Sociedad de Estadística e Investigación Operativa
Test (1999) *Vol.* **8**, *No.* 2, pp. 399–409

Discrete uniform mixtures via posterior means

Arjun K. Gupta*
Department of Mathematics and Statistics
Bowling Green State University, USA.
Jacek Wesołowski
Mathematical Institute
Warsaw University of Technology, Poland.

Abstract

In this paper, we are concerned with identification of a discrete uniform mixture by the posterior mean. An exact formula for a prior distribution is given. Also some examples featuring negative binomial, negative hypergeometric and beta-Pascal distributions are provided.

Key Words: Beta-Pascal distribution, discrete uniform distribution, identification of mixtures, mixture, negative binomial distribution, negative hypergeometric distribution, posterior mean.
AMS subject classification: primary 62H05, secondary 62F15.

1 Introduction

A basic question for a mixture model is its identifiability. Usually it is connected with one to one correspondence between the distributions of Y and X if the conditional distribution (mixture) $\mu_{X|Y}$ is given —see for example Titterington et al. (1985) and references given there. In this context discrete mixtures (we say that a mixture is discrete iff X and Y are discrete) were treated, for instance, in Patil and Bildikar (1966). This paper is devoted to the study of a special discrete mixture.

Another identifiability problem is connected with one to one correspondence between the regression function $m(x) = E(Y|X = x)$ (posterior mean) and the distribution of Y. In this scheme some discrete mixtures were considered in the literature:

- binomial and negative hypergeometric in Krishnaji (1974);
- binomial and Pascal in Korwar (1975);

*Correspondence to: Arjun K. Gupta, Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, USA. E-mail address: gupta@bgnet.bgsu.edu

Received: March 1996; Accepted: December 1998

- quasi-binomial in Korwar (1977);
- negative hypergeometric in Xekalaki (1981, 1983);
- binomial, Pascal and Poisson in Cacoullos and Papageorgiou (1983);
- hypergeometric and negative hypergeometric in Papageorgiou (1985);
- binomial and Pascal in Kyriakoussis and Papageorgiou (1991);
- power series in Sapatinas (1995) and Wesołowski (1995b);
- Poisson in Wesołowski (1996);
- negative binomial in Papageorgiou and Wesołowski (1997).

Some interesting general remarks on identifiability of discrete mixtures in this setting are given in Arnold et al. (1993). Other kinds of mixtures, which are not discrete, are also considered in such a context by many authors —for a review and some new results see, for example, Wesołowski (1995a).

Recall that a random variable X has a negative hypergeometric distribution $nhg(L, M, n)$ iff

$$P(X = k) = \frac{\binom{L+k-1}{k}\binom{M-L-k}{n-k}}{\binom{M}{n}}, \tag{1.1}$$

$k = 0, 1 \ldots, n$, where n is a non-negative integer and $0 < L < n + L \leq M$ —see Ord (1972, ch. 5). Problems of identifiability of the negative hypergeometric mixture

$$\mu_{X|Y} = nhg(L, N + Y - 1, Y)$$

for some natural numbers $1 \leq L < N$, by posterior means were studied in Papageorgiou (1985). The investigation was based on additional strong assumptions of infinite integrability of X and identifiability of a prior distribution by sequence of moments. The uniqueness result obtained there does not allow to reconstruct the prior distribution unless one is equipped with some a priori information or your guess is right. In Xekalaki (1981, 1983) characterization problems for negative hypergeometric mixtures involving linear regressions were considered.

In this paper we are interested in discrete uniform mixtures and the discrete uniform distribution (on a set $\{0, \ldots, n\}$) is a special case of (1.1) —$L = 1$ and $M = n + 1$. However we do not assume any additional integrability or identifiability conditions. Moreover we obtain an exact formula expressing the prior distribution in terms of any consistent posterior mean. Let us point out that the end point of its support is also determined by a property of the posterior mean. Also some necessary consistency conditions are obtained. This is done in Section 2. In Section 3 we show how the formula works by presenting examples of linear (4 examples) and non-linear (1 example) posterior means. Applying the formula derived in Section 2 we obtain exact prior distributions. In this way some members of the families of negative binomial, negative hypergeometric and beta-Pascal distributions have been identified.

This paper is a complement to an earlier investigation by the authors, Gupta and Wesolowski (1997), where the continuous uniform mixture $\mu_{X|Y}$ was treated. In that paper it was shown that linearity of $E(Y|X)$ characterizes priors as two kinds of beta or gamma distribution. Also a general identifiability result in the absolutely continuous case was derived there. An extension to beta mixtures has been obtained recently in Gupta and Wesołowski (1998).

2 Identification

Let $X = V_1$ and $Y = V_1 + V_2$, where V_1, V_2 are i.i.d. random variables with geometric distribution, i.e. $P(V_1 = k) = p(1-p)^k$, $k = 0, 1, \ldots$, where $p \in (0, 1)$. Then it is not difficult to observe that X is a discrete uniform mixture of Y of the form

$$\mu_{X|Y} = \mathcal{U}(\{0, \ldots, Y\}), \tag{2.1}$$

i.e.

$$P(X = k|Y = l) = \frac{1}{l+1} \quad 0 \le k \le l = 0, 1, \ldots$$

Observe that

$$E(Y|X) = X + \frac{1-p}{p}.$$

Observe that if the representation $X = V_1$ and $Y = V_1 + V_2$, where V_1, V_2 are some i.i.d. random variables, holds then the single condition (2.1) suffices to characterize the distribution of V_i, $i = 1, 2$, as geometric. Obvious calculations lead to the Cauchy functional equation for the pmf of V_1 then. Investigations of problems of such a kind go back to Patil and Seshadri (1964).

Here we drop off the underlying independence structure of (X, Y) and we are interested in the converse problem to the observation from the beginning of this section, i.e. we want to identify the discrete uniform mixture (2.1) through the posterior mean.

Theorem 2.1. *Assume that* (X, Y) *is a random vector such that* (2.1) *holds and* $\mathrm{supp}(Y) = \{0, 1, \ldots, N\}$, *where* $N \leq \infty$. *Then the joint distribution is uniquely determined by the posterior mean* $E(Y|X)$:

$$N = \inf\{j : m(j) = j,\ j = 0, 1, \ldots\},$$

where $m(j) = E(Y|X = j)$, $j = 0, 1, \ldots$ *and for any* $k \in \{0, 1, \ldots, N\}$,

$$P(Y = k) = \frac{(k+1)M(k)\prod_{j=0}^{k} d(j)}{\sum_{i=0}^{N}(i+1)M(i)\prod_{l=0}^{i} d(l)}, \tag{2.2}$$

where $M(k) = m(k+1) - m(k)$, $k < N$, $M(N) = m(N-1) - N + 1$ *if* $N < \infty$, *and*

$$d(j) = \begin{cases} 1 & j = 0, N, \\ \frac{m(j-1)-j+1}{m(j+1)-j} & j = 1, 2, \ldots, N-1. \end{cases}$$

Proof. The Bayes formula implies

$$m(k)\sum_{l=k}^{N} \frac{p_l}{l+1} = \sum_{l=k}^{N} \frac{l p_l}{l+1}, \tag{2.3}$$

where $p_l = P(Y = l)$, $l = 0, , 1 \ldots$ Let k_0 be the smallest natural number such that $m(k_0) = k_0$. The above equation implies immediately that $k_0 = N$ (if $\forall k,\ m(k) \neq k$, then $N = \infty$). Consider now the case $N = \infty$. All formulas given in the first part of the proof hold for any $k = 1, 2, \ldots,$ and the upper limit in all sums is ∞. Rewrite (2.3) as

$$[m(k) + 1]\sum_{l=k} \frac{p_l}{l+1} = \sum_{l=k} p_l.$$

Put $k - 1$ instead of k in the above equation and subtract one from the other to get

$$[m(k-1) - m(k)]\sum_{l=k} \frac{p_l}{l+1} + [m(k-1) + 1]\frac{p_{k-1}}{k} = p_{k-1}. \tag{2.4}$$

From the inequality

$$\frac{\sum_{l=k} \frac{lp_l}{l+1}}{\sum_{l=k} \frac{p_l}{l+1}} < \frac{\sum_{l=k+1} \frac{lp_l}{l+1}}{\sum_{l=k+1} \frac{p_l}{l+1}},$$

it follows that m is strictly increasing. Consequently (2.4) yields

$$\sum_{l=k} \frac{p_l}{l+1} = \frac{k-1-m(k-1)}{k[m(k-1)-m(k)]} p_{k-1}. \tag{2.5}$$

Next put $k+1$ instead of k in (2.5) and subtract one from the other to get

$$\frac{p_k}{k+1} = \frac{k-1-m(k-1)}{k[m(k-1)-m(k)]} p_{k-1} - \frac{k-m(k)}{(k+1)[m(k)-m(k+1)]} p_k.$$

Finally we obtain (recall that (2.3) implies $m(k) > k$)

$$p_k = \frac{k+1}{k} \frac{m(k+1)-m(k)}{m(k)-m(k-1)} \frac{m(k-1)-k+1}{m(k+1)-k} p_{k-1}. \tag{2.6}$$

The formula (2.2) follows from the above recurrence relation if we take into account the normalizing condition.

Now consider $N < \infty$. Similarly as above for $k = 1, \ldots, N-1$, the formula (2.6) holds. Observe that by (2.5)

$$p_N = (N+1) \frac{N-1-m(N-1)}{N[m(N-1)-m(N)]} p_{N-1}.$$

Consequently (2.6) yields

$$p_N = (N+1) \frac{m(N-1)-N+1}{m(1)-m(0)} \prod_{j=0}^{N-1} d(j) p_0$$

and once again p_0 is obtained by normalizing. □

Remark 2.1. Summing up the observations from the proof, if for the mixture (2.1) $\text{supp}(Y) = \{0, 1, \ldots, N\}$ then any consistent regression function m must be strictly increasing, $m(k) > k$, $k = 0, \ldots, N-1$ and $m(N) = N$.

Remark 2.2. If $\text{supp}(Y) \subset \{0, 1, \ldots, N\}$ for the mixture (2.1) and $m(k) = N$, $\forall k \in \text{supp}(Y)$, then by (2.4) $P(Y = N) = 1$.

3 Examples

This Section is an implementation of the theoretical result obtained in Theorem 2.1. We show that the formula (2.2) is computationally useful by introducing few examples. It appears that linear posterior mean characterizes some important distributions: negative binomial, negative hypergeometric and beta-Pascal (since the last one is not so widely known one can consult Ord (1972, ch. 5), for its definition and basic properties —see also Ex. 4 beneath). A negative binomial law appears also in a case of non-linear regression in the last example.

Example 3.1. Assume that (2.1) holds and

$$E(Y|X) = X + b. \tag{3.1}$$

Hence

$$\inf\{j : m(j) = j\} = \infty.$$

Then $M(j) = 1$, $d(j) = \frac{b}{b+1}$ for $j = 1, 2, \ldots$, and

$$\prod_{l=0}^{i} d(l) = \left(\frac{b}{b+1}\right)^i, \quad i = 0, 1, \ldots$$

Consequently

$$\sum_{i=0}^{\infty}(i+1)M(i)\prod_{l=1}^{i} d(l) = \sum_{i=0}^{\infty}(i+1)\left(\frac{b}{b+1}\right)^i = (b+1)^2.$$

Finally (2.2) implies

$$g(k) = P(Y = k) = (k+1)\frac{b^k}{(b+1)^{k+2}} \quad k = 0, 1, \ldots,$$

i.e. Y has a negative binomial distribution $nb(n = 2, p = 1/(1+b))$. Observe that if V_1, V_2 are geometric with parameter $1/(1+b)$ then $Y \stackrel{d}{=} V_1 + V_2$.

Example 3.2. For the mixture (2.1) with $E(Y|X) = (X + A)/2$, where A is a positive natural number, we have

$$\inf\{j : m(j) = j\} = A.$$

Hence $N = A$, $M(j) = 1/2$, and $d(j) = 1$, $j = 0, 1, \ldots, N$. Thus

$$\sum_{i=0}^{N}(i+1)M(i)\prod_{l=0}^{i}d(l) = \frac{(N+1)(N+2)}{4}$$

and by (2.2) Y has a negative hypergeometric distribution $nhg(2, N+2, N)$ (see (1.1)) and the joint distribution of (X, Y) is bivariate discrete uniform on the lattice triangle $\{(i,j) : 0 \le i \le j \le N\}$.

Example 3.3. More generally assume that for the mixture (2.1)

$$E(Y|X) = \frac{r-1}{r}X + \frac{A}{r}, \tag{3.2}$$

for some integers $r \ge 1$ and $A > 0$. Then $N = r$ and $M(j) = (r-1)/r$, $j = 0, 1, \ldots, N-1$, and $M(N) = 1/r$. Additionally $d(j) = (N+1-j)/(N+r-j)$, $j = 1, \ldots, N-1$, yield

$$\prod_{l=0}^{i}d(l) = \frac{\binom{N+r-2-i}{N-i}}{\binom{N+r-2}{N}}, \quad i = 1, 2, \ldots, N-1.$$

Hence

$$\sum_{i=0}^{N}(i+1)M(i)\prod_{l=0}^{i}d(l) = \frac{(N+r)(N+r-1)}{r^2}.$$

Consequently by (2.2) Y has a negative hypergeometric distribution $nhg(2, N+r, N)$ (see (1.1)).

Example 3.4. Take now the mixture (2.1) with

$$E(Y|X) = \frac{r+1}{r}X + \frac{L}{r}, \tag{3.3}$$

for some positive integers r and L. Then $N = \infty$ since $m(k) > k$, $\forall k = 0, 1, \ldots$ Additionally it is easily seen that $M(0) = M(j) = (r+1)/r$,

$$d(j) = \frac{L+j-1}{L+r+j+1}, \quad j = 1, 2, \ldots,$$

and

$$\prod_{l=0}^{i}d(l) = \frac{\binom{L+r+1}{L-1}}{\binom{L+r+i+1}{r+2}}, \quad i = 1, 2, \ldots$$

Consequently

$$\sum_{i=0}^{\infty}(i+1)M(i)\prod_{l=0}^{i}d(l)=\frac{(L+r+1)(L+r)}{r^2}.$$

Finally (2.1) implies

$$P(Y=k)=\frac{r}{r+2}\,\frac{(k+1)\binom{L+r-1}{r}}{\binom{L+r+k+1}{r+2}},\quad k=0,1,\ldots.$$

Recall, following Ord (1972, pp. 86), that the formula

$$p_i=\frac{A\binom{K+i-1}{i}\binom{A+B-1}{A}}{(A+K)\binom{K+A+B+i-1}{K+A}},\quad i=0,1,\ldots,$$

where A, B, K are some positive integers, defines the beta-Pascal $bP(A,B,K)$ distribution. Consequently in our example Y has the beta-Pascal $bP(r,L,2)$ distribution.

Remark 3.1. The family of distributions we obtained in Examples 3.1, 3.3 and 3.4 has an interesting continuity property: Consider the mixture (2.1) with $E(X)=b$; then $E(Y)=2b$. If (3.2) holds then $N=(r+1)b$. If (3.3) holds then $L=(r-1)b$. Consequently if the coefficient of X in (3.2) or (3.3) tends to 1 (which is the coefficient of X in (3.1)), i.e. $r\to\infty$, then also the second coefficient of (3.2) or (3.3) tends to the respective coefficient in (3.1). Additionally it is easily seen that

$$\lim_{r\to\infty}h_r=g=\lim_{r\to\infty}f_r,$$

where f_r, g and h_r, are pmf's of the prior distributions (beta-Pascal, negative binomial, negative hypergeometric) obtained in Examples 3.4, 3.1 and 3.3, respectively. (The negative binomial distribution is a limit of the negative hypergeometric and beta-Pascal —see Ord, 1972, pp. 89). A similar behaviour was observed in the case of linearity of posterior mean for the first and second type beta and gamma distribution for linear posterior mean in the continuous uniform mixture model in Gupta and Wesołowski (1997).

Example 3.5. Consider now an example of the uniform mixture (2.1) with non-linear posterior mean

$$E(Y|X) = X + \frac{X+5}{X+3}.$$

In this case it is easily seen that $N = \infty$. Additionally

$$M(k) = \frac{(k+2)(k+5)}{(k+3)(k+4)}, \quad k = 0, 1 \dots,$$

$$d(j) = \frac{(j+4)^2}{2(j+2)(j+5)}, \quad j = 1, 2, \dots$$

and consequently

$$\prod_{l=1}^{i} d(l) = \frac{5(i+3)(i+4)}{12(i+5)2^i}, \quad i = 1, 2, \dots$$

Hence

$$\sum_{i=0}^{\infty}(i+1)M(i)\prod_{l=1}^{i} d(l) = \frac{5}{12}\sum_{i=0}^{\infty}(i+1)(i+2)2^{-i} = \frac{20}{3}.$$

Finally applying (2.1) we arrive at

$$P(Y = k) = \frac{(k+1)(k+2)}{16}\left(\frac{1}{2}\right)^k, \quad k = 0, 1, \dots,$$

which is a negative binomial distribution $nb(n = 3, p = 1/2)$.

References

Arnold, B.C., E. Castillo and J.M. Sarabia (1993). Conditionally specified models: structure and inference. In *Multivariate Analysis: Future Directions*, 2 (C.M. Cuadras and C.R. Rao eds.) Elsevier, Amsterdam, 441-450.

Cacoullos, T. and H. Papageorgiou (1983). Characterizations of discrete distributions by a conditional distribution and a regression function. *Annals of the Institute of Statistical Mathematics*, **35**, 95-103.

Gupta, A.K. and J. Wesołowski (1997). Uniform mixtures via posterior means. *Annals of the Institute of Statistical Mathematics*, **49**, 171-180.

Gupta, A.K. and J. Wesołowski (1998). Regressional identifiability and identification for beta mixtures. Technical Report No. 98-13. Department of Mathematics and Statistics, Bowling Green State University, 1-10.

Korwar, R.M. (1975). On characterizing some discrete distributions by linear regression. *Communications in Statistics*, **4**, 1133-1147.

Korwar, R.M. (1977). On characterizing Lagrangian-Poisson and quasi-binomial distributions. *Communications in Statistics*, **6**, 1409-1416.

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

Kyriakoussis, A. and H. Papageorgiou (1991). Characterizations of logarithmic series distributions. *Statistica Neerlandica*, **45**, 1-8.

Ord, J.K. (1972). *Families of Frequency Distributions*. Hafner Publishing Company, New York.

Papageorgiou, H. (1985). On characterizing some discrete distributions by a conditional distribution and a regression function. *Biometrical Journal*, **27**, 473-479.

Papageorgiou, H. and J. Wesołowski (1997). Posterior mean identifies the prior distribution in NB and related models. *Statistics and Probability Letters*, **36**, 127-134.

Patil, G.P. and S. Bildikar (1966). Identifiability of countable mixtures of discrete probability distributions using methods of infinite matrices. *Proceedings of the Cambridge Philosophical Society*, **62**, 485-494.

Patil, G.P. and V. Seshadri (1964). Characterization theorems for some univariate probability distributions. *Journal of the Royal Statistical Society, B*, **26**, 286-292.

Sapatinas, T. (1995). Identifiability of mixtures of power series distributions and related characterizations. *Annals of the Institute of Statistical Mathematics*, **47**, 447-459.

Titterington, D.M., A.F.M. Smith, and U.E. Makov (1985). *Statistical Analysis of Finite Mixture Distributions*. Wiley, New York.

Wesołowski, J. (1995a). Bivariate distributions via a Pareto conditional distribution and a regression function. *Annals of the Institute of Statistical Mathematics*, **47**, 177-183.

Wesołowski, J. (1995b). Bivariate discrete measures via a power series conditional distribution and a regression function. *Journal of Multivariate Analysis*, **55**, 219-229.

Wesołowski, J. (1996). A new conditional specification of the bivariate Poisson conditionals distribution. *Statistica Neerlandica*, **50**, 390-393.

Xekalaki, E. (1981). A characterization of the negative hypergeometric distribution based on conditional expectation. In *Proceedings of the Sixth Conference on Probability Theory* (B. Bereanu, S. Grigorescu, M. Iosifescu, and T. Postelnicu eds.) Editura Academici Republicii Socialiste Romania, Bucuresti, 379-384.

Xekalaki, E. (1983). Infinite divisibility, completeness and regression properties of the univariate generalized Waring distribution. *Annals of the Institute of Statistical Mathematics*, **35**, 279-289.