{n-u}.$$

Consequently, (16) is proved. Q.E.D.

Remark 4 *Equation* (18) *is a (partial) Bell polynomial usually denoted by* $B_{n,u}(1^0, 2^1, 3^2, \ldots)$.

The combinatorial interpretation of Hoshino's (2005) construction will be investigated in the author's subsequent paper.

Proof of Theorem 8 We show the fact by the convergence of the pgf:

$$\lim_{n\to\infty} \mathrm{E}(z^{U_n}) = z\exp(\rho(z-1)). \tag{21}$$

The right hand side is the pgf of $X+1$.

The pgf of (16) is evaluated as

$$\begin{aligned}
\mathrm{E}(z^{U_n}) &= \sum_{u=1}^{n} (\rho+n)^{1-n} \rho^{u-1} z^u \binom{n-1}{u-1} n^{n-u} \\
&= z\left(\frac{\rho z + n}{\rho + n}\right)^{n-1} \\
&= z\left(1 + \frac{\rho(z-1)}{\rho+n}\right)^{n-1},
\end{aligned}$$

which converges to the right hand side of (21). Q.E.D.

References

[1] Aitchison, J. and Brown, J.A. (1957). The lognormal distribution. Cambridge University Press, Cambridge.

[2] Arratia, R., Barbour, A.D. and Tavarè, S. (2003). Logarithmic Combinatorial Structures: a Probabilistic Approach, European Mathematical Society Publishing House, Zürich.

[3] Bunge, J. and Fitzpatrick, M. (1993). Estimating the number of species: a review. Journal of the American Statistical Association, 88, 364–373.

[4] Consul, P.C. (1989). Generalized Poisson Distributions, Marcel Dekker, New York.

[5] Consul, P.C. and Jain, G.C. (1973). A generalization of the Poisson distribution. Technometrics, 15, 791–799.

[6] Consul, P.C. and Mittal, S.P. (1975). A new urn model with predetermined strategy, Biometrische Zeitschrift, 17, 67-75.

[7] Ewens, W.J. (1972). The sampling theory of selectively neutral alleles. Theoretical Population Biology, 3, 87–112.

[8] Good, I.J. (1953). The population frequencies of species and the estimation of population parameters. Biometrika, 40, 237–264.

[9] Gupta, P.L., Gupta, R.C. and Tripathi, R.C. (1996). Analysis of zero-adjusted count data. Computational Statistics and Data Analysis, 23, 207–218.

[10] Holla, M.S. (1966). On a Poisson-inverse Gaussian distribution. Metrika, 11, 115–121.

[11] Hoshino, N. (2002). On limiting random partition structure derived from the conditional inverse Gaussian-Poisson distribution. Technical Report CMU-CALD-02-100, School of Computer Science, Carnegie Mellon University.

[12] Hoshino, N. (2003). Random clustering based on the conditional inverse Gaussian-Poisson distribution. Journal of the Japan Statistical Society, 33, 105–117.

[13] Hoshino, N. (2004). Modeling strategy for the risk assessment of privacy. Proceedings of the 16th RAMP symposium, 133–148.

[14] Hoshino, N. (2005). Engen's extended negative binomial model revisited. Annals of the Institute of Statistical Mathematics, 57, 369–387.

[15] Hoshino, N. and Takemura, A. (1998). Relationship between logarithmic series model and other superpopulation models useful for microdata disclosure risk assessment. Journal of the Japan Statistical Society, 28, 2, 125–134.

[16] Janardan, K.G. (1975). Markov-Polya urn models with predetermined strategies - I. Gujarat Statistical Review, 2, 17–32.

[17] Johnson, N.L., Kotz, S. and Kemp, A.W. (1993). Univariate Discrete Distributions, 2nd ed., Wiley, New York.

[18] Kingman, J.F. (1978). Random partitions in population genetics. Proceedings of the Royal Society of London, A, 361, 1–20.

[19] Lambert, D. (1992). Zero inflated Poisson regression with an application to defects in manufacturing. Technometrics, 34, 1–14.

[20] Lehmann, E.L. (1991). Theory of point estimation. Wadsworth, California.

[21] Leroux, B.G. and Puterman, M.L. (1992). Maximum-penalized likelihood estimation for independent and Markov-dependent mixture models. Biometrics, 48, 545–558.

[22] Pitman, J. (2003). Poisson-Kingman partitions. Science and Statistics: A Festschrift for Terry Speed, Lecture Notes-Monograph Series 30, 1–34, Institute of Mathematical Statistics, Hayward, California. (available from http://arXiv.org/abs/math/0210396)

[23] Sibuya, M. (1993). A random clustering process. Annals of the Institute of Statistical Mathematics, 45, 459–465.

[24] Steutel, F.W. and van Harn, K. (2004). *Infinite Divisibility of Probability Distributions on the Real Line*. Marcel Dekker, New York.