Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 3, 265–278
DOI: 10.1214/11-BJPS139

Some Poisson mixtures distributions with a hyperscale parameter

Stéphane Laurent

Université catholique de Louvain

Abstract. We mainly investigate certain mixtures of Poisson distributions with a scale parameter in the mixing distribution. They help us to derive the bivariate Poisson mixtures arising from the prior and posterior predictive distributions in the semi-conjugate family defined by Laurent and Legrand (*ESAIM Probab. Stat.* (2011) DOI:10.1051/ps/2010018) for the "two Poisson samples" model, which contains in particular the reference prior when the parameter of interest is the ratio of the two Poisson rates.

1 Introduction

We firstly define certain families of univariate mixtures of Poisson distributions whose probability mass functions and probability generating functions involve the Gauss hypergeometric function or the Appell hypergeometric function in the more general case. Two interesting families are obtained by extending a family of Poisson mixtures by adding a scale parameter in the family of mixing distributions. We get in particular the "hyperscaled" Beta-negative binomial distributions. The "hyperscaling" on a Poisson mixture acts on its probability generating function by a linear change of variables; consequently the factorial moments of the Poisson mixtures we define have a simple expression. Actually all univariate Poisson mixtures we define are obtained by mixing some negative binomial distributions on their proportion parameter with a distribution defined from a Beta distribution or, in the more general case, with the distribution we name *Beta distribution of the third kind*. Except for this case, these Poisson mixtures can be straightforwardly simulated with any standard statistical software. These investigations are the object of Section 2.

In Section 3, the univariate Poisson mixtures help us to derive the predictive distributions corresponding to the semi-conjugate family of priors defined by Laurent and Legrand (2011) for the "two Poisson samples" model, and also the prior (but not the posterior) predictive distributions for a larger family we define (by adding a scale parameter for the prior on the ratio of the two Poisson rates). In particular, we get the posterior predictive distributions corresponding to the Berger &

Key words and phrases. Bayesian model, hypergeometric functions, Jeffreys prior, Poisson mixtures distributions, probability generating function, reference prior, semi-conjugate family of priors.
Received July 2010; accepted January 2011.

Bernardo reference prior for the ratio of the Poisson rates. Of course, the probabilistic results we give on these bivariate Poisson mixtures could be directly derived by naive calculations, but they are more easily and intuitively derived and more clearly expressed with the help of the univariate Poisson mixtures introduced in Section 2. As a by-product, we get a family of priors for the "one Poisson sample" model whose prior predictive distributions form the Beta-negative binomial family, thereby allowing more dispersion in the prior distributions as compared to the conjugate Gamma family.

We display below the main no(ta)tions and conventions of the paper.

Pochhammer symbol and hypergeometric functions. We use the Pochhammer symbol $(a)_n := a(a+1)\cdots(a+n-1)$ for ascending factorials, with $(a)_0 = 1$ by the empty product convention. Denoting by B the Beta function, the equality

$$B(c,d)(c)_m(d)_n = B(c+m, d+n)(c+d)_{m+n} \tag{1.1}$$

holds for all integers $m, n \geq 0$ and all real numbers $c, d > 0$. The *Gauss hypergeometric function* (see Slater, 1966) is, as usual, denoted by ${}_2F_1$. For complex parameters α, β, $\gamma \notin -\mathbb{N}$ and complex variable x with $|x| < 1$, it is defined as the sum of the absolute convergent series

$${}_2F_1(\alpha, \beta, \gamma; x) = \sum_{n=0}^{\infty} \frac{(\alpha)_n(\beta)_n}{(\gamma)_n} \frac{x^n}{n!},$$

and whenever $\Re(\alpha) > 0$ and $\Re(\gamma - \alpha) > 0$, its analytical continuation in the complex plane with the cut along $(1, +\infty)$ is given by the *Euler integral representation*:

$$\begin{aligned} B(\alpha, \gamma-\alpha){}_2F_1(\alpha; \beta, \gamma; x) &= \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-ux)^{-\beta}\,\mathrm{d}u \\ &= \int_0^{+\infty} \frac{z^{\alpha-1}(1+z)^{\beta-\gamma}}{(1+(1-x)z)^{\beta}}\,\mathrm{d}z, \end{aligned} \tag{1.2}$$

and it satisfies the relations

$$\begin{aligned} {}_2F_1(\alpha, \beta, \gamma; 1-y) &= y^{-\beta}{}_2F_1\left(\gamma-\alpha, \beta, \gamma; 1-\frac{1}{y}\right) \\ &= y^{-\alpha}{}_2F_1\left(\alpha, \gamma-\beta, \gamma; 1-\frac{1}{y}\right). \end{aligned} \tag{1.3}$$

The *Appell first hypergeometric function* (see Slater, 1966), shortly termed as Appell hypergeometric function in the present paper, is denoted by F_1. For complex parameters α, β, β', γ and complex variables x and y, it is defined for $|x| < 1$ and $|y| < 1$ as the sum of the absolute convergent double series

$$F_1(\alpha, \beta, \beta', \gamma; x, y) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(\alpha)_{m+n}(\beta)_m(\beta')_n}{(\gamma)_{m+n}} \frac{x^m y^n}{m!n!}.$$

Thus,

$$F_1(\alpha,\beta,\beta',\gamma;x,y)=\sum_{m=0}^{\infty}\frac{(\alpha)_m(\beta)_m}{(\gamma)_m}\,{}_2F_1(\alpha+m,\beta',\gamma+m;y)\frac{x^m}{m!}. \tag{1.4}$$

For $\Re(\alpha)>0$ and $\Re(\gamma-\alpha)>0$, the analytical continuation of the Appell hypergeometric function on $\{\Re(x),\Re(y)<1\}$ is given by the following so-called *Picard integral representation*:

$$\begin{aligned}
&B(\alpha,\gamma-\alpha)F_1(\alpha,\beta,\beta',\gamma;x,y)\\
&\quad=\int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-ux)^{-\beta}(1-uy)^{-\beta'}\,\mathrm{d}u \qquad (1.5)\\
&\quad=\int_0^{+\infty}\frac{z^{\alpha-1}(1+z)^{\beta+\beta'-\gamma}}{(1+(1-x)z)^{\beta}(1+(1-y)z)^{\beta'}}\,\mathrm{d}z.
\end{aligned}$$

Beta distributions of second and third kinds. The *Beta prime* distribution $\mathcal{B}'(c,d)$ with positive parameters c and d is defined as the distribution of the random variable $\psi:=\theta(1-\theta)^{-1}$ where the random variable θ has the usual Beta distribution $\mathcal{B}(c,d)$. Its density function is

$$\mathcal{B}'(\psi\mid c,d)=\frac{1}{B(c,d)}\frac{\psi^{c-1}}{(1+\psi)^{c+d}},\qquad \psi\geq 0. \tag{1.6}$$

Note that the distribution of ψ^{-1} is then $\mathcal{B}'(d,c)$, what we symbolically write as $\mathcal{B}'(c,d)^{-1}=\mathcal{B}'(d,c)$. The *Beta distribution of the second kind* $\mathcal{B}_2(c,d,\tau)$ with positive parameters c, d and τ is symbolically defined by $\mathcal{B}_2(c,d,\tau)=\tau\mathcal{B}'(c,d)$, which rigorously means, with the notations above, that this is the distribution of the random variable $\phi:=\tau\psi$. Its density function is then

$$\mathcal{B}_2(\phi\mid c,d,\tau)=\frac{\tau^{-c}}{B(c,d)}\frac{\phi^{c-1}}{(1+\phi/\tau)^{c+d}},\qquad \phi\geq 0. \tag{1.7}$$

Owing to $\mathcal{B}'(c,d)^{-1}=\mathcal{B}'(d,c)$, we obviously have $\mathcal{B}_2(c,d,\tau)^{-1}=\mathcal{B}_2(d,c,\tau^{-1})$. Kleiber and Kotz (2003) provide many details on the Beta distributions of the second kind. We finally define the *Beta distribution of the third kind* $\mathcal{B}_3(c,d,\kappa,\tau)$ with parameters $c>0$, $d>0$, $\tau>0$ and $\kappa\in\mathbb{R}$ as the distribution whose density function is

$$\mathcal{B}_3(\phi\mid c,d,\kappa,\tau)=\frac{1}{C_{c,d,\kappa,\tau}B(c,d)}\frac{\phi^{c-1}(1+\phi)^{-\kappa}}{(1+\phi/\tau)^{c+d-\kappa}},\qquad \phi\geq 0, \tag{1.8}$$

where, from (1.2) and (1.3),

$$\begin{aligned}
C_{c,d,\kappa,\tau}&={}_2F_1\left(c,c+d-\kappa,c+d;1-\frac{1}{\tau}\right)\\
&=\tau^{c+d-\kappa}{}_2F_1(d,c+d-\kappa,c+d;1-\tau).
\end{aligned}$$

It is easy to check that $\mathcal{B}_3(c,d,\kappa,\tau)^{-1} = \mathcal{B}_3(d,c,\kappa,\tau^{-1})$. With the terminology of Chen and Novick (1984), if $\phi \sim \mathcal{B}_3(c,d,\kappa,\tau)$, then the distribution of $\theta := \phi(1+\phi)^{-1}$ is a *four-parameter generalized Beta distribution*. It appears as the posterior distribution on the proportion parameter of the Bayesian binomial model whose prior on the odds parameter is a Beta distribution of the second kind.

Hyperscaled Poisson mixtures. When $\mathcal{P}\gamma$ denotes the mixture of the Poisson distributions $\mathcal{P}(\theta)$ with $\theta \sim \gamma(\mathrm{d}\theta)$ for some probability distribution γ on $(0,+\infty)$, then for a given $T > 0$, we denote by $T \star \mathcal{P}\gamma$ the mixture of Poisson distributions $\mathcal{P}(\theta T)$ with $\theta \sim \gamma(\mathrm{d}\theta)$. For instance, consider the well-known *Poisson–Gamma distribution* $\mathcal{PG}(a,b)$ with parameters $a, b > 0$ which is defined as the Poisson mixture with a Gamma distribution $\mathcal{G}(a,b)$ with rate parameter b, hence the hyperscaling on the Poisson–Gamma distribution $\mathcal{PG}(a,b)$ acts by division on b: one has $T \star \mathcal{PG}(a,b) = \mathcal{PG}(a,b/T)$. The probability mass function of $T \star \mathcal{PG}(a,b)$ is given by

$$T \star \mathcal{PG}(x \mid a,b) = \frac{(a)_x}{x!} \frac{b^a T^x}{(b+T)^{a+z}}, \qquad x \in \mathbb{N}, \tag{1.9}$$

where $\mathbb{N} = \{0, 1, \ldots\}$.

We generically use the letter “G” for denoting probability generating functions (p.g.f.). Knowing that $G_{\mathcal{P}(\theta)}(s) = e^{\theta(s-1)}$, we have the following equality

$$G_{\mathcal{P}\gamma}(s) = \int e^{\theta(s-1)} \gamma(\mathrm{d}\theta), \tag{1.10}$$

thereby giving an analytical continuation of $G_{\mathcal{P}\gamma}$ for $\Re(s) < 1$. Throughout this paper, it will be understood that we always consider this analytical continuation of the p.g.f. for any Poisson mixture distribution. The p.g.f. of the hyperscaled Poisson mixture $T \star \mathcal{P}\gamma$ is then given by the following elementary linear change of variables on the p.g.f. of $\mathcal{P}\gamma$:

$$G_{T\star\mathcal{P}\gamma}(s) = G_{\mathcal{P}\gamma}\big(1 - T(1-s)\big). \tag{1.11}$$

Consequently the nth factorial moment of $T \star \mathcal{P}\gamma$ equals T^n multiplied by the nth factorial moment of $\mathcal{P}\gamma$. Note also that (1.10) provides a link between $G_{\mathcal{P}\gamma}$ and the Laplace transform of γ, from what we can deduce that $\mathcal{P}\gamma$ uniquely determines γ. We will use the following expression of the p.g.f. of the Poisson–Gamma distribution $\mathcal{PG}(a,b)$:

$$G_{\mathcal{PG}(a,b)}(s) = \frac{b^a}{(1-s+b)^a}. \tag{1.12}$$

2 Some univariate Poisson mixtures

We shall define some families of Poisson mixtures (in fact, Poisson–Gamma mixtures). For each of them we provide the probability mass functions and the probability generating functions. The first family $\mathcal{PGB}_3$ of *Poisson–Gamma–Beta distributions of the third kind* contains the subfamily $\mathcal{PGB}_2$ of *Poisson–Gamma–Beta* *distributions of the second kind* which are nothing but the "hyperscaled" Beta-negative binomial distributions. We next define the family $\mathcal{PGIB}$ of *Poisson–Gamma-inverse Beta distributions* and its extension with a hyperscale parameter.

Poisson–Gamma–Beta distributions of the third kind. In view of the expression (1.9) of the probability masses of the Poisson–Gamma distribution, and the expression (1.8) of the density of the Beta distribution of the third kind, lemma below straightforwardly follows from Bayes' formula.

Lemma 2.1. *Let a, c, d, τ be positive numbers and κ a real number. Let ψ and x be random variables such that*

$$\psi \sim \mathcal{B}_3(c, d, \kappa, \tau) \quad \textit{and} \quad (x \mid \psi) \sim \mathcal{PG}(a, \psi).$$

Then $(\psi \mid x) \sim \mathcal{B}_3(c + a, d + x, \kappa + a + x, \tau)$.

We then define the *Poisson–Gamma–Beta distribution of the third kind* $\mathcal{PGB}_3(a, c, d, \kappa, \tau)$ with parameters $a, c, d, \tau > 0$ and $\kappa \in \mathbb{R}$ as the absolute distribution of x in lemma above. The probability mass it assigns at $x \in \mathbb{N}$ is then

$$\begin{aligned} &\mathcal{PGB}_3(x \mid a, c, d, \kappa, \tau) \\ &\quad = \frac{(a)_x}{x!} \frac{\mathcal{B}(c + a, d + x)}{\mathcal{B}(c, d)} \\ &\qquad \times \frac{{}_2F_1(c + a, c + d - \kappa, c + d + a + x; 1 - 1/\tau)}{{}_2F_1(c, c + d - \kappa, c + d; 1 - 1/\tau)}. \end{aligned} \tag{2.1}$$

In the particular case when $\kappa = 0$, we call this distribution the *Poisson–Gamma–Beta distribution of the second kind* and denote it by $\mathcal{PGB}_2(a, c, d, \tau)$. Its probability masses are then given by

$$\begin{aligned} &\mathcal{PGB}_2(x \mid a, c, d, \tau) \\ &\quad = \frac{(a)_x}{x!} \frac{\mathcal{B}(c + a, d + x)}{\mathcal{B}(c, d)} \\ &\qquad \times \tau^{-c} {}_2F_1\left(c + a, c + d, c + d + a + x; 1 - \frac{1}{\tau}\right), \qquad x \in \mathbb{N}. \end{aligned} \tag{2.2}$$

In case when $\tau = 1$, the Poisson–Gamma–Beta distribution of the second kind $\mathcal{PGB}_2(a, c, d, \tau)$ reduces to the well-known *Beta-negative binomial distribution* which we also call *Poisson–Gamma–Beta prime distribution* and we denote it by $\mathcal{PGB}'(a, c, d)$. Its probability masses are given by

$$\mathcal{PGB}'(x \mid a, c, d) = \frac{(a)_x}{x!} \frac{\mathcal{B}(c + a, d + x)}{\mathcal{B}(c, d)}, \qquad x \in \mathbb{N}.$$

This distribution is also known as a *type IV general hypergeometric distribution* and also named *generalized Waring distribution* (see Johnson, Kemp and Kotz, 2005). We notice that, with the notations given in the Introduction, the hyperscaling acts by division on the fourth parameter τ of $\mathcal{PGB}_2(a,c,d,\tau)$, that is:

$$T \star \mathcal{PGB}_2(a,c,d,\tau) = \mathcal{PGB}_2(a,c,d,\tau/T).$$

In particular $\mathcal{PGB}_2(a,c,d,\tau) = \tau^{-1} \star \mathcal{PGB}'(a,c,d)$, thus the $\mathcal{PGB}_2$ family is nothing but the family of hyperscaled Beta-negative binomial distributions.

Result 2.1. *The p.g.f.s in the $\mathcal{PGB}_3$ family are given by the following expressions.*

$$\begin{aligned}
&G_{\mathcal{PGB}'(a,c,d)}(s) \\
&\quad = \frac{B(c+a,d)}{B(c,d)}(1-s)^c {}_2F_1(c+a, c+d, c+d+a; s), \\
&G_{\mathcal{PGB}_2(a,c,d,\tau)}(s) \\
&\quad = \frac{B(c+a,d)}{B(c,d)}\left(\frac{1-s}{\tau}\right)^c \\
&\qquad \times {}_2F_1\left(c+a, c+d, c+d+a; 1-\frac{1-s}{\tau}\right), \\
&G_{\mathcal{PGB}_3(a,c,d,\kappa,\tau)}(s) \\
&\quad = \frac{B(c+a,d)}{B(c,d)}\frac{1}{(1-s)^a} \\
&\qquad \times \frac{F_1(c+a, a, c+d-\kappa, c+d+a; 1-1/(1-s), 1-1/\tau)}{{}_2F_1(c, c+d-\kappa, c+d; 1-1/\tau)}.
\end{aligned}$$

Proof. Using the expression (1.12) of $G_{\mathcal{PG}(a,\psi)}(s)$ and the expression (1.6) of $\mathcal{B}'(\psi \mid c,d)$, the integral representation (1.2) of ${}_2F_1$ straightforwardly yields

$$G_{\mathcal{PGB}'(a,c,d)}(s) = (1-s)^{-a}\frac{B(c+a,d)}{B(c,d)} {}_2F_1\left(c+a, a, c+d+a; 1-\frac{1}{1-s}\right),$$

which equals the announced expression due to identity (1.3). In the same way, the announced expression of $G_{\mathcal{PGB}_3(a,c,d,\kappa,\tau)}(s)$ is derived from the expression of $\mathcal{B}_3(\psi \mid c,d,\kappa,\tau)$ given by (1.8), and the integral representation (1.5) of the Appell hypergeometric function. The p.g.f. of $\mathcal{PGB}_2(a,c,d,\tau)$ is obtained from the p.g.f. of $\mathcal{PGB}'(a,c,d)$ with the help of the linear change of variables (1.11). □

Poisson–Gamma-inverse Beta distributions. Another straightforward application of Bayes' formula yields the following lemma.

Lemma 2.2. *Let a, c, d, τ be positive numbers. Let ψ and x be random variables such that*

$$\psi \sim \mathcal{B}'(c,d) \quad \textit{and} \quad (x \mid \psi) \sim \tau \star \mathcal{PG}(a, 1+\psi).$$

Then $(\psi \mid x) \sim \mathcal{B}_3(c, d+x, c+d-a, \tau+1)$.

In the lemma above, it is easy to see that $(1+\psi)^{-1}$ is distributed according to $\mathcal{B}(d,c)$. We then call *Poisson–Gamma-inverse Beta distribution* with parameters $a,c,d>0$ the absolute distribution of x in lemma above when $\tau=1$ and we denote it by $\mathcal{PGIB}(a,c,d)$. For an arbitrary $\tau>0$, the absolute distribution of x is then $\tau\star\mathcal{PGIB}(a,c,d)$ which we do not name. Using formula (1.1), we obtain that the probability mass of $\tau\star\mathcal{PGIB}(a,c,d)$ at $x\in\mathbb{N}$ is

$$\tau\star\mathcal{PGIB}(x\mid a,c,d)=\frac{\tau^x}{x!}\frac{(a)_x(d)_x}{(c+d)_x}{}_2F_1(d+x,a+x,c+d+x;-\tau). \tag{2.3}$$

In particular, one has $\tau\star\mathcal{PGIB}(c+d,c,d)=\tau\star\mathcal{PG}(d,1)$.

Result 2.2. *The probability generating function of $\tau\star\mathcal{PGIB}(a,c,d)$ is given by*

$$G_{\tau\star\mathcal{PGIB}(a,c,d)}(s)={}_2F_1\big(d,a,c+d;-\tau(1-s)\big).$$

Consequently the nth factorial moment of $\tau\star\mathcal{PGIB}(a,c,d)$ is $\tau^n\frac{(d)_n(a)_n}{(c+d)_n}$.

Proof. We know from (1.12) that the probability generating function of $\mathcal{PG}(a,1+\psi)$ is given by

$$G_{\mathcal{PG}(a,1+\psi)}(s)=(1-s)^{-a}\frac{(1+\psi)^a}{(1+(1+\psi)/(1-s))^a}=(2-s)^{-a}\frac{(1+\psi)^a}{(1+\psi/(2-s))^a}.$$

Hence, using the expression (1.6) of $\mathcal{B}'(\psi\mid c,d)$, the integral representation (1.2) and the identity (1.3) for ${}_2F_1$, we obtain

$$G_{\mathcal{PGIB}(a,c,d)}(s)=(2-s)^{-a}{}_2F_1\left(c,a,c+d;\frac{1-s}{2-s}\right)={}_2F_1(d,a,c+d;s-1).$$

The expression for $G_{\tau\star\mathcal{PGIB}(a,c,d)}(s)$ follows from the linear change of variables (1.11). The factorial moments derive from the equality

$$\frac{\mathrm{d}}{\mathrm{d}z}{}_2F_1(\alpha,\beta,\gamma;z)=\frac{\alpha\beta}{\gamma}{}_2F_1(\alpha+1,\beta+1,\gamma+1;z),$$

which easily follows from the power series representation of ${}_2F_1$. □

Using the factorial moments, we get that the mean of $\tau\star\mathcal{PGIB}(a,c,d)$ is $\tau ad(c+d)^{-1}$ and its variance is

$$\tau ad\left[\frac{1}{c+d}+\tau\frac{c(a+d+1)+d(d+1)}{(c+d)^2(c+d+1)}\right].$$

3 Some bivariate Poisson mixtures

Laurent and Legrand (2011) defined a natural semi-conjugate family of priors for the "two Poisson samples" model, which contains the Berger & Bernardo reference prior for the ratio of the two Poisson rates. The results we give in this section provide the prior and posterior predictive distributions for this family of priors, and also the prior predictive distributions for the larger family of priors defined in Lemma 3.1.

Throughout this section, we consider the statistical model given by two independent observations $x \sim \mathcal{P}(\lambda S)$ and $y \sim \mathcal{P}(\mu T)$ with unknown incidence rates λ, μ, and fixed "observation-opportunity sizes," or "sample sizes," S and T, and we denote by $\phi := \lambda/\mu$ the so-called relative risk. When μ and ϕ have independent prior distributions with $\mu \sim \mathcal{G}(a, b)$, then, as shown by Laurent and Legrand (2011), the conditional joint prior predictive distribution of (x, y) given ϕ is the bivariate Poisson–Gamma distribution having the marginal-conditional factorization

$$(y \mid \phi) \sim T \star \mathcal{PG}(a, b) \quad \text{and} \quad (x \mid y, \phi) \sim \phi S \star \mathcal{PG}(a + y, b + T).$$

Since the distribution of y does not depend on ϕ, Lemma 2.1 straightforwardly yields the following lemma.

Lemma 3.1. *For any positive numbers a, b, c, d, if the joint prior of (μ, ϕ) is defined by*

$$(\mu \mid \phi) \sim \mathcal{G}(a, b) \quad \textit{and} \quad \phi \sim \mathcal{B}_2(c, d, \rho),$$

then the joint posterior on (μ, ϕ) is given by

$$(\mu \mid \phi, x, y) \sim \mathcal{G}(a + x + y, b + \phi S + T)$$

and

$$(\phi \mid x, y) \sim \frac{T + b}{S} \times \mathcal{B}_3\left(c + x, d + a + y, a + x + y, \rho \frac{S}{b + T}\right).$$

The semi-conjugate family defined by Laurent and Legrand (2011) is the case when $\rho = (T + b)/S$, that is, when ϕ has the $\mathcal{B}_2(c + x, d + a + y, \rho)$ posterior distribution. Hereafter, we will simply call it *the semi-conjugate family*. More particularly, the case when $\rho = (T + b)/S$ and $a = c = 1/2$ and $b = d = 0$ corresponds to the Berger & Bernardo reference prior when ϕ is considered to be the parameter of interest (hereafter called the *ϕ-reference prior*). We refer to Bernardo (2005) for a review on reference priors.

Bailey distribution. The posterior predictive distributions will involve the bivariate discrete distribution we define now and we call it the *Bailey distribution* for the following reason. Bailey (1935) gave the following reduction formula:

$$
\begin{aligned}
F_1(\alpha, \beta, \beta', \beta+\beta'; 1-x, 1-y) &= x^{-\alpha}{}_2F_1\left(\alpha, \beta', \beta+\beta'; 1-\frac{y}{x}\right) \\
&= y^{-\alpha}{}_2F_1\left(\alpha, \beta, \beta+\beta'; 1-\frac{x}{y}\right),
\end{aligned}
\tag{3.1}
$$

as well as a reduction formula for $F_1(\alpha, \beta, \beta', \gamma; z, z)$. Note that (3.1) is a consequence of Result 2.1 by taking $\kappa = 0$. The following elementary equality is then nothing but a particular case of one or the other of these two reductions:

$$
\sum_{x=0}^{\infty}\sum_{y=0}^{\infty} \frac{(\alpha)_{x+y}(\beta)_x(\beta')_y}{x!y!(\beta+\beta')_{x+y}} z^{x+y} = (1-z)^{-\alpha}, \qquad |z| < 1.
$$

Then we define the bivariate *Bailey distribution* $\mathcal{B}ailey(a, c, d, \rho)$ with parameters $a, c, d, \rho > 0$ as the probability distribution on $\mathbb{N}^2$ whose probability masses are given by

$$
\begin{aligned}
\mathcal{B}ailey(x, y \mid a, c, d, \rho) = {}& \left(\frac{\rho}{1+\rho}\right)^a \frac{(a)_{x+y}(c)_x(d)_y}{x!y!(c+d)_{x+y}} \\
& \times \left(\frac{1}{1+\rho}\right)^{x+y}, \qquad x, y \in \mathbb{N}.
\end{aligned}
\tag{3.2}
$$

By the double power series representation of F_1, we easily see that the p.g.f. of $\mathcal{B}ailey(a, c, d, \rho)$ is given by

$$
G_{\mathcal{B}ailey(a,c,d,\rho)}(u, v) = (1-\theta)^a F_1(a, c, d, c+d; \theta u, \theta v) \qquad \text{with } \theta = (1+\rho)^{-1},
$$

which, from equality (3.1), reduces to

$$
G_{\mathcal{B}ailey(a,c,d,\rho)}(u, v) = \left(\frac{\rho}{1-v+\rho}\right)^a {}_2F_1\left(a, c, c+d, 1-\frac{1-u+\rho}{1-v+\rho}\right). \tag{3.3}
$$

As a by-product of our derivation of the posterior predictive distributions for the semi-conjugate family, we will see (in Result 3.2) that *the Bailey distribution $\mathcal{B}ailey(a, c, d, \rho)$ is a bivariate Poisson mixture whose first and second margins are $\rho^{-1} \star \mathcal{PGIB}(a, d, c)$ and $\rho^{-1} \star \mathcal{PGIB}(a, c, d)$*, respectively, and we will see that *the scalar hyperscaling acts on the fourth parameter ρ by division*, that is:

$$
(\tau M, \tau N) \star \mathcal{B}ailey(a, c, d, \rho) = (M, N) \star \mathcal{B}ailey(a, c, d, \rho/\tau), \tag{3.4}
$$

where we have extended in a obvious way our hyperscaling notation "$\star$" for bivariate Poisson mixtures. We note the following analogous of the linear change of variables (1.11) for bivariate Poisson mixtures:

$$
G_{(T,T')\star(\mathcal{P}\gamma\otimes\mathcal{P}\gamma')}(u, v) = G_{\mathcal{P}\gamma\otimes\mathcal{P}\gamma'}\big(1-T(1-u), 1-T'(1-v)\big). \tag{3.5}
$$

Predictive distributions for the semi-conjugate family. Our next Results 3.1 and 3.2 will straightforwardly yield that the predictive distributions corresponding to the semi-conjugate family are the following ones.

- The marginal prior predictive distributions are

$$x \sim (T+b) \star \mathcal{PGB}_2(a,d,c,b) \quad \text{and} \quad y \sim T \star \mathcal{PG}(a,b),$$

  and the conditional prior predictive distribution of x given y is

$$(x \mid y) \sim \mathcal{PGB}'(a+y,d,c).$$

- Denoting by x^* and y^* the "future observations" and by S^* and T^* the "future sample sizes," the marginal posterior predictive distributions are

$$(x^* \mid x,y) \sim \frac{S^*}{S} \star \mathcal{PGIB}(a+x+y,d+a+y,c+x)$$

  and

$$(y^* \mid x,y) \sim \frac{T^*}{T+b} \star \mathcal{PGIB}(a+x+y,c+x,d+a+y),$$

  and the joint posterior predictive distribution is

$$(x^*,y^* \mid x,y) \sim \left(\frac{S^*}{S},\frac{T^*}{T+b}\right) \star \mathcal{B}ailey(a+x+y,c+x,d+a+y,1),$$

  which simplifies to $\mathcal{B}ailey(a+x+y,c+x,d+a+y,R^{-1})$ in the case when $\frac{S^*}{S} = \frac{T^*}{T+b} =: R$.

The following result provides the prior predictive distributions for the larger family of priors defined in Lemma 3.1 by substituting t for T and τ for ρS.

Result 3.1. *Let $a,b,c,d,\tau,t>0$ be given numbers and consider a four-tuple of random variables (μ,ψ,x,y) whose distribution is defined by*:

- *$\mu \sim \mathcal{G}(a,b)$ is independent of $\psi \sim \mathcal{B}'(c,d)$;*
- *x and y are conditionally independent given (μ,ψ) and their conditional distributions are $(x \mid \mu,\psi) \sim \mathcal{P}(\mu\psi\tau)$ and $(y \mid \mu,\psi) \sim \mathcal{P}(\mu t)$.*

Then the marginal distributions of x and y are

$$y \sim t \star \mathcal{PG}(a,b) \quad \textit{and} \quad x \sim \tau \star \mathcal{PGB}_2(a,d,c,b),$$

the conditional distribution of x given y is

$$(x \mid y) \sim \tau \star \mathcal{PGB}_2(a+y,d,c,b+t),$$

and the joint probability generating function of x and y is given by

$$G(u,v) = \frac{B(d+a,c)}{B(d,c)} \left(\frac{b}{\tau(1-u)}\right)^a {}_2F_1\left(d+a,a,c+d+a;1-\frac{b+t(1-v)}{\tau(1-u)}\right).$$

Proof. Obviously, $y \sim t \star \mathcal{PG}(a, b)$. As the conditional distribution of $\lambda = \mu\psi$ given ψ is $\frac{1}{b} \times \mathcal{G}(a, \psi^{-1})$, one has $(x \mid \psi) \sim \frac{\tau}{b} \star \mathcal{PG}(a, \psi^{-1})$ and hence $x \sim \tau \star \mathcal{PGB}_2(a, d, c; b)$. Using Bayes' formula, we easily see that the conditional distribution of μ given (y, ψ) depends only on y and is $\mathcal{G}(a + y, b + t)$. Hence, the conditional distribution of x given (y, ψ) is $\tau\psi \star \mathcal{PG}(a + y, b + t) = \frac{\tau}{b+t} \star \mathcal{PG}(a + y, \psi^{-1})$. Since y and ψ are independent, because μ and ψ are independent and the conditional distribution $(y \mid \mu, \psi)$ does not depend on ψ, we then have $(x \mid y) \sim \tau \star \mathcal{PGB}_2(a + y, d, c, b + t)$. Therefore, we know from Result 2.1 that the p.g.f. of the conditional law of x given y is given by

$$G(u \mid y) = \frac{B(d + a + y, c)}{B(d, c)} \left(\frac{\tau(1 - u)}{b + t}\right)^d \times {}_2F_1\left(d + a + y, c + d, c + d + a + y; 1 - \frac{\tau(1 - u)}{b + t}\right),$$

therefrom, using formula (1.1), the expression of $t \star \mathcal{PG}(y \mid a, b)$ given by (1.9), and the series expansion (1.4) of the Appell hypergeometric function,

$$G(u, v) = \frac{B(d + a, c)}{B(d, c)} \left(\frac{b}{b + t}\right)^a \left(\frac{\tau(1 - u)}{b + t}\right)^d \times F_1\left(d + a, a, c + d, c + d + a; \frac{vt}{b + t}, 1 - \frac{\tau(1 - u)}{b + t}\right),$$

thereby yielding the announced expression owing to the Bailey reduction (3.1). Obviously, we also could have firstly derived the p.g.f. in the particular case $t = \tau = 1$, and then we would have derived the general case by applying the linear change of variables (3.5). □

The following result provides the posterior predictive distributions for the semi-conjugate family by substituting x for x^*, y for y^*, a for $a + x + y$, c for $c + x$, d for $d + a + y$, t for $T^*/(T + b)$, and τ for S^*/S. Unfortunately, the conditional distribution of x given y in the result below is not of the kind of distributions defined in this paper.

Result 3.2. *Let $a, c, d, t, \tau > 0$ be given and consider a four-tuple of random variables (μ, ψ, x, y) whose distribution is defined by*:

- $(\mu \mid \psi) \sim \mathcal{G}(a, 1 + \psi)$ *and* $\psi \sim \mathcal{B}'(c, d)$;
- *x and y are conditionally independent given (μ, ψ) and their conditional distributions are $(x \mid \mu, \psi) \sim \mathcal{P}(\mu\psi\tau)$ and $(y \mid \mu, \psi) \sim \mathcal{P}(\mu t)$.*

Then the marginal distributions of x and y are

$$x \sim \tau \star \mathcal{PGIB}(a, d, c) \quad \textit{and} \quad y \sim t \star \mathcal{PGIB}(a, c, d),$$

the joint distribution of x and y is

$$(x, y) \sim (\tau, t) \star \mathcal{B}ailey(a, c, d, 1),$$

and the probability mass assigned by this distribution at $x, y \in \mathbb{N}$ *is*

$$p(x, y) = \frac{(a)_{x+y}(c)_x(d)_y}{x!y!(c+d)_{x+y}} \frac{\tau^x t^y}{(t+1)^{a+x+y}} \times {}_2F_1\left(c+x, a+x+y, c+d+x+y; 1 - \frac{\tau+1}{t+1}\right).$$

One also has

$$(x, y) \sim (\tau/\rho, t/\rho) \star \mathcal{B}ailey(a, c, d, \rho^{-1})$$

whatever the value of $\rho > 0$ (*thereby showing* (3.4)). *The joint probability generating function of x and y is given by*

$$G_{(\tau,t)\star\mathcal{B}ailey(a,c,d,1)}(u, v) = [1 + t(1 - v)]^{-a} {}_2F_1\left(c, a, c+d, 1 - \frac{1 + \tau(1-u)}{1 + t(1-v)}\right).$$

Proof. One has $(y \mid \psi) \sim t \star \mathcal{PG}(a, 1+\psi)$ and $(x \mid \psi) \sim \psi\tau \star \mathcal{PG}(a, 1+\psi) = \tau \star \mathcal{PG}(a, 1+\psi^{-1})$, hence $y \sim t \star \mathcal{PGIB}(a, c, d)$ and $x \sim \tau \star \mathcal{PGIB}(a, d, c)$. One obtains $(\mu \mid y, \psi) \sim \mathcal{G}(a+y, t+1+\psi)$ by applying Bayes' formula, hence $(x \mid y, \psi) \sim \psi\tau \star \mathcal{PG}(a+y, t+1+\psi)$. This yields

$$p(x, y \mid \psi)\mathcal{B}'(\psi \mid c, d) = \frac{1}{B(c,d)} \frac{(a)_{x+y}}{x!y!} \frac{\tau^x t^y}{(t+1)^{a+x+y}} \times \frac{\psi^{c+x-1}(1+\psi)^{-(c+d-a)}}{(1 + \psi(\tau+1)/(t+1))^{a+x+y}}.$$

The expression of $p(x, y)$ is obtained by using formula (1.1) and the density function of $\mathcal{B}_3(c+x, d+y, c+d-a, \frac{\tau+1}{t+1})$ given by (1.8), and we see in view of (3.2) that $p(x, y) = \mathcal{B}ailey(x, y \mid a, c, d, \rho^{-1})$ when $t = \tau =: \rho$. Hence, one obviously has $(x, y) \sim (\tau/\rho, t/\rho) \star \mathcal{B}ailey(a, c, d, \rho^{-1})$ in the general case, thereby showing (3.4), and the expression of the p.g.f. is obtained by applying the linear change of variables (3.5) to the p.g.f. of $\mathcal{B}ailey(a, c, d, 1)$ given by (3.3). □

The covariance between x and y in lemma above is easy to derive with the help of the probability generating function; we find that

$$\mathrm{Cov}(x, y) = \frac{\tau t a c d(c+d-a)}{(c+d)^2(c+d+1)}.$$

In particular, in the context of the semi-conjugate family, we see that the posterior predictive covariance $\mathrm{Cov}(x^*, y^* \mid x, y)$ is always positive.

Comparison with the conjugate family. The natural conjugate family of priors for the "two Poisson samples" model is formed by the independent products of Gamma distributions on μ and λ. This family contains the Jeffreys prior which is the case when $\mu \sim \mathcal{G}(\frac{1}{2}, 0)$ and $\lambda \sim \mathcal{G}(\frac{1}{2}, 0)$, and, as noticed by Laurent and Legrand (2011), the Jeffreys prior and the ϕ-reference prior yield the same posterior on ϕ, but do not yield the same posterior predictive distributions. We note however that these posterior predictive distributions are close, because of

$$(\tau, t) \star \mathcal{B}ailey(a, c, d, 1) \approx \big(\tau \star \mathcal{PG}(c, 1)\big) \otimes \big(t \star \mathcal{PG}(d, 1)\big) \qquad \text{when } a \approx c + d,$$

with equality when $a = c + d$. This approximation follows from the two approximations $(a)_{x+y}/(c+d)_{x+y} \approx 1$ and

$$\begin{aligned} &{}_2F_1\left(c + x, a + x + y, c + d + x + y; 1 - \frac{\tau + 1}{t + 1}\right) \\ &\qquad \approx {}_2F_1\left(c + x, a + x + y, a + x + y; 1 - \frac{\tau + 1}{t + 1}\right) = \left(\frac{\tau + 1}{t + 1}\right)^{-c-x}. \end{aligned}$$

Thus, in the context of the semi-conjugate family, one obtains

$$(x^*, y^* \mid x, y) \approx \big(S^* \star \mathcal{PG}(c + x, S)\big) \otimes \big(T^* \star \mathcal{PG}(d + a + y, T + b)\big)$$

when $c + d$ is small, as in the case of the ϕ-reference prior for which $c + d = 0.5$, and moreover, in that case, the right member is exactly the posterior predictive distribution corresponding to the Jeffreys prior.

By-product: A family of priors for the Poisson model. The semi-conjugate family in the case when $T = 0$ in the "two Poisson samples" model yields a family of priors for the "one Poisson sample" model for which the prior predictive distributions form the Beta-negative binomial family. Precisely, we obtain the following result.

Consider the "one Poisson sample" model $x \sim \mathcal{P}(\lambda S)$ with known "sample size" S and unknown rate parameter λ. If the prior on λ has the distribution of the product $\mu\phi$ of two independent random variables $\mu \sim \mathcal{G}(a, S)$ and $\phi \sim \mathcal{B}'(c, d)$, then

- the prior predictive is $x \sim \mathcal{PGB}'(a, d, c)$;
- the posterior predictive is $(x^* \mid x) \sim \frac{S^*}{S} \star \mathcal{PGIB}(a + x, d + a, c + x)$;
- the posterior distribution $(\lambda \mid x)$ is the distribution of $\mu\theta$ where $\mu \sim \mathcal{G}(a + x, S)$ and $\theta \sim \mathcal{B}'(c + x, a + d)$ are independent random variables.

The prior and the posterior distributions of λ can be straightforwardly simulated but they are not analytically easy to handle. It can be shown that both the prior and the posterior densities of λ involve the Tricomi function. Certain distributions whose densities involve this function are studied by Fitzgerald (2000), but they do not cover the case above.

Acknowledgment

Research supported by IAP research network Grant P6/03 of the Belgian government (Belgian Science Policy).

References

Bailey, W. N. (1935). Appell's hypergeometric functions of two variables. In *Generalized Hypergeometric Series* 73–83, 99–101. Cambridge: Cambridge Univ. Press.

Bernardo, J. M. (2005). Reference analysis. In *Handbook of Statistics* **25** (D. K. Dey and C. R. Rao, eds.), 17–90. Amsterdam: Elsevier. MR2490522

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

Fitzgerald, D. L. (2000). Statistical aspects of Tricomi's function and modified Bessel functions of the second kind. *Stochastic Environmental Research and Risk Assessment* **14**, 139–158.

Johnson, N. L., Kemp, A. W. and Kotz, S. (2005). *Univariate Discrete Distributions*, 3rd ed. New York: Wiley. MR2163227

Kleiber, C. and Kotz, S. (2003). *Statistical Size Distributions in Economics and Actuarial Sciences*. Hoboken: Wiley. MR1994050

Laurent, S. and Legrand, C. (2011). A Bayesian framework for the ratio of two Poisson rates in the context of vaccine efficacy trials. *ESAIM: Probability and Statistics*. DOI:10.1051/ps/2010018.

Slater, L. J. (1966). *Generalized Hypergeometric Functions*. Cambridge: Cambridge Univ. Press. MR0201688

HEC-Management School
University of Liège
Rue Louvrex 14
4000 Liège
Belgium
E-mail: slaurent@ulg.ac.be