Bayesian Analysis (2009) **4**, Number 4, pp. 817–846

A Case for Robust Bayesian Priors with Applications to Clinical Trials

Jairo A. Fúquene*, John D. Cook† and Luis R. Pericchi‡

Abstract. Bayesian analysis is frequently confused with *conjugate* Bayesian analysis. This is particularly the case in the analysis of clinical trial data. Even though conjugate analysis is perceived to be simpler computationally (but see below, Berger's prior), the price to be paid is high: such analysis is not robust with respect to the prior, i.e. changing the prior may affect the conclusions without bound. Furthermore, conjugate Bayesian analysis is blind with respect to the potential conflict between the prior and the data. Robust priors, however, have bounded influence. The prior is discounted automatically when there are conflicts between prior information and data. In other words, conjugate priors may lead to a dogmatic analysis while robust priors promote self-criticism since prior and sample information are not on equal footing. The original proposal of robust priors was made by de-Finetti in the 1960's. However, the practice has not taken hold in important areas where the Bayesian approach is making definite advances such as in clinical trials where conjugate priors are ubiquitous.

We show here how the Bayesian analysis for simple binary binomial data, expressed in its exponential family form, is improved by employing Cauchy priors. This requires no undue computational cost, given the advances in computation and analytical approximations. Moreover, we introduce in the analysis of clinical trials a robust prior originally developed by J.O. Berger that we call Berger's prior. We implement specific choices of prior hyperparameters that give closed-form results when coupled with a normal log-odds likelihood. Berger's prior yields a robust analysis with no added computational complication compared to the conjugate analysis. We illustrate the results with famous textbook examples and also with a real data set and a prior obtained from a previous trial. On the formal side, we present a general and novel theorem, the "Polynomial Tails Comparison Theorem." This theorem establishes the analytical behavior of any likelihood function with tails bounded by a polynomial when used with priors with polynomial tails, such as Cauchy or Student's t. The advantages of the theorem are that the likelihood does not have to be a location family nor exponential family distribution and that the conditions are easily verifiable. The binomial likelihood can be handled as a direct corollary of the result. Next, we proceed to prove a striking result: the intrinsic prior to test a normal mean, obtained as an objective prior for hypothesis testing, is a limit of Berger's robust prior. This result is useful for assessments and for MCMC computations. We then generalize the theorem to prove that Berger's prior and intrinsic priors are robust with normal likelihoods. Finally, we apply the

*Institute of Statistics, School of Business Administration, University of Puerto Rico, San Juan, PR, mailto:jairo.a.fuquene@uprrp.edu

†Division of Quantitative Sciences, M.D. Anderson Cancer Center, University of Texas, Houston, TX, jdcook@mdanderson.org

‡Department of Mathematics, University of Puerto Rico, San Juan, PR, mailto:lrpericchi@uprrp.edu

 DOI:10.1214/09-BA431

results to a large clinical trial that took place in Venezuela, using prior information based on a previous clinical trial conducted in Finland.

Our main conclusion is that introducing the existing prior information in the form of a robust prior is more justifiable simultaneously for federal agencies, researchers, and other constituents because the prior information is coherently discarded when in conflict with the sample information.

Keywords: Berger's Prior, Clinical Trials, Exponential Family, Intrinsic Prior, Parametric Robust Priors, Polynomial Tails Comparison Theorem, Robust Priors

1 Introduction

In Bayesian statistics the selection of the family of prior distributions is crucial to the analysis of data because the conclusions depend on this selection. However, there is little analysis of clinical trials using non-conjugate priors. It is common to report an analysis using different conjugate priors: clinical, skeptical, and non-informative. The precision in these priors is important and sensitivity analyses regarding the priors is necessary. One approach to this problem is advocated by Greenhouse and Wasserman (1995) who compute bounds on posterior expectations over an ε-contaminated class of prior distributions. An alternative solution is proposed in Carlin and Louis (1996), where one re-specifies the prior and re-computes the result. These authors obtain fairly specific results for some restricted non-parametric classes of priors. Along the same line, another alternative is the "prior partitioning" of Carlin and Sargent (1996) which selects a suitably flexible class of priors (a non-parametric class whose members include a quasi-unimodal, a semi-parametric normal mixture class, and the fully parametric normal family) and identify the priors that lead to posterior conclusions of interest. These are (very few) proposals about what may be called "non-parametric" robustness to the prior. The proposals in this paper are "parametric" robust Bayesian analysis, quite distinct from the previous proposals. Some general results on parametric Bayesian robustness are in Dawid (1973), O'Hagan (1979), Pericchi and Sansó (1995). We believe that the main road forward for clinical trials is on the parametric side for three reasons. First, it is more natural to represent the information given by a previous trial in terms of parametric priors. More generally, parametric priors are easier to assess. Second, it is far more clear how to generalize a parametric robust analysis to hierarchical modeling than a non-parametric class of priors. Finally, non-parametric priors do not appear to have achieved a significant impact in practice. In Gelman et al. (2008) the authors take a somewhat similar point of view to ours. Their arguments are very applied while ours are more theoretical and so the papers are complementary.

Clinical trials are exceptional in at least two ways. First, there is often substantial "hard" prior data. Second, there are multiple parties overseeing the analysis: researchers, statisticians, regulatory bodies such as the FDA, data and safety monitoring boards, journal editors, etc. In this framework there are fundamental issues such as the following. How do you assess a prior from the prior data? How do you assess how relevant the previous data is to the current trial? By using prior information are we

enhancing the analysis or biasing it? Our key message in this paper is that robust priors are a better framework to get consensus in clinical trials for the following reason:

1. Prior information may be substantial about certain characteristics like location, scale, but it is very weak about the tails of prior distributions.
2. The tail size is crucial in the posterior inference when there is conflict between prior and sample.
3. The behavior of posterior inference under robust priors is superior because when the prior information is irrelevant for the case at hand, then the prior information is coherently and automatically discarded by Bayes' theorem.

Conjugate light-tailed priors do not have these features and may be called "dogmatic." See Berger (1985) for an authoritative discussion of these issues and our example in 6.1. Of course if all involved had unlimited time for several sensitivity analysis, the results using light tailed priors might be acceptable. Instead, we are suggesting that Bayes' theorem should be allowed to perform coherently the sensitivity analyses, and for that heavy tailed priors are required. A referee has pointed out that "a researcher who had carefully constructed a prior distribution that reflected substantial available information almost certainly would prefer for that information to be reflected in the posterior distribution or at least for prior/data conflict to be recognized and investigated". Certainly, if someone has a prior that they want included in the analysis, fine. But it need not be the only prior used. There's no harm in repeating an analysis with several priors, and in fact it is a recommended practice to do so. Furthermore, there are limits to how well someone can quantify their prior uncertainty, particularly far from the center of their estimate. It is hard to imagine that someone could say that their prior belief follows a normal distribution rather than a Student-t with say six degrees of freedom. If individuals cannot specify with fidelity the tail behavior of their subjective priors, the tail behavior should be determined by technical criteria such as robustness.

The popular normal/normal (N/N) or beta/binomial (B/B) conjugate analysis (see for example Spiegelhalter et al. (2004)) will be exposed in this article as non-robust. Workable (parametric) alternatives are available to the practitioner. For motivation consider: The posterior mean μ_n in the N/N and B/B models is (see next section) $\mu_n = (n_0 + n)^{-1}(n_0\mu + n\bar{X}_n)$. Thus the mean is a convex combination of the prior expectation, μ, and the data average, $\bar{X}_n$, and thus the prior has *unbounded influence*. For example, as the location prior/data conflict $|\mu - \bar{x}|$ grows, so does $|\mu_n - \bar{x}|$ and without bound. These considerations motivate the interest in non-conjugate models for Bayesian analysis of clinical trials, and more generally motivate heavy-tailed priors. (See the theorem in the next section.)

We may employ the following heuristic: Bayesian clinical trials are not better because they stop sampling earlier (although they often do) but because they stop intelligently, that is the stopping is conditional on the amount of evidence. Robust priors are not better because they have less influence (though this is true) but because they influence in a more intelligent way: the influence of the robust prior is a function of the potential

conflict between prior and sample information about the region where the parameters are most likely to live. (For more general measures of prior-sample conflict see for example, Evans and Moshonov (2006)). In this paper we show that the Cauchy prior is robust in two models for clinical trials. Pericchi and Smith (1992) considered the robustness of the Student-t prior in the Student-t/normal model. We consider as a particular case the Cauchy/normal (C/N) model for normal log-odds. Much less known, however, is the robust property of the Cauchy prior with the binomial likelihood and more generally for exponential family likelihoods. To prove the robustness of the Cauchy prior when coupled with a binomial likelihood, we prove a more general result that only requires a bound in the tail behavior of the likelihood. This novel theorem is easy to verify and is very general. Under these conditions, when the prior and the model are in conflict, then the prior acts "as if" it were uniform. In other words, the prior influences the analysis only when prior information and likelihood are in broad agreement. Otherwise Bayes' theorem effectively switches back to a uniform prior. In this paper we rely heavily on the fact that the binomial likelihood belongs to the exponential family (though the theorem is not limited to exponential family likelihoods) showing the robustness of the Cauchy prior in the Cauchy/binomial (C/B) model for binary data.

Cauchy priors do not lead to analytical closed-form results, but our next suggestion does. In his very influential book (Berger (1985)) Berger proposes a prior (called here "Berger's prior"). We use Berger's prior for clinical trials analysis, assuming a prior mean and scale suggested by previous data or by general features of the current trial. It turns out that this gives closed-form results when coupled with a normal log-odds likelihood. We show the robustness of Berger's prior for the Berger-prior/normal log-odds (BP/N) model, which makes it more attractive than both the Cauchy and conjugate priors. We also prove here a striking result: The intrinsic prior to test a normal mean of Berger and Pericchi (1996) which is obtained as an objective prior for hypothesis testing, is also a limit of Berger's robust prior. This result is useful for assessments and for MCMC computations. We then generalize the Polynomial Tails Comparison theorem to prove that Berger's prior and intrinsic priors are robust with normal likelihoods. We finally apply the results to massive clinical trial that took place in Venezuela, and the prior information is taken from a previous clinical trial in Finland.

Lastly we remark that the hierarchical model is **not** the solution for the lack of robustness of conjugate analysis. Quite to the contrary, the hierarchical model should use robust priors in the hierarchy to prevent unbounded and undesirable shrinkages. This is being studied in work in progress by M.E. Perez, and L.R. Pericchi.

This article is organized as follows. Section 2 is devoted to the Polynomial Tails Comparison Theorem. In Section 3 we review the prior specification and posterior moments of the C/B model. In Section 4 we examine the robustness of the Cauchy prior in the C/B posterior model. In the Sections 3 and 4 we show the application of the C/B model in a clinical trial. In Section 5 we describe the robustness of the C/N and BP/N models and prove that the intrinsic prior is a limit of Berger's priors. In Section 6 we prove the Generalized Polynomial Tails theorem and illustrate the results in a real and important clinical trial published in the New England Journal of Medicine. We make some closing remarks in Section 7.

2 The Polynomial Tails Comparison Theorem

The following theorem is decidedly useful and easy to apply when determining whether a prior is robust with respect to a likelihood.

For $\nu > 0$, define

$$t(\lambda; \mu, \nu) = \left(1 + \frac{(\lambda - \mu)^2}{\nu}\right)^{-(\nu+1)/2}.$$

Aside from a normalization constant that would cancel out in our calculations, $t(\lambda; \mu, \nu)$ is the PDF of a Student-t distribution with ν degrees of freedom centered at μ.

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

$$\int_{|\lambda|>m} f(\lambda)\, d\lambda = \mathcal{O}(m^{-\nu-1-\varepsilon}). \tag{1}$$

In the application we have in mind, f is a binomial likelihood function although the result is more general. For instance the latter holds for any ν in any likelihood with exponentially decreasing tails.

Denote by $\pi^T(\lambda \,|\, \text{data})$ and $\pi^U(\lambda \,|\, \text{data})$ the posterior densities employing the Student-t and the uniform prior densities respectively. Applying Bayes' rule to both densities, for any parameter value λ_0 the following ratio:

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

Theorem 2.1. *For fixed λ_0,*

$$\lim_{\mu \to \infty} \frac{\int_{-\infty}^{\infty} f(\lambda)\, t(\lambda; \mu, \nu)\, d\lambda}{t(\lambda_0; \mu, \nu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda} = 1. \tag{2}$$

Proof. We will show that

$$\lim_{\mu \to \infty} \frac{\int_{-\infty}^{\infty} f(\lambda)\, t(\lambda; \mu, \nu)\, d\lambda - t(\lambda_0; \mu, \nu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda}{t(\lambda_0; \mu, \nu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda} = 0. \tag{3}$$

Note that the numerator can be written as

$$\int_{-\infty}^{\infty} f(\lambda)\, (t(\lambda; \mu, \nu) - t(\lambda_0; \mu, \nu))\, d\lambda.$$

We break the region of integration in the numerator into two parts, $|\lambda| < \mu^k$ and $|\lambda| > \mu^k$, for some $0 < k < 1$ that we will pick later, and show that as $\mu \to \infty$ each integral goes to zero faster than the denominator.

First consider

$$\int_{|\lambda|<\mu^k} f(\lambda)\,(t(\lambda;\mu,\nu)-t(\lambda_0;\mu,\nu))\;d\lambda. \tag{4}$$

For every λ, there exists a ξ between λ and λ_0 such that

$$t(\lambda;\mu,\nu)-t(\lambda_0;\mu,\nu)=t'(\xi;\mu,\nu)(\lambda-\lambda_0)$$

by the mean value theorem. Since $\mu\to\infty$, we can assume $\mu>\mu^k>\lambda_0$.

$$\begin{aligned}
|t(\lambda;\mu,\nu)-t(\lambda_0;\mu,\nu)| &= |t'(\xi;\mu,\nu)(\lambda-\lambda_0)| \\
&= \frac{(\nu+1)|\lambda-\mu|\,|\lambda-\lambda_0|}{\nu\left(1+\frac{(\lambda-\mu)^2}{\nu}\right)^{(\nu+3)/2}} \\
&= \frac{\mathcal{O}(\mu^{1+k})}{\Omega(\mu^{\nu+3})} \\
&= \mathcal{O}(\mu^{k-\nu-2}).
\end{aligned}$$

[Here we use the familiar $\mathcal{O}$ notation and the less familiar Ω notation. Just as $f=\mathcal{O}(\mu^n)$ means that f is eventually bounded *above* by a constant multiple of μ^n, the notation $f=\Omega(\mu^n)$ means that f is eventually bounded *below* by a constant multiple of μ^n.]

As $\mu\to\infty$, the integral (4) goes to zero as $\mathcal{O}(\mu^{k-\nu-2})$. Since $t(\lambda_0;\mu,\nu)$ is $\Omega(\mu^{-\nu-1})$, the ratio of the integral (4) to $t(\lambda_0;\mu,\nu)$ is $\mathcal{O}(\mu^{k-1})$. Since $k<1$, this ratio goes to zero as $\mu\to\infty$.

Next consider the remaining integral,

$$\int_{|\lambda|>\mu^k} f(\lambda)\,(t(\lambda;\mu,\nu)-t(\lambda_0;\mu,\nu))\;d\lambda. \tag{5}$$

The term $t(\lambda;\mu,\nu)-t(\lambda_0;\mu,\nu)$ is bounded, and we assumed

$$\int_{|x|>m} f(\lambda)\,d\lambda=\mathcal{O}(m^{-\nu-1-\varepsilon}).$$

Therefore the integral (5) is $\mathcal{O}((\mu^k)^{-\nu-1-\varepsilon})=\mathcal{O}(\mu^{-k(\nu+1+\varepsilon)})$. Since $t(\lambda_0;\mu,\nu)$ is $\Omega(\mu^{-\nu-1})$, the ratio of the integral (5) to $t(\lambda_0;\mu,\nu)$ is of order $\mathcal{O}(\mu^{-k(\nu+1+\varepsilon/(\nu+1)})$. This term goes to zero as $\mu\to\infty$ provided $k>(\nu+1)/(\nu+1+\varepsilon)$. □

Note that in particular the theorem applies when f is the likelihood function of a binomial model with at least one success and one failure and $\nu=1$, i.e. a Cauchy prior.

3 The Binomial Likelihood with Conjugate and Cauchy Priors

Assume a sample of size n with $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$. The binomial likelihood in its explicit exponential family form is given by

$$f(X_+ \mid \lambda) \propto \exp\left\{X_+\lambda - n\log(1+e^{\lambda})\right\}, \tag{6}$$

where $X_+ = \sum_{i=1}^{n} X_i \sim \text{binomial}(n, \theta)$ is the number of success in n trials. Notice that for the binomial likelihood, it is enough to assume that there is at least one success and one failure, i.e. $0 < X_+ < n$, (for assumption (2.1) of the theorem of the previous section to be fulfilled for every $\nu \geq 1$), since then the binomial has exponentially decreasing tails. The natural parameter is the log-odds, $\lambda = \log(\theta/(1-\theta))$, which is the parameter to be modeled as a Cauchy variable later, for which one can make use of the theorem. If desired, a Student-t prior with more than one degree of freedom can be used, and all results apply as well. We employ the Cauchy for good use of "conservatism" regarding the treatment of prior information, a point shared with Gelman et al. (2008).

First we perform a conjugate analysis, expressing the beta(a, b) prior, after the transformation of the parameter θ to log-odds, as

$$\pi_B(\lambda) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{e^{\lambda}}{1+e^{\lambda}}\right)^{a}\left(\frac{1}{1+e^{\lambda}}\right)^{b} \qquad a, b > 0. \tag{7}$$

The cumulant generating function of the prior distribution $\pi_B(\lambda)$ is given by

$$K_{\lambda}(t) = -\log(\Gamma(a)\Gamma(b)) + \log(\Gamma(a+t)) + \log(\Gamma(b-t)), \tag{8}$$

hence $E_B(\lambda) = \Psi(a) - \Psi(b)$ and $V_B(\lambda) = \Psi'(a) + \Psi'(b)$, where $\Psi(\cdot)$ is the digamma function and $\Psi'(\cdot)$ is the trigamma function, that are extensively tabulated in for example Abramowitz and Stegun (1970). The posterior distribution of the B/B model is given by

$$\pi_B(\lambda \mid X_+) = K \times \exp\left\{(a+X_+)\lambda - (n+a+b)\log(1+e^{\lambda})\right\} \tag{9}$$

where $K = \dfrac{\Gamma(n+a+b)}{\Gamma(X_+ + a)\Gamma(n - X_+ + b)}$. On the other hand, one proposal for robust analysis for binomial data (see also the next sections for Berger's prior for an alternative) is to use a Cauchy prior for the natural parameter λ in order to achieve robustness with respect to the prior,

$$\pi_C(\lambda) = \frac{\beta}{\pi[\beta^2 + (\lambda - \alpha)^2]}, \tag{10}$$

with parameters of location and scale α and β respectively. The posterior distribution of the C/B model is

$$\pi_C(\lambda \mid X_+) = \frac{\exp\left\{X_+\lambda - n\log(1+e^{\lambda}) - \log\left(\beta^2 + (\lambda-\alpha)^2\right)\right\}}{m(X_+)},$$

where $m(X_+)$ is the predictive marginal. Notice that this posterior also belongs to the exponential family. One approach to the approximation of $m(X_+)$ is Laplace's method, refined by Tierney and Kadane (1986) for statistical applications given by $m(X_+) \approx \sqrt{2\pi}\hat{\sigma}n^{-1/2}\exp\{-nh(\hat{\lambda})\}$ where $-nh(\lambda) = \log(\pi_C(\lambda)f(X_+\,|\,\lambda))$, $\hat{\lambda}$ is the maximum of $-h(\hat{\lambda})$, and $\hat{\sigma} = [h^{''}(\lambda)]^{-1/2}|_{\lambda=\hat{\lambda}}$. The accuracy is of order $\mathcal{O}(n^{-1})$.

Example 3.1. A Textbook Clinical Trial Example. We apply the preceding approximation adapting an example considered in Spiegelhalter et al. (2004). Suppose that previous experience with similar compounds has suggested that a drug has a true response rate θ, between 0 and 0.4, with an expectation around 0.2. For normal distributions we know that $m \pm 2s$ includes just over 95% of the probability, so if we were assuming a normal prior we might estimate $m = 0.2$ and $s = 0.1$. However, the beta distributions with reasonably high a and b have approximately normal shape, so that $\theta \sim \text{beta}(a = 3,\ b = 12)$. Suppose that we test the drug on 20 additional patients and observe 16 positive responses ($X_+ = 16$). Then the likelihood of the experiment is $X_+ \sim \text{binomial}(n = 20,\ \theta)$ and the posterior in this case $\theta \mid X_+ \sim \text{beta}(a = 19,\ b = 16)$. Our proposal is to use a Cauchy prior in order to achieve robustness with respect to the prior, $\pi_C(\lambda)$, with the same parameters of location and scale of the beta prior. For this example the location and the scale are $\Psi(3) - \Psi(12) = -1.52$ and $\sqrt{\Psi'(3) + \Psi'(12)} = 0.69$ respectively. Figures 1 and 2 display a large discrepancy between the means of the prior information and the normalized likelihood (i.e. the posterior density using a uniform prior) of the data. In the B/B model the prior and the likelihood receive equal weight. The weight of the likelihood in the C/B posterior model is higher that in the B/B model. The form of the C/B model is much closer to the normalized likelihood.

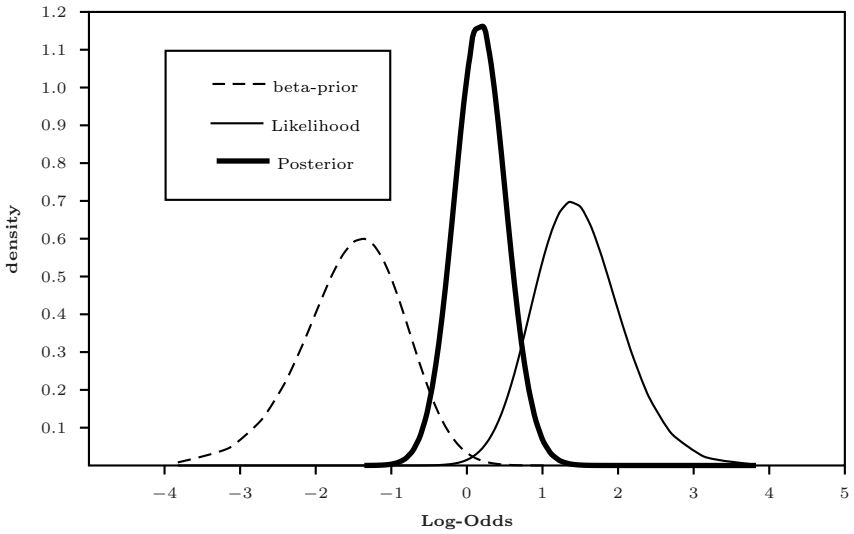

Figure 1: Beta prior, normalized binomial likelihood and B/B posterior model for the Example 1.

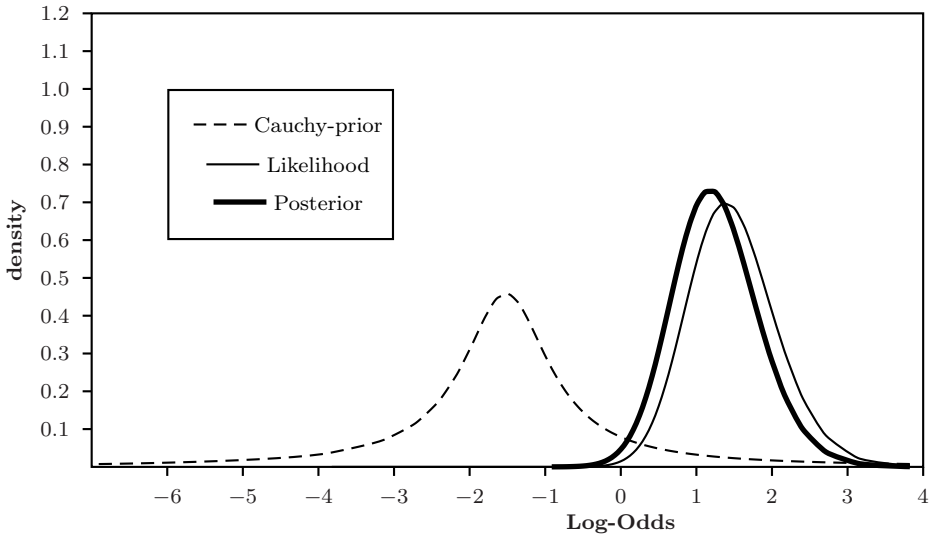

Figure 2: Cauchy prior, normalized binomial likelihood and C/B posterior model for the Example 1.

The posterior moments of the natural parameter of an exponential family are considered in Pericchi et al. (1993) and Gutierrez-Peña (1997). The cumulant generating function of the posterior, $\pi_B(\lambda \mid X_+)$, in the B/B model is

$$K_{\lambda \mid X_+}(t) = \log\left(\frac{\Gamma(X_+ + a + t)\Gamma(n - X_+ + b - t)}{\Gamma(X_+ + a)\Gamma(n - X_+ + b)}\right), \tag{11}$$

hence

$$E_B(\lambda \mid X_+) = \Psi(X_+ + a) - \Psi(n - X_+ + b) \tag{12}$$

$$V_B(\lambda \mid X_+) = \Psi^{'}(X_+ + a) + \Psi^{'}(n - X_+ + b). \tag{13}$$

In the C/B model, we need to calculate the approximation of $E_C(\lambda \mid X_+)$ and $V_C(\lambda \mid X_+)$. The posterior expectation $E_C(\lambda \mid X_+)$ involves the ratio of two integrals and the Laplace method can be used, as

$$\tilde{E}(\lambda \mid X_+) = \left(\frac{\sigma^*}{\hat{\sigma}}\right) \exp\left\{-n[h^*(\lambda^*) - h(\hat{\lambda})]\right\}, \tag{14}$$

where $-nh^*(\lambda) = \log(\lambda \pi_C(\lambda) f(X_+ \mid \lambda))$, λ^* is the maximum of $-h^*(\lambda)$ and $\sigma^* = [h^{*''}(\lambda)]^{-1/2}|_{\lambda=\lambda^*}$. The error in (14) is of order $\mathcal{O}(n^{-2})$ (see Tierney and Kadane (1986)). However, in (14) we must assume that λ does not change sign. Tierney et al. (1989) recommend to add a large constant c to λ, apply Laplace's method (14) and finally subtract the constant. We let $\tilde{E}_C(\lambda \mid X_+)$ and $\tilde{V}_C(\lambda \mid X_+)$ to denote approximate posterior expectation and posterior variance of the C/B model

$$\tilde{E}_C(\lambda \mid X_+) = \tilde{E}(c + \lambda \mid X_+) - c. \tag{15}$$

$$\tilde{V}_C(\lambda \mid X_+) = \tilde{E}((c + \lambda)^2 \mid X_+) - [\tilde{E}(c + \lambda \mid X_+)]^2. \tag{16}$$

For both functions $h(\hat{\lambda})$ and $h^*(\lambda)$ it is not possible to find the maximum analytically and then we use Newton Raphson algorithm. Here c is the value of λ such that $\pi_C(\lambda = c \mid X_+) \leq 0.5^{-4}$ and the starting value in the Newton-Raphson algorithm is the maximum likelihood estimator (MLE) of the natural parameter, $\hat{\lambda} = \log((\bar{X}_n)/(1-\bar{X}_n))$.

Result 3.1. *The posterior expectations for the C/B model and B/B satisfy the following:*

1. *Robust result:*
$$\lim_{\alpha \to \pm\infty} E_C(\lambda \mid X_+) \approx \hat{\lambda} + \frac{e^{2\hat{\lambda}} - 1}{2ne^{\hat{\lambda}}}. \tag{17}$$
2. *Non-robust result:*
$$\lim_{E_B(\lambda) \to \pm\infty} E_B(\lambda \mid X_+) \to \pm\infty. \tag{18}$$

respectively.

Proof. See the Appendix. Result 3.1 is a corollary of the Theorem 2.1. □

Note: the limit (17) is not equal to the MLE, but consistent with Theorem 2.1.

4 Computations with Cauchy Priors

We use weighted rejection sampling to compute the ("exact") posterior moments in the C/B model due to its simplicity and generality for simulating draws directly from the target density $\pi_C(\lambda \mid X_+)$ (see Smith and Gelfand (1992)). In the C/B model the *envelope function* is the Cauchy prior. The rejection method proceeds as follows:

1. Calculate $M = f(X_+ \mid \hat{\lambda})$.
2. Generate $\lambda_j \sim \pi_C(\lambda)$.
3. Generate $U_j \sim$ uniform(0,1).
4. If $MU_j\, \pi_C(\lambda_j) < f(X_+ \mid \lambda_j)\, \pi_C(\lambda_j)$, accept λ_j. Otherwise reject λ_j and go to Step 2.

It is clear that the Cauchy density is an envelope. Because it is simple to generate Cauchy distributed samples, the method is feasible. Using Monte Carlo methods and 10,000 random samples from $\pi_C(\lambda \mid X_+)$ we compute E_{sim} and V_{sim}. Results available from the authors show that the agreement between the Laplace approximations and the rejection algorithm is quite good for sample sizes bigger than $n = 10$. In Figures 3 to 5 we illustrate the striking qualitative difference of posterior moments, as a function of the discrepancy between prior and sample location $|\mu - \bar{x}|$. Figure 3 shows that the beta prior has an unbounded influence and it is not robust. Figures 4 and 5 display the qualitative forms of dependence of the posterior expectation and variance on the discrepancy between the prior location and the MLE using a Cauchy prior. Here $\hat{\lambda} = 0$ and a and b take various values with their sum fixed at 50. In Figures 4 and 5, the approximations (15) and (16) are shown as functions of the discrepancy. Note that (16) is non-monotonic in the discrepancy. The posterior expectation, $\tilde{E}_C(\lambda \mid X_+)$, is a function of the "information discount."

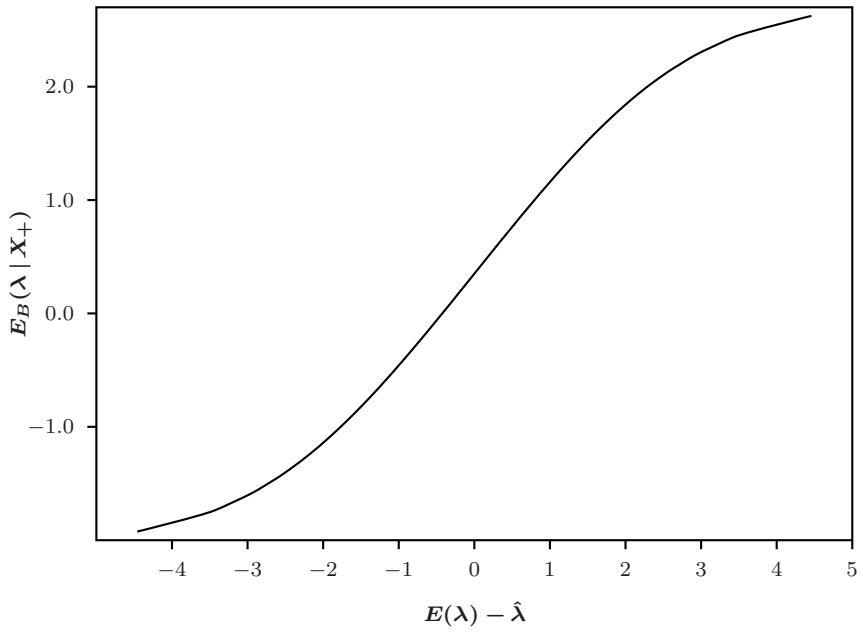

Figure 3: Behavior of the posterior expectation, $E_B(\lambda \mid X_+)$, in the B/B model for values $n = 10$, $\hat{\lambda} = 0$ and $a + b = 50$.

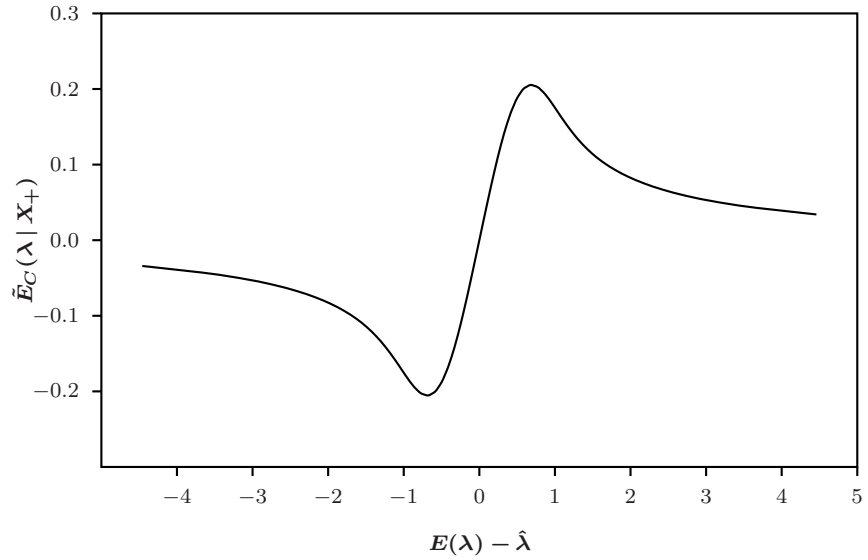

Figure 4: Behavior of the posterior expectation, $\tilde{E}_C(\lambda \mid X_+)$, in the C/B model for values $n = 50$, $\hat{\lambda} = 0$ and $a + b = 50$.

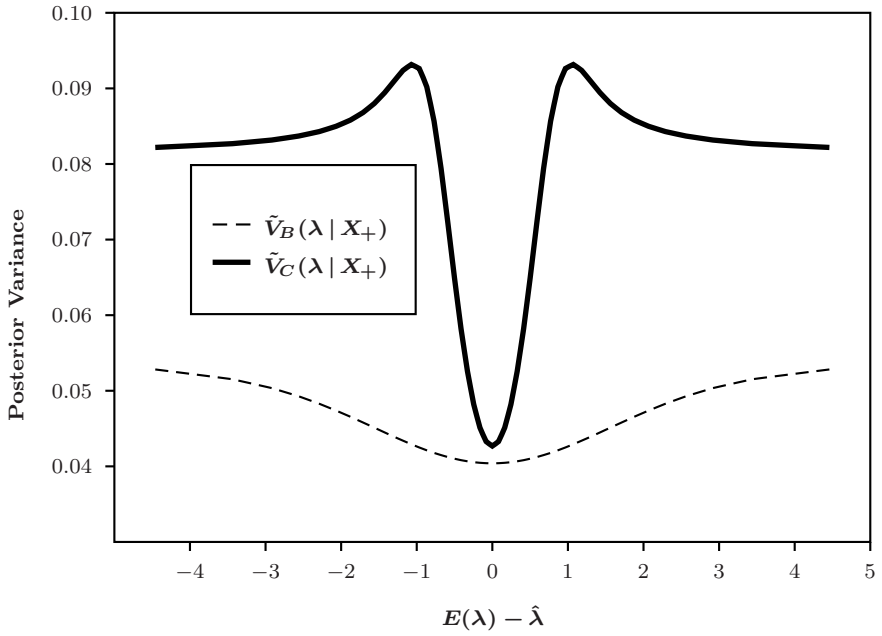

Figure 5: Behavior of the posterior variance, $V_B(\lambda \mid X_+)$ in the B/B and $\tilde{V}_C(\lambda \mid X_+)$ in the C/B for values $n = 50$, $\hat{\lambda} = 0$ and $a + b = 50$.

Example 4.1. Textbook Example (Continued): Moments and predictions for binary data.

$E_B(\lambda \mid X_+)$	$V_B(\lambda \mid X_+)$	$\tilde{E}_C(\lambda \mid X_+)$	$\tilde{V}_C(\lambda \mid X_+)$	M.L.E
0.18	0.12	1.26	0.33	1.39

Table 1: Posterior Expectation and Variance for the B/B and C/B models.

The estimate resulting from the C/B model is closer to the data because the MLE 0.8 of θ is closer to 0.77 than to 0.54, the estimate resulting from the B/B model.

In Table 1 there is a large difference between the values of the posterior mean (0.18) for the B/B model and the MLE. On the other hand, the results of the C/B model and the MLE $\hat{\lambda}$ are similar. The discrepancies between the expectations of the posterior models and the MLE are approximately 3.5 and 0.23 standard errors for B/B and C/B respectively. For the scale of θ that is the true response rate for a Bernoulli trials set, we know that the predictive mean of the total number of successes in m trials is $E(X_m) = mE(\theta \mid X_+)$. If we plan to treat 40 additional patients in the B/B model the predictive mean is $40 \times 0.54 \approx 22$, and in the C/B model is equal to $40 \times 0.77 \approx 31$. The estimate resulting from the C/B model is closer to the data because the MLE 0.8 of θ is closer to 0.77 than to 0.54, the estimate resulting from the B/B model. The beta prior is more "dogmatic" than the Cauchy prior leading to non-robust results. Bayesian analysis is not dogmatic in general, but *conjugate* Bayesian analysis can be. This is a major selling point of robust Bayesian methods.

5 Normal Log-Odds and Berger's Prior

An alternative to the binomial likelihood is the normal likelihood in the log-odds, see Spiegelhalter et al. (2004). Pericchi and Smith (1992) showed some aspects of the robustness of the Student-t prior for a normal location parameter and provided approximations to the posterior moments in the model Student-t/normal. The Cauchy prior, as a Student-t with one degree of freedom, can be used in this context as well. However, for normal log-odds there is a robust prior that leads to a closed-form posterior and moments, a sort of "best of both worlds." Bayesians have long come to terms with the disadvantages of procedures based on conjugate priors because of the desire for closed-form results. However, Berger (1985) proposed for comparison of several means a robust prior (called "Berger's prior" in this work) that gives closed-form results for coupled normal means. Berger's prior (BP) is similar to a Cauchy prior in the tails. Our proposal is an analysis based on Berger's prior that we call BP/N posterior model. In this work the location of Berger's prior, $\pi^{\mu}_{BP}(\lambda)$, is denoted by μ. This prior has the

following form.

$$\pi_{BP}(\lambda) = \int_0^1 N(\lambda \,|\, \mu; \frac{d+b}{2\nu} - d) \cdot \frac{1}{2\sqrt{\nu}}\, d\nu \tag{19}$$

Here $N(\lambda \,|\, \mu, \tau^2)$ denotes a normal density on the parameter λ with mean and variance μ, τ^2 respectively, which is well-defined whenever $b \geq d$. The hyper-parameters d and b have to be assessed (see the end of the section for alternative assessments). We set here $b = \beta^2$ (equal to the scale of the Cauchy) and $d = \sigma^2/n$.

Suppose that $X_1, \ldots, X_n \sim \text{normal}(\lambda, \sigma^2)$ where σ^2 is assumed known and λ is unknown, then the Berger's prior is

$$\pi_{BP}(\lambda) = \int_0^1 K \times \exp\left\{-\frac{n}{2}\left[\frac{2\nu(\lambda-\mu)^2}{\sigma^2(1-2\nu)+n\beta^2}\right]\right\} d\nu \tag{20}$$

where

$$K = \frac{\sqrt{n}}{\sqrt{4\pi(\sigma^2(1-2\nu)+n\beta^2)}}. \tag{21}$$

Result 5.1. *Suppose that $X_1, \ldots, X_n \sim$ normal(λ, σ^2) where σ^2 is assumed known and λ is unknown. The predictive distribution of the BP/N model is*

$$m(\bar{X}_n) = \frac{\sqrt{\sigma^2+n\beta^2}}{\sqrt{4\pi n}(\bar{X}_n-\mu)^2}\left[1 - \exp\left\{-\frac{n(\bar{X}_n-\mu)^2}{\sigma^2+n\beta^2}\right\}\right].$$

The posterior distribution of the BP/N model is

$$\pi_{BP}(\lambda \,|\, \bar{X}_n) = \frac{\pi_{BP}(\lambda)\exp\left\{-\frac{n(\bar{X}_n-\lambda)^2}{2\sigma^2}\right\}}{\frac{\sigma\sqrt{\sigma^2+n\beta^2}}{\sqrt{2}n(\bar{X}_n-\mu)^2}\left[1 - \exp\left\{-\frac{n(\bar{X}_n-\mu)^2}{\sigma^2+n\beta^2}\right\}\right]}. \tag{22}$$

The posterior expectation of the BP/N model

$$E_{BP}(\lambda \,|\, \bar{X}_n) = \bar{X}_n + \frac{2\sigma^2 n(\bar{X}_n-\mu)^2 - 2\sigma^2(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)}{n(\bar{X}_n-\mu)(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)}, \tag{23}$$

and the posterior variance of the BP/N model is

$$\begin{aligned} V_{BP}(\lambda \,|\, \bar{X}_n) = \frac{\sigma^2}{n} - \frac{\sigma^4}{n^2}\left\{\frac{4n^2(\bar{X}_n-\mu)^2 f(\bar{X}_n)}{(\sigma^2+n\beta^2)^2(f(\bar{X}_n)-1)^2}\right\} \\ + \frac{\sigma^4}{n^2}\left\{\frac{2(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)((\sigma^2+n\beta^2)(f(\bar{X}_n)-1)-n)}{(\sigma^2+n\beta^2)^2(f(\bar{X}_n)-1)^2(\bar{X}_n-\mu)^2}\right\}, \end{aligned} \tag{24}$$

where $f(\bar{X}_n) = \exp\left\{\dfrac{n(\bar{X}_n - \mu)^2}{\sigma^2 + n\beta^2}\right\}$.

Proof. See Appendix. □

The posterior expectation for the BP/N model satisfies the following

$$\lim_{\mu \to \pm\infty} E_{BP}(\lambda \,|\, \bar{X}_n) = \bar{X}_n; \ \lim_{\mu \to \bar{X}_n} E_{BP}(\lambda \,|\, \bar{X}_n) = \bar{X}_n. \tag{25}$$

This can be shown simply using L'Hôpital's rule on the expression of the posterior expectation (23) and proves the robustness of the Berger prior coupled with the normal log-odds (see also Berger (1985)). Also we have the following result for a Cauchy prior (as a corollary of the theorem):

$$\lim_{\alpha \to \pm\infty} E_{CN}(\lambda \,|\, \bar{X}_n) \approx \lim_{\alpha \to \pm\infty} \bar{X}_n - \frac{2\sigma^2(\bar{X}_n - \alpha)}{n(\beta^2) + (\bar{X}_n - \alpha)^2} = \bar{X}_n. \tag{26}$$

5.1 The Intrinsic Prior as the Limit of Berger's Prior

It is a striking sort of synthesis result that the limit (as $n \to \infty$, for $d = \sigma^2/n$) of Berger's prior is the intrinsic prior (Berger and Pericchi (1996)). Define $\eta = \lambda - \mu$ and recall the standard intrinsic prior for a normal location parameter:

$$\varphi(\eta) = \frac{1}{2\sqrt{\pi}} \frac{1 - \exp(-\eta^2)}{\eta^2}. \tag{27}$$

and extend it to a scale family by defining

$$\varphi(\eta; \sigma) = \frac{1}{\sigma} \varphi\left(\frac{\eta}{\sigma}\right).$$

Then

$$\lim_{d \to 0} \pi_{BP}(\eta; b, d) = \varphi(\eta, \sqrt{b})$$

as we will prove in the next section.

5.2 Bounds for Berger's prior

In this section we develop upper and lower bounds for the density of Berger's prior. We then use these bounds to prove that the limiting case of Berger's prior is the intrinsic prior. First, define

$$w(\nu; b, d) = \frac{1}{b + d - 2\nu d}.$$

We will suppress the dependence on b and d unless there is a need to be explicit. With this notation the integral defining Berger's prior becomes

$$\pi_B(\eta) = \frac{1}{2\sqrt{\pi}} \int_0^1 \exp\left(-\eta^2\, w(\nu)\, \nu\right)\ w(\nu)^{1/2}\, d\nu.$$

Next we multiply and divide by $(w\nu)'$ where the prime indicates the derivative with respect to ν. Then

$$\pi_B(\eta) = \frac{1}{2\sqrt{\pi}} \int_0^1 \exp\left(-\eta^2\, w(\nu)\, \nu\right)\ (w(\nu)\nu)' \frac{w(\nu)^{1/2}}{(w(\nu)\,\nu)'}\, d\nu.$$

Note that

$$\frac{w(\nu)^{1/2}}{(w(\nu)\,\nu)'} = \frac{(b+d-2\nu d)^{3/2}}{b+d}.$$

Therefore

$$k_1(b,d) \equiv \frac{(b-d)^{3/2}}{b+d} \le \frac{w(\nu)^{1/2}}{(w(\nu)\,\nu)'} \le (b+d)^{1/2} \equiv k_2(b,d).$$

It follows that

$$\begin{aligned}
\pi_{BP}(\eta) &= \frac{1}{2\sqrt{\pi}} \int_0^1 \exp\left(-\eta^2\, w(\nu)\, \nu\right)\ (w(\nu)\nu)' \frac{w(\nu)^{1/2}}{(w(\nu)\nu)'}\, d\nu. && (28)\\
&\le \frac{1}{2\sqrt{\pi}} \int_0^1 \exp\left(-\eta^2\, w(\nu)\, \nu\right)\ (w(\nu)\nu)' k_2(b,d)\, d\nu. && (29)\\
&= \frac{k_2(b,d)}{2\sqrt{\pi}\eta^2}\left(1 - \exp\left(-\frac{\eta^2}{b-d}\right)\right) && (30)
\end{aligned}$$

by computing the integral (29). Similarly, by applying the lower bound $k_1(b,d)$ in the integral (29) and reversing the direction of the inequality

$$\pi_{BP}(\eta) \ge \frac{k_1(b,d)}{2\sqrt{\pi}\eta^2}\left(1 - \exp\left(-\frac{\eta^2}{b-d}\right)\right).$$

To summarize,

$$k_1(b,d)\,\psi(\eta;b,d) \le \pi_B(\eta;b,d) \le k_2(b,d)\,\psi(\eta;b,d) \tag{31}$$

where

$$\psi(\eta;b,d) = \frac{1 - \exp\left(-\frac{\eta^2}{b-d}\right)}{2\sqrt{\pi}\eta^2}.$$

Note that as $d \to 0$ the terms $k_1(b,d)$ and $k_2(b,d)$ converge to $\sqrt{b}$. Therefore the upper and lower bounds on $\pi_{BP}(\eta;b,d)$ converge to the intrinsic prior scaled by $\sqrt{b}$.

Also, these bounds suggest that one could construct an efficient accept-reject algorithm for sampling from Berger's prior by using the intrinsic prior as a proposal density.

Notice that the intrinsic prior was obtained by a completely unrelated method (Berger and Pericchi (1996)). It was originally obtained as the implicit prior to which the arithmetic intrinsic Bayes factor converges. The intrinsic Bayes factor is derived within an approach to objective model selection. It is pleasant that it coheres with robust Bayesian reasoning. The intrinsic prior does not yield closed-form results with the normal likelihood. The next theorem generalizes the Polynomial Tails Comparison Theorem to prove that the intrinsic prior, as well as the Cauchy prior, are robust.

6 Generalized Polynomial Tails Comparison Theorem

We begin by reviewing the symbols $\mathcal{O}$, Ω, and Θ used to denote asymptotic order, extending the notation used in Section 2. We say $f(\lambda) = \mathcal{O}(g(\lambda))$ if there exist positive constants M and C such that for all $\lambda > M$,

$$f(\lambda) \leq Cg(\lambda).$$

Similarly, we say $f(\lambda) = \Omega(g(\lambda))$ if there exist positive constants M and C such that for all $\lambda > M$,

$$f(\lambda) \geq Cg(\lambda).$$

One could read $\mathcal{O}$ as "eventually bounded above by a multiple of" and Ω as "eventually bounded below by a multiple of." Finally, we say $f(\lambda) = \Theta(g(\lambda))$ if $f(\lambda) = \mathcal{O}(g(\lambda))$ and $f(\lambda) = \Omega(g(\lambda))$.

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

$$\int_{|\lambda|>m} f(\lambda)\, d\lambda = \mathcal{O}(m^{-d-\varepsilon}) \tag{32}$$

for positive constants d and ε. In particular, note that this condition is satisfied for the binomial likelihood in logistic form as long as there has been at least one success and at least one failure observed. The condition also applies for any likelihood function with exponentially decreasing tails.

Let $p(\lambda)$ be a continuous, symmetric distribution. (The assumption of symmetry is not essential, but the distributions we are most interested in are symmetric and the assumption simplifies the presentation.) We may extend p to a location-scale family as:

$$p(\lambda; \mu, \sigma) = \frac{1}{\sigma}\, p\left(\frac{\lambda - \mu}{\sigma}\right).$$

Assume that as $\lambda \to \infty$, $p(\lambda) = \Theta(\lambda^{-d})$ and that $p'(\lambda; \mu) = \mathcal{O}(\lambda^{-d-1})$ where p' is the derivative of p with respect to λ. We will show later that the Student-t family, Berger's prior, and the intrinsic prior all satisfy these two conditions.

We are now ready to state and prove the generalized polynomial tails comparison theorem. Denote by $\pi^P(\lambda \,|\, \text{data})$ and $\pi^U(\lambda \,|\, \text{data})$ the posterior densities employing the prior $p(\lambda \,|\, \mu, \sigma)$ and the uniform prior respectively. Applying Bayes' rule to both densities yields for any parameter value λ_0 the following ratio:

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

Theorem 6.1. *For fixed λ_0,*

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

Proof. Since our only assumptions on the prior $p(\lambda; \mu, \sigma)$ involve the asymptotic order of p and its derivative, and since these assumptions are not effected by a scaling factor σ, we may assume $\sigma = 1$ and drop σ from our notation.

We will show that

$$\lim_{\mu \to \infty} \frac{\int_{-\infty}^{\infty} f(\lambda) p(\lambda; \mu)\, d\lambda - p(\lambda_0; \mu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda}{p(\lambda_0; \mu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda} = 0. \tag{33}$$

Note that the numerator may be written as

$$\int_{-\infty}^{\infty} f(\lambda)(p(\lambda; \mu) - p(\lambda_0; \mu))\, d\lambda.$$

We break the region of integration in the numerator into two parts, $|\lambda| < \mu^k$ and $|\lambda| > \mu^k$, for some $0 < k < 1$ to be chosen later, and show that as $\mu \to \infty$ each integral goes to zero faster than the denominator.

First consider

$$\int_{|\lambda| < \mu^k} f(\lambda)(p(\lambda; \mu) - p(\lambda_0; \mu))\, d\lambda. \tag{34}$$

By the fact that $p(\lambda; \mu) = p(\lambda - \mu)$ and the mean value theorem we have

$$\left| \int_{|\lambda| < \mu^k} f(\lambda)(p(\lambda; \mu) - p(\lambda_0; \mu))\, d\lambda \right| \leq \int_{|\lambda| < \mu^k} f(\lambda) p'(\xi(\lambda)) |\lambda - \lambda_0|\, d\lambda$$

where $-\mu^k < \lambda, \lambda_0 < \mu^k$ and each $\xi(\lambda)$ is between $\lambda - \mu$ and $\lambda_0 - \mu$. Therefore $\xi(\lambda) = \mathcal{O}(\mu)$ and $p'(\xi(\lambda)) = \mathcal{O}(\mu^{-d-1})$. The term $|\lambda - \lambda_0|$ is $\mathcal{O}(\mu^k)$ and so the the integral (34) is $\mathcal{O}(\mu^{k-d-1})$. The denominator of (33) is $\Omega(\mu^d)$ and so the contribution of (34) to (33) is $\mathcal{O}(\mu^{k-d-1})/\Omega(\mu^{-d}) = \mathcal{O}(\mu^{k-1})$. Since $k < 1$, this term goes to zero as $\mu \to \infty$.

Next consider

$$\int_{|\lambda|>\mu^k} f(\lambda)(p(\lambda;\mu) - p(\lambda_0;\mu))\, d\lambda. \tag{35}$$

The term $(p(\lambda;\mu) - p(\lambda_0;\mu))$ is bounded and so by the assumption on the tails of the likelihood function f, the integral (35) is of order $\mathcal{O}((\mu^k)^{-d-\varepsilon}) = \mathcal{O}(\mu^{-k(d+\varepsilon)})$. Therefore the contribution of the integral (35) to the ratio (33) is $\mathcal{O}(\mu^{d-k(d+\varepsilon)})$ and so this term goes to 0 as $\mu \to \infty$ provided $k < d/(d+\varepsilon)$. □

Next we show that the Student-t family, the intrinsic prior, and Berger's prior all satisfy the conditions of the Generalized Polynomial Tail Comparison Theorem.

Clearly the tails of a Student-t distribution with ν degrees of freedom are $\Theta(\lambda^{-1-\nu})$. Also, the intrinsic prior and Berger's prior are clearly $\Theta(\lambda^{-2})$ in the tails. The derivative conditions remain to be demonstrated.

The density for a Student-t is proportional to

$$\left(1 + \frac{\lambda^2}{\nu}\right)^{-(\nu+1)/2}$$

and so its derivative is proportional to

$$-(1+\nu)\lambda \left(1 + \frac{\lambda^2}{\nu}\right)^{-(3+\nu)/2}$$

which is of order $\mathcal{O}(\lambda^{-2-\nu})$.

For the intrinsic prior, the asymptotic order is determined by the λ^{-2} term, the $1 - \exp(-\lambda^2)$ being essentially 1 in the tails. Therefore the asymptotic order of the derivative of the tails is λ^{-3}.

Showing that the derivative of Berger's prior has the necessary asymptotic order is more involved.

By differentiating inside the integral defining Berger's prior, we have

$$\frac{d}{d\lambda}\pi_B(\lambda) = -\frac{\lambda}{\sqrt{\pi}} \int_0^1 \exp(-w\lambda^2\nu) w^{3/2} \nu\, d\nu.$$

Next we multiply and divide by the derivative with respect to ν of $w\nu\lambda^2$ and define

$$M = \sup_{0\le\nu\le 1} \frac{1}{\sqrt{\pi}} \frac{w^{3/2}}{(w\nu)'} = \frac{1}{\sqrt{\pi(b+d)}}.$$

Then

$$|\pi'_{BP}(\lambda)| \le \frac{M}{\lambda} \int_0^1 \nu \exp(-w\lambda^2\nu)(w\lambda^2\nu)'\, d\nu.$$

Next, we integrate by parts, showing the the right hand side above equals

$$\frac{M}{\lambda}\left(\int_0^1 \exp(-w\lambda^2\nu)\,d\nu - \exp(-w\lambda^2)\right) \leq \frac{M}{\lambda}\left(\int_0^1 \exp(-w\lambda^2\nu)\,d\nu\right).$$

We can show that

$$\int_0^1 \exp(-w\lambda^2\nu)\,d\nu = \mathcal{O}(\lambda^{-2})$$

by an argument similar to that used to establish the bounds on the tails of $\pi_{BP}(\lambda)$. Therefore $\pi'_{BP}(\lambda) = \mathcal{O}(\lambda^{-3})$ and so π_{BP} satisfies the requirements of the Generalized Polynomial Tail Comparison Theorem.

Figures 6 and 7 display the qualitative forms of dependence of the posterior mean and variance on the discrepancy between the prior location parameter and the observed sample mean, for $n = 10$ and $\beta^2 = \sigma^2 = 1$. The posterior expectation and variance are shown as functions of the discrepancy $|\mu - \bar{X}_n|$. Figure 6 shows that the posterior expectations with a Cauchy prior and with Berger's prior are very similar. In both posterior models the posterior expectation has a bounded influence. On other hand, Figure 7 displays that the variances have the same qualitative form, but the variance with the Cauchy prior is smaller when μ tends to $\bar{X}_n$. We argue that the variance with Berger's prior is preferable than with the Cauchy in this example. Finally, if we consider a normal prior for this analysis then the posterior variance is constant in $|\mu - \bar{X}_n|$, and equal to 0.09.

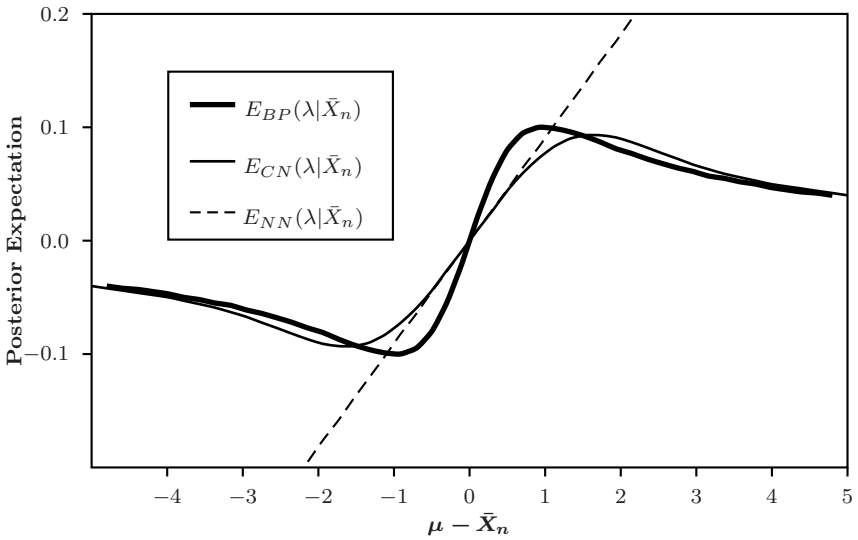

Figure 6: Behavior of the posterior expectation: $E_{BP}(\lambda|\bar{X}_n)$ in the BP/N, $E_{CN}(\lambda|\bar{X}_n)$ in the C/N and $E_{NN}(\lambda|\bar{X}_n)$ in the N/N model. For values $n = 10$, $\bar{X}_n = 0$ and $\beta = \sigma = 1$.

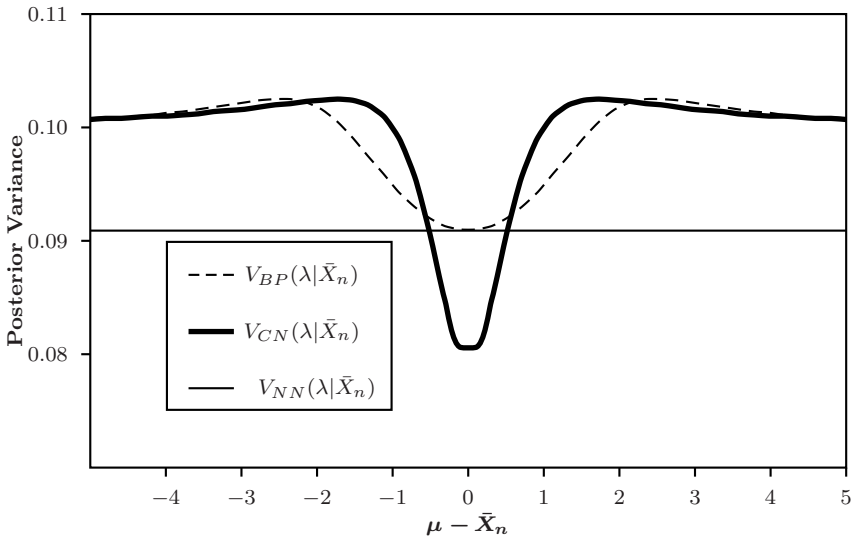

Figure 7: Behavior of the posterior variance: $V_{BP}(\lambda|\bar{X}_n)$ in the BP/N, $V_{CN}(\lambda|\bar{X}_n)$ in the C/N and $V_{NN}(\lambda|\bar{X}_n)$ in the N/N model. For values $n = 10$, $\bar{X}_n = 0$ and $\beta = \sigma = 1$.

Example 6.1. Application BP/N model for Example 3.1. In this example the Berger prior has $\mu = -1.52$ and $\beta = 0.63$. We must approximate the binomial likelihood by a normal distribution. For the likelihood (6), the Fisher information is $I_n(\lambda) = (ne^{\lambda}/(1+e^{\lambda})^2)$. In this example $\bar{X}_n \sim N(\log(0.8/(1-0.8)), (1+e^{1.38})^2/20e^{1.38})$, that is, $\bar{X}_n \sim N(1.38, 0.31)$. The posterior mean and variance of λ for the BP/N model are $E_{BP}(\lambda|\bar{X}_n) = 1.16$ and $V_{BP}(\lambda|\bar{X}_n) = 0.33$ respectively. These results are robust and very similar to the obtained with the Cauchy prior for the C/B model.

6.1 Application: BP/N and C/N model in a clinical trial

In this section we show application of the C/N and BP/N model in a clinical trial.

Example 6.2. Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine.
Reference: Pérez-Schael et al. (1997).
Study Design: Randomized, double blind, placebo-controlled trial.
Aim of Study: To compare rhesus rotaviruses-based quadrivalent vaccine (a new drug that is highly effective in preventing severe diarrhea in developed countries) and placebo.
Outcome measure: Over approximately 19 to 20 months, episodes of gastroenteritis were evaluated at the hospital among infants. The ratio of the odds of response (episode of gastroenteritis) following the new treatment to the odds of response on the conventional: $OR < 1$ therefore favors the new treatment.
Statistical Models: Approximate normal likelihood and normal prior for the logarithm of the odds ratio. In the Cauchy prior and Berger's prior the values of the

location parameters are the same with respect to normal prior. The scale is the same in the Cauchy and normal prior.
Prior Distribution: Was based on a published trial: Joensuu et al. (1997), where it is shown that in Finland the vaccine had a high success rate in preventing severe rotavirus diarrhea. In this trial the primary efficacy analysis was based on children of which 1128 received three doses of rhesus rotaviruses-based quadrivalent vaccine and 1145 placebo. 100 episodes of gastroenteritis were severe, 8 in rhesus rotaviruses-based quadrivalent recipients and 92 in placebo recipients:

		Vaccine	Placebo	
Event	Episode of gastroenteritis	8	92	100
	Non-episode of gastroenteritis	1120	1053	2173
		1128	1145	2273

Table 2: Episodes of gastroenteritis in the groups Vaccine and Placebo, Finland.

Loss function or demands: None specified.
Computation/software: Conjugate normal analysis and C/N and BP/N models.
Evidence from study: In this randomized, double-blind, placebo-controlled trial, 2207 infants received three oral of the rhesus rotaviruses-based quadrivalent vaccine. The following data show the episodes of gastroenteritis in Venezuela.

		Vaccine	Placebo	
Event	Episode of gastroenteritis	70	135	205
	Non-episode of gastroenteritis	1042	960	2002
		1112	1095	2207

Table 3: Episodes of gastroenteritis in the groups Vaccine and Placebo, Venezuela.

Results: We show the normal approximation for binary data for the log-odds with the approximate standard error recommended in Spiegelhalter et al. (2004) for 2×2 tables, following their suggestion of an standard error of the likelihood normal and N/N posterior model equal to $\sigma = 2$. In Table ?? the prior and likelihood have a standard deviation of $\sigma/\sqrt{n_0} = 0.36$ and $\sigma/\sqrt{n} = 0.15$ respectively. The posterior mean for the posterior model N/N is equal to $(n_0\mu + n\bar{X}_n)/(n_0 + n) = -0.99$. We see that the standard errors of the C/N and BP/N model with respect to the likelihood are equal. The influence of the equivalent number of observations in the posterior distribution ($n_0 + n = 31 + 178 = 209$, thus the likelihood can be thought to have around $178/31 \approx 6$ times as much information as the prior) over the standard error ($\sigma/\sqrt{n_0 + n}$) is very high in the N/N model. The data of the current experiment (data in the Venezuelan experiment) dominated the C/N and BP/N models, resulting in a posterior expectation much closer to the MLE.

	Location			Scale		
	Prior	Normalized likelihood	Posterior	Prior	Normalized likelihood	Posterior
N/N	-2.45	-0.73	-0.99	0.36	0.15	0.14
C/N	-2.45	-0.73	-0.76	0.36	0.15	0.15
BP/N	-2.45	-0.73	-0.76	0.36	0.15	0.15

Table 4: Exact and approximate moments of the N/N, C/N and BP/N models in the scale of log-odds.

The expectations of the BP/N and C/N posterior models and the MLE are approximately equal. We can see in Table 5 that N/N, C/N and BP/N posterior models are in favor of the vaccine (OR<1). However, the risk reduction in the N/N model is 63% (the estimated odds ratio is around $e^{-0.99} = 0.37$) and in the C/N and BP/N models is around 53% (in the normalized likelihood is 52%). The credible interval of the C/N and BP/N posterior model is closely related to the data in the trial. Finally, the

	OR	95% Credible Interval (OR Scale)
N/N	0.37	[0.28; 0.49]
C/N	0.47	[0.35; 0.63]
BP/N	0.47	[0.35; 0.63]
Normalized likelihood	0.48	[0.36; 0.65]

Table 5: Odds ratio and Credible Interval of the Posterior Model.

discrepancies between the expectation of the posterior models and the MLE are 1.86 standard errors for N/N and 0.2 for C/N and BP/N. This case dramatically illustrates the danger of assuming a conjugate prior as prior information in clinical trials. Figures 8 and 9 show the posterior distributions obtained in the conjugate analysis and non-conjugate analysis. We see that the prior distribution receives more weight in the N/N model. The posterior model C/N is very similar to the normalized likelihood. For the Figure 9 the posterior distributions for the C/N and BP/N model are almost the same. The results in the N/N model are suspect because the mean posterior is far from the likelihood given the conflict between the Finnish and the Venezuelan data. Incidentally, the researchers concluded that the Finnish and the Venezuelan responses were qualitatively different given the different levels of exposure of the children to the virus. In short, the robust analyses are giving a sensible answer while the conjugate analysis myopically insists that Finland and Venezuela are quite similar in respect to children's responses. On the other hand, if the two cases were indeed similar, without a drastic conflict on responses, then the robust analyses would give answers quite similar to the conjugate analysis, with conclusions with high precision. In other words, the use of robust priors makes Bayesian responses adaptive to potential conflicts between current data and previous trials.

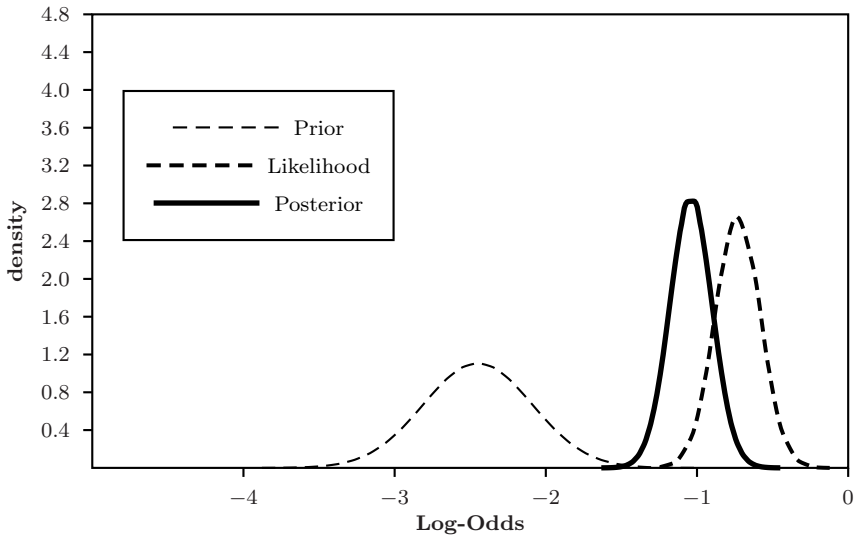

Figure 8: Prior(Finland), normalized likelihood (Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the N/N model.

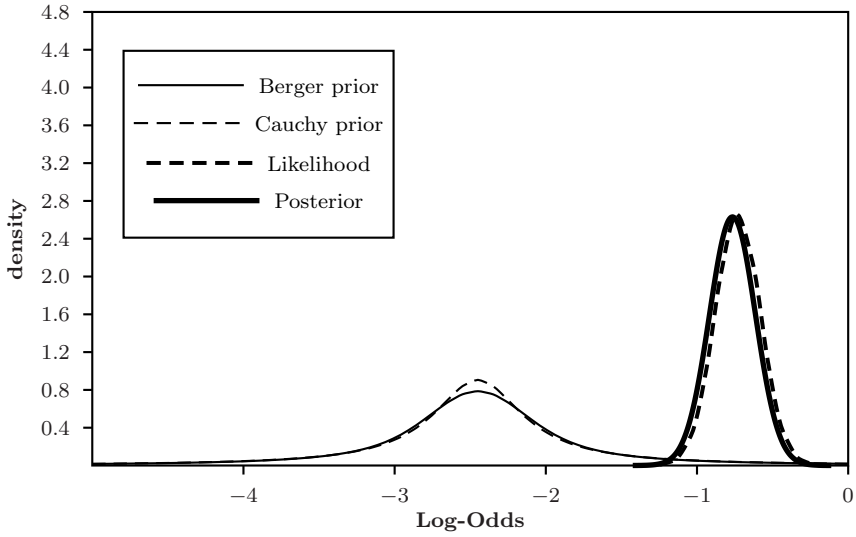

Figure 9: Prior(Finland), normalized likelihood(Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the C/N and BP/N model.

7 Conclusions

The issues discussed in this paper have led us to the following conclusions: 1). The Cauchy prior in the Cauchy/binomial model is robust but the beta prior in the conjugate beta/binomial model for inference on the log-odds is not. We can use the Cauchy/binomial model in clinical trials making a robust prediction in binary data. 2). Simulation of the moments in the Cauchy/binomial model reveals that the approximation performs well over a range of $n \geq 10$. Furthermore, we can use rejection sampling with either large or small sample sizes for exact results. 3) Berger's prior is very useful in clinical trials for a robust estimation since it gives closed-form exact results (when the normal log-odds likelihood is employed), and at the same time does not have the defects of conjugate priors. It can be argued that besides computational convenience it is superior to the Cauchy as a robust prior, because the posterior variance does not decrease as much as with the Cauchy, when the assessed priors scales are equal or close, see Figure 7. Berger's prior seems more cautious. 4). In more complex situations, with several different centers that are modeled with a hierarchical structure, the use of robust priors may be even more important. This will be explored elsewhere. 5). The use of prior information in terms of robust (and non-conjugate) priors will be much more acceptable to both researchers and regulatory agencies, because the prior can not dominate the likelihood when the data conflict with the prior. Remember the archetypal criticism of "Bayesian" analysis: "With Bayes, you can get the results you want, by changing your prior!" This should say instead: "With *conjugate* Bayes, you can get the results you want, by changing your prior!"

1 Proofs of Results 3.1

1.1 Cauchy Prior

Proof. Invoking the Polynomial Tails Comparison Theorem, we can use the uniform prior instead of the Cauchy prior when $\alpha \to \pm\infty$ for the binomial likelihood, (assuming that $0 < X_+ < n$) so the generating function for the C/B model is

$$\lim_{\alpha\to\pm\infty} E_C(e^{t\lambda}|X_+) = \frac{\int_{-\infty}^{\infty} \exp\left\{X_+\lambda - n\log(1+e^{\lambda}) + t\lambda\right\} d\lambda}{\int_{-\infty}^{\infty} \exp\left\{X_+\lambda - n\log(1+e^{\lambda})\right\} d\lambda}, \tag{36}$$

after of the transformation $\lambda = \log(\theta/(1-\theta))$, (36) is

$$\lim_{\alpha\to\pm\infty} E_C(e^{t\lambda}|X_+) = \frac{\Gamma(X_+ + t)\Gamma(n - X_+ - t)}{\Gamma(X_+)\Gamma(n - X_+)}, \tag{37}$$

hence

$$\lim_{\alpha\to\pm\infty} E_C(\lambda|X_+) = \Psi(X_+) - \Psi(n - X_+). \tag{38}$$

The approximation of the Digamma function (see Abramowitz and Stegun (1970)) is

$$\Psi(z) \approx \log(z) - \frac{1}{2z} - \mathcal{O}(z^{-2}), \tag{39}$$

hence

$$\lim_{\alpha\to\pm\infty} E_C(\lambda|X_+) \approx \log\left(\frac{\bar{X}_n}{1-\bar{X}_n}\right) - \frac{1}{2X_+} + \frac{1}{2(n-X_+)} - \mathcal{O}(X_+^{-2}) + \mathcal{O}((n-X_+)^{-2}). \tag{40}$$

Now, we show that the limit in (36) exists: consider the following functions of real variable with positive real values, defined by the equations

$$F(\lambda,t) = \frac{\exp\{(X_+ + t)\lambda\}}{(1+e^\lambda)^n}; f(\lambda) = \frac{\exp(X_+\lambda)}{(1+e^\lambda)^n}; \tau(\lambda) = \frac{1}{\beta^2+\lambda^2}, \tag{41}$$

where $X_+, n \in \mathbb{N}$; $n \geq 2$, $X_+ \geq 1$ and β is a positive constant. We prove that the convolutions of $F * \tau$ and $f * \tau$, defined respectively by the equations

$$\int_{\mathbb{R}} F(\lambda)\tau(\alpha-\lambda)d\lambda = \int_{-\infty}^{\infty} \frac{\exp\{X_+\lambda - n\log(1+e^\lambda) + t\lambda\}}{\beta^2+(\lambda-\alpha)^2}\, d\lambda \tag{42}$$

$$\int_{\mathbb{R}} f(\lambda)\tau(\alpha-\lambda)d\lambda = \int_{-\infty}^{\infty} \frac{\exp\{X_+\lambda - n\log(1+e^\lambda)\}}{\beta^2+(\lambda-\alpha)^2}\, d\lambda \tag{43}$$

are finite. For $\lambda \in (-\infty,\infty)$, we have

$$\begin{aligned} |F(\lambda)g(\alpha-\lambda)| &= \left|\frac{\exp\{(X_+ + t)\lambda\}}{(1+e^\lambda)^n(\beta^2+(\alpha-\lambda)^2)}\right| \\ &\leq \frac{|\exp(X_+ + t)\lambda|}{|\exp(n\lambda)|}\,\beta^{-2} \\ &\leq \frac{\exp\{(t-s)|\lambda|\}}{\beta^2} \\ &= g(\lambda) \end{aligned}$$

Since $|F(\lambda)g(\alpha-\lambda)|$ is dominated by the function $g(\lambda)$, and g belongs to $L^1(\mathbb{R})$, if $t-s \leq 0$ (where $s = n - X_+ \geq 1$). Therefore $F * \tau < \infty$. A similar argument shows

$$|f(\lambda)g(\alpha-\lambda)| \leq \frac{\exp\{-s|\lambda|\}}{\beta^2} \tag{44}$$

and thus $f * \tau < \infty$. □

1.2 Conjugate Prior

Proof. We have $E_B(\lambda) \to \infty$ as $a \to \infty$ and $E_B(\lambda) \to -\infty$ as $b \to \infty$, the approximation of the posterior expectation for the conjugate beta/binomial model is

$$\begin{aligned} E_B(\lambda|X_+) &\approx \log\left(\frac{n\bar{X}_n + a}{n(1-\bar{X}_n)+b}\right) - \frac{1}{2(n\bar{X}_n+a)} + \frac{1}{2(n(1-\bar{X}_n)+b)} \\ &- \mathcal{O}((n\bar{X}_n+a)^2) + \mathcal{O}((n(1-\bar{X}_n)+b)^2) \end{aligned}$$

and $E_B(\lambda|X_+) \to \infty$ as $a \to \infty$ and $E_B(\lambda|X_+) \to -\infty$ as $b \to \infty$. □

2 Proof of Result 5.1 Berger Prior

Proof. We make the change of variable $\eta = \lambda - \mu$. With the normal likelihood

$$f(\bar{X}_n \mid \eta) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{n}{2\sigma^2}(\eta - (\bar{X}_n - \mu))^2\right\}, \tag{45}$$

it follows that the predictive density satisfies the relation

$$m(\bar{X}_n) = \int_0^1 \int_{-\infty}^{\infty} K \times \exp\left\{-\frac{n}{2}K_2\right\} d\eta\, d\nu, \tag{46}$$

where

$$K_2 = \left[\frac{2\nu\eta^2}{\sigma^2(1-2\nu) + n\beta^2} + \frac{1}{\sigma^2}(\eta - (\bar{X}_n - \mu))^2\right]. \tag{47}$$

The method of completing the square tell us that

$$K_2 = \left[\eta - \frac{(\bar{X}_n - \mu)(\sigma^2(1-2\nu) + n\beta^2)}{\sigma^2 + n\beta^2}\right]^2 \frac{\sigma^2 + n\beta^2}{\sigma^2(1-2\nu) + n\beta^2} + \frac{2\nu(\bar{X}_n - \mu)^2}{\sigma^2 + n\beta^2}. \tag{48}$$

The generating-function of the posterior distribution (22) is given by

$$E_{BP}(e^{t\eta} \mid \bar{X}_n) = \frac{\int_0^1 \int_{-\infty}^{\infty} K \times \exp\left\{-\frac{n}{2}K_3\right\} d\eta\, d\nu}{\int_0^1 \int_{-\infty}^{\infty} K \times \exp\left\{-\frac{n}{2}K_2\right\} d\eta\, d\nu}, \tag{49}$$

where

$$K_3 = \left[\frac{2\nu\eta^2}{\sigma^2(1-2\nu) + n\beta^2} + \frac{1}{\sigma^2}(\eta - (\bar{X}_n - \mu))^2 - \frac{2t}{n}(\eta + \mu)\right]. \tag{50}$$

Hence, the cumulant-generating function of the posterior distribution (22) is given by

$$K_{\eta|\bar{X}_n} \propto \log\left[1 - \exp\left\{-\frac{n}{\sigma^2 + n\beta^2}\left(\bar{X}_n - \mu + \frac{t}{n}\right)^2\right\}\right] - 2\log\left(\bar{X}_n - \mu + \frac{t}{n}\right) + \frac{n}{2}\left(\bar{X}_n - \mu + \frac{t}{n}\right)^2 + t\mu. \tag{51}$$

□

References

Abramowitz, M. and Stegun, I. (1970). *Handbook of Mathematical Functions. National Bureau of Standards*, volume 46. Applied Mathematics Series. 823, 841

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, second edition. 819, 820, 829, 831

Berger, J. O. and Pericchi, L. R. (1996). "The Intrinsic Bayes Factor for Model Selection and Prediction." *JASA*, 91: 112–115. 820, 831, 833

Carlin, B. P. and Louis, T. A. (1996). "Identifying Prior Distributions that Produce Specific Decisions, with Application to Monitoring Clinical Trials." In *Bayesian Analisys in Statistics and Econometrics: Essays in Honor of Arnold Zellner*, 493–503. New York: Wiley. 818

Carlin, B. P. and Sargent, D. J. (1996). "Robust Bayesian Approaches for clinical trial monitoring." *Statistics in Medicine*, 15: 1093–1106. 818

Dawid, A. P. (1973). "Posterior expectations for large observations." *Biometrika*, 60: 664–667. 818

Evans, M. and Moshonov, H. (2006). "Checking for Prior-Data Conflict." *Bayesian Analysis*, 1: 893–914. 820

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). "A weakly informative default prior distribution for logistic and other regression models." *Annals of Applied Statistics*, 2: 1360–1383. 818, 823

Greenhouse, J. B. and Wasserman, L. A. (1995). "Robust Bayesian Methods for Monitoring Clinical Trials." *Statistics in Medicine*, 14: 1379–1391. 818

Gutierrez-Peña, E. (1997). "Moments for the Canonical Parameter of an Exponential Family Under a Conjugate Distribution." *Biometrika*, 84: 727–732. 825

Joensuu, J., Koskenniemi, E., Pang, X.-L., and Vesikari, T. (1997). "Randomised placebo-controlled trial of rhesus-human reassortant rotavirus vaccine for prevention of severe rotavirus gastroenteritis." *The Lancet*, 350: 1205–1209. 838

O'Hagan, A. (1979). "On outlier rejection phenomena in Bayes inference." *JRSSB*, B, 41: 358–367. 818

Pérez-Schael, I., Guntiñas, M. J., Pérez, M., Pagone, V., Rojas, A. M., González, R., Cunto, W., Hoshino, Y., and Kapikian, A. Z. (1997). "Efficacy of the Rhesus Rotavirus-Based Quadrivalent Vaccine in Infants and Young Children in Venezuela." *The New England Journal of Medicine*, 337: 1181–1187. 837

Pericchi, L. R. and Sansó, B. (1995). "A note on bounded influence in Bayesian analysis." *Biometrika*, B, 82(1): 223–225. 818

Pericchi, L. R., Sansó, B., and Smith, A. F. M. (1993). "Posterior Cumulant Relationships in Bayesian Inference Involving the Exponential Family." *JASA*, 88: 1419–1426. 825

Pericchi, L. R. and Smith, A. F. M. (1992). "Exact and Approximate Posterior Moments for a Normal Location Parameter." *JRSSB*, 54: 793–804. 820, 829

Smith, A. F. M. and Gelfand, A. E. (1992). "Bayesian Statistics Without Tears: A Sampling-Resampling Perspective." *The American Statistician*, 46: 84–88. 826

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. London: Wiley. 819, 824, 829, 838

Tierney, L. and Kadane, J. B. (1986). "Accurate Approximations for Posterior Moments and Marginal Densities." *JASA*, 81: 82–86. 824, 825

Tierney, L., Kass, R. E., and Kadane, J. B. (1989). "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions." *JASA*, 84: 710–716. 825

Funding

National Science Foundation (DMS-0604896 to LRP).

Acknowledgments

We thank Dr. María Egleé Pérez for helpful comments and several suggestions. JF was supported by Institute of Statistics, College of Business Administration, University of Puerto Rico-RRP and by M D Anderson Cancer Center. LRP was in sabbatical leave by The University of Puerto Rico-RRP. Detailed comments by referees and editors were most useful in preparing a final version.