Author's Accepted Manuscript

www.elsevier.com/locate/jspi

Bivariate prior distributions via branching exchangeable sequences

S.J. Hatjispyros, Theodoros Nicoleris, Stephen G. Walker

PII: S0378-3758(07)00322-9
DOI: doi:10.1016/j.jspi.2007.06.034
Reference: JSPI 3406

To appear in: *Journal of Statistical Planning and Inference*

Received date: 14 November 2006
Revised date: 30 May 2007
Accepted date: 14 June 2007

Cite this article as: S.J. Hatjispyros, Theodoros Nicoleris and Stephen G. Walker, Bivariate prior distributions via branching exchangeable sequences, *Journal of Statistical Planning and Inference* (2007), doi:10.1016/j.jspi.2007.06.034

Bivariate Prior Distributions via Branching Exchangeable Sequences

BY
S. J. HATJISPYROS †, THEODOROS NICOLERIS †
AND STEPHEN G. WALKER ‡

University of the Aegean, Greece †, and University of Kent, UK ‡

Abstract

This paper provides a novel approach to constructing bivariate prior distributions. The idea is based on the notion of partial exchangeability. In particular, in a simple extension of the exchangeable sequence, we create two dependent exchangeable sequences via a branching mechanism. This implies the existence of a bivariate prior distribution.

KEYWORDS: Bivariate prior distribution; Exchangeability; Partial exchangeability.

1. Introduction. This paper is concerned with the construction of bivariate prior distributions. In the univariate case, for which there is a vast catalogue of distributions, it is possible to appeal to de Finetti's Representation Theorem for an exchangeable sequence.

Our approach for the bivariate setting requires an extension to the exchangeable sequence. We achieve this via a branching exchangeable sequence. The idea being that an exchangeable sequence branches into two conditionally independent exchangeable sequences at a particular point. Hence, there are two exchangeable sequences which are dependent due to the coincident initial parts they share in common. The amount they share in common obviously contributes to the extent of the dependence. We believe this to be the most natural and simple extension to an exchangeable sequence and hence appropriate for a bivariate prior construction.

Walker and Muliere (2003) introduced a bivariate Dirichlet process based on this idea. However, in this case, the marginal distributions are both Dirichlet processes with the same parameter. The aim in this paper is to develop the idea and to describe how it is feasible to have different marginal distributions in a parametric setting.

To set the scene, consider a parametric family of distributions $F(\cdot|\theta)$, with $\theta \in \Theta$. If $Z_1, Z_2, \ldots$ are exchangeable and conditional on θ are independent and identically distributed from $F(\cdot|\theta)$, then it is well known that there exists a prior distribution on Θ, say $\pi(\theta)$, such that for all n,

$$\mathrm{P}(Z_1 \le z_1, \ldots, Z_n \le z_n) = \int_\Theta \prod_{i=1}^n F(z_i|\theta)\, \pi(\theta)\, \mathrm{d}\theta.$$

We now discuss our idea for the bivariate prior distribution.

If we have two independent exchangeable sequences $Z_1^{(1)}, Z_2^{(1)}, \ldots$ and $Z_1^{(2)}, Z_2^{(2)}, \ldots$, which have the same distribution, then we have clearly created a bivariate distribution on $\Theta \times \Theta$ but with no dependence between the two marginals. To create a dependence in the bivariate distribution we will create a dependence between the two exchangeable sequences, as must be done. The simplest way we can see how to do this, with ease of interpretation, is to force a length, say r, of each sequence to coincide. Obviously, the larger r is, the greater the dependence between the two sequences, which will manifest itself in a dependence in the bivariate distribution on $\Theta \times \Theta$. Clearly also, $r = 0$ corresponds to the independent case.

So, consider the first r variables of an exchangeable sequence

$$Z_1, \ldots, Z_r, Z_{r+1}, \ldots.$$

At the integer r a branching occurs to give two dependent exchangeable sequences

$$\mathbf{Z}^{(1)} = \{Z_1^{(1)}, \ldots, Z_r^{(1)}, Z_{r+1}^{(1)}, Z_{r+2}^{(1)}, \ldots\}$$

and

$$\mathbf{Z}^{(2)} = \{Z_1^{(2)}, \ldots, Z_r^{(2)}, Z_{r+1}^{(2)}, Z_{r+2}^{(2)}, \ldots\},$$

with $Z_i = Z_i^{(1)} = Z_i^{(2)}$, for $i = 1, \ldots, r$, where the parts

$$Z_{r+1}^{(1)}, Z_{r+2}^{(1)}, \ldots$$

and

$$Z_{r+1}^{(2)}, Z_{r+2}^{(2)}, \ldots$$

are conditionally independent given $Z_1, \ldots, Z_r$. The first sequence $\mathbf{Z}^{(1)}$ is independent and identically distributed from $F(\cdot|\theta_1)$, given θ_1, and $\mathbf{Z}^{(2)}$ is independent and identically distributed from $F(\cdot|\theta_2)$, given θ_2. If we now let $X_i = Z_{i+r}^{(1)}$ for $i \ge 1$ and $Y_j = Z_{j+r}^{(2)}$ for $j \ge 1$, then

$$\mathrm{P}(X_i \le x_i, i = 1, \ldots, n; Y_j \le y_j, j = 1, \ldots, m | Z_1 = z_1, \ldots, Z_r = z_r)$$

$$= \int_{\Theta^2} \prod_{i=1}^n F(x_i|\theta_1) \prod_{j=1}^m F(y_j|\theta_2)\, \pi(\theta_1|\mathbf{z})\, \pi(\theta_2|\mathbf{z})\, \mathrm{d}\theta_1 \mathrm{d}\theta_2.$$

Hence, the bivariate prior distribution, for the conditionally dependent sequences $X_1, X_2 \ldots; Y_1, Y_2 \ldots$, which will become the observables, is given by

$$\pi(\theta_1, \theta_2) = \int \pi(\theta_1|\mathbf{z})\, \pi(\theta_2|\mathbf{z})\, m(\mathbf{z})\, \mathrm{d}\mathbf{z},$$

where $\mathbf{z} = \{z_1, \ldots, z_r\}$,

$$\pi(\theta_l|\mathbf{z}) = \left\{ \prod_{i=1}^{r} f(z_i|\theta_l) \right\} \pi(\theta_l)/m(\mathbf{z}),$$

for $l = 1, 2$, and

$$m(\mathbf{z}) = \int \prod_{i=1}^{r} f(z_i|\theta)\, \pi(\theta)\, \mathrm{d}\theta.$$

There are many bivariate prior distributions which can be formed through a representation of the above type. However, we are providing motivation for a particular mixture representation for $\pi(\theta_1, \theta_2)$. The particular mixture involved is derived from notions of partial exchangeability and the idea of creating dependence via two exchangeable sequences having a common branch.

To illustrate, we briefly consider what happens if we ask for a bivariate normal distribution. So $\pi(\theta_1)$ and $\pi(\theta_2)$ will be taken to be standard normal and $f(z|\theta)$ being normal with mean θ and known variance σ^2. Some elementary work yields that $\pi(\theta_1, \theta_2)$ are bivariate standard normal with correlation

$$\frac{r}{\sigma^2 + r}.$$

Hence the role of r is quite explicit. This example also points out a particular issue. It would seem the correlation can only take discrete values, the smallest being 0 and the next smallest being $1/(1 + \sigma^2)$. This is true if one restricts r to be an integer; however, having obtained the form of the bivariate distribution one can readily see that r can be extended to the non-negative real line. This extension will typically be the case so ultimately there will be no discrete dependence structures.

As we have so far described things, the prior distribution of θ_1 would be the same as for θ_2, that is $\pi(\theta)$. This paper is about allowing for a more general bivariate prior in which the marginal distribution of θ_1 is not necessarily the same as for θ_2.

Describing the layout of the paper, in Section 2 we detail the construction behind the bivariate prior distributions for which marginal prior distributions are different. Section 3 describes a bivariate beta distribution and

Section 4 a bivariate exponential distribution. In Section 5 we present a means by which to undertake posterior analysis via sampling based methods. Section 6 contains a simple illustrative example involving a simulated data set and the estimation of a ROC curve. Finally, Section 7 contains a brief discussion.

2. Bivariate prior distributions. As we saw in the previous section, if the marginal prior distributions are to be the same then we take $Z_i^{(1)} = Z_i^{(2)}$ for $i = 1, \ldots, r$. For a different marginal prior we will use a deterministic relation between them. We would like marginally for $Z_i^{(1)}$ to be distributed as

$$F_1(\cdot) = \int F(\cdot|\theta)\, \pi(\theta|\lambda_1)\, \mathrm{d}\theta$$

and for $Z_i^{(2)}$ to be distributed as

$$F_2(\cdot) = \int F(\cdot|\theta)\, \pi(\theta|\lambda_2)\, \mathrm{d}\theta,$$

where (λ_1, λ_2) are fixed hyper-parameters for the two marginal prior distributions. We can achieve this without difficulty. There are two cases to consider, the continuous and discrete cases.

2.1 The continuous case. For this case we start by taking

$$Z_1^{(1)} \sim F_1$$

and then

$$Z_2^{(1)}|[Z_1^{(1)} = z] \sim \int F(\cdot|\theta)\, \pi(\theta|z, \lambda_1)\, \mathrm{d}\theta,$$

where $\pi(\theta|z, \lambda_1)$ is the conditional distribution of θ given $Z_1^{(1)} = z$. In general, take

$$Z_i^{(1)}|[Z_{i-1}^{(1)} = z_{i-1}, \ldots, Z_1^{(1)} = z_1] \sim \int F(\cdot|\theta)\, \pi(\theta|z_1, \ldots, z_{i-1}, \lambda_1)\, \mathrm{d}\theta.$$

To create the dependence we will take, for $1 \le i \le r$,

$$Z_i^{(2)} = F_2^{-1}\left(F_1(Z_i^{(1)})\right).$$

This gives us the initial sequences with correct marginal distributions. A bivariate distribution is constructed due to the connection between the initial part of the sequences. Then $Z_{r+1}^{(1)}, \ldots$ and $Z_{r+1}^{(2)}, \ldots$ are conditionally independent, given $Z_1^{(1)}, \ldots, Z_r^{(1)}$. Obviously, if $F_1 \equiv F_2$, then $Z_i^{(1)} = Z_i^{(2)}$.

The data will be the two conditionally independent sequences $X_i = Z_{r+i}^{(1)}$ and $Y_i = Z_{r+i}^{(2)}$ for $i = 1, 2, \ldots$. The role of $Z_i^{(1)}$ and $Z_i^{(2)}$, for $i = 1, \ldots, r$, is as two (deterministically related) latent processes which make the two data sets conditionally independent and unconditionally dependent. In fact the two sequences are then referred to as being partially exchangeable in the sense of de Finetti which implies the existence of a bivariate prior distribution.

The bivariate distribution can be understood in the following illuminating way. Take

$$\theta_1 \sim \pi(\cdot|\lambda_1)$$

then take

$$Z_1^{(1)}, \ldots, Z_r^{(1)}|\theta_1 \sim F(\cdot|\theta_1).$$

Then transform to obtain $Z_i^{(2)} = F_2^{-1}(F_1(Z_i^{(1)}))$ for $i = 1, \ldots, r$. We now take as θ_2 the random variable which conditionally on $Z_1^{(2)}, \ldots, Z_r^{(2)}$ is represented as

$$\theta_2|Z_1^{(2)}, \ldots, Z_r^{(2)} \sim \pi(\cdot|Z_1^{(2)}, \ldots, Z_r^{(2)}, \lambda_2),$$

the conditional distribution of θ given observations $Z_1^{(2)}, \ldots, Z_r^{(2)}$. Now define,

$$\mathbf{Z}^{(1)} = \{Z_1^{(1)}, \ldots, Z_r^{(1)}\} \text{ and } \mathbf{Z}^{(2)} = \{Z_1^{(2)}, \ldots, Z_r^{(2)}\}.$$

From here we have the joint prior density given by

$$\pi(\theta_1, \theta_2) = \int \pi(\theta_1|\mathbf{z}^{(1)})\, \pi(\theta_2|\mathbf{z}^{(2)})\, m_1(\mathbf{z}^{(1)})\, \mathrm{d}\mathbf{z}^{(1)},$$

where

$$m_1(\mathbf{z}^{(1)}) = \int \prod_{i=1}^{r} f(z_i^{(1)}|\theta_1)\, \pi(\theta_1|\lambda_1)\, \mathrm{d}\theta_1.$$

The marginal distributions can be obtained as:

$$\pi(\theta_1) = \int \pi(\theta_1|\mathbf{z}^{(1)})\, m_1(\mathbf{z}^{(1)})\, \mathrm{d}\mathbf{z}^{(1)} = \pi(\theta_1|\lambda_1)$$

and

$$\begin{aligned}\pi(\theta_2) &= \int \pi(\theta_2|\mathbf{z}^{(2)})\, m_1(\mathbf{z}^{(1)})\, \mathrm{d}\mathbf{z}^{(1)} \\ &= \int \pi(\theta_2|\mathbf{z}^{(2)})\, m_2(\mathbf{z}^{(2)})\, \mathrm{d}\mathbf{z}^{(2)} \ . \\ &= \pi(\theta_2|\lambda_2).\end{aligned}$$

The nonparametric prior of Walker and Muliere (2003) took $F_1 = F_2$ and hence marginally the prior has the same distribution.

To understand the correlation we will start by assuming that $\lambda = \lambda_1 = \lambda_2$. In this case we have

$$\pi(\theta_2|Z_1, \ldots, Z_r, \lambda)$$

with the Z_i being independently distributed from $F(\cdot|\theta_1)$ and $\theta_1 \sim \pi(\cdot|\lambda)$. Using the consistency result of Doob (1949), we have that

$$\pi(\theta_2|Z_1, \ldots, Z_r, \lambda) \to \delta_{\theta_1},$$

where δ_θ denotes a point mass of 1 at θ. So, as $r \to \infty$, the joint distribution for (θ_1, θ_2) accumulates along the line $\theta_1 = \theta_2$. Obviously, for $r = 0$, θ_1 and θ_2 are independent. When $\lambda_1 \neq \lambda_2$, then instead we have the prior accumulating along the line $(\theta_1, \xi(\theta_1))$, where $\xi(\theta)$ is given by

$$F\left(F_2^{-1}(F_1(\cdot))|\xi(\theta)\right) = F(\cdot|\theta).$$

We will return to r later.

2.2 The discrete case. We will now examine the case where the observations come from two discrete distributions. In this case the inverse of a distribution F, denoted here by $F^-(\cdot)$, is interpreted as:

$$F^-(y) = \inf\{x : F(x) \geq y\}$$

where $y \in (0, 1)$. This is a more challenging case since the strict monotonicity of the two distributions does not hold.

One may be tempted to define the deterministic relation between the latent variables as the natural analogue of the continuous one; namely, $Z_i^{(2)} = F_2^-(F_1(Z_i^{(1)}))$ for $i = 1, \ldots, r$. A shortcoming, though, is that $Z_i^{(2)}$, for $i = 1, \ldots, r$, is not distributed as F_2. So, in what follows we present an elegant and intuitive way of creating the two latent processes in the discrete case. This idea can be used, as before, to generate two exchangeable sequences of variables with predetermined marginals.

As before, the bivariate distribution can be understood in the following way. Take

$$\theta_1 \sim \pi(\cdot|\lambda_1)$$

and let $\mathbf{U} = (U_1, \ldots, U_r)$ be a random sample of size r with the U_i's being independent and identically distributed as $U_i \sim U(0, 1)$. Take

$$Z_i^{(1)} = F^-(U_i\,|\theta_1), \quad i = 1, \ldots, r,$$

and then take θ_2 as the random variable which conditionally on $Z_1^{(2)}, \ldots, Z_r^{(2)}$ has distribution

$$\theta_2|Z_1^{(2)}, \ldots, Z_r^{(2)} \sim \pi(\cdot|Z_1^{(2)}, \ldots, Z_r^{(2)}, \lambda_2),$$

the conditional distribution of θ given observations $Z_1^{(2)}, \ldots, Z_r^{(2)}$, where

$$Z_i^{(2)} = F^-(U_i\,|\theta), \;\; i = 1, \ldots, r.$$

Note that these sequences are correlated through their common "link" which is the sample values from the uniform distribution. Clearly, unconditionally, $Z_i^{(1)} \sim F_1$ and $Z_i^{(2)} \sim F_2$.

The joint distribution of $\mathbf{Z^{(1)}}$ is given by

$$P(\mathbf{Z^{(1)}} = \mathbf{z^{(1)}}) = \int \prod_{i=1}^{r} f(z_i^{(1)}|\theta_1)\, \pi(\theta_1|\lambda_1)\, \mathrm{d}\theta_1.$$

Similarly, for $\mathbf{Z^{(2)}}$. From here we have the joint prior density

$$\pi(\theta_1, \theta_2) = \int_{[\mathbf{u}\in[0,1]^r]} \pi(\theta_1|\mathbf{z^{(1)}}(\mathbf{u}))\, \pi(\theta_2|\mathbf{z^{(2)}}(\mathbf{u}))\, \mathrm{d}\mathbf{u}.$$

The marginal distributions can be easily evaluated,

$$\begin{aligned} \pi(\theta_1) &= \textstyle\int_{\mathbf{u}\in[0,1]^r} \pi(\theta_1|\mathbf{z^{(1)}}(\mathbf{u}))\, \mathrm{d}\mathbf{u} \\ &= \textstyle\sum_{\{\mathbf{z^{(1)}}\}} \pi(\theta_1|\mathbf{z^{(1)}}) P(\mathbf{u} : \mathbf{Z^{(1)}}(\mathbf{u}) = \mathbf{z^{(1)}}) \\ &= \textstyle\sum_{\{\mathbf{z^{(1)}}\}} \pi(\theta_1|\mathbf{z^{(1)}}) P(\mathbf{Z^{(1)}} = \mathbf{z^{(1)}}) \\ &= \pi(\theta_1|\lambda_1) \end{aligned}$$

Similarly, $\pi(\theta_2) = \pi(\theta_2|\lambda_2)$. The approach used for the discrete case could equally have been used in the continuous case as well.

3. A bivariate beta distribution. Here we describe a bivariate beta prior distribution based on ideas presented in Section 2.1 with $r = 2$, though we can do the general r case it is more complicated. Briefly, the method is as follows: We generate two latent sequences which are dependent through the sharing of a common part. Each sequence is indexed through a parameter which is drawn, independently of the other, from a fixed marginal prior distribution. The bivariate prior distribution is then obtained as the marginal of the product between the joint unconditional distribution of the two latent sequences and the joint "latent" posterior of the parameters.

Let $\theta_1 \sim \text{Beta}(\alpha_1, \beta_1)$ and $\theta_2 \sim \text{Beta}(\alpha_2, \beta_2)$; also for $i = 1, 2$ let $U_i \sim U(0, 1)$ be two independent uniform random variables. Consider

$$Z_i^{(1)} = F^{-}(U_i \,|\theta_1) \text{ and } Z_i^{(2)} = F^{-}(U_i \,|\theta_2), \quad i = 1, 2,$$

where F is taken as the distribution of a Bernoulli random variable. It is convenient to introduce

$$S_1 = Z_1^{(1)} + Z_2^{(1)} \sim \text{Bin}(2, \theta_1) \text{ and } S_2 = Z_1^{(2)} + Z_2^{(2)} \sim \text{Bin}(2, \theta_2).$$

The joint distribution of (S_1, S_2), conditionally on (θ_1, θ_2), is given by

$$P(S_1 = i\,, S_2 = j \mid \theta_1,\ \theta_2) = P\left(Z_1^{(1)} + Z_2^{(1)} = i\,, Z_1^{(2)} + Z_2^{(2)} = j \mid \theta_1,\ \theta_2\right)$$

$$= \sum_{(\kappa_1,\, \kappa_2)} P(Z_1^{(1)} = i - \kappa_1\,, Z_1^{(2)} = j - \kappa_2) P(Z_2^{(1)} = \kappa_1\,, Z_2^{(2)} = \kappa_2).$$

Assuming, without loss of generality, that $\theta_1 \leq \theta_2$ the distribution can be summarized as follows;

$$P(S_1 = i\,, S_2 = j \mid \theta_1,\ \theta_2) = C(2-i, j-i) C(2, i) \theta_1^i (1-\theta_2)^{2-j} (\theta_2 - \theta_1)^{j-i}$$

for $0 \leq i \leq j \leq 2$ and $P(S_1 = i\,, S_2 = j \mid \theta_1,\ \theta_2) = 0$ for $i > j$. Here $C(i, j) = i!/(j!(i-j)!)$.

The next step is to calculate the joint(unconditional on (θ_1, θ_2)) distribution of (S_1, S_2).

Now, for $0 \leq i \leq j \leq 2$, $P(S_1 = i\,, S_2 = j)$ is given by

$$\int_0^1 \int_0^1 P(S_1 = i\,, S_2 = j \mid \theta_1,\ \theta_2) \, \text{Beta}(\theta_1|\alpha_1, \beta_1) \, \text{Beta}(\theta_2|\alpha_2, \beta_2) \mathrm{d}\theta_1 \mathrm{d}\theta_2$$

$$= \tfrac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\,\Gamma(\beta_1)} \tfrac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\,\Gamma(\beta_2)} C(2-i, j-i) C(2, i)$$

$$\times \int_0^1 \int_0^1 \theta_1^{\alpha_1+i-1} (1-\theta_1)^{\beta_1-1} \theta_2^{\alpha_2-1} (1-\theta_2)^{\beta_2-j+1} (\theta-2-\theta_1)^{j-i} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2.$$

Denoting for the moment by D all the terms which do not depend on $(\theta_1,\ \theta_2)$, the above expression becomes

$$D \int_0^1 \int_0^1 \theta_1^{i+\alpha_1-1} (1-\theta_1)^{\beta_1-1} \theta^{\alpha_2-1} (1-\theta_2)^{\beta_2-j+1} (\theta_2 - \theta_1)^{j-i} \, \mathrm{d}\theta_1 \mathrm{d}\theta_2$$

$$= D \textstyle\sum_{m=0}^{j-i} (-1)^{j-i-m} C(j-i, m) \left\{ \tfrac{\Gamma(\alpha_1+j-m)\,\Gamma(\beta_1)}{\Gamma(\alpha_1+\beta_1+j-m)} \right\} \left\{ \tfrac{\Gamma(\alpha_2+m)\,\Gamma(\beta_2-j+2)}{\Gamma(\alpha_2+\beta_2+m-j+2)} \right\}.$$

Therefore, the prior $\pi(\theta_1, \theta_2)$ is given by

$$\sum_{i=0}^{2} \sum_{j=0}^{2} \pi(\theta_1|i)\, \pi(\theta_2|j)\, P(S_1 = i\ , S_2 = j)$$

$$= \sum_{i=0}^{2} \sum_{j=i}^{2} \mathrm{Be}(\theta_1|\alpha_1 + i, \beta_1 + 2 - i)\, \mathrm{Be}(\theta_2|\alpha_2 + j, \beta_2 + 2 - j)\, P(S_1 = i\ , S_2 = j)$$

$$= \sum_{i=0}^{2} \sum_{j=i}^{2} \sum_{m=0}^{j-i} (-1)^{j-i-m} C(2-i, j-i) C(2, i) C(j-i, m) \tfrac{\Gamma(\alpha_1+\beta_1+2)}{\Gamma(\alpha_1+i)\,\Gamma(\beta_1+2-i)}$$

$$\times \tfrac{\Gamma(\alpha_2+\beta_2+2)}{\Gamma(\alpha_2+j)\,\Gamma(\beta_2+2-j)} \tfrac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\,\Gamma(\beta_1)} \tfrac{\Gamma(\alpha_2+\beta_2)}{\Gamma(\alpha_2)\,\Gamma(\beta_2)} \tfrac{\Gamma(\alpha_1+j-m)\,\Gamma(\beta_1)}{\Gamma(\alpha_1+\beta_1+j-m)} \tfrac{\Gamma(\alpha_2+m)\,\Gamma(\beta_2+2-j)}{\Gamma(\alpha_2+\beta_2+m-j+2)}$$

$$\times \theta_1^{\alpha_1+i-1} (1-\theta_1)^{\beta_1+1-i}\ \theta_2^{\alpha_2+j-1} (1-\theta_2)^{\beta_2+1-j}.$$

This is our bivariate beta distribution.

Some important points to notice are the following: the dependence between the parameters in the bivariate prior is actually generated through the dependence between (S_1, S_2) since these parameters are conditionally independent on those variables. The distribution of the (S_1, S_2) can be thought of as a bivariate predictive constructed through the latent variable sequences. The bivariate prior is then created by expressing it as posterior $\times$ predictive and then integrating appropriately.

We will now confirm that the marginal density of θ_1 is $\mathrm{Beta}(\alpha_1, \beta_1)$. The density can be written as

$$\begin{aligned} \pi(\theta_1) &= \tfrac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\,\Gamma(\beta_1)} \theta_1^{\alpha_1-1} (1-\theta_1)^{\beta_1+1} \\ &\quad \times \sum_{i=0}^{2} C(2, i) \left(\tfrac{\theta_1}{1-\theta_1}\right)^{i}\ A(i) \tfrac{\Gamma(\alpha_1+\beta_1+2)\Gamma(\alpha_2+\beta_2)\Gamma(\beta_1)}{\Gamma(\alpha_1+i)\,\Gamma(\beta_1+2-i)\Gamma(\alpha_2)\,\Gamma(\beta_2)}, \end{aligned}$$

where $A(i)$ is given by

$$\sum_{j=i}^{2} \sum_{m=0}^{j-i} (-1)^{j-i-m} \frac{C(2-i, j-i) C(j-i, m) \Gamma(\alpha_1 + j - m) \Gamma(\alpha_2 + m)\, \Gamma(\beta_2 + 2 - j)}{\Gamma(\alpha_1 + \beta_1 + j - m) \Gamma(\alpha_2 + \beta_2 + m - j + 2)}.$$

For the result to be true it suffices to show that for $i = 0, 1, 2$,

$$A(i) = \left(\frac{\Gamma(\alpha_1 + \beta_1 + 2) \Gamma(\alpha_2 + \beta_2) \Gamma(\beta_1)}{\Gamma(\alpha_1 + i)\, \Gamma(\beta_1 + 2 - i) \Gamma(\alpha_2)\, \Gamma(\beta_2)} \right)^{-1}.$$

For the case $i = 2$ it is trivially true. For $i = 1$ we have that $A(1)$ is given by

$$\sum_{j=1}^{2}\sum_{m=0}^{j-1}(-1)^{j-1-m}\frac{C(1,j-1)C(j-1,m)\Gamma(\alpha_1+j-m)\Gamma(\alpha_2+m)\,\Gamma(\beta_2+2-j)}{\Gamma(\alpha_1+\beta_1+j-m)\Gamma(\alpha_2+\beta_2+m-j+2)}$$

$$= \frac{\Gamma(\alpha_1+1)\,\Gamma(\beta_1+1)\Gamma(\alpha_2)\,\Gamma(\beta_2)}{\Gamma(\alpha_1+\beta_1+2)\Gamma(\alpha_2+\beta_2)\Gamma(\beta_1)}\left\{\frac{(\beta_2+\alpha_2)(\alpha_1+\beta_1+1)}{\beta_1(\beta_2+\alpha_2)} - \frac{\alpha_1+1}{\beta_1}\right\}$$

$$= \frac{\Gamma(\alpha_1+1)\,\Gamma(\beta_1+1)\Gamma(\alpha_2)\,\Gamma(\beta_2)}{\Gamma(\alpha_1+\beta_1+2)\Gamma(\alpha_2+\beta_2)\Gamma(\beta_1)}$$

as required. Similarly, $A(0)$ is given by

$$\frac{\Gamma(\alpha_1)\,\Gamma(\alpha_2)\,\Gamma(\beta_2)}{\Gamma(\alpha_1+\beta_1+2)\Gamma(\alpha_2+\beta_2)}\left\{\beta_1(\alpha_1+\beta_1+1)+\alpha_1(\alpha_1+1)-\alpha_1(\alpha_1+\beta_1+1)\right\}$$

$$= \frac{\Gamma(\alpha_1)\,\Gamma(\beta_1+2)\Gamma(\alpha_2)\,\Gamma(\beta_2)}{\Gamma(\alpha_1+\beta_1+2)\Gamma(\alpha_2+\beta_2)\Gamma(\beta_1)}.$$

Therefore,

$$\begin{aligned}\pi(\theta_1) &= \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\,\Gamma(\beta_1)}\theta_1^{\alpha_1-1}(1-\theta_1)^{\beta_1+1}\sum_{i=0}^{2}C(2,i)\left(\frac{\theta_1}{1-\theta_1}\right)^i \\ &= \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\,\Gamma(\beta_1)}\theta_1^{\alpha_1-1}(1-\theta_1)^{\beta_1-1}.\end{aligned}$$

In a similar fashion we can show that $\theta_2 \sim \text{Beta}(\alpha_2, \beta_2)$.

The posterior distribution for the θ's rely on sufficient statistics, when they exist, of dimension 1, or at least of dimension smaller than r. One such example is the Bernoulli model with beta prior model, examined earlier. The question that arises is whether we can use sufficiency to our advantage and instead of generating an independent and identically distributed sample of size r from the uniform distribution, as outlined in Section 2.1, and then separately each $Z_i^{(1)}$ in $\mathbf{Z^{(1)}}$, we can use the sufficiency of S_1 and obtain the same model by first generating a *single* uniform variable and then sampling S_1 directly from a $\text{Bin}(r, \theta)$. Here, in expanded form,

$$S_1 = \sum_{i=1}^{r} Z_i^{(1)}.$$

A similar procedure would be followed for

$$S_2 = \sum_{i=1}^{r} Z_i^{(2)}.$$

Thus consider the following model: Let $\theta_1 \sim \mathrm{Beta}(\alpha_1, \beta_1)$ and $F(\cdot|\theta)$ denote the distribution function of a $\mathrm{Bin}(r, \theta)$ random variable. Then generate $U \sim U(0, 1)$ and let $S_1 = F^{-}(U|\theta_1)$ so we have that

$$S_1|\theta_1 \sim \mathrm{Bin}(r, \theta_1)$$

with posterior

$$\theta_1|[S_1 = s_1] \sim \mathrm{Beta}(\alpha_1 + s_1, \beta_1 + r - s_1).$$

Next, we take as θ_2 the random variable which conditionally on S_2 is represented as

$$\theta_2|[S_2 = s_2] \sim \pi(\cdot|S_2 = s_2, \alpha_2, \beta_2) = \mathrm{Beta}(\alpha_2 + s_2, \beta_2 + r - s_2),$$

the posterior distribution for θ given S_2, where $S_2 = F^{-}(U|\theta)$, with prior $\theta \sim \mathrm{Beta}(\alpha_2, \beta_2)$.

As it turns out the sufficiency approach for constructing the model does not produce the same model as in the previous method. Specifically, in the binomial-beta case examined previously, the joint conditional distribution $P(S_1 = i\ , S_2 = j \mid \theta_1,\ \theta_2)$ is not the same for the two approaches. For instance, we have found that when $r = 2$ and when $1 - \sqrt{1 - \theta_2^2} < \theta_1 \leq \theta_2$, then in the single uniform/binomial generation approach, $P(S_1 = 0\ , S_2 = 2 \mid \theta_1,\ \theta_2) = 0$ whereas in the r Bernoulli sequences generation we have that $P(S_1 = 1\ , S_2 = 1 \mid \theta_1,\ \theta_2) = (\theta_2 - \theta_1)^2$. This obviously has an effect on the bivariate distribution. Notice that the bivariate beta distribution is a mixture of independent beta distributions mixed by the distribution of (S_1, S_2). This means that in the first approach there are beta components present in the mixture which are absent in the second approach simply because $P(S_1, S_2) = 0$. The correlation between (θ_1, θ_2) is also affected. For instance, when the marginals are $\mathrm{Beta}(1, 1)$ and $\mathrm{Beta}(2, 1)$, then the correlation in one case is 0.316 and in the other case is 0.367.

4. A bivariate exponential distribution. Here we study a model based on the sufficiency idea, a Poisson model with exponential prior. We could use the more general gamma prior but for simplicity will use the exponential. So consider the following model:

$$\theta_1 \sim \exp(\lambda_1).$$

Let $F(\cdot|\theta)$ denote the distribution function of a $\mathrm{Poisson}(r\theta)$ random variable. Generate $U \sim U(0, 1)$ and let $S_1 = F^{-}(U|\theta_1)$. So we have that

$$S_1|\theta_1 \sim \mathrm{Poisson}(r\theta_1)$$

with posterior

$$\theta_1|[S_1 = s_1] \sim \text{Gamma}(s_1 + 1, r + \lambda_1).$$

We can obviously in this case consider non-integer values for r. We will return and comment on this point in the discussion section at the end of the paper.

Finally, we take as θ_2 the random variable which conditionally on S_2 is represented as

$$\theta_2|[S_2 = s_2] \sim \pi(\cdot|S_2 = s_2, \lambda_2) = \text{Gamma}(s_2 + 1, r + \lambda_2),$$

the posterior distribution for θ given S_2, where $S_2 = F^-(U|\theta)$, with prior for $\theta \sim \exp(\lambda_2)$.

For $i = 1, 2$, the (unconditional) distribution of S_i is given as

$$p(s_i) = \frac{\lambda_i \Gamma(s_i + 1) r^{s_i}}{(s_i)!(\lambda_i + r)^{s_i}} = \left(\frac{\lambda_i}{\lambda_i + r}\right)\left(\frac{r}{\lambda_i + r}\right)^{s_i} = (1 - \rho_i)\rho_i^{s_i}, \quad s_i = 0, 1, 2 \ldots$$

where $\rho_i = r/(\lambda_i + r)$. Therefore, $S_i \sim \text{Geom}(\rho_i)$. This is a Poisson-Gamma distribution in the case where the mixing distribution is $\exp(\lambda_i)$.

It is evident that the pair S_1 and S_2 can equivalently be obtained, directly, through the following alternative sampling scheme. Generate a point from a $U(0, 1)$ and subsequently generate two variables S_1 and S_2 each following the geometric distribution with respective parameters ρ_1 and ρ_2. Now,

$$\text{corr}(\theta_1, \theta_2) = \frac{\text{cov}(S_1, S_2)}{(\text{Var}(S_1) + \text{E}(S_1) + 1)^{1/2}(\text{Var}(S_2) + \text{E}(S_2) + 1)^{1/2}}.$$

To be able to calculate the correlation and the joint density of (θ_1, θ_2) we need to compute the joint distribution of (S_1, S_2).

In what follows we will compute explicitly the joint distribution of (S_1, S_2) and then provide explicit expressions for the correlation and the joint density of (θ_1, θ_2) in the cases: i) $\rho_2 = \rho_1^2$, ii) $\rho_2 = \rho_1$.

Proposition 4.1 Let $U \sim U(0, 1)$ and consider the random variables $S_1 = F_1^-(U)$ and $S_2 = F_2^-(U)$, where F_1 and F_2 are two geometric distribution functions with respective parameters ρ_1 and ρ_2. Assume without loss of generality that $\rho_1 \geq \rho_2$. Then the joint distribution of (S_1, S_2) is given as

$$P(S_1 = j\ , S_2 = k) = \begin{cases} \rho_2^k - \rho_1^{\lfloor ka \rfloor + 1} & j = \lfloor ka \rfloor \\ (1-\rho_1)\rho_1^j & \lfloor ka \rfloor < j \leq \lfloor (k+1)a \rfloor - 1 \\ \rho_1^{\lfloor (k+1)a \rfloor} - \rho_2^{k+1} & j = \lfloor (k+1)a \rfloor \\ 0 & \text{otherwise,} \end{cases}$$

where $a \in \mathcal{R},\ a \geq 1$ such that $\rho_2 = \rho_1^a$, and for any $x \in \mathcal{R}$, $\lfloor x \rfloor$ is the integer part of x. The proof is given in the Appendix.

Case 1. ($a = 2$, i.e. $\rho_2 = \rho_1^2$). Before we proceed we remark that in our calculations, depending on convenience, we will interchange between ρ_2 and ρ_1^2. From Proposition 4.1 we have that

$$P(S_1 = j\ , S_2 = k) = \begin{cases} \rho_2^k(1-\rho_1) & j = 2k \\ (1-\rho_1)\rho_2^k\rho_1 & j = 2k+1 \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 1 for a graphical representation of F_1 and F_2. The correlation between θ_1 and θ_2 is given in the following proposition.

Proposition 4.2 Under the model described in Case 1,

$$\mathrm{corr}(\theta_1, \theta_2) = \frac{2\rho_2(1-\rho_1)}{1-\rho_2}.$$

The proof is given in the Appendix.

Now, $\pi(\theta_1, \theta_2)$ is given by

$\int_0^1 \pi(\theta_1|S_1(u))\, \pi(\theta_2|S_2(u))\, \mathrm{d}u$

$= \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} \pi(\theta_1|j)\, \pi(\theta_2|k)\, P(S_1 = j\ , S_2 = k)$

$= \sum_{k=0}^{\infty} \rho_2^k(1-\rho_1)\ \mathrm{Ga}(\theta_2|k+1,\ r/\rho_2)\ \{\mathrm{Ga}(\theta_1|2k+1,\ r/\rho_1) + \rho_1 \mathrm{Ga}(\theta_1|2k+2,\ r/\rho_1)\}.$

The marginal distribution of θ_1 is

$$\sum_{k=0}^{\infty} \rho_2^k (1-\rho_1)\{\mathrm{Ga}(\theta_1|2k+1,\ r/\rho_1) + \rho_1 \mathrm{Ga}(\theta_1|2k+2,\ r/\rho_1)\}$$

$$= (1-\rho_1)\exp\{-(r\theta_1/\rho_1)\} \sum_{k=0}^{\infty} \rho_2^k \left\{ (r/\rho_1)^{2k+1} \frac{\theta_1^{2k}}{\Gamma(2k+1)} + \rho_1 (r/\rho_1)^{2k+2} \frac{\theta_1^{2k+1}}{\Gamma(2k+2)} \right\}$$

$$= \left(\frac{1-\rho_1}{\rho_1}\right) r \exp\{-(r\theta_1/\rho_1)\} \sum_{k=0}^{\infty} \left\{ r^{2k} \frac{\theta_1^{2k}}{\Gamma(2k+1)} + r^{2k+1} \frac{\theta_1^{2k+1}}{\Gamma(2k+2)} \right\}$$

$$= \left(\frac{1-\rho_1}{\rho_1}\right) r \exp\{-(r\theta_1/\rho_1)\} \sum_{k=0}^{\infty} \frac{(r\theta_1)^k}{\Gamma(k+1)}$$

$$= \lambda_1 \exp\{-\lambda_1 \theta_1\}.$$

Therefore, $\theta_1 \sim \exp(\lambda_1)$.

Similarly, it can be shown that $\theta_2 \sim \exp(\lambda_2)$. Therefore, the joint density of (θ_1, θ_2) is a bivariate exponential distribution with

$$\mathrm{corr}(\theta_1, \theta_2) = \frac{2\rho_2(1-\rho_1)}{1-\rho_2}.$$

Case 2. ($a = 1$, i.e. $\rho_2 = \rho_1$). This is the case where the marginals of the bivariate prior distribution are the same. Therefore, we consider the following model,

$$\theta_1 \sim \exp(\lambda)$$

with

$$S|\theta_1 \sim \mathrm{Poisson}(r\theta_1)$$

and

$$\theta_2|[S=s] \sim \mathrm{Gamma}(s+1, r+\lambda).$$

Here in expanded form, $S = \sum_{i=1}^{r} Z_i$. Note we obtain the same model whether we sample S from a Poisson$(r\theta_1)$, or separately each Z_i from a Poisson(θ_1). In this case we can compute the correlation directly,

$$\begin{aligned}\mathrm{corr}(\theta_1, \theta_2) &= \tfrac{\mathrm{Var}\{\mathrm{E}(\theta|S)\}}{\mathrm{Var}(\theta)} \\ &= \tfrac{\mathrm{Var}(S)}{\mathrm{Var}(S)+\mathrm{E}(S+1)} \\ &= \tfrac{r}{r+\lambda}.\end{aligned}$$

The prior is given by

$$\pi(\theta_1,\theta_2) = \sum_{s=0}^{\infty} \lambda \exp(-\lambda\theta_1) \left\{ \frac{(r\theta_1)^s \exp(-r\theta_1)}{s!} \right\} \left\{ \frac{(r+\lambda)^{s+1}\theta_2^s \exp(-(r+\lambda)\theta_2}{\Gamma(s+1)} \right\}$$

$$= (r+\lambda)\exp\{-(r+\lambda)(\theta_1+\theta_2)\} \sum_{s=0}^{\infty} \frac{(r(r+\lambda)\theta_1\theta_2)^s}{(s!)^2}.$$

If we set $\rho = r/(r+\lambda)$ then the above becomes

$$\pi(\theta_1,\theta_2) = \frac{\lambda^2}{1-\rho} \exp\left\{ \frac{-\lambda(\theta_1+\theta_2)}{1-\rho} \right\} \sum_{s=0}^{\infty} \frac{(\rho\lambda^2\theta_1\theta_2)^s}{(1-\rho)^{2s}(s!)^2}$$

$$= \frac{\lambda^2}{1-\rho} \exp\left\{ \frac{-\lambda(\theta_1+\theta_2)}{1-\rho} \right\} I_0\left(\frac{2(\rho\lambda^2\theta_1\theta_2)^{1/2}}{1-\rho} \right),$$

where

$$I_0(x) = \sum_{k=0}^{\infty} (x/2)^{2k}/(k!)^2$$

is the modified Bessel function of the first kind, of order zero.

The above distribution is a member of the bivariate exponential distribution with parameters $\lambda_1 = \lambda_2 = \lambda$ and $\rho = r/(r+\lambda)$ which is a special case of the bivariate gamma distribution. For more on these bivariate distributions, see Downton (1970) and Iliopoulos (2002). In Figure 2 we plot graphs of the bivariate exponential prior with $\lambda = \lambda_1 = \lambda_2 = 2.3$, $\rho = \rho_1 = \rho_2$ and $a = 1$. From top left going clockwise we have $(\rho = 0.3, r = 1)$, $(\rho = 0.6, r = 4)$, $(\rho = 0.8, r = 9)$, and $(\rho = 0.95, r = 44)$. These graphs demonstrate the convergence to the density with all the mass concentrated on $\theta_1 = \theta_2$.

To sample from the posterior we need the following full conditionals,

1. $\theta_1 | S, \mathbf{X^n} \sim \text{Gamma}(S + \sum_{i=1}^{n} X_i + 1, r + n + \lambda)$
2. $\theta_2 | S, \mathbf{Y^m} \sim \text{Gamma}(S + \sum_{i=1}^{m} Y_i + 1, r + m + \lambda)$
3. $$\mathrm{P}(S = s|\theta_1,\theta_2) = \frac{\{\theta_1\theta_2\, r(r+\lambda)\}^s}{I_0\left\{2\sqrt{\theta_1\theta_2\, r(r+\lambda)}\right\} s!^2}.$$

Distribution 3. is known as the Bessel distribution with parameters $\nu = 0$ and $a = 2\{\theta_1\theta_2 r(r+\lambda)\}^{1/2}$. For more details on the Bessel distribution see Yuan and Kalbfleisch (2000). All theses distributions are easy to sample.

5. Posterior analysis. We will refer to the set of latent variables as

$$\mathbf{Z^r} = \{Z_1^{(1)}, \ldots, Z_r^{(1)}, Z_1^{(2)}, \ldots, Z_r^{(2)}\}.$$

Given $\mathbf{Z^r}$, a priori θ_1 and θ_2 are independently distributed. Thus, the posterior updating can be done separately for each θ_1 and θ_2 conditionally on $\mathbf{Z^r}$. The data are such that given θ_1, $\mathbf{X^n} = (X_1, \ldots, X_n)$ are independent and identically distributed observations from $F(\cdot|\theta_1)$, and that given θ_2, $\mathbf{Y^m} = (Y_1, \ldots, Y_m)$ are independent and identically distributed observations from $F(\cdot|\theta_2)$. Therefore, given these facts, it is clear that we can take advantage of sampling methods, such as the Gibbs sampler and, more generally, MCMC (Smith and Roberts, 1993; Tierney, 1994).

It is quite obvious then that given $\mathbf{Z^r}$, we have the conditional posterior distribution for θ_1 being given, up to proportionality, by

$$\pi(\theta_1|\mathbf{Z^r}, \mathbf{X^n}) \propto \prod_{i=1}^{r} f(Z_i^{(1)}|\theta_1) \prod_{i=1}^{n} f(X_i|\theta_1)\, \pi(\theta_1|\lambda_1)$$

and the conditional posterior distribution for θ_2 being, up to proportionality, given by

$$\pi(\theta_2|\mathbf{Z^r}, \mathbf{Y^m}) \propto \prod_{i=1}^{r} f(Z_i^{(2)}|\theta_2) \prod_{j=1}^{m} f(Y_j|\theta_2)\, \pi(\theta_2|\lambda_2).$$

Thus it is seen that the latent data contribute in a nice way to the likelihood function and if $f(\cdot|\theta)$ and $\pi(\theta)$ form a conjugate pair then we have conditional conjugacy.

Finally, to complete the picture in order to undertake a Gibbs sampler, we require the full conditional distribution for $\mathbf{Z^{(1)}}$. This is given up to proportionality by

$$f(\mathbf{Z^{(1)}}|\theta_1, \theta_2) \propto \frac{f(\mathbf{Z^{(1)}}|\theta_1)\, f(\mathbf{Z^{(2)}}|\theta_2)}{m_1(\mathbf{Z^{(1)}})}.$$

It is perhaps always going to be easier to sample the $Z_i^{(1)}$'s individually rather than together. The conditional density for this is given by

$$f(Z_i^{(1)}|Z_{-i}^{(1)}, \theta_1, \theta_2) \propto \frac{f(Z_i^{(1)}|\theta_1)\, f(Z_i^{(2)}|\theta_2)}{m(Z_i^{(1)}|Z_{-i}^{(1)})}.$$

Typically these distributions will be straightforward to sample.

6. Application in the estimation of the ROC curve. The examination of summary measures which are used for discriminating between two distributions has always been a matter of interest to statisticians. A familiar situation is, for example, when one wants to assess the accuracy of a diagnostic marker when measurements arise from a healthy and a diseased population.

A useful statistical tool for the comparison of these two populations is the Receiver Operating Characteristic (ROC) curve and the associated area under this curve. To first introduce some necessary notation, let $\mathbf{X^n} = (X_1, \ldots, X_n)$ and $\mathbf{Y^m} = (Y_1, \ldots, Y_m)$ be two sequences of observations which represent random samples from two distributions, say, F_1 and F_2. The observations could be measurements from two groups and are possibly closely related. In our diagnostic marker situation, F_1 could represent the healthy individuals and F_2 the diseased ones.

Using terms from our medical example, the ROC curve is a plot of the true positive rates (the probability that the marker shows high values when in group $\{D =$ diseased$\}$) and the false positive rates (the probability that the marker shows high values when in group $\{H =$ healthy$\}$). In other words, it is the plot of the sensitivity against the 1-specificity, or

$$(1 - F_1(c), 1 - F_2(c)), \; -\infty \leq c \leq \infty.$$

When F_1 is absolutely continuous there is a closed form expression for the ROC curve, given as the plot of

$$\mathrm{ROC}(t) = 1 - F_2\left\{F_1^{-1}(1-t)\right\}, \;\; 0 \leq t \leq 1.$$

On the other hand, when F_1 is discrete, the ROC curve is a set of points which, by convention, are joined by straight line segments.

If one, or both, of the F_1 and F_2 are unknown we could estimate them by their empirical distribution functions and then plot the empirical ROC curve; that is, the graph of

$$(1 - F_{1n}(c_i), 1 - F_{2n}(c_i))$$

for a choice of k cut-off values $\{c_1, \ldots, c_k\}$. See, for example, Hsieh and Turnbull (1996) and Nakas et al. (2003).

This indicates how close the two distributions are or how well the diagnostic marker can distinguish between the two distributions F_1 and F_2. A frequently used summary measure of diagnostic accuracy is the area under the curve(AUC) given by

$$\mathrm{AUC} = P(X < Y) + (1/2)P(X = Y).$$

The parametric area under the curve (PAUC) is given by

$$\text{PAUC}(\theta_1, \theta_2) = P(X < Y|\theta_1, \theta_2) + (1/2)P(X = Y|\theta_1, \theta_2)$$

where θ_1 and θ_2 are points drawn from the posterior distribution of (θ_1, θ_2).

Here we present and discuss a simulated-data example using the model outlined in Section 4 and Case 2. We simulated data using $\theta_1 = 1$ and $\theta_2 = 2$ and then drew independent samples $\mathbf{X^n} = (X_1, \ldots, X_n)$, where $X_i|\theta_1 \sim \text{Poisson}(r\theta_1), \quad i = 1, \ldots, n$, and $\mathbf{Y^n} = (Y_1, \ldots, Y_n)$, where $Y_j|\theta_2 \sim \text{Poisson}(r\theta_2), \quad j = 1, \ldots, n$. We fixed the values of λ to be 1 and so the relation between ρ and r is given by $\rho = r/(r+1)$.

The results are summarised in Figure 3 and Figure 4 where various histograms of the PAUC are drawn for various values of n and r.
We present some interesting remarks:

1. In Figure 3, for small $\rho\,(= 0.3)$, that is when θ_1 and θ_2 are almost uncorrelated the PAUC tends to be centered around its true value which is 0.711 (see Figure 3a). As n increases the PAUC becomes more concentrated around this value (see Figures 3b and 3c).

2. In Figure 4, for large $\rho\,(= 0.95)$, that is when θ_1 is approximately equal to θ_2, and for relatively small n, the prior influence is considerable and the PAUC tends to concentrate around 0.5 which is the value of the PAUC when $\theta_1 = \theta_2$ (see Figure 4a). As n gets larger the sample becomes more dominant and again we have concentration around 0.711 (see Figure 4b and 4c).

So the role of r is purely determining the correlation between θ_1 and θ_2,and as mentioned in Section 1, it is possible to allow r here to take non-integer values.

7. Discussion. These results are encouraging. We have presented easy to work with and well motivated bivariate prior distributions by exploiting the ideas of de Finetti and partial exchangeability. We believe the bivariate prior distributions have a naturally interpretable dependence relation and the degree of correlation is understandable via a single parameter which is easy to understand via thoughts on prior sample sizes. Moreover, the latent variable approach makes sampling from posterior distributions quite easy via a Gibbs sampler.

Here we include some further ideas. One is the possibility of introducing a negatively correlated prior distribution by taking

$$Z_j^{(1)} = F_1^{-1}(U_j)$$

with

$$Z_j^{(2)} = F_2^{-1}(1 - U_j)$$

for all $j = 1, \ldots, r$. This is worth exploring, as is the idea of choosing different parametric families $F_1(\cdot|\theta_1)$ and $F_2(\cdot|\theta_2)$. If the families are different then the prior families could also be different in order to preserve conjugacy.

On the other hand, extending our ideas to a p-variate prior distributions seems problematical to us. One idea could be to construct a random tree with p nodes and so each pair of variates have a random common branch. This would also need to be explored.

Finally, we discuss the correlation associated with our bivariate models. The discreteness of r implies a discrete range the correlation. We believe however that the simple interpretation of r, as a sample size for which the two branches share a common stem, compensates for this to a large degree.

Note also that in the Poisson-exponential case, we could actually extend the r to the positive reals by considering the sufficient statistic based model. This of course gives us a full range for the correlation with no gaps. We are not sure of when this can happen for other models but will be applicable for the class of convolution-closed exponential family.

Acknowledgements. The research of the third author is financed by an Advanced Research Fellowship from the U.K. Engineering and Physical Science Research Council.

References

Doob, J.L. (1949). Application of the theory of martingales. In *Le Calcul des Probabilités et ses Applications* 23–27. Colloques Internationaux du Centre National de la Recherche Scientifique, Paris.

Downton F. (1970). Bivariate exponential distributions in reliability theory. *J. Roy. Statist. Soc. B* **32**, 408–417.

Hsieh, F. and Turnbull, B.W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve . *Ann. Statist.* **24**, 25–40.

Iliopoulos, G. (2003). Estimation of parametric functions in Downton's bivariate exponential distributions. *J. Statist. Plann. Inference* **117**, 169–184.

NAKAS, C., YIANNOUTSOS, C.T., BOSCH, R.J. and MOYSSIADIS, C. (2003). Assessment of diagnostic markers by goodness-of-fit tests. *Statist. Med.* **22**, 2503–2513.

WALKER, S.G. and MULIERE, P. (2003). A bivariate Dirichlet process. *Statist. Probab. Lett.* **64**, 1–7.

YUAN, L. and KALBFLEISCH, J.D. (2000). On the Bessel distribution and related problems. *Ann. Inst. Statist. Math.* **90**, 577–588.

Appendix

Proof of Proposition 4.1 For any given set of parameters ρ_1, ρ_2 there exists an $a \in \mathcal{R}$, $a \geq 1$, such that $\rho_2 = \rho_1^a$. For any given $s_2 = 0, 1 \dots$ let

$$j(s_2) = \sup\{j : F_1(j) \leq F_2(s_2)\} = \sup\{j : 1 - \rho_1^{j+1} \leq 1 - \rho_2^{s_2+1}\}$$

$$= \sup\{j : \rho_1^{j+1} \geq \rho_1^{a(s_2+1)}\} \Rightarrow j(s_2) = \lfloor (s_2+1)a \rfloor - 1.$$

The number $j(s_2)$ indicates how "dense" F_1 is in F_2. Namely, how many values of F_1 can be found between two successive values of F_2. For any $j, k = 0, 1 \dots$,

$$P(S_1 = j\ , S_2 = k) = P(u : S_1(u) = j, S_2(u) = k)$$

$$= \begin{cases} F_1(\lfloor ka \rfloor) - F_2(k) & j = \lfloor ka \rfloor \\ P(S_1 = j) & \lfloor ka \rfloor < j \leq j(k) = \lfloor (k+1)a \rfloor - 1 \\ F_2(k) - F_1(j(k)) & j = j(k) + 1 = \lfloor (k+1)a \rfloor \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \rho_2^k - \rho_1^{\lfloor ka \rfloor + 1} & j = \lfloor ka \rfloor \\ (1-\rho_1)\rho_1^j & \lfloor ka \rfloor < j \leq \lfloor (k+1)a \rfloor - 1 \\ \rho_1^{\lfloor (k+1)a \rfloor} - \rho_2^{k+1} & j = \lfloor (k+1)a \rfloor \\ 0 & \text{otherwise.} \end{cases}$$

Proof of Proposition 4.2 Now

$$\mathrm{E}(S_1S_2) = \sum_{k=0}^{\infty} k\{2k\rho_2^k(1-\rho_1)+(2k+1)(1-\rho_1)\rho_2^k\rho_1\} = \sum_{k=0}^{\infty} k(1-\rho_1)\rho_2^k\{2k(1+\rho_1)+\rho_1\}$$

$$= 2(1-\rho_1)(1+\rho_1)\sum_{k=0}^{\infty} k^2\rho_2^k + (1-\rho_1)\rho_1\sum_{k=0}^{\infty} k\rho_2^k$$

$$= \frac{2(1-\rho_1)(1+\rho_1)}{(1-\rho_2)}\left\{\frac{\rho_2(1+\rho_2)}{(1-\rho_2)^2}\right\} + \frac{(1-\rho_1)\rho_1\rho_2}{(1-\rho_2)^2}$$

and

$$\mathrm{Cov}(S_1, S_2) = \frac{2\rho_2(1+\rho_2)}{(1-\rho_2)^2} + \frac{(1-\rho_1)\rho_1\rho_2}{(1-\rho_2)^2} - \frac{\rho_1\rho_2}{(1-\rho_1)(1-\rho_2)} = \frac{2\rho_2}{(1-\rho_2)^2}$$

with

$$\mathrm{Var}(S_i) + \mathrm{E}(S_i) + 1 = 1/(1-\rho_i)$$

for $i = 1, 2$. Therefore,

$$\mathrm{corr}(\theta_1, \theta_2) = \frac{2\rho_2(1-\rho_1)}{1-\rho_2}.$$

Notice that $\mathrm{corr}(\theta_1, \theta_2) \leq 1$ since $\rho_1(2\rho_1 - 1) \leq 1$.

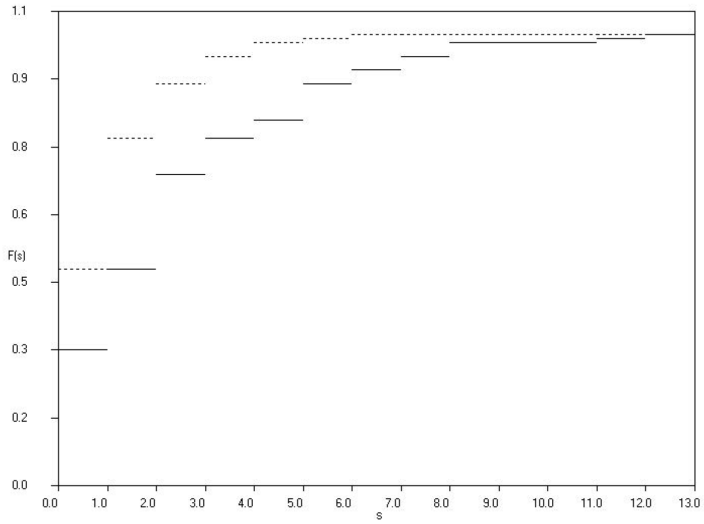


Figure 1: Graphs of $F_1(s)$ (----) and $F_2(s)$ (—), $\rho_2 = \rho_1^2$, $\rho_1 = 0.70$

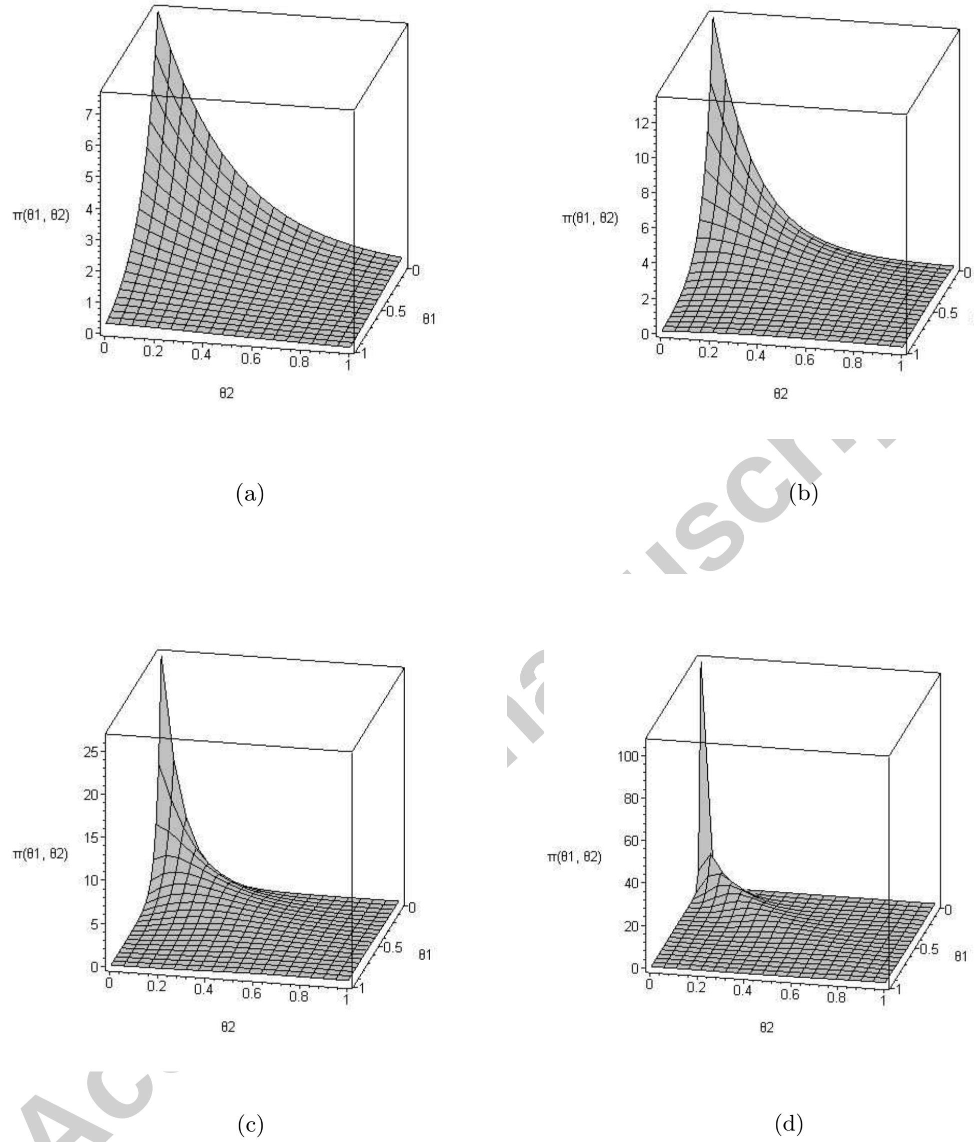
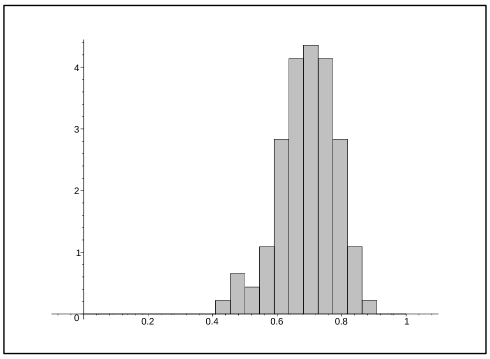
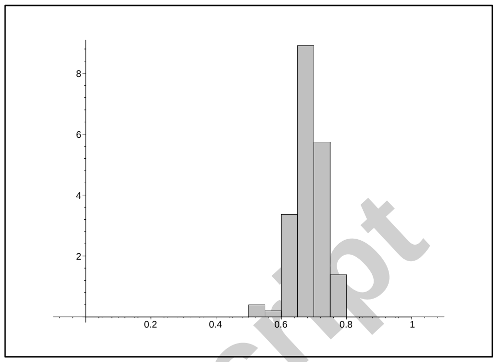


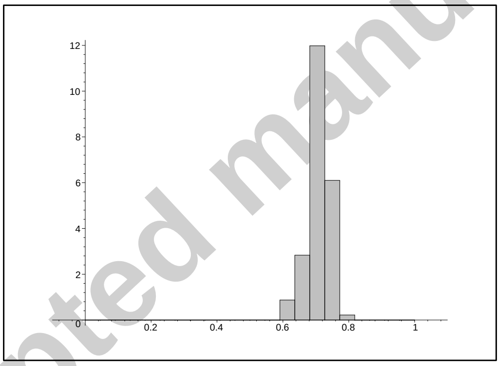
Figure 2: In all figures $\lambda = 2.3$, and, θ_1, θ_1 in $[0,1]$. In figure (a) we have $\rho = 0.30$ ($r = 1$), in figure (b) we have $\rho = 0.60$ ($r = 4$), in figure (c) we have $\rho = 0.80$ ($r = 9$), in figure (d) we have $\rho = 0.95$ ($r = 44$)



(a)

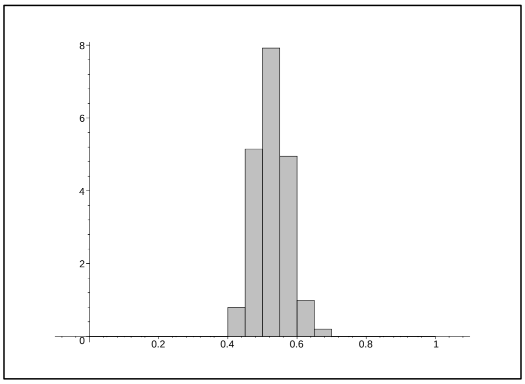


(b)

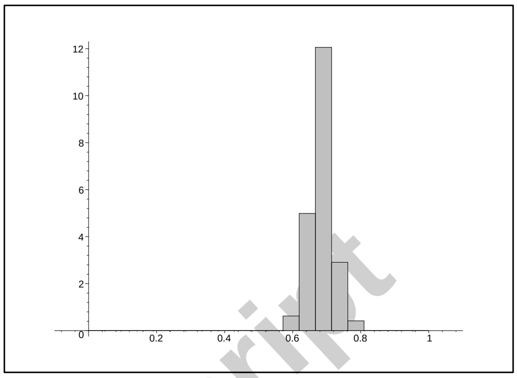


(c)

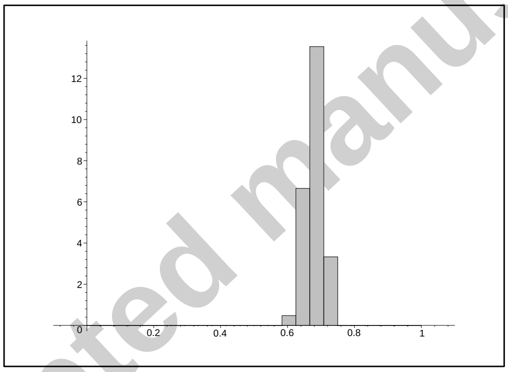
Figure 3: In all figures $\lambda = 2.3$, $r = 1$ ($\rho = 0.3$), in figures (a), (b) and (c) we take $m = n = 10$, 50, 100 respectively.



(a)



(b)



(c)

Figure 4: In all figures $\lambda = 2.3$, $r = 44$ ($\rho = 0.95$), in figures (a), (b) and (c) we take $m = n = 10$, 50, 100 respectively.