Ann. Inst. Statist. Math.
Vol. 49, No. 1, 171–180 (1997)

UNIFORM MIXTURES VIA POSTERIOR MEANS

ARJUN K. GUPTA[1] AND JACEK WESOLOWSKI[2]

[1] *Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403-0221, U.S.A.*
[2] *Mathematical Institute, Warsaw University of Technology, Plac Politechniki 1, 00-661 Warsaw, Poland*

(Received June 5, 1995; revised February 19, 1996)

Abstract. The problem of identification of uniform mixtures via posterior means is studied. For linear posterior means a complete solution is given. It determines a family of prior distributions involving beta of both kinds and gamma. Identifiability via any consistent posterior mean is also investigated.

Key words and phrases: Mixtures, posterior mean, identification, characterization of probability distribution, uniform conditional distribution, conditional moments, linearity of regression, gamma distribution, beta type distributions.

1. Introduction

Let (X, Y) be a random vector and let $\mu_{X|Y}$ be the conditional distribution of X given Y. In such a setting X (or its distribution μ_X) is called a mixture with respect to Y (or its distribution μ_Y).

An important aspect of theory of mixtures is a question of identifiability of the model. Originally it was considered as the problem of one to one correspondence between μ_X and μ_Y. Foundations of such an approach are given in Teicher (1961), Barndorff-Nielsen (1965) and Patil and Bildikar (1966), where many specific examples are also provided.

Another identifiability question arises when, instead of marginal distribution, the posterior mean $E(Y \mid X)$ is taken into account. Beginning in mid-seventies many papers were devoted to this subject, see for example: Krishnaji (1974), Korwar (1975), Cacoullos and Papageorgiou (1983, 1984), Papageorgiou (1984*a*, 1984*b*, 1985), Arnold *et al.* (1993), Wesolowski (1995*a*, 1995*b*). These papers are concerned with two problems:

- *Identification of a mixture*: What is the joint distribution of (X, Y) for some given posterior mean?
- *Identifiability of a mixture*: Is the joint distribution of (X, Y) uniquely determined by any (consistent) posterior mean?

If the answer to the second question is positive then in many cases the first one may be settled by educated guessing. However, in general, these are separate problems. Sometimes to solve the second one some additional information is necessary (as it is in the case we consider here; see also Papageorgiou (1984*b*, 1985), where infinite integrability and moment identifiable distributions were additionally assumed).

In this paper we are interested in uniform mixtures. In Section 2 we are concerned with the identification problem involving any linear posterior mean. The family of probability distributions arising as the solution of the problem consists of beta of both kinds and gamma with parameters depending on the coefficients of linear form defining the regression function. The family has an intriguing continuity property allowing to approximate gamma density by beta type densities. Also a special form of the second posterior moment is studied. Section 3 is devoted to the identifiability question under an additional assumption that priors are absolutely continuous. Let us emphasize that we do not make use of identifiability of a mixture via the posterior distribution which was often applied in earlier papers—see for example Cacoullos and Papageorgiou (1984). (Recall that Teicher (1961) pointed out that identification of some uniform mixtures by posterior distribution is not possible.) Instead we are interested directly in deriving the distribution of the prior. Obviously the joint distribution is identified if the prior is known. In the closing Section 4 a uniform mixtures with the homoscedasticity of the posterior distribution is considered.

2. Identification

Let $X = V_1$ and $Y = V_1 + V_2$, where V_1, V_2 are i.i.d. exponential r.v.'s with the mean b. Then it is not difficult to observe that X is a uniform mixture of Y of the form

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

i.e. the conditional density $f_{X|Y}$ of X given Y has the form

$$f_{X|Y=y}(x) = \begin{cases} y^{-1} & 0 < x < y, \\ 0 & \text{otherwise.} \end{cases}$$

The first two posterior moments have the forms

$$\begin{aligned} E(Y \mid X) &= X + b, \\ E(Y^2 \mid X) &= X^2 + 2bX + 2b^2. \end{aligned} \tag{2.2}$$

Here we are interested in the converse of the above observations without assuming the summation scheme.

At first let us consider the question of identification of the uniform mixture (2.1) via a general linear form of the posterior mean

$$E(Y \mid X) = aX + b, \tag{2.3}$$

where a and b are some real constants. When $\mu_{X|Y}$ is known it suffices to know the prior distribution to determine the bivariate measure.

Such problems, involving linearity of regression, were discussed for other mixtures: binomial and Pascal in Korwar (1975), quasi-binomial in Korwar (1977), normal in Cacoullos and Papageorgiou (1984), Pareto in Wesolowski (1995a) and second kind beta in Wesolowski (1994).

From (2.1) observe that $E(X) = 0$ iff $E(Y) = 0$ iff $P(X = Y = 0) = 1$. Hence to get rid off that degenerate case we will assume that $E(X)$ is non-zero.

THEOREM 2.1. *Let (X, Y) be a random vector satisfying (2.1) and (2.3) for some constants a and b. Assume additionally that $0 < E(X) < \infty$. Then $b > 0$ and only the following cases are possible*:

(i) $1 < a < 2$ *and Y has a second kind beta distribution with the density*

$$f_a(y) = ab^{-2}y\left[1 + \frac{(a-1)}{b}y\right]^{(2a-1)/(1-a)} I(y > 0); \tag{2.4}$$

(ii) $a = 1$ *and Y has a gamma distribution with the density*

$$g(y) = b^{-2}ye^{-y/b}I(y > 0); \tag{2.5}$$

(iii) $0 < a < 1$ *and Y has a first kind beta distribution with the density*

$$h_a(y) = f_a(y)I\left(y < \frac{b}{1-a}\right); \tag{2.6}$$

(iv) $a = 0$ *and* $P(Y = b) = 1$.

PROOF. Linearity of regression (2.3) via integrability assumption implies

$$E(Ye^{-tX}) = E(E(Y \mid X)e^{-tX}) = b\phi_X(t) - a\phi'_X(t) \tag{2.7}$$

for any $t \geq 0$, where ϕ_X denotes the Laplace transform of X. On the other hand observe that the r.v. $(1 - e^{-tX})/X$ is integrable for any $t \geq 0$ since it is bounded by t. Consequently (2.1) yields

$$tE(Ye^{-tX}) = 1 - \phi_Y(t), \qquad t \geq 0, \tag{2.8}$$

where ϕ_Y is the Laplace transform of Y. Again from (2.1) we have

$$t\phi_X(t) = E\left(\frac{1 - e^{-tY}}{Y}\right), \qquad t \geq 0. \tag{2.9}$$

Denote the function on the right hand side of (2.9) by H. We want to compute derivative of ϕ_X. To this end we must know that differentiation on the right hand side of (2.9) is possible. Consider for any small $|h|$ the expression

$$\left|E\left(\frac{(1 - e^{-(t+h)Y})/Y}{h}\right) - E\left(\frac{(1 - e^{-tY})/Y}{h}\right) - \phi_Y(t)\right|$$
$$= \left|E\left(e^{-tY}\frac{-e^{-hY} + 1 - hY}{hY}\right)\right| \leq \frac{|h|}{2}E(Y).$$

For $a \in (0,1)$ the solution of (2.11) has the form

$$H(t) = \frac{\int_0^t u^{a/(1-a)} e^{b/(1-a)u} du}{t^{a/(1-a)} e^{b/(1-a)t}(1-a)}, \qquad t \geq 0.$$

Consequently

$$\phi_Y(t) = \frac{1}{1-a}\left[1 - \frac{a+bt}{(1-a)t} \frac{\int_0^t u^{a/(1-a)} e^{b/(1-a)u} du}{t^{a/(1-a)} e^{b/(1-a)t}}\right], \qquad t \geq 0,$$

which is the Laplace transform of the first kind beta density given in (2.6).

Now take $a = 0$. Then (2.11) has a simple form

$$H'(t) + bH(t) = 1, \qquad t \geq 0.$$

Hence $H(t) = b^{-1} + ce^{-bt}$, which yields $\phi_Y(t) = e^{-bt}$, $t \geq 0$. Thus $Y = b$ a.s.

Finally observe that for $a < 0$ the equation (2.11) has no solution valid in the whole interval $(0, \infty)$ since in this case the function $G(u) = u^{a/(1-a)} e^{b/(1-a)u}$, $u > 0$, is not integrable in the intervals $(0, \epsilon_1)$ and (ϵ_2, ∞) for any positive reals ϵ_1, ϵ_2. □

Remark 1. From the viewpoint of Bayesian estimation theory the following immediate corollary of Theorem 2.1 may be of some interest: *The Bayes estimator $\hat{m}$ of $E(Y)$ based on an observation of the uniform mixture* (2.1) *under the quadratic loss function has the form $\hat{m}(X) = X + b$ iff the prior distribution is gamma with the scale parameter b and the shape parameter* 2.

Remark 2. Observe that the joint distribution of (X, Y) determined by (ii) of Theorem 2.1 (also in Theorem 2.2 beneath) is a special case of multivariate gamma law discussed thoroughly in Mathai and Moschopoulos (1992).

Remark 3. The part (ii) of Theorem 2.1 may be restated as follows: *If X/Y, Y are independent, $X/Y \stackrel{d}{=} U$, where U has the uniform distribution on* $(0,1)$, *and $E(Y - X \mid X) = b$ then X and $Y - X$ are i.i.d. exponential r.v.'s.* In this context we would like to recall Kotz and Steutel (1988) characterization of the exponential law, in which also independence and uniform distribution are involved: *If X, $Y - X$ are i.i.d. r.v.'s and $X \stackrel{d}{=} UY$, where U is as above and additionally independent of Y, then X is exponential.*

Remark 4. The family of densities we arrive at in Theorem 2.1 has an interesting continuity property. It is easily seen that for any positive x

$$\lim_{a \to 1^-} h_a(x) = g(x) = \lim_{a \to 1^+} f_a(x).$$

It means that the gamma distribution with density (2.5) may be approximated by first or second kind beta. An interesting question is if such approximation involving beta type distributions are also possible for gamma distributions with the shape parameter different from 2 (including exponential distribution). It seems that the scale parameter of the gamma law is not important in such problems.

Remark 5. Let us point out that for $a = b = 1/2$ the joint distribution of (X, Y), obtained from Theorem 2.1, is uniform in the triangle $\{(x, y) : 0 < x < y < 1\}$.

Remark 6. The distributions of X's in respective cases of Theorem 2.1 can be immediately derived:

(i) X has a Pareto type distribution with the density:

$$f_X(x) = \frac{1}{b}\left(1 + \frac{a-1}{b}x\right)^{a/(1-a)} I(x > 0);$$

(ii) X has an exponential distribution with the density:

$$f_X(x) = \frac{1}{b}e^{-x/b}I(x > 0);$$

(iii) X has a first kind beta distribution with the density:

$$f_X(x) = \frac{1}{b}\left(1 - \frac{1-a}{b}x\right)^{a/(1-a)} I\left(0 < x < \frac{b}{1-a}\right);$$

(iv) X has a uniform distribution with the density:

$$f_X(x) = \frac{1}{b}I(0 < x < b).$$

Now let us turn our attention to the second posterior moment. Here we want only to indicate that such a characteristic may also determine the uniform mixture. Therefore we consider only one special form of the second posterior moment.

THEOREM 2.2. *Let (X, Y) be a random vector satisfying* (2.1) *and* (2.2) *for some positive constant b. Assume additionally that X is square integrable. Then $(X, Y) \stackrel{d}{=} (V_1, V_1 + V_2)$, where V_1, V_2 are i.i.d. exponential r.v.'s with the mean b.*

PROOF. Without loss of generality we may take $b = 1$. Consider (2.2). Applying similar argument as in the proof of Theorem 2.1 we easily get the identity

$$\text{(2.13)} \qquad \frac{mt^2}{2} = (1 + t + t^2)E\left(\frac{1 - e^{-tY}}{Y}\right) - t(1 + t)\phi_Y(t), \qquad t \geq 0,$$

where $m = E(Y)$. Then rewrite (2.13) as

$$H'(t)t(1 + t) - H(t)(1 + t + t^2) = -\frac{mt^2}{2}, \qquad t \geq 0$$

(the function H is defined in the proof of Theorem 2.1). This is a first order ordinary differential equation and its solution is of the form

$$H(t) = \frac{mt}{2(1+t)} + c\frac{t}{1+t}e^t, \qquad t \geq 0,$$

where c is a constant. Consequently

$$\phi_Y(t) = \frac{1}{(1+t)^2}\left(\frac{m}{2} + ce^t(1+t+t^2)\right), \qquad t \geq 0.$$

Since ϕ_Y is the Laplace transform, we must have $c = 0$ and $m = 2$. Consequently Y is a gamma r.v. with the mean 2. Now the final result follows by (2.1). □

If instead of (2.2) it is allowed that the second posterior moment has a general quadratic form, i.e.

$$E(Y^2 \mid X) = aX^2 + bX + c,$$

where a, b, c are some constants then under (2.1), similarly as in the proof of Theorem 2.2, it follows that

$$mt^2 = (a-1)t^2H''(t) - (2a+bt)tH'(t) + (2a+bt+ct^2)H(t), \qquad t \geq 0.$$

However its solution seems to be far more complicated than the earlier one (under $a = 1$ it becomes a first order equation) and we do not follow this path here. Instead more general approach is proposed in Remark 11, Section 3.

3. Identifiability

Here we are concerned with the problem of unique determination of the uniform mixture (2.1) by any consistent posterior mean. Since in such an approach Laplace transform does not work well, to be able to work with densities we assume additionally that the prior is absolutely continuous.

The main result of this section gives

THEOREM 3.1. *Let (X, Y) be a random vector satisfying* (2.1). *Assume additionally that Y is absolutely continuous with the support* $[0, T]$, $(0 < T \leq \infty)$. *Then T, the prior distribution and, consequently, the mixture are uniquely determined by the regression function $m(x) = E(Y \mid X = x)$, $x \in (0, T)$.*

PROOF. From the Bayes formula

$$m(x)\int_x^T y^{-1}f(y)dy = \int_x^T f(y)dy, \qquad x \in (0,T), \tag{3.1}$$

where f denotes the density of Y. Hence it is easily seen that

$$T = \inf\{x : m(x) = x, x > 0\},$$

$m(x) > x$, $x > 0$, and m is differentiable. Upon taking derivatives of both sides of (3.1) we get

$$xm'(x)\int_x^T y^{-1}f(y)dy = (m(x)-x)f(x), \qquad x \in (0,T). \tag{3.2}$$

Now divide (3.2) by $xm'(x)$ (observe that (3.1) implies that m is strictly increasing on $(0,T)$) and then differentiate both sides to obtain

$$\frac{m'(x)}{x-m(x)} = \frac{\left[\frac{m(x)-x}{xm'(x)}f(x)\right]'}{\frac{m(x)-x}{xm'(x)}f(x)}, \qquad x \in (0,T).$$

Hence the density of the prior is given by the formula

$$f(x) = c\frac{xm'(x)}{x-m(x)}\exp\left(\int\frac{dx}{x-m(x)}\right), \qquad x \subset (0,T), \tag{3.3}$$

where c is a norming constant. □

Remark 7. Observe that we have got not only the uniqueness result but also for given m we can calculate the density of Y by formula (3.3). For example taking $m(x) = x+1$ (i.e. $m'(x) = 1$) we obtain $T = \infty$ and by (3.3) the prior has the gamma density (2.5) with $b = 1$.

Remark 8. As another example consider the regression function of the form

$$m(x) = \frac{2(x^2+x+1)}{3(x+1)}, \qquad x \in (0,T).$$

Hence $T = 1$ and simple calculations lead to

$$m'(x) = \frac{2x(x+2)}{3(x+1)^2}, \qquad m(x) - x = -\frac{(x+2)(x-1)}{3(x+1)}, \qquad x \in (0,1).$$

Consequently the formula (3.3) implies $f(x) = 3x^2 I(0 < x < 1)$, which means that Y has a power function distribution.

Remark 9. Observe that in the course of the proof of Theorem 3.1 we obtained the following necessary consistency conditions for the regression function m:

- m has to be strictly increasing in $(0,T)$;
- $m(x) > x$, $x \in (0,T)$;
- m has to be differentiable;
- $m(T) = T$.

Remark 10. If a linear posterior mean $m(x) = ax + b$, $x \in (0, T)$, is considered in the setting of Theorem 3.1 then $a \geq 2$ is also possible. (It is not allowed in Theorem 2.1, where integrability assumption is imposed.) In this case Y has the density (2.4), but its mean does not exist (hence X is not integrable).

Remark 11. Similar argument as in the proof given above may be applied if instead of the posterior mean we are interested in other posterior characteristics of the form $E(G(Y) \mid X)$: Assume that the function $n(x) = E(G(Y) \mid X = x)$, $x \in (0, T)$, is known. Additionally let $G(Y)$ be integrable. Then mimicking the proof of Theorem 3.1 we obtain the following formula for the density of Y

$$f(x) = c\frac{x|n'(x)|}{|n(x) - G(x)|} \exp\left(-\int \frac{n'(x)}{G(x) - n(x)} dx\right), \qquad x \in (0, T).$$

4. Homoscedasticity

We end our considerations with another approach to the initial example of Section 2: $X = V_1$, $Y = V_1 + V_2$, where V_1 and V_2 are i.i.d. exponential with the mean b. Instead of considering conditional moments we observe that the posterior distribution is homoscedastic, i.e.

$$\text{Var}(Y \mid X) = b^2. \tag{4.1}$$

As a converse we obtain an analogue of Theorem 2.2, but this time we need additional smoothness condition, which enables us to use methods developed throughout Section 3. On the other hand integrability condition can be omitted.

THEOREM 4.1. *Let (X, Y) be a random vector satisfying* (2.1) *and* (4.1) *for some positive constant b. Assume additionally that Y is absolutely continuous with the support $[0, T]$ $(0 < T \leq \infty)$. Then $(X, Y) \stackrel{d}{=} (V_1, V_1 + V_2)$, where V_1, V_2 are i.i.d. exponential r.v.'s with the mean b.*

PROOF. Similarly as in the proof of Theorem 3.1 we easily get

$$\int_x^T y^2 g(y)dy \int_x^T g(y)dy - \left(\int_x^T yg(y)dy\right)^2 = b^2\left(\int_x^T g(y)dy\right)^2, \tag{4.2}$$

for any $T > x > 0$, where $g(y) = y^{-1}f(y)$, $T > y > 0$, and f is the density of Y. Then upon taking derivatives in (4.2), then canceling g at both sides and differentiating the resulting equation once again, one gets

$$b^2 g'' = g.$$

Hence $g(x) = K \exp(-x/b) I(0 < x < T)$ which, upon inserting it in (4.2), yields $T = \infty$. ⊔

Acknowledgements

We are indebted to one of the referees whose simulating question is answered in Theorem 4.1. Other referees' remarks helped to improve the presentation of the material. We appreciate also suggestions made by Barry C. Arnold, which lead to Remark 10.

References

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

Barndorff-Nielsen, O. (1965). Identifiability of mixtures of exponential families, *J. Math. Anal. Appl.*, **12**, 115–121.

Cacoullos, T. and Papageorgiou, H. (1983). Characterizations of discrete distributions by a conditional distribution and a regression function, *Ann. Inst. Statist. Math.*, **35**, 95–103.

Cacoullos, T. and Papageorgiou, H. (1984). Characterizations of mixtures of continuous distributions by their posterior means, *Scand. Actuar. J.*, **8**, 23–30.

Korwar, R. M. (1975). On characterizing some discrete distributions by linear regression, *Comm. Statist.*, **4**, 1133–1147.

Korwar, R. M. (1977). On characterizing Lagrangian-Poisson and quasi-binomial distributions, *Comm. Statist.*, **6**, 1409–1416.

Kotz, S. and Steutel, F. W. (1988). Note on a characterization of exponential distributions, *Statist. Probab. Lett.*, **6**, 201–203.

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

Mathai, M. A. and Moschopoulos, P. G. (1992). A form of multivariate gamma distribution, *Ann. Inst. Statist. Math.*, **44**, 97–106.

Papageorgiou, H. (1984*a*). Characterizations of multinomial and negative multinomial mixtures by regression, *Austral. J. Statist.*, **26**, 25–29.

Papageorgiou, H. (1984*b*). Characterizations of continuous binomial and negative binomial mixtures, *Biometrical J.*, **26**, 795–798.

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

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

Teicher, H. (1961). Identifiability of mixtures, *Ann. Math. Statist.*, **32**, 244–248.

Wesolowski, J. (1994). *Bivariate distributions via a second kind beta conditional distribution* and a regression function, *Journal of Mathematical Sciences* (to appear).

Wesolowski, J. (1995*a*). Bivariate distributions via a Pareto conditional distribution and a regression function, *Ann. Inst. Statist. Math.*, **47**, 177–183.

Wesolowski, J. (1995*b*). Bivariate discrete measures via power series conditional distribution and a regression function, *J. Multivariate Anal.*, **55**, 219–229.