Supplemental Material for

"Bayesian Kernel Mixtures for Counts"

Antonio Canale* & David B. Dunson†

A. PROOF OF THEOREMS

To prove Lemma 1 we have first to ensure the existence of at least one element in $\mathcal{L}$ for every $p \in \mathcal{C}$. This is done in the following Lemma.

Lemma. *For every count measure $p_0 \in \mathcal{C}$ and rounding function $g(\cdot)$ defined in (3), there exists at least one $f_0 \in \mathcal{L}$ such that $g(f_0) = p_0$.*

Proof. The lemma is trivially proved by defining f_0 as a step function of the form

$$f_0(x) = \frac{p_0(0)}{a_1 - b} 1\!\!1_{[b,a_1)}(x) + \sum_{h=1}^{\infty} \frac{p_0(j)}{a_{j+1} - a_h} 1\!\!1_{[a_j,a_{j+1})}(x),$$

where $1\!\!1_A(x)$ is 1 iff $x \in A$ and b is an arbitrary number such that (b, a_1) is in the domain of f. □

B. ALGEBRAIC DETAILS TO CENTER THE ROUNDED MIXTURE OF GAUSSIANS PRIOR

Assuming the prior specified in Section 2.2, with

$$P_0 = N(\mu; \mu_0, \kappa\tau^{-1}) Ga(\tau; \nu/2, \nu/2)$$

*Dip. Scienze Statistiche, Università di Padova, 35121 Padova, Italy (*canale@stat.unipd.it*)
†Dept. Statistical Science, Duke University, Durham, NC 27708, USA (*dunson@stat.duke.edu*)

we have

$$
\begin{aligned}
\mathrm{E}\{F(a_j)\} &= \mathrm{E}\left\{\sum_{h=1}^{\infty} \pi_h \Phi(a_j;\mu_h,\tau_h^{-1})\right\} \\
&= \int_{R\times R^+} \Phi(a_j;\mu,\tau^{-1}) N(\mu;\mu_0,\tau^{-1}\kappa) Ga(\tau;\nu/2,\nu/2)\, d\mu\, d\tau \\
&= \int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{a_j} N(y^*;\mu,\tau^{-1}) N(\mu;\mu_0,\tau^{-1}\kappa) Ga(\tau;\nu/2,\nu/2)\, d\mu\, d\tau\, dy^* \qquad \text{(A.1)}
\end{aligned}
$$

Marginalizing out μ from (A.1) we get

$$
\mathrm{E}\left\{F(a_j)\right\} = \int_{-\infty}^{a_j}\int_0^{\infty} N(y^*;\mu_0,(\kappa+1)/\tau) Ga(\tau;\nu/2,\nu/2)\, d\tau\, dy^*.
$$

while marginalizing out τ we obtain

$$
\mathrm{E}\left\{F(a_j)\right\} = \int_{-\infty}^{a_j} t_\nu(y^*;\mu_0,\kappa+1) dy^*
$$

that gives equation (5).

To obtain equation (6) we need to compute the second moment of $F_D(a_j, a_{j+1})$ as

$$
\begin{aligned}
\mathrm{E}\left\{F_D(a_j,a_{j+1})^2\right\} =& \mathrm{E}\left\{\left(\sum_{h=1}^{\infty} \pi_h \Phi_D(a_j,a_{j+1};\mu_h,\tau_h^{-1})\right)^2\right\} \\
=& \sum_{h=1}^{\infty} \mathrm{E}\left\{\left(\pi_h \Phi_D(a_j,a_{j+1};\mu_h,\tau_h^{-1})\right)^2\right\} + \\
&+ 2\sum_{k\neq l} \mathrm{E}\left\{\pi_k\pi_l \Phi_D(a_j,a_{j+1};\mu_k,\tau_k^{-1})\Phi_D(a_j,a_{j+1};\mu_l,\tau_l^{-1})\right\} \\
=& \sum_{h=1}^{\infty} \mathrm{E}\left\{\pi_h^2\right\} \mathrm{E}\left\{\Phi_D(a_j,a_{j+1};\mu_h,\tau_h^{-1})^2\right\} + \\
&+ 2\sum_{k\neq l} \mathrm{E}\left\{\pi_k\pi_l\right\} \mathrm{E}\left\{\Phi_D(a_j,a_{j+1};\mu_k,\tau_k^{-1})\Phi_D(a_j,a_{j+1};\mu_l,\tau_l^{-1})\right\}.
\end{aligned}
$$

Using the stick-breaking construction of the π_h and the results on the variance of the beta distribution we have

$$\mathrm{E}\{F_D(a_j, a_{j+1})^2\} = \frac{1}{\alpha+1}\mathrm{E}\left\{\Phi_D(a_j, a_{j+1}; \mu, \tau^{-1})^2\right\} + \frac{\alpha}{\alpha+1}\mathrm{E}\left\{\Phi_D(a_j, a_{j+1}; \mu, \tau^{-1})\right\}^2$$

where the expectations are with respect to $(\mu, \tau) \sim P_0$. This leads to

$$\mathrm{Var}\{p(j)\} = \frac{1}{\alpha+1}\left[\mathrm{E}\left\{\Phi_D(a_j, a_{j+1}; \mu, \tau^{-1})^2\right\} - \mathrm{E}\left\{\Phi_D(a_j, a_{j+1}; \mu, \tau^{-1})\right\}^2\right]$$

and hence to equation (6).

C. SIMULATION STUDY RESULTS

The plots in Figure 7-10 report the results for the empirical coverage of 95% credible intervals for the $p(j)$s for all scenarios and sample size used in Section 2.5.

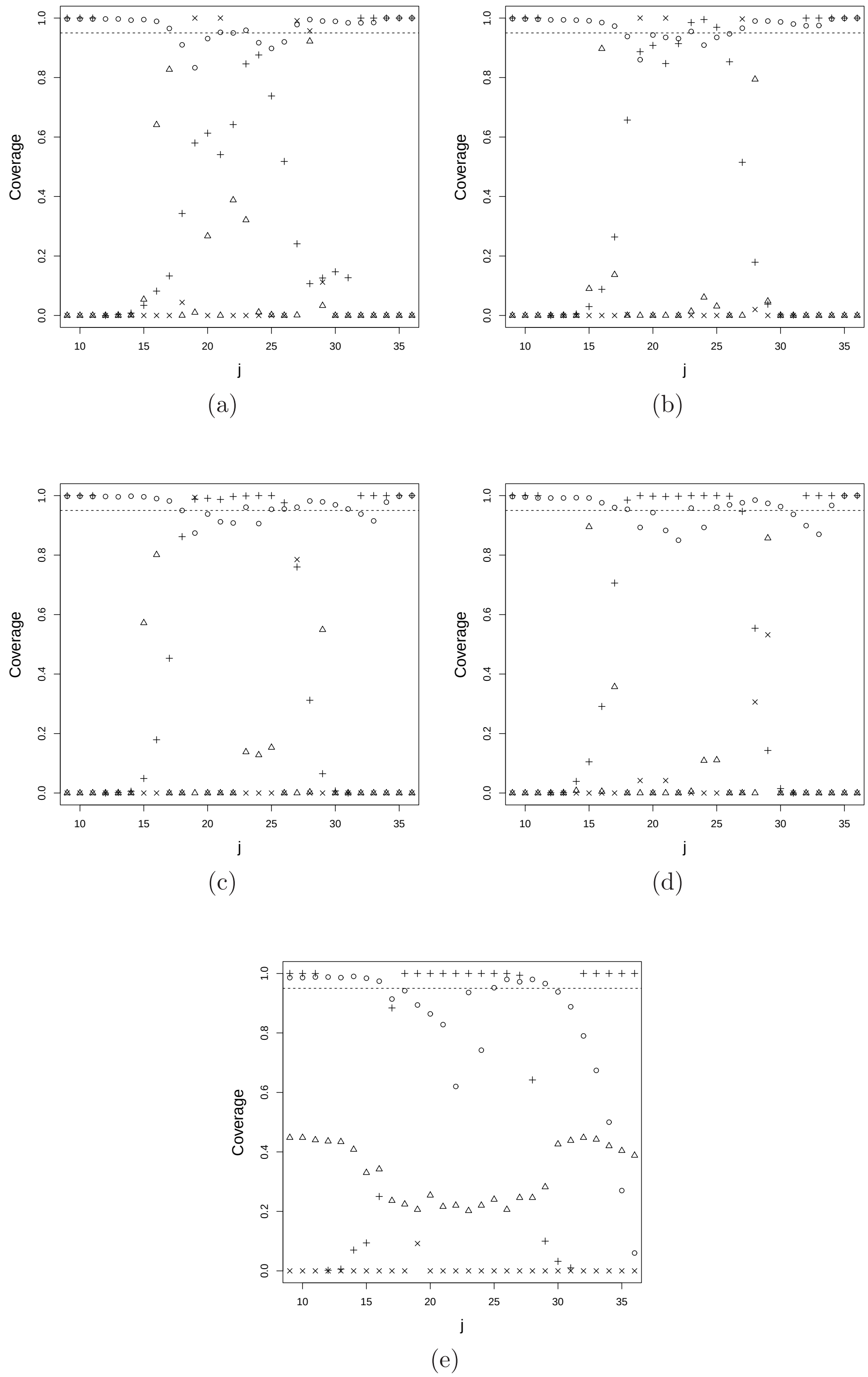


Figure 7: Coverage of 95% credible intervals for $p(j)$ under the first scenario. Points represent the RMG method, cross-shaped dots the DP with $\alpha = 1$, triangles the DP with $\alpha \sim Ga(1, 1)$ and x-shaped dots the DPM of Poisson. Sample size increases from top to bottom, namely $n = 10, 25, 50, 100, 300$.

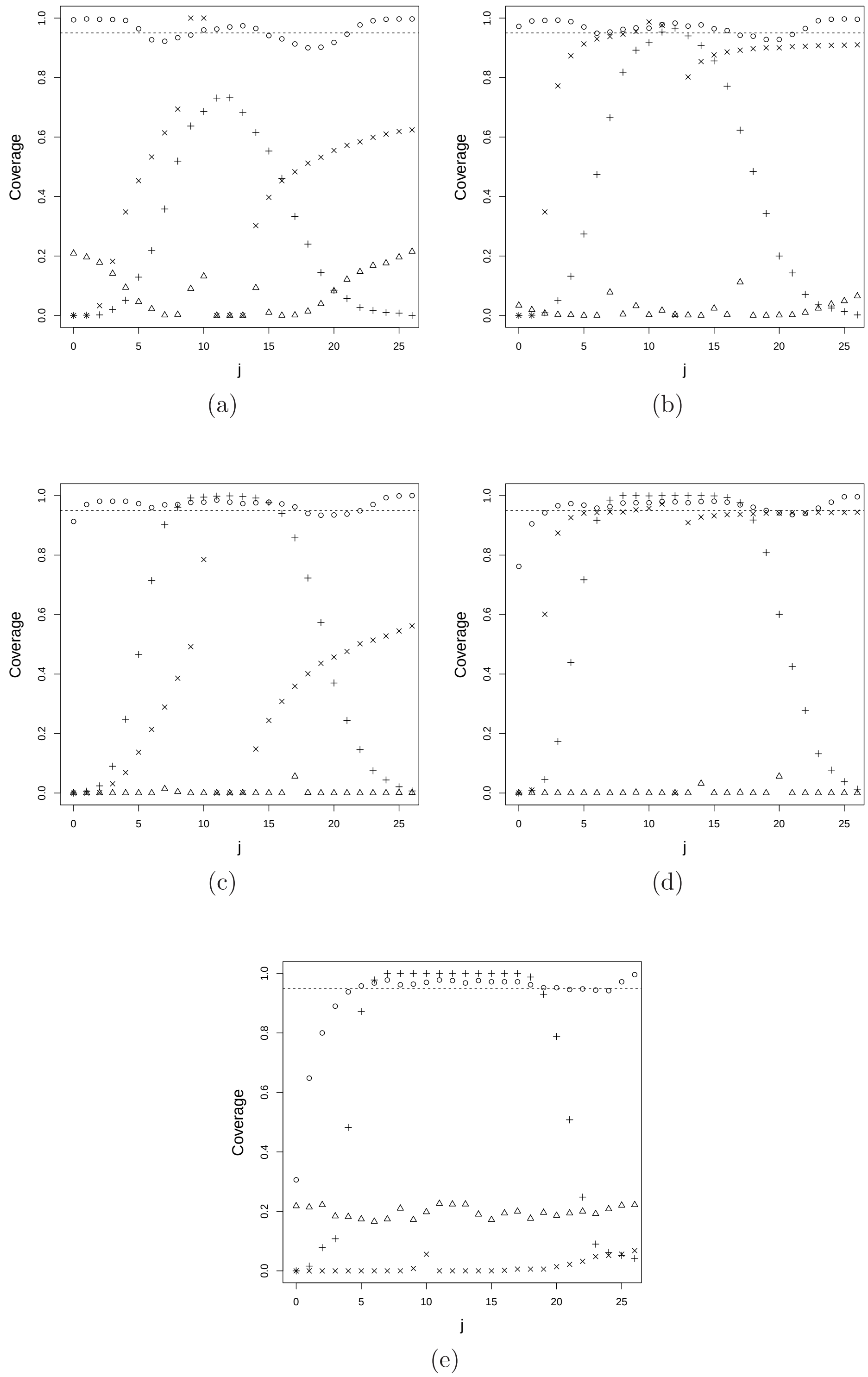
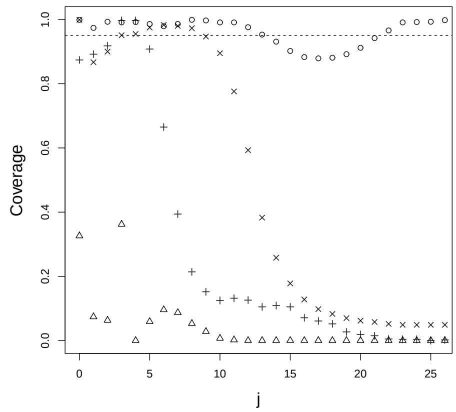
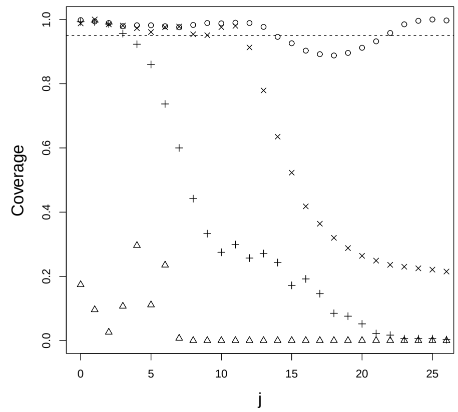


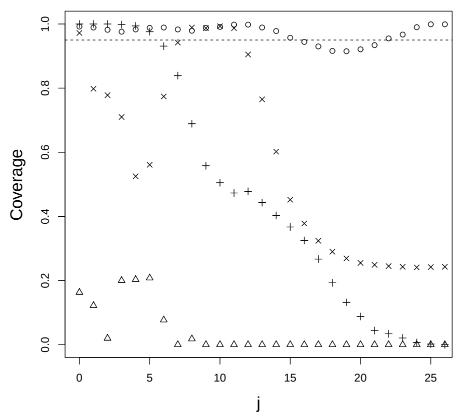
Figure 8: Coverage of 95% credible intervals for $p(j)$ under the second scenario. Points represent the RMG method, cross-shaped dots the DP with $\alpha = 1$, triangles the DP with $\alpha \sim Ga(1,1)$ and x-shaped dots the DPM of Poisson. Sample size increases from top to bottom, namely $n = 10, 25, 50, 100, 300$.



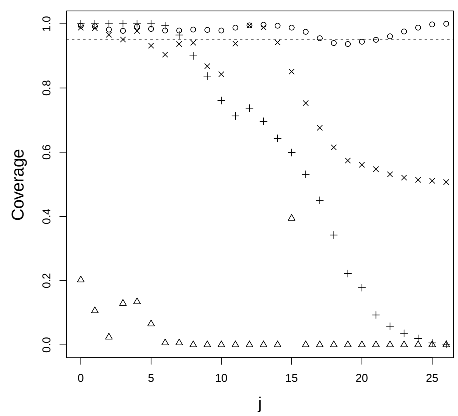
(a)



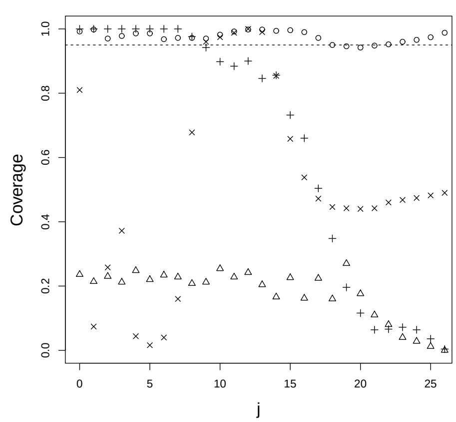
(b)



(c)

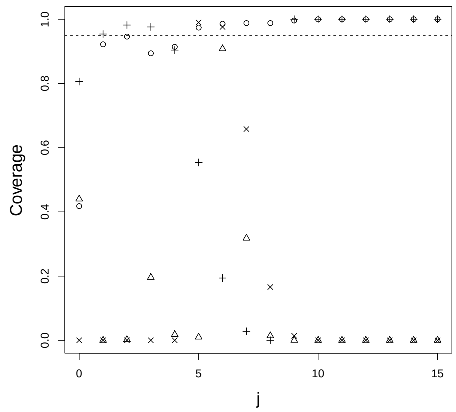


(d)

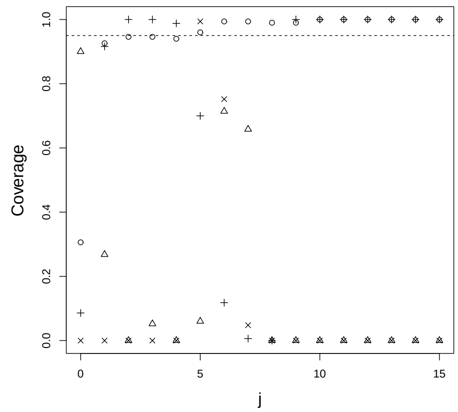


(e)

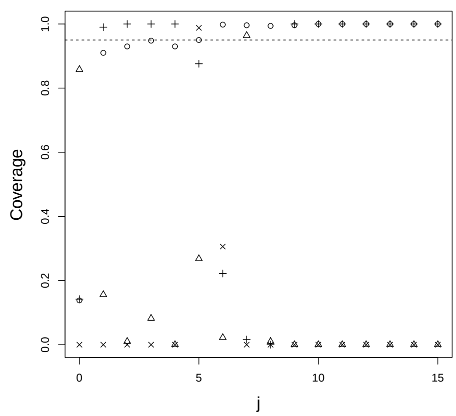
Figure 9: Coverage of 95% credible intervals for $p(j)$ under the third scenario. Points represent the RMG method, cross-shaped dots the DP with $\alpha = 1$, triangles the DP with $\alpha \sim Ga(1,1)$ and x-shaped dots the DPM of Poisson. Sample size increases from top to bottom, namely $n = 10, 25, 50, 100, 300$.



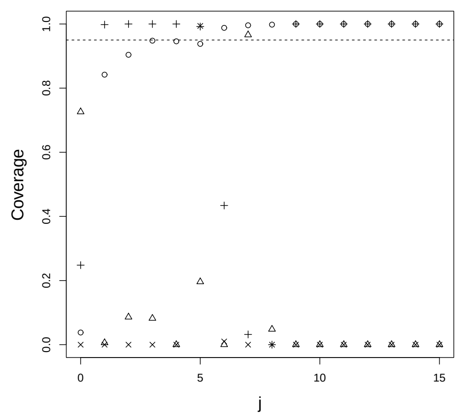
(a)



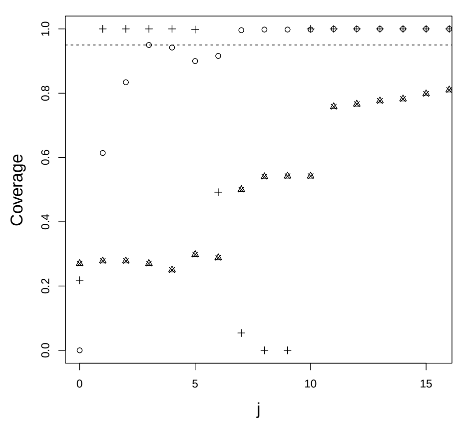
(b)



(c)



(d)



(e)

Figure 10: Coverage of 95% credible intervals for $p(j)$ under the fourth scenario. Points represent the RMG method, cross-shaped dots the DP with $\alpha = 1$, triangles the DP with $\alpha \sim Ga(1, 1)$ and x-shaped dots the DPM of Poisson. Sample size increases from top to bottom, namely $n = 10, 25, 50, 100, 300$.