Brazilian Journal of Probability and Statistics
2012, Vol. 26, No. 3, 279–287
DOI: 10.1214/10-BJPS134

A note on Bayesian robustness for count data

Jairo A. Fúquene and Moises Delgado

University of Puerto Rico

Abstract. The usual Bayesian approach for count data is Gamma/Poisson conjugate analysis. However, in this conjugate analysis the influence of the prior distribution could be dominant even when prior and likelihood are in conflict. Our proposal is an analysis based on the Cauchy prior for natural parameter in exponential families. In this work, we show that the Cauchy/Poisson posterior model is a robust model for count data in contrast with the usual conjugate Bayesian approach Gamma/Poisson model. We use the polynomial tails comparison theorem given in (*Bayesian Anal.* **4** (2009) 817–843) that gives easy-to-check conditions to ensure prior robustness. In short, this means that when the location of the prior and the bulk of the mass of the likelihood get further apart (a situation of conflict between prior and likelihood information), Bayes theorem will cause the posterior distribution to discount the prior information. Finally, we analyze artificial data sets to investigate the robustness of the Cauchy/Poisson model.

1 Introduction

In recent years, the Bayesian robustness methods have been very important in developments of Bayesian Analysis. We can find in the literature different proposals about robust priors. For example, in Dawid (1973), O'Hagan (1979), Berger (1985), Pericchi and Smith (1992) and Gelman et al. (2008), robust priors for location parameters are studied; however, for the Poisson likelihood there is no previously known clear results in Bayesian robustness. In Fúquene, Cook and Pericchi (2009), the Cauchy and Berger's robust heavy-tailed priors are considered and several mathematical results are presented. These authors obtain specific results for the Binomial and Normal likelihoods with applications to clinical trials. On the other hand, the proposal in this paper is to show the robustness of the Cauchy prior for the Poisson likelihood. Robust priors have bounded influence, in other words the prior is discounted automatically when there are conflicts between prior information and data. An important term in this paper is "The posterior mean is robust with respect to the prior" which is explained in the following definition (Fúquene (2009)):

Definition 1.1. Let λ be a random variable with prior distribution, $\pi(\lambda)$, with location parameter μ. The posterior mean is robust with respect to the prior, if and

Key words and phrases. Exponential family, polynomial tails comparison theorem, robust priors, Cauchy/Poisson model.

Received May 2010; accepted November 2010.

only if, the posterior mean remains bounded as $\mu \to +\infty$ or $\mu \to -\infty$. That is, the posterior mean is robust if there exists a constant M such that $-M < E(\lambda|\mathbf{y}) < M$.

Fúquene, Cook and Pericchi (2009) present a novel result, the polynomial tails comparison theorem, which gives easy-to-check conditions to ensure prior robustness for the natural parameter in exponential families. We can use this result for the Poisson likelihood, because the likelihood does not have to be location/scale. In this work, we show the robustness of the cauchy prior for the Cauchy/Poisson posterior model in contrast with the usual conjugate approach Gamma/Poisson posterior model. The paper proceeds as follows: in Section 2, we give a background of the Cauchy/Poisson and Gamma/Poisson posterior models. In Section 3, we study the prior specification and posterior moments of the Cauchy/Poisson model. In Section 4, we analyze artificial data sets to investigate the robustness of the Cauchy/Poisson model. We make some closing concluding remarks in Section 5.

2 The Poisson likelihood with conjugate and Cauchy priors

The Poisson likelihood arises in the study of data taking the form of counts. In words, let a sample of size n, $X_1, \ldots, X_n \sim \text{Poisson}(\theta)$. The Poisson distribution in the exponential family form is

$$f(\bar{X}_n|\lambda) \propto \exp(n\bar{X}_n\lambda - ne^{\lambda}), \tag{2.1}$$

where $\bar{X}_n = \sum_{i=1}^{n} X_i$ and the natural parameter is given by $\lambda = \log(\theta)$. The maximum likelihood estimator (MLE) of the natural parameter is $\hat{\lambda} = \log(\bar{X}_n)$. Now we perform a conjugate analysis, and express the Gamma(α, β) prior, after of the transformation of the parameter θ to $\lambda = \log(\theta)$, as

$$p_G(\lambda) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \exp(\lambda\alpha - \beta e^{\lambda}), \qquad \alpha, \beta > 0. \tag{2.2}$$

The cumulant generating function of the prior distribution $p_G(\lambda)$ is given by $E_G(e^{t\lambda}) \propto \log(\Gamma(\alpha + t)) - t\log(\Gamma(\beta))$, hence

$$E_G(\lambda) = \Psi(\alpha) - \log(\beta); \qquad V_G(\lambda) = \Psi'(\alpha), \tag{2.3}$$

where $\Psi(\cdot)$ is the digamma function and $\Psi'(\cdot)$ is the trigamma function (see Abramowitz and Stegun (1992)). The posterior distribution for the Gamma/Poisson model is given by

$$f_{\text{GP}}(\lambda|\bar{X}_n) = \frac{(\beta + n)^{\alpha + n\bar{X}_n}}{\Gamma(\alpha + n\bar{X}_n)} \exp\{(\alpha + n\bar{X}_n)\lambda - (\beta + n)e^{\lambda}\}; \quad \alpha, \beta > 0. \tag{2.4}$$

We have the cumulant generation function of the Gamma/Poisson model in closed form give as $E_{\text{GP}}(e^{t\lambda}|\bar{X}_n) \propto \log(\Gamma(\alpha + n\bar{X}_n + t)) - t\log(\Gamma(\beta + n))$, hence the posterior expectation and variance are given by

$$E_{\text{GP}}(\lambda|\bar{X}_n) = \Psi(\alpha + n\bar{X}_n) - \log(\beta + n);$$
$$V_{\text{GP}}(\lambda|\bar{X}_n) = \Psi'(\alpha + n\bar{X}_n). \tag{2.5}$$

On the other hand, we consider a Cauchy prior for the natural parameter $\lambda = \log(\theta)$

$$p_C(\lambda) = \frac{\beta}{\pi[\beta^2 + (\lambda - \nu)^2]}, \tag{2.6}$$

with parameters of location and scale ν and β, respectively, the posterior distribution of the C/P model is

$$f_{\text{CP}}(\lambda|\bar{X}_n) = \frac{\exp\{n\bar{X}_n\lambda - ne^{\lambda} - \log(\beta^2 + (\lambda - \nu)^2)\}}{p(\bar{X}_n)}, \tag{2.7}$$

where $p(\bar{X}_n)$ is the predictive marginal. Approaches to the approximation of $p(\bar{X}_n)$ are the Laplace's method, the rejection method and Markov chain Monte Carlo (MCMC) methods. We can see that the posterior (2.7) has the form

$$f_{\text{CP}}(\lambda|\bar{X}_n) = \exp\{\theta y - nM(\theta) + \rho(\theta) - c(y)\} \tag{2.8}$$

where $c(y) = \log(p(\bar{X}_n))$, $\rho(\theta) = \log(p_C(\lambda))$, $\theta y = n\bar{X}_n\lambda$ and $M(\theta) = e^{\lambda}$. Pericchi, Sanso and Smith (1993) show that posterior distributions that have the form (2.8) belong to the exponential family.

3 Computations with Cauchy and conjugate priors

Because the Cauchy/Poisson model has only one parameter, we can use the rejection method to find the posterior moments of this model (see Gamerman and Lopes (2006)). It is clear that the Cauchy density is an envelope, and it is simple to generate Cauchy distributed samples, so the method is well defined and feasible. The rejection method proceeds as follows:

1. Calculate $M = f(\bar{X}_n|\hat{\lambda})$.
2. Generate $\lambda_j \sim p_C(\lambda)$.
3. Generate $U_j \sim \text{uniform}(0, 1)$.
4. If $\text{MU}_j\, p_C(\lambda_j) < f(\bar{X}_n|\lambda_j)p_C(\lambda_j)$, accept λ_j. Otherwise reject λ_j and go to step 2.
5. Return to step 1 and repeat, until the desired sample $\{\lambda_j, j = 1, \dots, 10{,}000\}$ is obtained. The members in this sample will then be random variables from $f_{\text{CP}}(\lambda|\bar{X}_n)$.

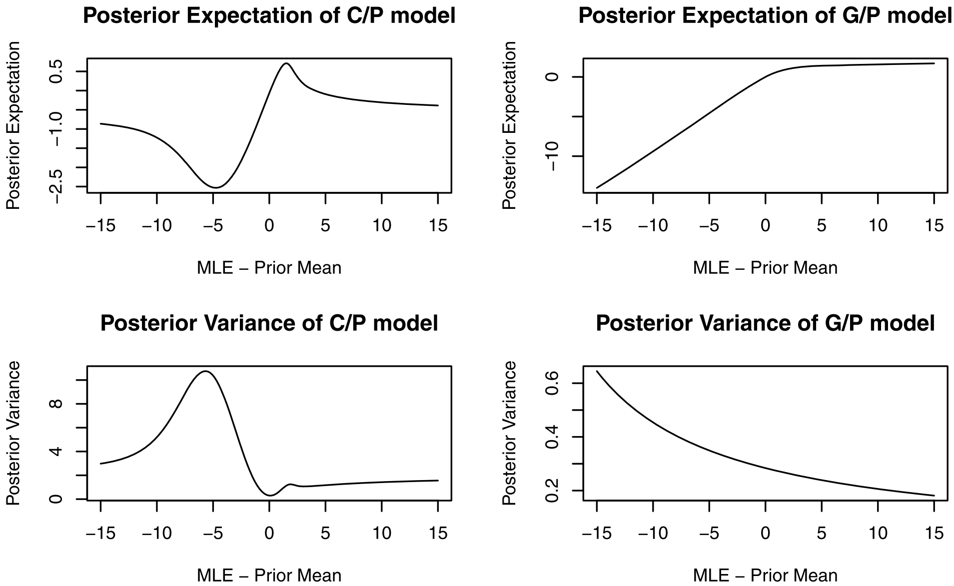

Figure 1 *Posterior expectation and posterior variance of the Cauchy/Poisson* (*C/P*) *and Gamma/Poisson* (*G/P*) *posterior models*.

We use Monte Carlo Methods for the posterior moments of the Cauchy/Poisson model.

In Figure 1, the MLE is kept fixed at $\log(\bar{X}_n) = 0$ and the prior location is moved to create a conflict between data and prior. With a Gamma prior the posterior expectation is unbounded. We can see that the posterior expectation of the Cauchy/Poisson posterior model is bounded. In other words, when prior and likelihood information are in conflict, the posterior expectation of the Cauchy/Poisson tends to MLE. In contrast to the Gamma/Poisson model, the posterior variance with the Cauchy prior is *not* monotonic in the conflict between the MLE and prior location.

Now, we show the Polynomial Tails Comparison theorem presented in Fúquene, Cook and Pericchi (2009). In order to decide if a Cauchy prior is robust with respect to a Poisson likelihood, the following theorem is useful and easy to apply.

Let $f(\lambda)$ be any likelihood function such that as $|\lambda| \to \infty$

$$\int_{|\lambda|>m} f(\lambda)\, d\lambda = \mathcal{O}(m^{-2-\varepsilon}). \tag{3.1}$$

For this paper, f is a Poisson distribution. Define

$$c(\lambda; \mu) = \frac{b}{\pi(b^2 + (\lambda - \mu)^2)} \tag{3.2}$$

for some $b > 0$. This is the Cauchy PDF with center μ and scale b. Denote by $\pi^C(\lambda|\text{data})$ and $\pi^U(\lambda|\text{data})$ the posterior densities employing the Cauchy and the Uniform prior densities respectively. Applying Bayes rule to both densities, it yields for any parameter value λ_0 the following ratio:

$$\frac{\pi^C(\lambda_0|\text{data})}{\pi^U(\lambda_0|\text{data})} = \frac{\int_{-\infty}^{\infty} f(\lambda)c(\lambda;\mu)\,d\lambda}{c(\lambda_0;\mu)\int_{-\infty}^{\infty} f(\lambda)\,d\lambda}.$$

Theorem 3.1. *For fixed* λ_0,

$$\lim_{\mu\to\infty} \frac{\int_{-\infty}^{\infty} f(\lambda)c(\lambda;\mu)\,d\lambda}{c(\lambda_0;\mu)\int_{-\infty}^{\infty} f(\lambda)\,d\lambda} = 1. \tag{3.3}$$

In other words, when there is a conflict between prior information and the sample information, the Cauchy prior effectively becomes an uniform prior, and in this precise sense the prior information is discounted. We need that the Poisson likelihood to be of order $(m^{-2-\varepsilon})$ in order to use Theorem 3.1.

Therefore, let $m > 0$ be such that $\forall\lambda > m$, $\frac{\lambda}{n} + \bar{X}_n\lambda < \exp(\lambda)$, so we have that

$$\exp(\lambda + n\bar{X}_n\lambda) < \exp(n\exp(\lambda)),$$
$$\frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))} < \frac{1}{\exp(\lambda)}$$

hence

$$\int_m^{\infty} \frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))}\,d\lambda < \int_m^{\infty} \frac{1}{\exp(\lambda)}\,d\lambda = -\frac{1}{\exp(\lambda)}\bigg|_m^{\infty} = \frac{1}{\exp(m)}.$$

Furthermore,

$$\lim_{m\to\infty} m^{2+\varepsilon}\int_m^{\infty} \frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))}\,d\lambda = \lim_{m\to\infty} \frac{m^{2+\varepsilon}}{\exp(m)} = 0. \tag{3.4}$$

On the other hand, for $\lambda < m < 0$

$$\int_{-\infty}^{m} \frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))}\,d\lambda < \int_{-\infty}^{m} \exp(n\bar{X}_n\lambda)\,d\lambda = \frac{1}{n\bar{X}_n}\exp(nx\lambda)\bigg|_{\infty}^{m} = \frac{\exp(n\bar{X}_n m)}{n\bar{X}_n}.$$

Furthermore,

$$\lim_{m\to\infty} m^{2+\varepsilon}\int_{-\infty}^{m} \frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))}\,d\lambda = \lim_{m\to\infty} \frac{m^{2+\varepsilon}\exp(n\bar{X}_n m)}{n\bar{X}_n} = 0. \tag{3.5}$$

From (3.4) and (3.5), when $m \to \infty$,

$$\int_{|\lambda|>m}^{\infty} \frac{\exp(n\bar{X}_n\lambda)}{\exp(n\exp(\lambda))}\,d\lambda = \mathcal{O}\left(\frac{1}{m^{2+\varepsilon}}\right).$$

Because of the Poisson likelihood is of order $(m^{-2-\varepsilon})$, we can use the polynomial comparison theorem to find the behavior of the posterior expectation of the Cauchy/Poisson model. Hence, we have the following result.

Corollary 3.1. *The posterior expectations for the Cauchy/Poisson and Gamma/Poisson posterior models satisfy the following*:

1. *Robust result*:

$$\lim_{\nu\to\pm\infty} E_{\mathrm{CP}}(\lambda|X_+) \approx \hat{\lambda} - \frac{1}{2ne^{\hat{\lambda}}}. \tag{3.6}$$

2. *Non-robust result*:

$$\lim_{E_G(\lambda)\to\pm\infty} E_{\mathrm{GP}}(\lambda|X_+) \to \pm\infty, \tag{3.7}$$

respectively.

Note: *the limit* (3.6) *is approximately equal to the MLE.*

Proof. Given the polynomial tails comparison theorem, we can use the uniform prior instead of the Cauchy prior when $\nu \to \pm\infty$ for the Poisson likelihood, the generating function for the Cauchy/Poisson model is

$$\lim_{\nu\to\pm\infty} E_{\mathrm{CP}}(e^{t\lambda}|\bar{X}_n) = \frac{\int_{-\infty}^{\infty} \exp\{n\bar{X}_n\lambda - ne^{\lambda} + t\lambda\}\,d\lambda}{\int_{-\infty}^{\infty} \exp\{n\bar{X}_n\lambda - ne^{\lambda}\}\,d\lambda}, \tag{3.8}$$

after of the transformation $\lambda = \log(\theta/(1-\theta))$, (3.8) is

$$\lim_{\nu\to\pm\infty} E_{\mathrm{CP}}(e^{t\lambda}|\bar{X}_n) = \frac{\Gamma(n\bar{X}_n + t)}{n^t\Gamma(n\bar{X}_n)}, \tag{3.9}$$

hence

$$\lim_{\nu\to\pm\infty} E_{\mathrm{CP}}(\lambda|\bar{X}_n) = \Psi(n\bar{X}_n) - \log(n), \tag{3.10}$$

the approximation of the Digamma function (Abramowitz and Stegun (1992)) is

$$\Psi(z) \approx \log(z) - \frac{1}{2z} - \mathcal{O}(z^{-2}), \tag{3.11}$$

hence

$$\lim_{\nu\to\pm\infty} E_{\mathrm{CP}}(\lambda|\bar{X}_n) \approx \log(n\bar{X}_n) - \frac{1}{2n\bar{X}_n} - \mathcal{O}((n\bar{X}_n)^{-2}) - \log(n). \tag{3.12}$$

With the Gamma prior, $E_G(\lambda) \approx \log(\alpha/\beta) - 1/2\alpha - \mathcal{O}(\alpha^{-2})$. We can see that $E_G(\lambda) \to \infty$ as $\alpha \to \infty$ and $E_G(\lambda) \to -\infty$ as $\beta \to \infty$, the approximation of the posterior expectation for the conjugate G/P model is

$$E_{\mathrm{GP}}(\lambda|\bar{X}_n) \approx \log(\alpha + n\bar{X}_n) - \frac{1}{2(\alpha + n\bar{X}_n)} - \mathcal{O}\big((\alpha + n\bar{X}_n)^{-2}\big) - \log(\beta + n)$$

and $E_{\mathrm{GP}}(\lambda|\bar{X}_n) \to \infty$ as $\alpha \to \infty$ and $E_{\mathrm{GP}}(\lambda|\bar{X}_n) \to -\infty$ as $\beta \to \infty$. □

4 Illustration

This example is taken from the blog called "Introduction to Bayesian Thinking" of Jim Albert (available electronically at http://learnbayes.blogspot.com). Suppose we are interested in learning about the proportion of official at-bats that are home runs, called the home run rate, λ, for Derek Jeter[1] before the start of the 2004 season. Suppose our prior beliefs are that the median is equal to 0.05 and the 90th percentile is equal to 0.081. On the other hand, the likelihood information is based in the number of at-bats and home runs hit by Jeter in the 2004 season.

We can obtain this data in jeter2004 contained in the LearnBayes package available from the Comprehensive R Archive Network at http://CRAN.R-project.org (R Development Core Team (2010)). We have with this information that Jeter obtains $\sum_{i=1}^{n} X_i = 23$ home runs in $n = 643$ at-bats.

Here the two priors that match this information are Cauchy and Gamma. For the Gamma prior, the parameters are $\alpha = 6$ and $\beta = 113.5$, in the Log-Odds scale the expectation and scale of the Gamma prior are respectively, $\Psi(6) - \log(113.5) = -3.02$ and $\sqrt{\Psi'(6)} = 0.42$. For the Cauchy prior, the location is the same as in Gamma prior and the scale can be calculated as $\beta = (\log(0.05) + 3.02)/\tan(\pi(0.9 - 1/2)) = 0.16$.

Figure 2 displays the Cauchy/Poisson and Gamma/Poisson posterior models. In this figure, we can see that the posterior and likelihood are very similar. In other words, when the prior and likelihood information are consistent for the Poisson

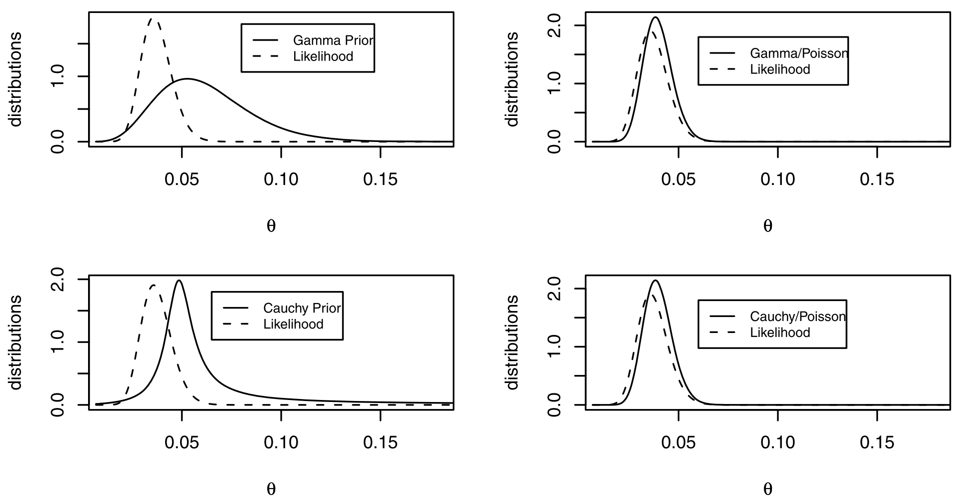

Figure 2 *Cauchy/Poisson and Gamma/Poisson when prior and likelihood are consistent.*

[1]Derek Sanderson Jeter is an American professional baseball player considered to be one of the best players of his generation.

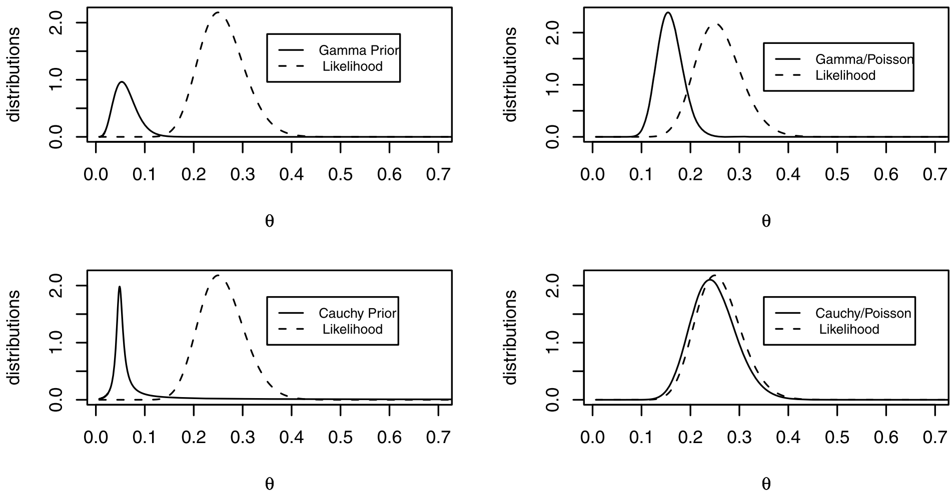

Figure 3 *Cauchy/Poisson and Gamma/Poisson when prior and likelihood are in conflict.*

likelihood the results are approximately equal with either Cauchy or Gamma priors.

On the other hand, suppose that Jeter during the 2004 season hits 30 home runs in 120 at-bats. In this case, Figure 3 displays the situation of conflict between prior and likelihood information. We can see that the Cauchy/Poisson model is more related with the sample data. In contrast, the weight of the Gamma prior is higher than in Cauchy/Poisson model. Figure 3 illustrates how the weight of the prior in the conjugate case is very high when prior and likelihood are in conflict.

5 Concluding remarks

(1) The Cauchy prior in the Cauchy/Poisson model is robust but the Gamma prior in the conjugate Cauchy/Poisson model for the inference of the Log-Odds is not.

(2) We can use the rejection method to calculate easily the posterior moments of the Cauchy/Poisson model.

(3) This approach has major application to several areas including for example the Poisson model with extra variation in Bayesian methods for ecology or in a poisson model parameterized in terms of rate and exposure.

(4) Finally, the use of a robust cauchy prior in the Cauchy/Poisson model with a hierarchical structure may be even more important, recent results of robust hierarchical models for a normal likelihood are shown in Perez and Pericchi (2009).

Acknowledgments

J. A. Fúquene is supported by PII—School of Business Administration, UPR-RRP. We thanks to the associated editor and referee by detailed comments which were very useful in preparing the last version of this paper.

References

Abramowitz, M. and Stegun, I. A. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. New York: Dover. MR1225604

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer-Verlag. MR0804611

Dawid, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60**, 664–667. MR0336889

Fúquene, J. (2009). Robust Bayesian priors in clinical trials: An R package for practitioners. *Biometric Brazilian Journal* **27**, 627–643.

Fúquene, J. A., Cook, J. D. and Pericchi, L. R. (2009). A case for robust Bayesian priors with applications to clinical trials. *Bayesian Analysis* **4**, 817–843. MR2570090

Gelman, A., Jakulin, A., Pittau, M. G. and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics* **2**, 1360–1383. MR2655663

O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society, Ser. B* **41**, 358–367. MR0557598

Perez, M. E. and Pericchi, L. R. (2009). The case for a fully robust hierarchical Bayesian analysis of clinical trials. Technical Report, Wharton School of Business.

Pericchi, L. R., Sanso, B. and Smith, A. F. M. (1993). Posterior cumulant relationships in Bayesian inference involving the exponential family. *Journal of the American Statistical Association* **88**, 1419–1426. MR1245379

Pericchi, L. R. and Smith, A. F. M. (1992). Exact and approximate posterior moments for a normal location parameter. *Journal of the Royal Statistical Society, Ser. B* **54**, 793–804. MR1185223

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available at http://www.R-project.org.

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd ed. London: Chapman & Hall. MR2260716

Institute of Statistics
School of Business Administration
University of Puerto Rico
Río Piedras Campus
PO Box 23332
San Juan
Puerto Rico 00931-3332
USA
E-mail: jairo.a.fuquene@uprrp.edu

Department of Mathematics
College of Natural Sciences
University of Puerto Rico
Río Piedras Campus
PO Box 23355
San Juan
Puerto Rico 00931-3355
USA
E-mail: moises.r.delgado@uprrp.edu