Bayesian estimation and prediction for certain type of mixtures

Aziz LMoudden[a,b] & Éric Marchand[a], [1]
a Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1
b Département de mathématiques, Université du Québec à Montréal, Montréal, Qc, CANADA, H2X 3Y7 e-mails: eric.marchand@usherbrooke.ca; aziz.lmoudden@usherbrooke.ca

Summary

For two vast families of mixture distributions and a given prior, we provide unified representations of posterior and predictive distributions. Model applications presented include bivariate mixtures of Gamma distributions labelled as Kibble-type, non-central Chi-square and F distributions, the distribution of R^2 in multiple regression, variance mixture of normal distributions, and mixtures of location-scale exponential distributions including the multivariate Lomax distribution. An emphasis is also placed on analytical representations and the relationships with a host of existing distributions and several hypergeometric functions of one or two variables.

AMS 2010 subject classifications: 62F15; 62E15; 62C10

Keywords and phrases: Bayesian analysis; Bivariate gamma; Kibble; Lomax; Mixtures, Non-central distributions; Posterior distribution; Predictive distribution.

1. Introduction

Mixture models are ubiquitous in probability and statistics. Such models, whether they are finite mixture models, mixtures of Poisson, exponential, gamma or normal distributions, etc., are quite useful and appealing for best representing data and heterogeneous environments. As well, distributional properties of mixture models are often quite elegant and instructive. However, analytical and numerical challenges are present and well documented, namely in terms of likelihood-based and Bayesian inference.

It is also the case that several familiar distributions are representable in terms of mixtures, and that such representation facilitates the derivation of various statistical properties and approaches to statistical inference. Prominent examples include the noncentral chi-square, Beta, and Fisher distributions, which arise typically in relationship to quadratic forms in normal linear models. Other important examples include the distribution of the square of a multiple correlation coefficient (R^2) in a standard multiple regression linear model with normally distributed errors, as well as the vast class of univariate or bivariate Gamma mixtures which include the Kibble distribution (see Example 1.1).

We consider mixture models for summary statistics $X, Y \in \mathbb{R}^d$ which are of one of the following two types:

$$\text{Type I}: \; X|K \sim f_K \text{ with } K|\theta \sim g_\theta, \quad Y|J \sim q_J \text{ with } J|\theta \sim h_\theta\,; \tag{1.1}$$

$$\text{Type II}: \; X|K,\theta \sim f_{\theta,K} \text{ with } K \sim g, \quad Y|J,\theta \sim q_{\theta,J} \text{ with } J \sim h\,; \tag{1.2}$$

[1]September 7, 2020

The classification of these types will be adhered to and seems to be a natural way to present the various expressions and examples that make up this paper. In the above models, the mixing variables K, J will typically be either a discrete or continuous univariate distribution, but more generally $K, J \in \mathcal{K}$ with $\mathcal{K}$ a measurable space. The parameter $\theta \in \mathbb{R}^p$ is unknown and a prior density will be assumed. Otherwise the densities $f_k, f_{\theta,k}, q_j,$ and $q_{J,\theta}$ will be assumed known (except for the value of θ) and absolutely continuous with respect to a $\sigma-$finite measure ν. Similarly, the densities g and g_θ for K, as well as h and h_θ for J, will be assumed to be known and absolutely continuous with respect to a $\sigma-$finite measure μ. Examples will include both discrete and continuous mixing for K and J, as well as discrete and continuous models for the conditional distributions of X and Y.

With analytical expressions (Theorem 3.1 and Theorem 4.2) for Bayesian posterior distributions for θ based on X, as well as Bayesian predictive densities for Y based on X, we are particularly interested in eliciting the general structures driving these Bayesian solutions and illustrating the wide breadth of applications. In the illustrations, the solutions involve mixtures themselves and bring into play special functions such as Hypergeometric and Appell which are readily available in mathematical packages. They are also "exact' and do not rely on stochastic approximations (e.g., MCMC) which are omnipresent in Bayesian computing.

Much focus is placed on identifying concise representations and informative connections between distributions. Model applications that we present include bivariate mixtures of Gamma distributions that we label as Kibble-type, non-central Chi-square and F distributions, the distribution of R^2 in multiple regression, variance mixture of normal distributions, and mixtures of location-scale exponential distributions including the multivariate Lomax distribution. For bivariate gamma mixtures, which we focus on in Section 4, we also consider bivariate prior distributions with dependence structures, in particular as occurring under an order restriction on the parameters. The posterior and predictive distribution decompositions in this work follow familiar paths, but the analytical representations provided are nevertheless unified, useful and insightful, and lead to simplifications in Bayesian posterior analyses. This is also the case where we purposely exploit the mixture representation of a familiar distribution.

Here is a first illustration of Type I and Type II mixtures, which will be studied more generally in Sections 3 and 4 respectively, and which borrows notations presented in Section 2.

Example 1.1. *The Kibble bivariate gamma distribution (Wicksell, 1933; Kibble, 1941) for $X = (X_1, X_2)$, admits the following mixture representation:*

$$X_i|K \overset{i.d}{=} G(\nu + K, \frac{1-\rho}{\lambda_i}), i = 1, 2, \text{ with } K \overset{d}{=} NB(\nu, 1-\rho)\,. \tag{1.3}$$

Here $\nu, \lambda_1, \lambda_2 > 0$, and $\rho \in [0, 1)$. Such a distribution originates for instance in describing the joint distribution of sample variances generated from a bivariate normal distribution with correlation coefficient ρ. The cases $\rho = 0$ reduce to $X_i \overset{i.d}{=} G(\nu, \frac{1}{\lambda_i}), i = 1, 2$. Now, observe that we have a Type I mixture if $\theta = \nu$ with known $\lambda_1, \lambda_2, \rho$, and a Type II mixture for $\theta = (\lambda_1, \lambda_2)$; or again $\theta = (\frac{1-\rho}{\lambda_1}, \frac{1-\rho}{\lambda_2})$; with known ν, ρ. A third type of mixture arises when ρ is unknown, but will not be further addressed in this paper. The mixing representation of the Kibble distribution has been exploited in a critical fashion for Bayesian analysis by Iliopoulos et al. (2005).

2. Notations and definitions

Here are some notations and definitions used throughout concerning some special functions and various distributions that appear below, related either to the model, the mixing variables, the prior, the posterior, or the predictive distribution. We will write $X \sim f$ to indicate that the random variable X has density (or p.m.f.) f. We will also write $f \sim D$ to signify that density f is associated with distribution D. We will the use the symbol $\stackrel{d}{=}$ to represent equality in distribution and $\stackrel{i.d}{=}$ for equality in distribution with independent components.

In what follows, we denote $\mathrm{P}(\theta)$ as the Poisson distribution with mean θ and p.m.f. (or density) $\frac{e^{-\theta}\theta^n}{n!}\,\mathbb{I}_{\mathbb{N}}(n)$. We write $N \stackrel{d}{=} NB(r,\xi)$, with $r > 0$ and $\xi \in (0,1)$, to denote a negative binomial distribution with p.m.f. (or density) $\frac{(r)_n}{n!}\,\xi^r(1-\xi)^n\,\mathbb{I}_{\mathbb{N}}(n)$.

Throughout the paper, we define for positive real numbers $a_1,\ldots,a_p,b_1,\ldots,b_q$, and $z \in \mathbb{R}$, the generalized hypergeometric function as

$$ {}_pF_q(a_1,...,a_p;b_1,...,b_q;z) = \sum_{k=0}^{\infty} \frac{\prod_{i=1}^{p}(a_i)_k}{\prod_{j=1}^{q}(b_j)_k}\frac{z^k}{k!}\,, $$

with $(\alpha)_m = \frac{\Gamma(\alpha+m)}{\Gamma(\alpha)}$ the Pochhammer function defined here for $\alpha > 0, m \in \mathbb{N}$. We write $N \stackrel{d}{=} \mathrm{Hyp}(a_1,\ldots a_p;b_1,\ldots,b_q;\lambda)$, for $a_i > 0, b_j > 0$, to denote a generalized hypergeometric distribution (e.g., Johnson et al., 1995) with p.m.f.

$$ \mathrm{P}(N=n) \;=\; \frac{1}{{}_pF_q(a_1,...,a_p;b_1,...,b_q;\lambda)}\,\frac{\prod_{i=1}^{p}(a_i)_n}{\prod_{j=1}^{q}(b_j)_k}\frac{\lambda^n}{n!}\;\mathbb{I}_{\mathbb{N}}(n)\,. $$

For these distributions, we will have positive a_i's and b_j's, and λ will be non negative and in the radius of convergence of the ${}_pF_q$ function. With such a notation for instance, we could write $X \stackrel{d}{=} \mathrm{Hyp}(-;-;\theta)$ for $X \stackrel{d}{=} \mathrm{P}(\theta)$, and $X \stackrel{d}{=} \mathrm{Hyp}(r;-;1-\xi)$ for $X \stackrel{d}{=} \mathrm{NB}(r,\xi)$.

We denote $\mathrm{G}(a,b)$, $\mathrm{B}(a,b)$, and $\mathrm{B2}(a,b,\sigma)$, for $a,b,\sigma > 0$, as Gamma, Beta, and Beta type II distributions, and with densities $\frac{(bt)^a e^{-bt}}{t\Gamma(a)}\,I_{(0,\infty)}(t)$, $\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,t^{a-1}\,(1-t)^{b-1}\,I_{(0,1)}(t)$, and $\frac{\Gamma(a+b)\,\sigma^b}{\Gamma(a)\,\Gamma(b)}\,\frac{t^{a-1}}{(\sigma+t)^{a+b}}\,I_{(0,\infty)}(t)$, respectively. The latter Beta type II family includes Pareto distributions on $\mathbb{R}_+$ for $a = 1$.

The Kummer distribution of type II, denoted $K2(a,b,c,\sigma)$ for parameters $a,c,\sigma > 0$ and $b \in \mathbb{R}$, is taken with density

$$ \frac{\sigma^b}{\Gamma(a)\,\psi(a,1-b,c)}\,\frac{t^{a-1}}{(t+\sigma)^{a+b}}\,e^{-ct/\sigma}\,I_{(0,\infty)}(t)\,, $$

where ψ is the confluent hypergeometric function of type II defined for $\gamma_1,\gamma_3 > 0$ and $\gamma_2 \in \mathbb{R}$ as: $\psi(\gamma_1,\gamma_2,\gamma_3) = \frac{1}{\Gamma(\gamma_1)}\int_{\mathbb{R}_+} t^{\gamma_1-1}\,(1+t)^{\gamma_2-\gamma_1-1}\,e^{-\gamma_3 t}\,dt$. This class of distributions include Gamma distributions with choices $b = -a$. The class can also be extended to include cases $c = 0, b > 0$ which correspond to $\mathrm{B2}(a,b,\sigma)$ distributions.

We will denote McKay's bivariate gamma distribution with parameters $\gamma_1,\gamma_2,\gamma_3(>0)$ as $Z = (Z_1,Z_2) \stackrel{d}{=} \mathrm{M}(\gamma_1,\gamma_2,\gamma_3)$ with p.d.f.

$$ \frac{{\gamma_3}^{\gamma_1+\gamma_2}}{\Gamma(\gamma_1)\,\Gamma(\gamma_2)}\,z_1^{\gamma_1-1}\,(z_2-z_1)^{\gamma_2-1}\,e^{-\gamma_3 z_2}\,\mathbb{I}_{(0,\infty)}(z_1)\,\mathbb{I}_{(z_1,\infty)}(z_2)\,. $$

The distribution has a long history (e.g., McKay, 1934) and is a benchmark bivariate distribution to model durations that are ordered, It is easy to verify that the marginals are distributed as

$Z_1 \stackrel{d}{=} \mathrm{G}(\gamma_1, \gamma_3)$ and $Z_2 \stackrel{d}{=} \mathrm{G}(\gamma_1 + \gamma_2, \gamma_3)$, and that Z_1 and $Z_2 - Z_1$ are independently distributed, with $Z_2 - Z_1 \stackrel{d}{=} \mathrm{G}(\gamma_2, \gamma_3)$. A generalization will be presented in Section 4.

3. Type I mixtures

We begin with Type I mixtures, providing general representations for the posterior distribution $\theta|x$, as well as the predictive distribution $Y|x$, and following up with various examples and observations.

Theorem 3.1. *Let $X, Y|\theta$ be conditionally independent distributed as in (1.1) and let π have prior density for θ with respect to $\sigma-$finite measure τ. Let π_k and f'_x be the densities given by*

$$\pi_k(u) \;=\; \frac{\pi(u)\, g_u(k)}{m_\pi(k)} \;\; \textit{and} \;\; f'_x(k) \;=\; \frac{f_k(x)\, m_\pi(k)}{\int_{\mathcal{K}} f_k(x)\, m_\pi(k)\, d\mu(k)}\,,\; u \in \Theta, x \in \mathbb{R}^d,$$

with m_π the density given by

$$m_\pi(k) = \int_\Theta g_\theta(k)\, \pi(\theta)\, d\tau(\theta)\,,\; k \in \mathcal{K}.$$

Then,

(a) *The posterior distribution of $U \stackrel{d}{=} \theta|x$ admits the mixture representation:*

$$U|K' \sim \pi_{K'}\,,\; K' \sim f'_x\,; \tag{3.4}$$

(b) *The Bayes predictive density of Y admits the representation:*

$$Y|J' \sim q_{J'}\,, J' \sim h'_x, \tag{3.5}$$

with

$$h'_x(j') \;=\; \int_{\mathcal{K}} q_\pi(j'|k')\, f'_x(k')\, d\mu(k')\,, \;\; \textit{and}\; q_\pi(j'|\, k') = \int_\Theta h_\theta(j')\, \pi_{k'}(\theta)\, d\tau(\theta)\,. \tag{3.6}$$

Proof. (a) We have indeed

$$\begin{aligned} \pi(\theta|x) \;&\propto\; \int_{\mathcal{K}} f_k(x)\, g_\theta(k)\, \pi(\theta)\, d\mu(k) \\ &\propto\; \int_{\mathcal{K}} f_k(x)\, m_\pi(k)\, \pi_k(\theta)\, d\mu(k) \propto \int_{\mathcal{K}} f'_k(x)\, \pi_k(\theta)\, d\mu(k). \end{aligned}$$

(b) The predictive density of Y, i.e. the conditional density of $Y|x$, is given by:

$$\begin{aligned} q_\pi(y|x) \;&=\; \int_\Theta q(y|\theta)\, \pi(\theta|x)\, d\tau(\theta) \\ &=\; \int_\Theta \left(\int_{\mathcal{K}} q_{j'}(y)\, h_\theta(j')\, d\mu(j') \right) \left(\int_{\mathcal{K}} \pi_{k'}(\theta)\, f'_x(k')\, d\mu(k') \right) d\tau(\theta) \\ &=\; \int_{\mathcal{K}} q_{j'}(y) \left(\int_{\mathcal{K}} f'_x(k') \int_\Theta h_\theta(j')\, \pi_{k'}(\theta)\, d\tau(\theta)\; d\mu(k') \right) d\mu(j')\,, \end{aligned}$$

where we have used (3.4). This establishes the result. □

Remark 3.1. *The posterior and predictive distribution representations of Theorem 3.1 are particularly appealing. Indeed, observe that posterior distribution (3.4) mixes the $\pi'_{k'}s$ which correspond to the posterior density of $\theta|X = k'$ as if one had actually observed $K = k'$. Moreover, the mixing density f'_x for K' is a weighted version of the marginal density m_π for K with weight proportional to $f_k(x)$.*

The predictive distribution for Y mixes the same densities $q_{J'}$ as the model density for Y, with the prior mixing density h_θ replaced by the posterior mixing density h'_x. Furthermore, this mixing density is itself a mixture of the predictive (or conditional) densities $q_\pi(\cdot|k')$ of J as if one had observed $K = k'$.

The following examples concern posterior and predictive distribution illustrations of Theorem 3.1.

Example 3.2. *In model (1.1), consider Poisson mixing $K|\theta \stackrel{d}{=} P(\theta)$ with a $G(a, b)$ prior for θ. From this familiar set-up, we obtain $\pi_{k'}$ as a $G(a + k', 1 + b)$ density and m_π as a $NB(a, b/(1 + b))$ p.m.f. Following (3.4), the posterior distribution $\pi(\theta|x)$ is a mixture of the above π_k's with mixing density on $\mathbb{N}$ given by*

$$f'_x(k) \propto \frac{(a)_k}{k!} \frac{b^a}{(1+b)^{a+k}} f_k(x) .$$

Now, consider the cases: **(i)** $f_k \stackrel{d}{=} G(p/2 + k, 1/2)$, **(ii)** $f_k \stackrel{d}{=} B(p/2 + k, q/2)$ *and* **(iii)** $f_k \stackrel{d}{=} B2(p/2 + k, q/2)$. *In the context of model (1.1), case* **(i)** *corresponds to a non-central chi-square distribution with p degrees of freedom and non-centrality parameter 2θ ($\chi^2_p(2\theta)$), case* **(ii)** *to a non-central Beta distribution with shape parameters $p/2$, $q/2$, and non-centrality parameter 2θ, and case* **(iii)** *to the density of $\frac{p}{q}F$, with F distributed as a non-central F distribution with degrees of freedom p, q and non-centrality parameter 2θ. The latter two cases are essentially equivalent though and related by the fact that $\frac{X}{1-X}|\theta$ in* **(ii)** *is distributed as $X|\theta$ is in* **(iii)**.

For **(i)**, *we obtain*

$$f'_x \sim Hyp\left(a; \frac{p}{2}; \frac{x}{2(1+b)}\right) . \tag{3.7}$$

Observe that the above generalized hypergeometric distribution reduces to a $P(\frac{x}{2(1+b)})$ distribution for $a = p/2$. The posterior expectation can be evaluated with the help its mixture representation and standard calculations involving the above p.m.f. One obtains

$$\begin{aligned} \mathbb{E}(\theta|x) &= \mathbb{E}^{K'|x}\{\mathbb{E}(\theta|K', x)\} \\ &= \mathbb{E}_{f'_x}(\frac{a+K'}{1+b}) \qquad (3.8) \\ &= \frac{a}{1+b}\left(1 + \frac{x}{p(1+b)} \frac{{}_1F_1(a+1; p/2+1; \frac{x}{2(1+b)})}{{}_1F_1(a; p/2; \frac{x}{2(1+b)})}\right) , \qquad (3.9) \end{aligned}$$

with the case $a = p/2$ simplifying to $\mathbb{E}(\theta|x) = \frac{p/2}{1+b} + \frac{x/2}{(1+b)^2}$ as noted by Saxena and Alam (1982).

For the two other cases, we obtain the mixing power series densities:

$$f'_x \sim Hyp\left(a, \frac{p+q}{2}; \frac{p}{2}; \frac{x}{1+b}\right) \text{ for } (\mathbf{ii}), \text{ and } f'_x \sim Hyp\left(a, \frac{p+q}{2}; \frac{p}{2}; \frac{x}{(1+x)(1+b)}\right) \text{ for } (\mathbf{iii}). \tag{3.10}$$

Observe that the case $a = p/2$ reduces to a $NB(\frac{p+q}{2}, 1 - \frac{x}{1+b})$ distribution for **(ii)** *and a $NB(\frac{p+q}{2}, 1 - \frac{x}{(1+x)(1+b)})$ distribution for* **(iii)**. *The posterior expectation may be computed from (3.8) with simplifications occurring for $a = p/2$, yielding for instance in* **(ii)** *: $\mathbb{E}(\theta|x) = \frac{p(1+b)+qx}{2(1+b-x)}$. Finally, we point out that the above posterior distribution representation applies as well for the improper prior choice $\pi(\theta) = \theta^{a-1}\,\mathbb{I}_{(0,\infty)}(\theta)$ (i.e., $b = 0$), with m_π not a p.m.f., but given by $m_\pi(k) \propto \frac{(a)_k}{k!}$.*

Example 3.3. *Turning now to predictive densities with the same set-up as in Example 3.2, we consider $Y|\theta$ distributed identically to $X|\theta$ (i.e, $q_j = f_j$ and $g_\theta = h_\theta$). For $Y|\theta \stackrel{d}{=} \chi^2_p(2\theta)$ (i.e., case* **(i)***), Theorem 3.1 tells us that the Bayes predictive density of Y admits the mixture representation: $Y|J' \stackrel{d}{=} G(p/2+J', 2)$ with $J' \sim h'_x$ given in (3.6). The latter admits itself the mixture representation*

$$J'|K' \stackrel{d}{=} NB(a + K', \frac{1+b}{2+b}),\ K' \sim f'_x \text{ as in (3.7)}, \tag{3.11}$$

with the former being the Bayes predictive density of $J' \stackrel{d}{=} P(\theta)$ based on $K' \stackrel{d}{=} P(\theta)$ and prior $\theta \stackrel{d}{=} G(a, b)$. Alternatively represented, the mixing p.m.f. for J' may be expressed as

$$\begin{aligned} h'_x(j') &= \sum_{k'} \frac{(a+k')_{j'}}{j'!} (\frac{1+b}{2+b})^{a+k'} (\frac{1}{2+b})^{j'} \frac{(a)_{k'}}{k'!\,(\frac{p}{2})_{k'}} \frac{\left(\frac{x}{2(1+b)}\right)^{k'}}{{}_1F_1(a; p/2; \frac{x}{2(1+b)})} \\ &= \sum_{k'} \frac{(a+j')_{k'}}{j'!} (\frac{1+b}{2+b})^{a} (\frac{1}{2+b})^{j'} \frac{(a)_{j'}}{k'!\,(\frac{p}{2})_{k'}} \frac{\left(\frac{x}{2(2+b)}\right)^{k'}}{{}_1F_1(a; p/2; \frac{x}{2(1+b)})} \\ &= \frac{(a)_{j'}}{j'!} (\frac{1+b}{2+b})^{a} (\frac{1}{2+b})^{j'} \frac{{}_1F_1(a+j'; p/2; \frac{x}{2(2+b)})}{{}_1F_1(a; p/2; \frac{x}{2(1+b)})}, \end{aligned}$$

which also can be viewed directly as a weighted $NB(a, \frac{1+b}{2+b})$ p.m.f.

The non-central Beta and Fisher distributions are similar. For instance, in the former case with $X|\theta$ and $Y|\theta$ identically distributed with $Y|J \stackrel{d}{=} B(p/2 + J, q/2)$ and $J \stackrel{d}{=} P(\theta)$, predictive densities associated with $G(a, b)$ priors are also distributed as mixtures

$$Y|J' \stackrel{d}{=} B(p/2 + J', q/2),\ \text{with } J'|K' \stackrel{d}{=} NB(a + K', \frac{1+b}{2+b}),\ K' \sim f'_x \text{ as in (3.10)}.$$

For the mixing p.m.f. of J', a development as above yields the expression:

$$h'_x(j') = \frac{(a)_{j'}}{j'!} (\frac{1+b}{2+b})^{a} (\frac{1}{2+b})^{j'} \frac{{}_2F_1(a+j', (p+q)/2; p/2; \frac{x}{2+b})}{{}_2F_1(a, (p+q)/2; p/2; \frac{x}{1+b})}.$$

Example 3.4. *A doubly non-central F distribution is a type I mixture (e.g., Bulgren, 1971) with $X \in \mathbb{R}_+$ and $K = (K_1, K_2) \in \mathbb{N}^2$ admitting the representation*

$$X|K \stackrel{d}{=} B2\left(\frac{\nu_1}{2} + K_1, \frac{\nu_2}{2} + K_2, \frac{\nu_2}{\nu_1}\right) \text{ with } K_1, K_2|\theta \stackrel{i.d}{=} P(\frac{\theta_1}{2}) \text{ and } P(\frac{\theta_2}{2}). \tag{3.12}$$

Such a distribution arises naturally as a multiple of the ratio of two independent noncentral chi-squared distributions, and reduces to a non-central F for $\theta_2 = 0$. Consider now the application of Theorem 3.1 for the prior $\theta_i \stackrel{i.d}{=} G(a_i, b_i), i = 1, 2$, with $a_1, a_2, b_1, b_2 > 0$, where we have the familiar

representations of $\pi_k(\theta_1,\theta_2)$ as the joint density of independent $G(a_i+k_i,1+b_i)\,,i=1,2,$ distributions, and of $m_\pi(k_1,k_2)$ as the joint p.m.f. of independent $NB(a_i,\frac{b_i}{1+b_i})\,,i=1,2,$ distributions. Representation (3.4) tells us that the posterior distribution of (θ_1,θ_2) is a mixture of the π_K's with mixing density

$$\begin{aligned} f'_x(k) &\propto f_k(x)\, m_\pi(k) \\ &\propto \frac{(\frac{\nu_1+\nu_2}{2})_{k_1+k_2}\,(a_1)_{k_1}\,(a_2)_{k_2}}{(\frac{\nu_1}{2})_{k_1}\,(\frac{\nu_2}{2})_{k_2}\,k_1!\,k_2!}\;(\beta_1)^{k_1}\;(\beta_2)^{k_2}\;I_{\mathbb{N}^2}(k_1,k_2) \\ \Rightarrow f'_x(k) &= \frac{(\frac{\nu_1+\nu_2}{2})_{k_1+k_2}\,(a_1)_{k_1}\,(a_2)_{k_2}}{F_2(\frac{\nu_1+\nu_2}{2};a_1,a_2;\frac{\nu_1}{2},\frac{\nu_2}{2};\beta_1,\beta_2)}\;\frac{1}{(\frac{\nu_1}{2})_{k_1}\,(\frac{\nu_2}{2})_{k_2}\,k_1!\,k_2!}\;(\beta_1)^{k_1}\,(\beta_2)^{k_2}\;I_{\mathbb{N}^2}(k), \end{aligned} \tag{3.13}$$

with $\beta_1=\frac{x}{(\frac{\nu_2}{\nu_1}+x)\,(1+b_1)}$, $\beta_2=\frac{\nu_2/\nu_1}{(\frac{\nu_2}{\nu_1}+x)\,(1+b_2)}$, and where F_2 is the Appell function of the second kind given by

$$F_2(\gamma_1;\gamma_2,\gamma_3;\gamma_4,\gamma_5;w,z)\;=\;\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\gamma_1)_{m+n}\,(\gamma_2)_m\,(\gamma_3)_n}{m!\,n!\,(\gamma_4)_m\,(\gamma_5)_n}\,w^m\,z^n\,.$$

The bivariate p.m.f. f'_x in (3.13) and how it is arisen here are of interest. It is a bivariate power series p.m.f. generated by the coefficients of the Appell F_2 function and will be bona fide p.m.f. for $\beta_1,\beta_2>0$ and $\beta_1+\beta_2<1$. Appell's F_1 function appears in a similar way, again in a Bayesian framework, as a bivariate discrete distribution called Bailey by Laurent (2012) (see also Jones & Marchand, 2019 for another derivation). For the particular case $a_i=\nu_i/2, i=1,2,$ the p.m.f. in (3.13) simplifies and the corresponding random pair (K'_1,K'_2), admits the stochastic representation: $K'_1\sim NB(a_1+a_2,1-p_1)$ and $K'_2|K'_1\sim NB(a_1+a_2+K'_1,p_2)$ with $p_1=\beta_1/(1-\beta_2)$ and $p_2=1-\beta_2$.

Turning to the predictive density, consider $Y|\theta$ distributed as $X|\theta$ and the same prior on θ as above. Similarly to Example 3.3, which shares the same Poisson mixing and gamma prior, we obtain from Theorem 3.1 and (3.5), that the predictive density for Y admits the representation

$$Y|J'\stackrel{d}{=}B2\left(\frac{\nu_1}{2}+J'_1,\frac{\nu_2}{2}+J'_2,\frac{\nu_2}{\nu_1}\right),\;\text{with } J'_i|K'\stackrel{i.d}{=}NB(a+K'_i,\frac{1+b_i}{2+b_i}),\;K'\sim f'_x \text{ as in (3.13)}.$$

We point out that, if the distribution of $Y|\theta$ is of the same type as that of $X|\theta$ (in 3.12) with associated degrees of freedom ν'_1 and ν'_2 instead of ν_1 and ν_2, the only change in the previous expression is to replace the ν_i's by the ν'_i's for the distribution of $Y|J'$. When applicable, a similar observation applies to the other examples of this section.

Example 3.5. *Consider univariate Gamma mixtures with $X|K\stackrel{d}{=}G(\alpha,K)$ in model (1.1), with the mixing distribution $K|\theta\stackrel{d}{=}G(a,\theta)$ and prior $\theta\stackrel{d}{=}G(c,b)$, with known α,a,b,c. Theorem 3.1 applies and tells us that the posterior distribution $\theta|x$ is a mixture of $\pi_{K'}\sim G(a+c,K'+b)$ making use of a familiar posterior distribution for gamma models with gamma priors. In evaluating the mixing density f'_x of K' given in (3.4), it is easy to verify that $m_\pi\sim B2(a,c,b)$ and one thus obtains*

$$f'_x(k)\propto m_\pi(k)\,f_k(x)\propto\frac{k^{a+\alpha-1}}{(k+b)^{a+c}}\,e^{-xk}\,\mathbb{I}_{(0,\infty)}(x)\,,$$

which corresponds to a Kummer $K2(a+\alpha,c-\alpha,bx,b)$ distribution as defined in Section 2.

Now consider the Bayesian predictive density for the Gamma mixture $Y|J\stackrel{d}{=}G(\alpha',J)$ and $J|\theta\stackrel{d}{=}G(a',\theta)$, which includes the particular case of identically distributed $Y|\theta$ and $X|\theta$ for $\alpha'=\alpha$ and

$a' = a$. Observe that the density $q_\pi(\cdot|k')$ is that of $J|K = k'$ for the set-up $J|\theta \stackrel{d}{=} G(a', \theta)$, $K|\theta \stackrel{d}{=} G(a, \theta)$ independently distributed, and with $\theta \stackrel{d}{=} G(c, b)$. A calculation (e.g., Aitchison & Dunsmore, 1975) yields $q_\pi(\cdot|k') \stackrel{d}{=} B2(a', a + c, b + k')$. With the above, it follows from Theorem 3.1 that the predictive distribution for Y admits the representation:

$$Y|J' \stackrel{d}{=} G(\alpha', J')\,, \text{ with } J'|K' \stackrel{d}{=} B2(a', a + c, b + K') \text{ and } K' \stackrel{d}{=} K2(a + \alpha, c - \alpha, bx, b)\,.$$

An alternative representation comes from simply evaluating the marginal density h'_x of J'. A calculation gives:

$$h'_x(j') = \frac{\Gamma(a' + a + c)}{\Gamma(a')\,\Gamma(a + c)} \frac{b^{c-\alpha}\, j'^{(a'-1)}}{(j' + b)^{a'+c-\alpha}} \frac{\psi\,(a + \alpha, 1 + \alpha - a' - c, jx + bx)}{\psi\,(a + \alpha, 1 + \alpha - c, bx)} \mathbb{I}_{(0,\infty)}(j')\,,$$

with ψ the confluent hypergeometric function of type II as defined in Section 2.

The final application of Theorem 3.1 concerns the coefficient of determination in a standard multiple regression context.

Example 3.6. *Consider a coefficient of determination $X = R_1^2$, or square of a multiple correlation coefficient, that arises for a sample of size n_1 from $Z = (Z_1, \ldots, Z_m)^\top$, with $n_1 > m > 1$, and the regression of Z_1 based on $Z_2, \ldots, Z_m$. For more details on the underlying distributional theory, see for instance Muirhead (1982). It is well known that the distribution of X is a Type I mixture (1.1) with*

$$X|K \stackrel{d}{=} B\left(\frac{m-1}{2} + K, \frac{n_1 - m + 1}{2}\right)\,, \text{ with } K|\theta \stackrel{d}{=} NB\left(\frac{n_1 - 1}{2}, 1 - \theta\right)\,,$$

with $\theta = \rho^2$ is the theoretical squared multiple correlation coefficient. As in Marchand (2001), a convenient prior on θ is a Beta prior and it leads along with the negative binomial distributed $K|\theta$ to a conjugate posterior. Specifically, we obtain for $\theta \stackrel{d}{=} B(a, b)$ the posterior density $\pi_K \sim B\,(a + K, b + (n_1 - 1)/2)$, as well as the marginal p.m.f.

$$\begin{aligned} m_\pi(k) &= \int_{(0,1)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, \theta^{a-1}(1-\theta)^{b-1}\, \frac{\Gamma(\frac{n_1-1}{2} + k)}{k!\,\Gamma(\frac{n_1-1}{2})}\, \theta^k\,(1-\theta)^{(n_1-1)/2}\, d\theta \\ &= \frac{(\frac{n_1-1}{2})_k\,(a)_k}{k!\,(a + b + \frac{n_1-1}{2})_k}\, \frac{(b)_{(n_1-1)/2}}{(a+b)_{(n_1-1)/2}}\,, \text{ for } k \in \mathbb{N}\,. \end{aligned}$$

Theorem 3.1 tells us that the posterior distribution $\theta|x$ is a mixture of the π_k's with mixing

$$\begin{aligned} f'_x(k) &\propto f_k(x)\, m_\pi(k) \\ &\propto \frac{(\frac{n_1-1}{2})_k}{(\frac{m-1}{2})_k}\, x^k\, \frac{(\frac{n_1-1}{2})_k\,(a)_k}{k!\,(a + b + \frac{n_1-1}{2})_k} \\ \Rightarrow f'_x &\sim Hyp\left(\frac{n_1-1}{2}, \frac{n_1-1}{2}, a; \frac{m-1}{2}, a + b + \frac{n_1-1}{2}; x\right)\,. \end{aligned} \tag{3.14}$$

The result, which we have derived from the general context of Theorem 3.1 was obtained by Marchand (2001) in this specific set-up. In doing so, he defined such Beta mixtures as HyperBeta and also provided several graphs of prior-posterior densities for varying prior parameters a, b, sample size n, and observed values of X.

For the predictive density of a future $Y = R_2^2$ distributed as X but allowing for a possibly different sample size n_2, expression (3.5) tells us that such a predictive density admits the mixture representation:

$$Y|J' \stackrel{d}{=} B(\frac{m-1}{2} + J', \frac{n_2 - m + 1}{2}), J'|k' \sim q_\pi(\cdot|k'), \text{ and } K' \sim f'_x \text{ as in } (3.14), \text{with} \tag{3.15}$$

$q_\pi(\cdot|k')$ the predictive density for $J'|\theta \stackrel{d}{=} NB(\frac{n_2-1}{2}, 1-\theta)$ based on $K'|\theta \stackrel{d}{=} NB(\frac{n_1-1}{2}, 1-\theta)$ and prior $\theta \stackrel{d}{=} B(a,b)$. An evaluation of (3.6) yields

$$\begin{aligned} q_\pi(j'|k') &= \int_0^1 \frac{(\frac{n_2-1}{2})_{j'}}{j'!}\, \theta^{j'}\, (1-\theta)^{(n_2-1)/2}\, \frac{\Gamma(a+k'+b+\frac{n_1-1}{2})}{\Gamma(a+k')\,\Gamma(b+\frac{n_1-1}{2})}\, \theta^{a+k'-1}(1-\theta)^{b+(n_1-1)/2-1} d\theta \\ &= \frac{\Gamma(b+(n_1+n_2)/2-1)\,\Gamma(a+k'+b+\frac{n_1-1}{2})}{\Gamma(a+k'+b+(n_1+n_2)/2-1)\,\Gamma(b+\frac{n_1-1}{2})}\, \frac{(\frac{n_2-1}{2})_{j'}\,(a+k')_{j'}}{j'!\,(a+k'+b+(n_1+n_2)/2-1)_{j'}}, \text{ for } j' \in \mathbb{N}. \end{aligned}$$

It is worth noting that the above belongs to the class of p.m.f.'s of the form $p(n) = \frac{1}{{}_2F_1(a,b;c,1)} \frac{(a)_n\,(b)_n}{(c)_n\, n!}\, \mathbb{I}_{\mathbb{N}}(n)$ with $a, b, c > 0$, $c > a + b$, and as ${}_2F_1(a,b;c;1) = \frac{\Gamma(c)\,\Gamma(c-a-b)}{\Gamma(c-a)\,\Gamma(c-b)}$.

As a complement to the above, we briefly illustrate the usefulness of the predictive density (3.15) for visualizing how concordant or discordant the sample coefficients of determination $X = R_1^2$ and $Y = R_2^2$ can be when arising from studies set under the same conditions. Such a comparison also relates to the degree to which a second study replicates or not the results of an initial study when focusing on the R^2 summaries. Consider Figure 1 where both posterior densities for ρ^2 and predictive densities for $Y = R_2^2$ are drawn for $X = R_1^2 = 0.25, 0.50$, a B$(3,5)$ prior, and sample sizes $n_1 = 27, n_2 = 31$. The variability in the Bayesian predictive distributions is indeed quite significant, stemming of course from posterior uncertainty as well as model uncertainty. Further illustrating and for further interpretation of Figure 1, we record the expectations and standard deviations under the predictive densities: $\mathbb{E}(R_2^2|R_1^2 = 0.25) \approx 0.323$, $\sigma(R_2^2|R_1^2 = 0.25) \approx 0.159$, $\mathbb{E}(R_2^2|R_1^2 = 0.50) \approx 0.431$, and $\sigma(R_2^2|R_1^2 = 0.50), \approx 0.170$, as well as the selected probabilities $\mathbb{P}(R_2^2 > 0.40|R_1^2 = 0.25) \approx 0.310$ and $\mathbb{P}(R_2^2 < 0.10|R_1^2 = 0.25) \approx 0.087$.

4. Type II mixtures

We consider now Type II mixtures, provide general representations for the posterior distribution $\theta|x$, as well as the predictive distribution $Y|x$, and continue with various examples and observations.

Theorem 4.2. *Let $X, Y|\theta$ be conditionally independent distributed as in (1.2) and let π have prior density for θ with respect to $\sigma-$finite measure τ. Let $\pi_{k,x}$ and $g_{\pi,x}$ be respectively the conditional distribution densities of $\theta|x,k$ and $k|x$ given by*

$$\pi_{k,x}(\theta) = \frac{\pi(\theta)\, f_{\theta,k}(x)}{m_{\pi,k}(x)} \text{ and } g_{\pi,x}(k) = \frac{g(k)\, m_{\pi,k}(x)}{\int_{\mathcal{K}} g(k)\, m_{\pi,k}(x)\, d\mu(k)}, \theta \in \Theta, x \in \mathbb{R}^d, k \in \mathcal{K},$$

with $m_{\pi,k}$ given by

$$m_{\pi,k}(x) = \int_\Theta f_{\theta,k}(x)\, \pi(\theta)\, d\tau(\theta).$$

Then,

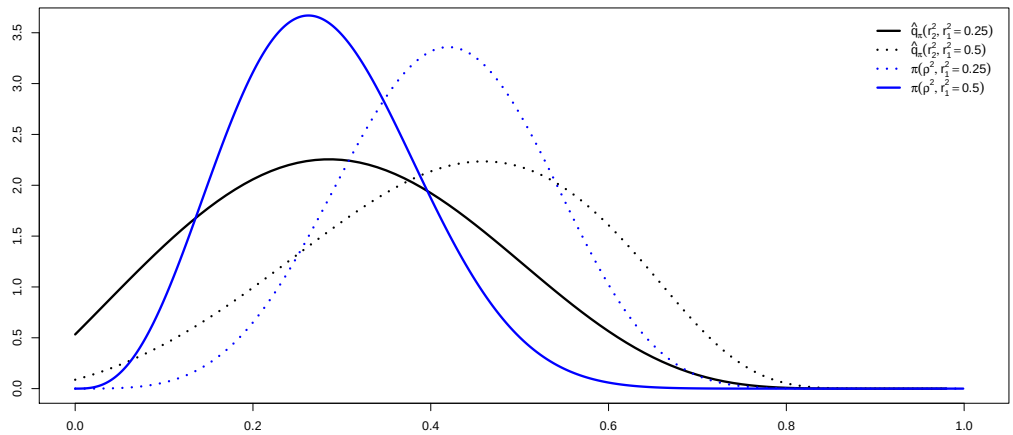

Figure 1: Bayesian predictive densities q_π for a future R_2^2 ($n_2 = 31$) based on two values of R_1^2 ($n_1 = 27$), and a B$(3, 5)$ prior.

(a) *The posterior distribution* $U \stackrel{d}{=} \theta|x$ *admits the mixture representation:*

$$U|K' \sim \pi_{K',x}\,,\; K' \sim g_{\pi,x}\,; \tag{4.16}$$

(b) *The Bayes predictive density of* Y *admits the representation:*

$$Y|J', K' \sim q_\pi(\cdot|J', K')\,, J' \sim h\,,\; K' \sim g_{\pi,x} \text{ independent}\,, \tag{4.17}$$

with

$$q_\pi(y|\, j', k') = \int_\Theta q_{\theta,j'}(y)\, \pi_{k',x}(\theta)\, d\tau(\theta)\,. \tag{4.18}$$

An alternative equivalent representation is:

$$Y|J' \sim q'_{x,J'}\,, J' \sim h,$$

with

$$q'_{x,j'}(y) \;=\; \int_{\mathcal{K}} q_\pi(y|j', k')\, g_{\pi,x}(k')\, d\mu(k')\,.$$

Proof. (a) We have indeed

$$\pi(\theta|x) \;=\; \int_{\mathcal{K}} \pi(\theta|x, k)\, g_{\pi,x}(k)\, d\mu(k)\,,$$

with

$$\pi(\theta|x, k) \;=\; \frac{\pi(\theta)\, f_{\theta,k}(x)\, g(k)}{\int_\Theta \pi(\theta)\, f_{\theta,k}(x)\, g(k)\, d\tau(\theta)} \;=\; \pi_{k,x}(\theta)\,,$$

$$\text{and } g_{\pi,x}(k) \;\propto\; \int_\Theta f_{\theta,k}(x)\, g(k)\, \pi(\theta)\, d\tau(\theta) \;=\; g(k)\, m_{\pi,k}(x)\,,$$

which establishes (4.16).

(b) The Bayes predictive density of Y, i.e. the conditional density of $Y|x$, is given by:

$$\begin{aligned}
q_\pi(y|x) &= \int_\Theta q(y|\theta)\,\pi(\theta|x)\,d\tau(\theta) \\
&= \int_\Theta \left(\int_\mathcal{K} q_{\theta,j'}(y)\,h(j')d\mu(j') \right) \pi(\theta|x)\,d\tau(\theta) \\
&= \int_\Theta \left(\int_\mathcal{K} q_{\theta,j'}(y)\,h(j')d\mu(j') \int_\mathcal{K} \pi_{k',x}(\theta)\,g_{\pi,x}(k')\,d\mu(k') \right) d\tau(\theta) \\
&= \int_\mathcal{K}\int_\mathcal{K} \left(\int_\Theta q_{\theta,j'}(y)\,\pi_{k',x}(\theta)\,d\tau(\theta) \right) g_{\pi,x}(k')\,h(j')\,d\mu(k')\,d\mu(j') \\
&= \int_\mathcal{K}\int_\mathcal{K} q_\pi(y|j',k')\,g_{\pi,x}(k')\,h(j')\,d\mu(k')\,d\mu(j')\,,
\end{aligned}$$

where we have used (4.16). This establishes the result. □

Remark 4.2. *Again, the posterior and predictive distributions have noteworthy representations. Namely, while the model distribution for $X|\theta$ mixes the $f_{\theta,k}$'s with mixing marginal density g of K, the posterior distribution (4.16) mixes the $\pi_{k',x}$'s, which represents for fixed k', x the posterior density of θ as if one had actually observed as well $K = k'$, with mixing density $g_{\pi,x}$ which is the marginal posterior distribution of K. Similarly, the density $m_{\pi,k'}$ is the marginal density of X as if $K = k'$ had been observed.*

In the second representation of the predictive distribution for Y, we have a mixture with the same mixing density h as of the model distribution of Y. The distributions mixed are themselves mixtures of the $q_\pi(\cdot|j',k')$ densities, which represent Bayes predictive densities for Y as if one had actually observed $K = k'$ and $J = j'$. With the first representation, we have bivariate mixing of the densities $q_\pi(\cdot|J',K')$ with independently distributed components J' and K'.

Example 4.7. *As a first illustration of the above Theorem, consider variance mixture of normal distributions with summaries $X|\theta, K \stackrel{d}{=} N_d(\theta, KI_d)$ with $K \sim g$ on $\mathbb{R}_+$, as well as $Y|\theta, J \stackrel{d}{=} N_d(\theta, JI_d)$ with $J \sim h$ on $\mathbb{R}_+$. Such a class includes not only finite mixtures with the mixing variables K and J discrete with finite support, but also several familiar distributions such as multivariate Student, Logistic and Laplace, among others.*

Now, consider the prior $\theta \stackrel{d}{=} N_d(\mu, \tau^2 I_d)$ with known μ, τ^2. In applying Theorem 4.2 for the posterior distribution of θ, it is well known that

$$\begin{aligned}
\pi_{k,x} &\sim N_d\left(x - \frac{k}{\tau^2+k}(x-\mu), \frac{\tau^2 k}{\tau^2+k} I_d \right)\,, \\
\text{and } \; m_{\pi,k} &\sim N_d\left(\mu, (\tau^2+k)\,I_d\right).
\end{aligned}$$

Theorem 4.2 thus implies that $U \stackrel{d}{=} \theta|x$ admits the representation

$$U|K' \sim \pi_{K',x} \text{ with } K' \sim g_{\pi,x}(k) \propto \frac{g(k)}{(k+\tau^2)^{d/2}}\; e^{-\frac{\|x-\mu\|^2}{2(k+\tau^2)}}\,. \tag{4.19}$$

Alternatively, we can express the above in terms of $Z = \frac{\theta - x}{\tau}$, $R = \frac{K'}{\tau^2 + K'}$, and $\beta = \frac{\mu - x}{\tau}$, yielding the mixture representation

$$Z|R \stackrel{d}{=} N_d\left(\beta R, RI_d\right), \;\; \text{with } R \sim g_{R,x}(r) \propto g\left(\frac{\tau^2 r}{1-r}\right)(1-r)^{d/2-2}\; e^{-\beta^2(1-r)/2}\,\mathbb{I}_{(0,1)}(r)\,,$$

$g_{R,x}$ being the density of R. The above posterior distribution for Z is an example of a variance-mean mixture of normal distributions popularized by Barndorff-Nielson et al. (1982), and much studied since in terms of theoretical and practical issues. For the improper prior $\pi(\theta) = 1$, it is not too difficult to evaluate directly the posterior, but Theorem 4.2 also applies. Indeed in this case, we have $\pi_{k,x} \sim N_d(x, kI_d)\,, m_{\pi,k}(x) = 1$, and the posterior distribution $U \stackrel{d}{=} \theta|x$ admits the representation of $U|K' \stackrel{d}{=} N_d(x, KI_d)$ with $K' \sim g$, which is a variance (only) mixture of normal distributions.

For the predictive distribution of Y, one can verify the Bayesian normal model predictive density calculation

$$q_\pi(\cdot|J', K') \sim N_d\left(x - \frac{K'}{\tau^2 + K'}(x-\mu)\,,\, (\frac{\tau^2\, K'}{\tau^2 + K'}) + J')I_d\right)\,.$$

With this, it follows from Theorem 4.2 that the Bayes predictive density of Y is a mixture of normal densities $q_\pi(\cdot|J', K')$ with $J' \sim h$ and $K' \sim g_{\pi,x}$ given in (4.19). Observe that the result applies if either X or Y is normally distributed by taking the corresponding mixing variable K or J to be degenerate. Finally, for the improper prior $\pi(\theta) = 1$, the above becomes $q_\pi(\cdot|J', K') \sim N_d(x, (J' + K')I_d)$ with $J' \sim h$ and $K' \sim g$, a result given by Kubokawa et al. (2015).

The next example deals with a mixture of exponential distributions (denoted $\mathcal{E}$) model which has garnered much over the years much interest in reliability.

Example 4.8. *We now consider here the following mixture of exponential distributions model:*

$$Z_1, \ldots, Z_n|K \stackrel{i.d}{=} \mathcal{E}(\mu, \frac{K}{\sigma})\,,\, K \sim g\,, \tag{4.20}$$

with $n \geq 2$, and with unknown location and scale parameters μ and σ. Here, the densities being mixed are exponential distributions such that $X_i - \mu|K \stackrel{d}{=} G(1, \frac{K}{\sigma})$. The above is a type II mixture and we set $\theta = (\mu, \rho)$ with $\rho = 1/\sigma$. Quite generally, model (4.20) is representative of n system components with a common environment, the case $n = 2$ having been put forth by Lindley & Singpurwalla (1986). A notable choice of g occurs for $K \stackrel{d}{=} G(\alpha, 1)$ with known α, and leads to the multivariate Lomax distribution (e.g., Nayak, 1987) which has generated much interest over the years.

It is reasonable to consider the summary (S, R) with $S = \sum_{i=1}^n (Z_i - Z_{(1)})$, $R = Z_{(1)} = \min_i Z_i$ (e.g., Petropoulos, 2017). The pair (S, R) is, namely, a jointly sufficient and complete statistic for degenerate K. Conditional on K, S and R are independently distributed as:

$$S|K \stackrel{d}{=} G(n-1, \rho K) \;\; \text{and} \;\; R|K \stackrel{d}{=} \mathcal{E}(\mu, nk\rho)\,. \tag{4.21}$$

We thus have a Type II mixture for $X = (S, R)$ with

$$f_{\theta,k}(x) \;=\; \frac{n(\rho k)^n}{\Gamma(n-1)}\, s^{n-2}\, e^{-\rho k\{s + n(r-\mu)\}}\; \mathbb{I}_{\mathbb{R}_+}(s)\, \mathbb{I}_{\mathbb{R}_+}(r-\mu)\,. \tag{4.22}$$

Consider now a gamma prior for ρ and a flat prior for μ so that

$$\pi(\rho, \mu) \;\propto\; \rho^{a-1}\, e^{-b\rho}\, \mathbb{I}_{\mathbb{R}_+}(\rho)\, \mathbb{I}_{\mathbb{R}}(\mu)\,, a, b > 0\,. \tag{4.23}$$

In fact, the posterior distribution exists as well for $b = 0$, and $a \in (2-n, 0]$ for $n > 2$. Evaluating the density $\pi_{k,x}$ of Theorem 4.2. we obtain

$$\pi_{k,x}(\rho, \mu) \;\propto\; \{\rho^{a+n-2}\, e^{-\rho(ks+b)}\}\, \{\rho kn\, e^{-\rho kn(r-\mu)}\}\; \mathbb{I}_{\mathbb{R}_+}(\rho)\, \mathbb{I}_{\mathbb{R}_+}(r-\mu)\,, \tag{4.24}$$

which implies the posterior distribution representation:

$$\rho|K,x \stackrel{d}{=} G(a+n-1, Ks+b)\,,\ \omega|\rho,K,x \stackrel{d}{=} \mathcal{E}(0,\rho Kn)\,, \tag{4.25}$$

where we have set $\omega = r-\mu$. Now, evaluating the marginal density[2] *$m_{\pi,k}$ of $X=(S,R)$, we have from (4.22) and (4.23):*

$$\begin{aligned} m_{\pi,k}(x) &= \int_\Theta \pi(\theta)\, f_{\theta,k}(x)\, d\tau(\theta) \\ &\propto k^n \int_{\mathbb{R}_+} \rho^{a+n-1}\, e^{-\rho(b+ks)} \{\int_{-\infty}^{r} e^{-\rho nk(r-\mu)} d\mu\}\, d\rho \\ &\propto \frac{k^{n-1}}{(\frac{b}{s}+k)^{a+n-1}}\,, \end{aligned}$$

as a function of $k \in \mathcal{K}$. Finally, putting the above elements together, we have for $X|K,\theta$ as in (4.21) with $K \sim g$ and prior density (4.23), the following representation for the posterior distribution of $(\rho, \omega = r-\mu)$:

$$\rho|K',x \stackrel{d}{=} G(a+n-1, K's+b)\,,\ \omega|\rho,K',x \stackrel{d}{=} \mathcal{E}(0,\rho K'n) \ \text{ with } K' \sim g_{\pi,x}(k) \propto \frac{g(k)\, k^{n-1}}{(\frac{b}{s}+k)^{a+n-1}}\,.$$

From this, we obtain for $g \stackrel{d}{=} G(\alpha,\beta)$ the Kummer density

$$g_{\pi,x} \stackrel{d}{=} K2(\alpha+n-1, a+1-\alpha, \frac{b\beta}{s}, \frac{b}{s})\,, \tag{4.26}$$

with $\beta = 1$ corresponding to the multivariate Lomax case. Alternatively, for an inverse Gamma density g with density proportional to $k^{-\alpha-1}\, e^{-\beta/k}\, \mathbb{I}_{(0,\infty)}(k)$, we obtain that the mixing density of K' is such that $\frac{1}{K'} \stackrel{d}{=} K2(a+\alpha+2, n-\alpha-3, \beta s/b, s/b)$.

Now, for deriving a predictive density via Theorem 4.2, based on $X=(S,R)$ distributed as above in (4.21) with $n=n_1$, consider predicting $Y=(S',R')$ such that:

$$S'|J \stackrel{d}{=} G(n_2-1,\rho J) \ \text{ independent of } \ R'|J \stackrel{d}{=} \mathcal{E}(\mu, n_2\rho J)\,,\ \text{ with } J \sim h\,. \tag{4.27}$$

These statistics arise naturally, just as the case for $X=(S,R)$, as jointly complete and sufficient conditional for the observables $W_1,\ldots,W_{n_2}|J \stackrel{i.d}{=} \mathcal{E}(\mu, J\rho)$ with $S' = \sum_{i=1}^{n_2}(W_i - W_{(1)})$ and $R' = W_{(1)}$. Following Theorem 4.2, based on a prior density π for $\theta=(\rho,\mu)$, we have that the Bayes predictive density of Y admits representation (4.17), which is $Y|J',K' \sim q_\pi(\cdot|J',K')$ with $J' \sim h$ and $K' \sim g_{\pi,x}$. There thus remains and we now seek to evaluate

$$q_\pi(s',r'|j',k') = \int_{\mathbb{R}^2} q_{\theta,j'}(s',r')\, \pi_{k',x}(\rho,\mu)\, d\rho\, d\mu\,, \tag{4.28}$$

with

$$q_{\theta,j'}(s',r') = \frac{s'^{(n_2-2)}\, e^{-\rho j' s'}}{\Gamma(n_2-1)}\, (\rho j')^{n_2}\, n_2\, e^{-n_2 j'\rho(r'-\mu)}\, \mathbb{I}_{\mathbb{R}_+}(s')\, \mathbb{I}_{(\mu,\infty)}(r')$$

[2] As a function of x, it is not a true density here as we have used an improper prior.

and

$$\pi_{k',x}(\rho,\mu) \;=\; \frac{\rho^{a+n_1-2}\, e^{-\rho(k's+b)}}{\Gamma(a+n_1-1)}\,(k's+b)^{a+n_1-1}\,(\rho n_1 k')\, e^{-\rho k' n_1 (r-\mu)}\, \mathbb{I}_{\mathbb{R}_+}(\rho)\, \mathbb{I}_{(-\infty,r)}(\mu)\,.$$

After some calculations, setting $A_0 = \frac{n_1 n_2 k' j'^{n_2}\, s'^{(n_2-2)}\, (k's+b)^{a+n_1-1}}{\Gamma(n_2-1)\,\Gamma(a+n_1-1)}$ *and* $B_0 = j's' + k's + b + n_2 j' r' + n_1 k' r$*, one obtains*

$$\begin{aligned} q_\pi(s',r'|j',k') &= A_0 \int_0^\infty \rho^{a+n_1+n_2-1}\, e^{-\rho B_0} \int_{-\infty}^{r\wedge r'} e^{\rho\mu(n_2 j' + n_1 k')} d\rho\, d\mu \\ &= \frac{A_0}{(n_2 j' + n_1 k')}\, \frac{\Gamma(a+n_1+n_2-1)}{\{j's' \,+\, k's \,+\, b \,+\, c(r')\,\}^{a+n_1+n_2-1}}\; \mathbb{I}_{\mathbb{R}_+}(s')\; \mathbb{I}_{\mathbb{R}}(r')\,, \end{aligned}$$

with $c(r') = |r'-r|\,\big\{n_2 j'\, I_{(r,\infty)}(r') + n_1 k'\, I_{(-\infty,r)}(r')\,\big\}$*. This provides an explicit expression for the sought-after (4.28), but a more appealing decomposition of the distribution of* (S',R') *under* $q_\pi(\cdot|j',k')$*, easily derived from the above, is:*

$$\begin{aligned} S'|R',J',K' &\overset{d}{=} B2(n_2-1, a+n_1, \frac{K's+b+c(r')}{J'}) \qquad (4.29) \\ R'|J',K' &\sim p\,\frac{A_1^{a+n_1-1}(a+n_1-1)}{(A_1+|r'-r|)^{a+n_1}}\, I_{(r,\infty)}(r') \;+\; (1-p)\,\frac{A_2^{a+n_1-1}(a+n_1-1)}{(A_2+|r'-r|)^{a+n_1}}\, I_{(-\infty,r)}(r')\,, \end{aligned}$$

with $p = n_1 K'/(n_1 K' + n_2 J')$*,* $A_1 = \frac{K's+b}{n_2 J'}$*, and* $A_2 = \frac{K's+b}{n_1 K'}$*. Finally, we point out that the above distribution for* R' *can be described as bilateral Pareto since, conditional on* J',K'*:*

$$R'-r|R'>r \overset{d}{=} B2(1, a+n_1-1, A_1)\,,\; r-R'|R'<r \overset{d}{=} B2(1,a+n_1-1,A_2) \;\; \textit{with}\; p=\mathbb{P}(R'>r)\,. \qquad (4.30)$$

Summarizing for the multivariate Lomax case with $K \overset{d}{=} G(\alpha,1)$*, the predictive density of* (S',R') *given* (s,r) *and prior* π *in (4.23) may be generated as:*

$$J' \sim h\,,\; K' \overset{d}{=} K2(\alpha+n_1-1, a+1, \frac{b}{s}, \frac{b}{s}),$$

and then $R'|J',K'$ *and* $S'|R',J',K'$ *as the bilateral Pareto and Beta type II respectively defined in (4.30) and (4.29) respectively.*

There are many other potential applications of Theorem 4.2, including mixtures of Gamma(T,θ) distributions with $T \sim g$, but we will focus hereafter on bivariate gamma mixtures, which we define as follows and which is inspired by the Kibble distribution mixture representation (see Example 1.1).

Definition 4.1. *A Kibble-type bivariate gamma mixture with parameters* $\nu_1,\nu_2,\theta_1,\theta_2$ *(all* >0*) and mixing distribution* g*, which we denote* $X=(X_1,X_2) \overset{d}{=} G2(\nu_1,\nu_2,\theta_1,\theta_2,g)$*, is defined through the representation:*

$$X_i|K \overset{i.d}{=} G(\nu_i+K,\theta_i), i=1,2, \;\; \textit{with}\; K\sim g\,.$$

In the above definition, one recovers the original Kibble distribution in (1.3) for $\nu_1=\nu_2=\nu, \theta_i = \frac{1-\rho}{\lambda_i}$, and $K \overset{d}{=} \mathrm{NB}(\nu, 1-\rho)$, but many other interesting cases are generated including Poisson mixtures with $K \overset{d}{=} \mathrm{P}(\lambda)$ and continuous mixing as well. Of course, one requires g such that $\mathbb{P}(K > -\min\{\nu_1,\nu_2\}) = 1$. An application of Theorem 4.2 to such bivariate gamma mixtures yields the following.

Corollary 4.1. *Consider type II mixtures as in (1.2) with* $X|\theta \stackrel{d}{=} G2(\nu_1, \nu_2, \theta_1, \theta_2, g)$ *and* $Y|\theta \stackrel{d}{=} G2(\nu_1', \nu_2', \theta_1, \theta_2, h)$*, and prior distribution* $\theta_i \stackrel{i.d}{=} G(a_i, b_i), i = 1, 2$*, Then,*

(a) *The posterior distribution of* θ_1, θ_2 *based on* x *is* $G2(\nu_1 + a_1, \nu_2 + a_2, b_1 + x_1, b_2 + x_2, g_{\pi,x})$ *with* $g_{\pi,x}$ *a density (or p.m.f.) such that*

$$g_{\pi,x}(k) \propto g(k) \frac{(\nu_1 + a_1)_k (\nu_2 + a_2)_k}{(\nu_1)_k (\nu_2)_k} \left(\frac{x_1 x_2}{(b_1 + x_1)(b_2 + x_2)} \right)^k . \tag{4.31}$$

(b) *The predictive distribution for* Y *admits the representation*

$$Y|J', K' \stackrel{i.d}{=} B2(\nu_i' + J', a_i + \nu_i + K', b_i + x_i), i = 1, 2, \text{ with } J' \sim h,\ K' \sim g_{\pi,x} \text{ independent.} \tag{4.32}$$

Proof. (a) We have $X_i|k' \stackrel{i.d}{=} \mathrm{G}(\nu_i + k', \theta_i)$ and $\theta_i \stackrel{i.d}{=} \mathrm{G}(a_i, b_i)$, which implies that $\pi_{k',x}$ is the density of $\theta_i|k', x \stackrel{i.d}{=} \mathrm{G}(a_i + \nu_i + k', b_i + x_i)$. It thus follows from (4.16) that the posterior distribution of θ_1, θ_2 is $\mathrm{G2}(\nu_1 + a_1, \nu_2 + a_2, b_1 + x_1, b_2 + x_2, g_{\pi,x})$. Finally, a direct calculation yields

$$m_{\pi,k}(x) \propto \frac{(\nu_1 + a_1)_k (\nu_2 + a_2)_k}{(\nu_1)_k (\nu_2)_k} \left(\frac{x_1 x_2}{(b_1 + x_1)(b_2 + x_2)} \right)^k , \tag{4.33}$$

as a function of k, and the result follows.

(b) This follows from (4.17), (4.31), and the evaluation of the predictive density of $Y_i|J = j' \stackrel{i.d}{=} \mathrm{G}(\nu_i' + j', \theta_i)$ based on $X_i|K = k' \stackrel{i.d}{=} \mathrm{G}(\nu_i + k', \theta_i)$, which has independently distributed B2 marginals as given in (4.32) and previously encountered in Example 3.5. □

Remark 4.3. *Corollary 4.1 applies quite generally with respect to the type of mixing, and include "degenerate" cases with* $\mathbb{P}_g(K = 0) = 1$ *and-or* $\mathbb{P}_h(J = 0) = 1$ *corresponding to independent components of* X *and* Y *respectively. The results of Corollary 4.1 apply as well to some choices of the* a_i*'s and* b_j*'s that lead to an improper prior, as long as the densities* $\pi_{K,x}$ *exist. These include for instance the choices* $b_1 = b_2 = 0$*, with* a_1, a_2 *such that* $\mathbb{P}(K > -\max\{\nu_1 + a_1, \nu_2 + a_2\}) = 1$*. Observe as well that the choice* $a_1 = a_2 = b_1 = b_2 = 0$*, which is typically taken as the usual non-informative prior, leads to the simplification* $g_{\pi,x} \equiv g$ *in (4.31) and (4.32).*

Example 4.9. *Corollary 4.1 applies for any mixing distributions* g *and* h*, such as Poisson mixing:* **(a)** $g \sim P(\lambda_X)$ *and* $h \sim P(\lambda_Y)$*, and negative binomial mixing:* **(b)** $g \sim NB(r_X, \xi_X)$ *and* $h \sim NB(r_Y, \xi_Y)$*, with only the mixing density* $g_{\pi,x}$ *varying. We obtain*

$$g_{\pi,x} \sim Hyp\left(\nu_1 + a_1, \nu_2 + a_2; \nu_1, \nu_2; \frac{\lambda_X x_1 x_2}{(b_1 + x_1)(b_2 + x_2)} \right) ,$$

for the Poisson case **(a)***, and*

$$g_{\pi,x} \sim Hyp\left(r_X, \nu_1 + a_1, \nu_2 + a_2; \nu_1, \nu_2; \frac{(1 - \xi_X) x_1 x_2}{(b_1 + x_1)(b_2 + x_2)} \right) ,$$

for the negative binomial case **(b)***. Case* **(b)** *reduces and applies to the original Kibble distribution, as in Example 1.1, for* $\nu_1 = \nu_2 = r_X$*,* $\nu_1' = \nu_2' = r_Y, \xi_X = \xi_Y = 1 - \rho$*,* $\theta_i = \frac{1-\rho}{\lambda_i}, i = 1, 2$*. Finally, it is immediate to extend the example to cases where the density* $g_{\pi,x}$ *is that of a generalized hypergeometric distribution, obtaining* $g_{\pi,x} \sim Hyp(a_1, \ldots, a_p, \nu_1 + a_1, \nu_2 + a_2; b_1, \ldots, b_q, \nu_1, \nu_2; \frac{\lambda x_1 x_2}{(b_1+x_1)(b_2+x_2)})$*.*

Our next example and application of Theorem 4.2 relates again to Kibble-type bivariate gamma mixtures, but with a prior on (θ_1, θ_2) adapted to situations where there exists an order restriction. Many such situations are plausible. An interesting case concerns $M|M|1$ queues and the estimation of the arrival (λ) and service (μ) rates for independent gamma observables, and in particular the ratio $\rho = \lambda/\mu$, whenever a steady-state distribution exists, i.e. $\lambda < \mu$ (Armero & Bayarri, 1994). In doing so, it seems natural to consider a McKay distribution as a prior. Another interesting prior choice, which was analyzed by Armero & Bayarri (1994) for independent gamma observables, is to start with two independent gamma distributions for θ_1 and θ_2 truncated to the ordering set $\theta_2 \geq \theta_1$. Before presenting how Theorem 4.2 applies, we catalog some information about a generalized version of the McKay distribution which contains both the above-mentioned prior choices, which will appear below in the posterior analyses, and which is of interest on its own.

Definition 4.2. *The generalized McKay distribution with parameters $\vec{\gamma} = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)$, with $\gamma_1 > 0$, $\gamma_2 > 0$, $\gamma_4 > 0$, $\gamma_3 > \gamma_5 \geq 0$ and $\gamma_1 + \gamma_2 + \gamma_4 > 1$, denoted $MG(\vec{\gamma})$, is a bivariate distribution with p.d.f.*

$$f_{Z_1,Z_2}(z_1, z_2) \;=\; \frac{1}{C(\vec{\gamma})}\; z_1^{\gamma_1-1}\,(z_2 - z_1)^{\gamma_2-1}\, z_2{}^{\gamma_4-1}\, e^{-\gamma_3 z_2 - \gamma_5 z_1}\, \mathbb{I}_{(0,\infty)}(z_1)\, \mathbb{I}_{(z_1,\infty)}(z_2)\,, \tag{4.34}$$

$$\textit{and}\;\; C(\vec{\gamma}) \;=\; \frac{\Gamma(\gamma_1)\Gamma(\gamma_2)}{\Gamma(\gamma_1+\gamma_2)}\, \frac{\Gamma(\gamma_1+\gamma_2+\gamma_4-1)}{\gamma_3^{\gamma_1+\gamma_2+\gamma_4-1}}\; {}_2F_1(\gamma_1, \gamma_1+\gamma_2+\gamma_4-1, \gamma_1+\gamma_2; -\frac{\gamma_5}{\gamma_3})\,. \tag{4.35}$$

Observe that the above distribution reduces to a McKay distribution ($\mathrm{M}(\gamma_1, \gamma_2, \gamma_3)$) for $\gamma_4 = 1$ and $\gamma_5 = 0$. As well, the case $\gamma_2 = 1$ with $\gamma_5 > 0$ is generated by two independently distributed $G(\gamma_1, \gamma_5)$ and $G(\gamma_4, \gamma_3)$ truncated to the region $z_2 > z_1$. We have collected some technical observations and properties, including the computation of the normalization constant $C(\vec{\gamma})$, of such distributions in the Appendix.

Example 4.10. *Consider the application of Theorem 4.2 for $X|\theta \stackrel{d}{=} G2(\nu_1, \nu_2, \theta_1, \theta_2, g)$ and $Y|\theta \stackrel{d}{=} G2(\nu_1', \nu_2', \theta_1, \theta_2, h)$ combined with the prior $\theta \stackrel{d}{=} MG(\vec{\gamma})$. From expressions below for the posterior density of θ and the predictive density for Y, we will be able to extract expressions for Gamma observables with g and h degenerate, as well as for both McKay prior and the truncated Gamma prior considered by Armero & Bayarri (1994). Combining the density of $\theta \stackrel{d}{=} MG(\vec{\gamma})$ with the likelihood $f_{\theta,k}(x) \propto \theta_1^{\nu_1+k}\, e^{-\theta_1 x_1} \theta_2^{\nu_2+k}\, e^{-\theta_2 x_2}$, we obtain that*

$$\begin{aligned} \pi_{k,x}(\theta) \;\sim\; MG(\vec{\gamma}_{k,x}),\;\; \textit{with}\; \vec{\gamma}_{k,x} = (\gamma_1', \gamma_2', \gamma_3', \gamma_4', \gamma_5'), \\ \textit{and}\; \gamma_1' = \gamma_1 + \nu_1 + k, \gamma_2' = \gamma_2, \gamma_3' = x_2 + \gamma_3, \gamma_4' = \gamma_4 + \nu_2 + k, \gamma_5' = x_1 + \gamma_5\,. \end{aligned} \tag{4.36}$$

We also have

$$\begin{aligned} m_{\pi,k}(x) \;&=\; \frac{1}{C(\vec{\gamma})}\, \frac{x_1^{\nu_1+k-1}\, x_2^{\nu_2+k-1}}{\Gamma(\nu_1+k)\,\Gamma(\nu_2+k)} \int_0^\infty \int_{\theta_1}^\infty \theta_1^{\gamma_1'-1}\,(\theta_2-\theta_1)^{\gamma_2'-1}\,\theta^{\gamma_4'-1}\, e^{-\left(\theta_2\gamma_3' + \gamma_5'\theta_1\right)}\, d\theta_2 d\theta_1 \\ &=\; \frac{C(\vec{\gamma}_{k,x})}{C(\vec{\gamma})}\, \frac{x_1^{\nu_1+k-1}\, x_2^{\nu_2+k-1}}{\Gamma(\nu_1+k)\,\Gamma(\nu_2+k)}\,, \end{aligned}$$

which implies that

$$g_{\pi,x}(k) \;\propto\; g(k)\, \frac{(x_1\, x_2\,)^k}{\Gamma(\nu_1+k)\,\Gamma(\nu_2+k)}\; C(\vec{\gamma}_{k,x})\,, \tag{4.37}$$

with $C(\cdot)$ defined in (4.35) and $\vec{\gamma}_{k,x}$ given in (4.36). From (3.4), we therefore infer that the posterior distribution $U \stackrel{d}{=} \pi|x$ admits the representation

$$U|K' \stackrel{d}{=} MG(\vec{\gamma}_{K',x}) \;\; \text{with } K' \sim g_{\pi,x}\,, \tag{4.38}$$

as given by (4.36) and (4.37). As mentioned above, the above representation applies to McKay distributed priors (taking $\gamma_4 = 1, \gamma_5 = 0$), as well as the truncated Gamma priors of Armero & Bayarri (taking $\gamma_2 = 1, \gamma_5 > 0$). However, despite its conciseness, this last representation (4.38) remains rather unwieldy due to the $g_{\pi,x}$ density term.

Degenerate case of the mixtures G2, i.e., with $\mathbb{P}(K = 0) = 1$ in Definition 4.1, are not particularly interesting when the joint prior distribution for θ_1 and θ_2 factorizes into independent components (e.g., as in Corollary 4.1), since the inference problem separates completely into two univariate independent problems. It is also not admittedly the focus of the mixture models considered in this paper. However, for non-independent components θ_1 and θ_2 such as the situation here, the Bayesian analysis is also interesting in the degenerate case. And, the above results applied to $X_i \stackrel{i.d}{=} G(\nu_i, \theta_i)$ yield immediately the posterior $\theta|x \sim MG(\vec{\gamma}_{0,x})$. This further connects to the McKay prior by taking $\gamma_4 = 1$ and $\gamma_5 = 0$, and to the Armero & Bayarri prior by taking $\gamma_2 = 1$ and $\gamma_5 > 0$. Furthermore, the latter work considers the change of variables $T_1 = \theta_1/\theta_2, T_2 = \theta_2$ for reasons of ease of interpretation, and properties of these transformed variables are briefly outlined in the Appendix below.

Turning now to the Bayesian predictive distribution for $Y|\theta$, still based on $X|\theta \stackrel{d}{=} G2(\nu_1, \nu_2, \theta_1, \theta_2, g)$ and prior $\theta \stackrel{d}{=} MG(\vec{\gamma})$, we have for (4.18), and making use of (4.36),

$$\begin{aligned} q_\pi(y|j', k') &= \int_0^\infty \int_{\theta_1}^\infty \prod_{i=1}^2 \frac{y_i^{\nu_i'+j'-1} e^{-\theta_i y_i}}{\Gamma(\nu_i' + j')} \, \frac{\theta_i^{(\nu_i'+j')}}{C(\vec{\gamma}_{k',x})} \, \theta_1^{\gamma_1'-1} \, (\theta_2 - \theta_1)^{\gamma_2'-1} \, \theta^{\gamma_4'-1} \, e^{-\left(\theta_2\gamma_3' + \gamma_5'\theta_1\right)} \, d\theta_2 d\theta_1 \\ &= \frac{y_1^{\nu_1'+j'-1} \, y_2^{\nu_2'+j'-1}}{\Gamma(\nu_1' + j) \, \Gamma(\nu_2' + j)} \, \frac{C(\vec{\gamma}_{k',j',x,y})}{C(\vec{\gamma}_{k',x})} \end{aligned} \tag{4.39}$$

for $y > 0$, and with $\vec{\gamma}_{k',j',x,y} = (\gamma_1'', \gamma_2'', \gamma_3'', \gamma_4'', \gamma_5'')$, $\gamma_1'' = \gamma_1 + \nu_1 + \nu_1' + k' + j', \gamma_2'' = \gamma_2, \gamma_3'' = \gamma_3 + x_2 + y_2, \gamma_4'' = \gamma_4 + \nu_2 + \nu_2' + k' + j'$, and $\gamma_5'' = \gamma_5 + x_1 + y_1$. We thus infer that the predictive distribution of Y admits representation (4.17) with $q_\pi(\cdot|J', K')$ given in (4.39), and $g_{\pi,x}$ given in (4.37). Finally, the same remarks as above applies for the degenerate case with $\mathbb{P}(K = 0) = \mathbb{P}(J = 0) = 1$, In particular, the Bayesian predictive density for Y is given by $q_\pi(\cdot|0, 0)$, i.e., expression (4.39) with $j' = k' = 0$.

We now turn to a further illustration for models and priors that are bivariate gamma mixtures. A motivation for this is the flexibility to have prior components that are non-independent in addressing inference for the Kibble distribution parameters or more generally for Kibble type gamma mixtures (Definition 4.1). But, in contrast to the previous example, the dependence among the prior components arises not through an order constraint, but by the presence of a hierarchical commonality.

Corollary 4.2. *Consider type II mixtures as in (1.2) with $X|\theta \stackrel{d}{=} G2(\nu_1, \nu_2, \theta_1, \theta_2, g)$ and $Y|\theta \stackrel{d}{=} G2(\nu_1', \nu_2', \theta_1, \theta_2, h)$, and prior distribution $\theta \stackrel{d}{=} G2(a_1, a_2, b_1, b_2, \epsilon)$. Then,*

(a) *The posterior distribution of θ based on x is $G2(a_1+\nu_1, a_2+\nu_2, b_1+x_1, b_2+x_2, \epsilon_x)$, with the mixing density (or p.m.f.) ϵ_x is that of $S =^d K' + L'$ where (K', L') has joint density (or p.m.f.) given by*

$$p_{K',L'}(k,l) \propto \frac{(a_1+\nu_1)_{k+l}\,(a_2+\nu_2)_{k+l}}{(a_1)_l\,(a_2)_l\,(\nu_1)_k\,(\nu_2)_k}\,\frac{(x_1x_2)^k\,(b_1b_2)^l}{\{(b_1+x_1)(b_2+x_2)\}^{k+l}}\;g(k)\;\epsilon(l)\,. \tag{4.40}$$

(b) *The predictive distribution for $Y = (Y_1, Y_2)$ admits the representation*

$$Y_i|J', K' \overset{i.d}{=} B2(\nu_i' + J', a_i+\nu_i+K', b_i+x_i), i=1,2, \text{ with } J' \sim h,\ K' \sim g_{\pi,x} \text{ independent,}$$

$$\begin{aligned} \textit{with} \qquad & g_{\pi,x}(k) \propto \frac{g(k)\,(x_1x_2)^k}{\{(x_1+b_1)(x_2+b_2)\}^k}\,\frac{I(k,x)}{(\nu_1)_k\,(\nu_2)_k}\,, \\ \textit{and} \qquad & I(k,x) = \int_{\mathcal{K}} \frac{(a_1+\nu_1)_{k+l}\,(a_2+\nu_2)_{k+l}}{(a_1)_1\,(a_2)_l}\left(\frac{b_1b_2}{(x_1+b_1)(x_2+b_2)}\right)^l d\epsilon(l)\,. \end{aligned} \tag{4.41}$$

Proof. (a) From Theorem 4.2, we have

$$\pi_{k,x}(\theta) \propto \int_0^\infty \frac{\theta_1^{a_1+\nu_1+k+l-1}\,\theta_2^{a_2+\nu_2+k+l-1}}{\Gamma(a_1+l)\,\Gamma(a_2+l)}\,e^{-\theta_1(x_1+b_1)-\theta_2(x_2+b_2)}\,(b_1b_2)^l\,d\epsilon(l)\,,$$

which is the density of a $G2\left(a_1+\nu_1+k, a_2+\nu_2+k, x_1+b_1, x_2+b_2, \epsilon'_{x,k}\right)$ distribution having mixing density

$$\epsilon'_{x,k}(l) = \frac{1}{I(k,x)}\,\frac{(a_1+\nu_1)_{k+l}\,(a_2+\nu_2)_{k+l}}{(a_1)_1\,(a_2)_l}\left(\frac{b_1b_2}{(x_1+b_1)(x_2+b_2)}\right)^l \epsilon(l)\,.$$

This along with Theorem 4.2 tell us that the posterior distribution $U \overset{d}{=} \theta|x$ admits the representation

$$U_i|K', L' \overset{i.d}{=} \mathrm{G}(\alpha_i+\nu_i+S, x_i+\beta_i)\,, i=1,2,$$

with $S =^d K' + L'$, $L'|K' = k' \sim \epsilon'_{x,k'}$, and $K' \sim g_{\pi,x}$. Finally, we have by collecting terms and with an interchange in the order of integration:

$$\begin{aligned} p_{K',L'}(k,l) &\propto \epsilon'_{x,k}(l)\,g(k)\,m_{\pi,k}(x) \\ &\propto \epsilon'_{x,k}(l)\,g(k)\int_\Theta f_{\theta,k}(x)\,\pi(\theta)\,d\tau(\theta) \\ &\propto \frac{\epsilon'_{x,k}(l)\,g(k)\,(x_1x_2)^k}{\Gamma(\nu_1+k)\,\Gamma(\nu_2+k)}\int_{\mathcal{K}}\int_\Theta \frac{\theta_1^{a_1+\nu_1+k+l-1}\,\theta_2^{a_2+\nu_2+k+l-1}}{\Gamma(a_1+l)\Gamma(a_2+l)}\,e^{-\theta_1(x_1+b_1)-\theta_2(x_2+b_2)}\,(b_1b_2)^l\,d\epsilon(l) \\ &\propto \epsilon'_{x,k}(l)\,g(k)\,\frac{(x_1x_2)^k}{\{(x_1+b_1)(x_2+b_2)\}^k}\,\frac{I(k,x)}{\Gamma(\nu_1+k)\,\Gamma(\nu_2+k)}\,, \end{aligned} \tag{4.42}$$

which is indeed (4.40).

(b) As in the proof of part (b) of Corollary 4.1, the result follows with the evaluation $g_x(k) \propto g(k)\,m_{\pi,k}(x)$ obtained in (4.42) and yielding (4.41). □

Remark 4.4. *The results of Corollary 4.2 are quite general and many particular cases follow, namely depending on the choices of mixing densities for the model and the prior. For degenerate mixing under the prior, i.e.,* $\mathbb{P}_\epsilon(L=0)=1$*, one recovers Corollary 4.1. For independent Gamma distributed observables with* $\mathbb{P}_g(K=0)=1$*, Corollary 4.2 applies for the posterior distribution, but with* $S =^d L'$ *and* L' *having density* $p_{K',L'}(0,l)$ *in (4.40). In this last case, Poisson or negative binomial distributed* L' *leads to results similar to those of Example 4.9. Finally, again with* K *degenerate at* 0*, the predictive distribution in part* **(b)** *of Corollary 4.2 applies with* $\mathbb{P}(K'=0)=1$ *under* $g_{\pi,x}$*.*

Example 4.11. *We consider Kibble models and priors as a particular application of Corollary 4.2. Namely, take* $\nu_1=\nu_2$*,* $a_1=a_2$*, as well as* $K \stackrel{d}{=} NB(\nu_1, 1-\rho)$*,* $L \stackrel{d}{=} NB(a_1, 1-\rho_\theta)$ *as the mixing variables associated with* $X|\theta$ *and* θ *respectively. With p.m.f.'s* $g(k) = \frac{(\nu_1)_k}{k!}\,(1-\rho)^{\nu_1}\,\rho^k\,\mathbb{I}_{\mathbb{N}}(k)$ *and* $\epsilon(l) = \frac{(a_1)_l}{l!}\,(1-\rho)^{a_1}\,\rho^l\,\mathbb{I}_{\mathbb{N}}(l)$*, we obtain in applying part* **(a)** *of Corollary 4.2 that*

$$p_{K',L'}(k,l) \;=\; \frac{1}{F_4(a_1+\nu_1, a_1+\nu_1; a_1, \nu_1; \omega_1, \omega_2)}\;\frac{(a_1+\nu_1)_{k+l}\,(a_1+\nu_1)_{k+l}}{(a_1)_l\,(\nu_1)_k\,l!\,k!}\;\omega_1^k\,\omega_2^l\,\mathbb{I}_{\mathbb{N}}(k)\,\mathbb{I}_{\mathbb{N}}(l)\,,$$

with $\omega_1 = \frac{\rho\,x_1\,x_2}{(x_1+b_1)(x_2+b_2)}$*,* $\omega_2 = \frac{\rho_\theta\,b_1\,b_2}{(x_1+b_1)(x_2+b_2)}$*, and where* F_4 *is the Appell function of the fourth kind given by*

$$F_4(\gamma_1,\gamma_2;\gamma_3,\gamma_4;w,z) \;=\; \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{(\gamma_1)_{m+n}\,(\gamma_2)_{m+n}}{m!\,n!\,(\gamma_3)_n\,(\gamma_4)_m}\,w^m\,z^n\,.$$

Part **(b)** *of Corollary 4.2 applies directly for the Bayesian predictive distribution with* $J' \stackrel{d}{=} NB(a_1, 1-\rho_\theta)$ *and* $g_{\pi,x}$ *derived from (4.41) or equivalently as the marginal p.m.f. of* K' *for the* $p_{K',L'}$*, which is given by*

$$g_{\pi,x}(k) \;=\; \frac{{}_2F_1(a_1+\nu_1+k, a_1+\nu_1+k; a_1; \omega_2)}{F_4(a_1+\nu_1, a_1+\nu_1; a_1, \nu_1; \omega_1, \omega_2)}\;\frac{(a_1+\nu_1)_k\,(a_1+\nu_1)_k}{(\nu_1)_k\,k!}\;\omega_1^k\,\mathbb{I}_{\mathbb{N}}(k)\,. \tag{4.43}$$

Remark 4.5. *The posterior distribution in Corollary 4.1 is conveniently expressed in terms of a G2 mixture distribution with mixing variable* $S = K' + L'$ *based on joint density (or p.m.f.) given in (4.40). The previous example gives a more precise form for a negative binomial mixture. The properties of* S *remain to be explored in general or for other specific choices of* g *and* ϵ*. In terms of the posterior expectation, it follows readily from the above that:*

$$\mathbb{E}(\theta|x) \;=\; \mathbb{E}^{S|x}\left\{\mathbb{E}(\theta|x,S)\right\} \;=\; \left(\frac{a_1+\nu_1+\mathbb{E}(S|x)}{x_1+b_1}\,,\;\frac{a_2+\nu_2+\mathbb{E}(S|x)}{x_2+b_2}\right)\,.$$

In the context of Example 4.11, one obtains with standard manipulations that

$$\mathbb{E}(S|x) \;=\; \mathbb{E}(K') \;+\; \mathbb{E}(L') \;=\; \frac{(a_1\;+\;\nu_1)^2\,(\omega_1 A_2/\nu_1 + \omega_2 A_3/a_1)}{A_1}\;,$$

with $A_1 = F_4(a_1+\nu_1, a_1+\nu_1; a_1, \nu_1; \omega_1, \omega_2)$*,* $A_2 = F_4(a_1+\nu_1+1, a_1+\nu_1+1; a_1, \nu_1+1; \omega_1, \omega_2)$*, and* $A_3 = F_4(a_1+\nu_1+1, a_1+\nu_1+1; a_1+1, \nu_1; \omega_1, \omega_2)$*.*

Appendix

Here are some observations on the generalized McKay distribution $MG(\vec{\gamma})$, with p.d.f. given in (4.34).

A. With the transformation $(Z_1, Z_2) \to (T_1 = Z_1/Z_2, T_2 = Z_2)$, one obtains the density

$$f_{T_1,T_2}(t_1,t_2) \;=\; \frac{1}{C(\vec{\gamma})}\, t_1^{\gamma_1-1}\,(1-t_1)^{\gamma_2-1}\, t_2^{\gamma_1+\gamma_2+\gamma_4-2}\, e^{-t_2(\gamma_3+\gamma_5 t_1)}\; \mathbb{I}_{(0,1)}(t_1)\, \mathbb{I}_{(0,\infty)}(t_2)\,. \tag{4.44}$$

From this, it follows immediately that

$$T_2|T_1 \stackrel{d}{=} \mathrm{G}(\gamma_1+\gamma_2+\gamma_4-1, \gamma_3+\gamma_5 T_1)\,,$$

and then using the value of $C(\vec{\gamma})$ given in (4.35), we obtain the marginal density

$$f_{T_1}(t_1) \;=\; \frac{\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1)\,\Gamma(\gamma_2)}\, \frac{t_1^{\gamma_1-1}\,(1-t_1)^{\gamma_2-1}}{{}_2F_1(\gamma_1, \gamma_1+\gamma_2+\gamma_4-1, \gamma_1+\gamma_2; -\frac{\gamma_5}{\gamma_3})}\, \left(\frac{\gamma_3}{\gamma_3+\gamma_5 t_1}\right)^{\gamma_1+\gamma_2+\gamma_4-1}\; \mathbb{I}_{(0,1)}(t_1)\,.$$

Observe that we have independence between T_1 and T_2 iff $\gamma_5 = 0$, which includes the McKay case, and that $T_1 \stackrel{d}{=} \mathrm{B}(\gamma_1, \gamma_2)$ whenever $\gamma_5 = 0$. We point out that, for $\gamma_4 = 1$, the above density for T_1 matches a generalized Beta distribution given by Chen & Novick (1984) with parameters $\gamma_1, \gamma_2, \lambda = 1 + \frac{\gamma_5}{\gamma_3}$, and p.d.f. $\lambda^{\gamma_1}\, \frac{\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1)\Gamma(\gamma_2)}\, \frac{t_1^{\gamma_1-1}\,(1-t_1)^{\gamma_2-1}}{(1-(1-\lambda)t_1)^{\gamma_1+\gamma_2}}\, \mathbb{I}_{(0,1)}(t_1)$. More generally, the density has appeared in Exton (1976), as well as in Armero & Bayarri (1994) where it is referred to as a Gauss hypergeometric distribution.

B. The given expression for the normalization constant $C(\vec{\gamma})$ can be checked by integrating (4.44) via an expansion of the term $e^{-\gamma_5 t_1 t_2}$ in its MacLaurin series, and interchanging integrals and sum.

C. The marginals densities of Z, which are gamma distributed for the McKay distribution, are given as follows:

$$\begin{aligned} f_{Z_1}(z_1) &= \frac{1}{C(\vec{\gamma})} \int_{z_1}^{\infty} f_{Z_1,Z_2}(z_1,z_2)\, dz_2 \\ &= \frac{1}{C(\vec{\gamma})}\, z_1^{\gamma_1-1}\, e^{-\gamma_5 z_1} \int_0^{\infty} u^{\gamma_2-1}\,(u+z_1)^{\gamma_4-1}\, e^{-\gamma_3(z_1+u)}\, du \\ &= \frac{\Gamma(\gamma_2)}{C(\vec{\gamma})}\, z_1^{\gamma_1+\gamma_2+\gamma_4-1}\, e^{-(\gamma_5+\gamma_3)\, z_1}\; \psi(\gamma_2, \gamma_2+\gamma_4, \gamma_3\, z_1)\,, \end{aligned}$$

for $z_1 > 0$, using the identity $\int_0^\infty t^{a-1}\,(\Delta+t)^{b-a-1}\, e^{-ct}\, dt \;=\; \Delta^{b-1}\,\Gamma(a)\,\psi(a,b,c\Delta)$ which holds for $a, c > 0$ and $b \in \mathbb{R}$. Observe that $Z_1 \stackrel{d}{=} \mathrm{G}(\gamma_1, \gamma_3+\gamma_5)$ for $\gamma_4 = 1$, as $\psi(a, a+1, c) \;=\; c^{-a}$ for $a, c > 0$, and that the additional specification $\gamma_5 = 0$ leads us back to the McKay case.

For the marginal density of Z_2, we have

$$\begin{aligned} f_{Z_2}(z_2) &= \frac{1}{C(\vec{\gamma})}\, z_2^{\gamma_4-1}\, e^{-\gamma_3 z_2} \int_0^{z_2} z_1^{\gamma_1-1}\,(z_2-z_1)^{\gamma_2-1}\, e^{-\gamma_5 z_1}\, dz_1 \\ &= \frac{1}{C(\vec{\gamma})}\, z_2^{\gamma_1+\gamma_2+\gamma_4-1}\, e^{-\gamma_3 z_2} \int_0^1 u^{\gamma_1-1}\,(1-u)^{\gamma_2-1}\, e^{-\gamma_5 z_2 u}\, du \\ &= \frac{z_2^{\gamma_1+\gamma_2+\gamma_4-1}\, e^{-\gamma_3 z_2}}{\Gamma(\gamma_1+\gamma_2+\gamma_4-1)}\, \frac{\gamma_3^{\gamma_1+\gamma_2+\gamma_4-1}\; {}_1F_1\,(\gamma_1; \gamma_1+\gamma_2; -\gamma_5 z_2)}{{}_2F_1\left(\gamma_1, \gamma_1+\gamma_2+\gamma_4-1; \gamma_1+\gamma_2; -\frac{\gamma_5}{\gamma_3}\right)}\,, \end{aligned}$$

for $z_2 > 0$, and with the change of variables $z_1 = z_2 u$, as well as the confluent hypergeometric identity $\int_0^1 u^{a-1}(1-u)^{b-1}\, e^{-cu}\, du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\, {}_1F_1(a; a+b; -c)$, $a, b, c > 0$. Observe that $Z_2 \stackrel{d}{=} \mathrm{G}(\gamma_1 + \gamma_2 + \gamma_4 - 1, \gamma_3)$ for $\gamma_5 = 0$, while $f_{Z_2}(z_2) = \frac{(\gamma_3^{\gamma_2})\,(\gamma_3+\gamma_5)^{\gamma_1}}{\Gamma(\gamma_1+\gamma_2)}\, z_2^{\gamma_1+\gamma_2}\, e^{-\gamma_3 z_2}\, {}_1F_1(\gamma_1, \gamma_1 + \gamma_2; -\gamma_5 z_2)\, \mathbb{I}_{(0,\infty)}(z_2)$ for $\gamma_4 = 1$ using the identity ${}_1F_0(a; -; -c) = (1+c)^{-a}$ for $a > 0$ and $c \in (-1, 1)$.

D. The moments are readily expressible in terms of the constant of normalization $C(\vec{\gamma})$. For instance, it is easy to obtain the mixed moments

$$\mathbb{E}(Z_1^m\, Z_2^n) = \frac{(\gamma_1)_m\,(\gamma_2)_n\,(\gamma_1+\gamma_2+\gamma_4-1)_s}{(\gamma_1+\gamma_2)_s\,\gamma_3^s}\, \frac{{}_2F_1(\gamma_1+m, \gamma_1+\gamma_2+\gamma_4+s-1, \gamma_1+\gamma_2+s; -\frac{\gamma_5}{\gamma_3})}{{}_2F_1(\gamma_1, \gamma_1+\gamma_2+\gamma_4-1, \gamma_1+\gamma_2; -\frac{\gamma_5}{\gamma_3})}\,,$$

for $m, n \in \mathbb{N}$ and $s = m + n$.

E. For McKay distributed (Z_1, Z_2), the variables $U_1 = Z_1$ and $U_2 = Z_2 - Z_1$ are independently distributed. For the general case, it is easy to verify that we obtain conditional distributions that are Kummer type II, namely:

$$U_2|U_1 \stackrel{d}{=} \mathrm{K2}(\gamma_2\,,\, 1-\gamma_2-\gamma_4\,,\, \gamma_3\, U_1\,,\, U_1) \text{ and } U_1|U_2 \stackrel{d}{=} \mathrm{K2}(\gamma_1\,,\, 1-\gamma_2-\gamma_4\,,\, (\gamma_3+\gamma_5)\, U_2\,,\, U_2)\,,$$

with independence iff $\gamma_4 = 0$.

Concluding remarks

For two common types of mixtures, we have provided analytical, novel and practical representations of Bayesian posterior and predictive distributions corresponding to a given prior distribution. By doing so, we have added to the catalogue of such Bayesian solutions which require little computation except recourse to some special but familiar functions and distributions. Specific model and prior distribution choices have been put to the forefront for illustrative purposes and for their practicality. These include inference for common non-central distributions (chi-square, Beta, Fisher), the distribution of the coefficient of determination R^2 in multiple regression models, mixtures of normal and exponential distributions, and bivariate gamma mixtures including the Kibble distribution.

Acknowledgements

Éric Marchand gratefully acknowledges the research support from the Natural Sciences and Engineering Research Council of Canada.

References

[1] Aitchison, J. & Dunsmore, I.R. (1975). Statistical Prediction Analysis. Cambridge University Press.

[2] Armero, C. & Bayarri, M. J. (1994). Prior assessments for prediction in queues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **43**, 139–153.

[3] Amaral, M.A. & Dunsmore, I.R. (1980). Optimal estimates of predictive distributions. *Biometrika*, **67**, 685-689.

[4] Barndorff-Nielsen, O.E., Kent, J. & Sørensen, M. (1982). Normal variance-mean mixtures and z-distributions. *International Statistical Review*, **50**, 145–159

[5] Bulgren, W. G. (1971). On representations of the doubly non-central F distribution. *Journal of the American Statistical Association*, **66**, 184–186.

[6] Chen, J.J. & Novick, M.R. (1984). Bayesian analysis for binomial models with generalized Beta prior distributions. *Journal of Educational and Behavioral Statistics*, **9**, 163–175.

[7] Exton, H. (1976). Multiple Hypergeometric Functions and Applications. Horwood.

[8] Hamza, M. & Vallois, P. (2016). On Kummer's distribution of type two and a generalized beta distribution. *Statistics & Probability Letters*, **118**, 60–69.

[9] Iliopoulos, G., Karlis, D. & Ntzoufras, I. (2005). Bayesian estimation in Kibble's bivariate gamma distribution. *Canadian Journal of Statistics*, **33**, 571-589.

[10] Johnson, N.L., Kotz, S. & Balakrishnan, N. (2000). Continuous multivariate distributions, *Volume 1: Models and Applications*, 2nd edition, John Wiley & Sons, New York.

[11] Johnson, N.L., Kotz, S. & Kemp, A. (1995). Discrete univariate distributions, 2nd edition, John Wiley & Sons, New York.

[12] Jones, M. C. & Marchand, É. (2019). Multivariate discrete distributions via sums and shares. *Journal of Multivariate Analysis*, **171**, 83-93.

[13] Kibble, W. F. (1941). A two-variate gamma-type distribution. *Sankhya*, **5**, 137–150.

[14] Kubokawa, T., Marchand, É. & Strawderman, W.E. (2017). A unified approach to estimation of noncentrality parameters, the multiple correlation coefficient, and mixture models. *Mathematical Methods of Statistics*, **26**, 134-148.

[15] Kubokawa, T., Marchand, É. & Strawderman, W.E. (2015). On predictive density estimation for location families under integrated squared error loss. *Journal of Multivariate Analysis*, **142**, 57–74.

[16] Landriault, D., Moutanabbir, K. & Willmot, G. (2015). A note on order statistics in the mixed Erlang case. *Statistics & Probability Letters*, **106**, 13–18.

[17] Laurent, S. (2012). Some Poisson mixtures distributions with a hyperscale parameter. *Brazilian Journal of Probability and Statistics*, **26**, 265–278.

[18] Lindley, D. & Singpurwalla (1986). Multivariate distributions for the life lengths of a system sharing a common environment. *Journal of Applied Probability*, **23**, 418–431.

[19] Marchand, É. (2001). Point estimation of the coefficient of determination. *Statistics & Decisions*, **19**, 137–154.

[20] McKay, A.T. (1934). Sampling from batches. *Supplement to the Journal of the Royal Statistical Society*, **1**, 207–216.

[21] Muirhead, R.J. (1982). Aspects of multivariate statistical theory. Wiley, New York.

[22] Nayak, T.P. (1987). Multivariate Lomax Distribution: Properties and Usefulness in Reliability Theory. *Journal of Applied Probability*, **24**, 170–177.

[23] Ong, S. H. & Lee, P. A, (1979). The noncentral negative binomial distribution. *Biometrical Journal*, **7**, 611–627.

[24] Patnaik, P.B. (1949). The noncentral chi-square and F distributions and their applications. *Biometrika*, **36**, 202–232.

[25] Petropoulos, C. (2017). Estimation of the order restricted scale parameters for two populations from the Lomax distribution. *Metrika*, **80**, 483–502.

[26] Saxena, K.M. & Alam, K. (1982). Estimation of the non-centrality parameter of a chi-squared distribution. *Annals of Statistics*, **10**, 1012–1016.

[27] Tiwari, R., Chib, S. & Jammalamadaka, S. R. (1989). Bayes estimation of the multiple correlation coefficient. *Communications in Statistics: Theory and Methods*, **18**, 1401–1413.

[28] Wicksell, S.D. (1933). On correlation functions of type III. *Biometrika*, **25**, 121–133.