# A Case for Robust Bayesian Priors with applications to Clinical Trials

**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. On the other hand, robust priors 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, after expressing in its exponentially family form, is improved by employing Cauchy priors. This requires no undue computational cost, given the advances in computation and analytical approximations. Moreover, we also 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 hyperparmeters that give closed-form results when coupled with a normal log-odds likelihood. Berger's prior yields the superior robust analysis with no added computational complication compared to the conjugate analysis. We illustrate the results with famous textbook examples and with a real data set and a prior obtained from a previous trial. On the formal side, we use a general and novel theorem, called 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. For Berger's prior robustness can be established directly since the exact expressions for posterior moments are known.

# A Case for Robust Bayesian Priors with applications to Clinical Trials

Jairo Fúquene[1], John D. Cook[2] and Luis Raúl Pericchi[3]. *

[1] Institute of Statistics, College of Business Administration, University of Puerto Rico, Rio Piedras.

[2] Department of Biostatistics, M.D.Anderson Cancer Center, University of Texas.

[3] Department of Mathematics. University of Puerto Rico, Rio Piedras.

**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. On the other hand, robust priors 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, after expressing in its exponentially family form, is improved by employing Cauchy priors. This requires no undue computational cost, given the advances in computation and analytical approximations. Moreover, we also 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 hyperparmeters that give closed-form results when coupled with a normal log-odds likelihood. Berger's prior yields the superior robust

*(Corresponding author) PO Box 23355. San Juan, PR 00931, USA. Tel. 1-787-528-1235; FAX 1-787-281-0651 luarpr@gmail.com

analysis with no added computational complication compared to the conjugate analysis. We illustrate the results with famous textbook examples and with a real data set and a prior obtained from a previous trial. On the formal side, we use a general and novel theorem, called 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. For Berger's prior robustness can be established directly since the exact expressions for posterior moments are known.
**Keywords:** *Berger's Prior, Clinical Trials, Exponential Family, Parametric Robust Priors, Polynomial Tails Comparison Theorem, Robust Priors.*

# 1 Introduction

In Bayesian statistics the selection of the prior family of 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. A common solution is to report the analysis using different priors: clinical, skeptical, and non-informative. The precision in these priors is important and sensitivity analysis regarding the priors is necessary. One approach to this problem is advocated by Greenhouse and Wasserman [1], who compute bounds on posterior expectations over an $\epsilon$-contaminated class of prior distributions. An alternative solution proposed in Carlin and Louis [2], where one re-specifies the prior and re-computes the result. These authors obtain fairly specific results for some restricted non-parametric priors classes. Along the same line, another alternative is the "prior partitioning" of Carlin and Sargent [3], who select 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 (very few) proposals, are on what can 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 [4], O'Hagan [5], Pericchi and Sansó [6]. 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 robustness. 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-parametrical class of priors. Finally, non-parametric priors do not appear to have achieved a significant impact

in practice, so give a parametric robustness a chance!

The popular Normal/Normal (N/N) or Beta/Binomial (B/B) conjugate analysis (see for example Spiegelhalter et al. [7].) will be exposed in this article as non-robust. Workable (parametric) alternatives are available to the practitioner. For motivation consider: The posterior mean $\mu_n$ 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, $\mu$, 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.)

However, to avoid confusion 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. Robust priors are not better because they influence the inference less (although they often do) 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 the parameters are most likely to live. In this paper we show that the Cauchy prior is robust in two posterior models for clinical trials. Pericchi and Smith [8] 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 proved 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 use strongly 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 [9], 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. 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 J.D. Cook, M.E. Perez, and L.R. Pericchi.

This article is organized as follows. Section 2 is devoted to present the Polynomial Tails Comparison Theorem. In section 3 we review the prior specification and posterior moments of the C/B model. In the 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 model in simulated and real clinical trials. We make some closing remarks in section 6.

# 2 The Polynomial Tails Comparison Theorem

In order to decide if a prior is robust with respect to a likelihood, the following theorem (see Fúquene et al. [10]) is decidedly useful and easy to apply.

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 $\nu$ degrees of freedom centered at $\mu$.

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

$$\int_{|\lambda|>m} f(\lambda)\, d\lambda = \mathcal{O}(m^{-\nu-1-\varepsilon}). \tag{2.1}$$

In the application we have in mind, $f$ is a Binomial likelihood function although the result is much more general. For instance the latter holds for any $\nu$ 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, yield for any parameter value $\lambda_0$ the following ratio:

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

**Theorem 2.1.** *For fixed $\lambda_0$,*

$$\lim_{\mu\to\infty} \frac{\int_{-\infty}^{\infty} f(\lambda)\, t(\lambda; \mu, \nu)\, d\lambda}{t(\lambda_0; \mu, \nu) \int_{-\infty}^{\infty} f(\lambda)\, d\lambda} = 1. \tag{2.2}$$

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

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

$$f(X_+|\lambda) \propto \exp\left\{X_+\lambda - n\log(1+e^{\lambda})\right\}, \tag{3.1}$$

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 can make use of the theorem.

First we perform a conjugate analysis, and express the Beta$(a, b)$ prior, after of the transformation of the parameter $\theta$ to log-odds, as

$$\pi_B(\lambda) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\left(\frac{e^{\lambda}}{1+e^{\lambda}}\right)^{a}\left(\frac{1}{1+e^{\lambda}}\right)^{b} \qquad a, b > 0. \tag{3.2}$$

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

$$K_{\lambda}(t) = -\log(\Gamma(a)\Gamma(b)) + \log(\Gamma(a+t)) + \log(\Gamma(b-t)), \tag{3.3}$$

hence

$$E_B(\lambda) = \Psi(a) - \Psi(b); \qquad V_B(\lambda) = \Psi^{'}(a) + \Psi^{'}(b), \tag{3.4}$$

where $\Psi(\cdot)$ is Digamma function and $\Psi^{'}(\cdot)$ is Trigamma function, that are extensively tabulated in for example Abramowitz and Stegun [11]. The posterior distribution of the B/B model is given by

$$\pi_B(\lambda \mid X_+) = K \times \exp\left\{(a+X_+)\lambda - (n+a+b)\log(1+e^{\lambda})\right\} \tag{3.5}$$

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 next sections for Berger's prior for an alternative), is to use a Cauchy prior for the natural parameter $\lambda$ in order to achieve robustness with respect to the prior,

$$\pi_C(\lambda) = \frac{\beta}{\pi[\beta^2 + (\lambda - \alpha)^2]}, \tag{3.6}$$

with parameters of localization and scale $\alpha$ and $\beta$ 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 the exponential family. One approach to the approximation of $m(X_+)$ is the Laplace's method, refined by Tierney and Kadane [12] 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. [7]. Suppose that previous experience with similar compounds has suggested that a drug has a true response rate $\theta$, 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 localization and scale whit in the Beta prior. For this example the localization 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.

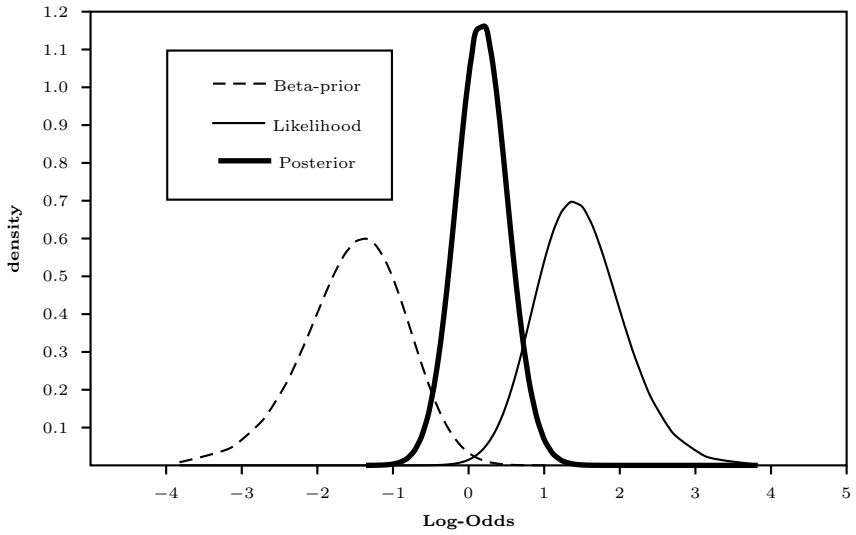


Figure 1: Beta prior, normalized binomial likelihood and B/B posterior model for the Example 1.

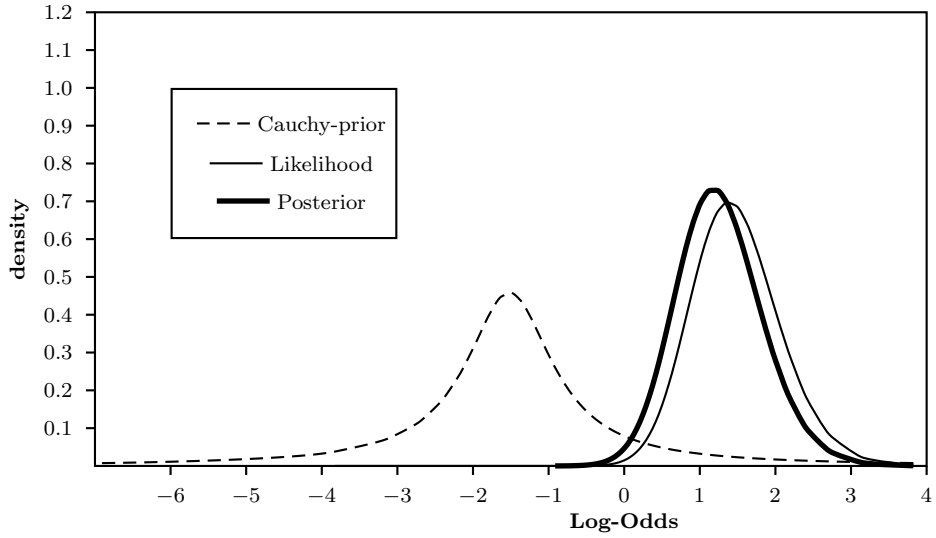


Figure 2: Cauchy prior, normalized binomial likelihood and C/B posterior model for the Example 1.

The posterior moments of the natural parameter of an exponential family are consider in Pericchi et al. [13] and Gutierrez-Pena [14]. The cumulant-generating function of the posterior, $\pi_B(\lambda \mid X_+)$, in the B/B model is

$$K_{\lambda|X_+}(t) = \log\left(\frac{\Gamma(X_+ + a + t)\Gamma(n - X_+ + b - t)}{\Gamma(X_+ + a)\Gamma(n - X_+ + b)}\right), \tag{3.7}$$

hence

$$E_B(\lambda|X_+) = \Psi(X_+ + a) - \Psi(n - X_+ + b) \tag{3.8}$$

$$V_B(\lambda|X_+) = \Psi^{'}(X_+ + a) + \Psi^{'}(n - X_+ + b). \tag{3.9}$$

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

$$\tilde{E}(\lambda|X_+) = \left(\frac{\sigma^*}{\hat{\sigma}}\right)\exp\left\{-n[h^*(\lambda^*) - h(\hat{\lambda})]\right\}, \tag{3.10}$$

where $-nh^*(\lambda) = \log(\lambda\pi_C(\lambda)f(X_+|\lambda))$, $\lambda^*$ is the maximum of $-h^*(\lambda)$ and $\sigma^* = [h^{*''}(\lambda)]^{-1/2}|_{\lambda=\lambda^*}$. The error in (3.10) is of order $\mathcal{O}(n^{-2})$ (see Tierney and Kadane [12]). However, in (3.10) we must assume that $\lambda$ does not change sign. Tierney et al. [15] recommend to add a large constant $c$ to $\lambda$, apply Laplace's method (3.10) and finally subtract the constant. We let $\tilde{E}_C(\lambda|X_+)$ and $\tilde{V}_C(\lambda|X_+)$ to denote approximate posterior expectation and posterior variance of the C/B model

$$\tilde{E}_C(\lambda|X_+) = \tilde{E}(c + \lambda|X_+) - c. \tag{3.11}$$

$$\tilde{V}_C(\lambda|X_+) = \tilde{E}((c + \lambda)^2|X_+) - [\tilde{E}(c + \lambda|X_+)]^2. \tag{3.12}$$

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 $\lambda$ 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, $\hat{\lambda} = \log((\bar{X}_n)/(1 - \bar{X}_n))$.

# 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 [16]). In the C/B model the *envelope function* is the Cauchy prior. The rejection method proceeds as follows:

1. Calculate $M = f(X_+|\hat{\lambda})$.
2. Generate $\lambda_j \sim \pi_C(\lambda)$.
3. Generate $U_j \sim$ Uniform(0,1).
4. If $MU_j\,\pi_C(\lambda_j) < f(X_+|\lambda_j)\,\pi_C(\lambda_j)$, accept $\lambda_j$; otherwise, reject $\lambda_j$. Go to step 2.

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. 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 addition, the approximations (3.11) and (3.12) are shown as functions of the discrepancy. Note that (3.12) is non-monotonic in the discrepancy. The posterior expectation, $\tilde{E}_C(\lambda|X_+)$, is a function of the "information discount."

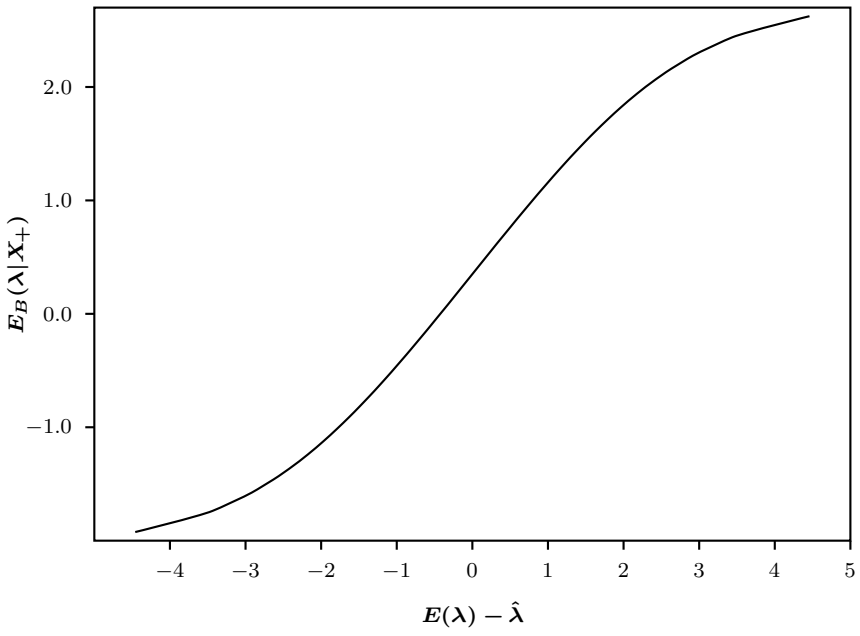


Figure 3: Behavior of the posterior expectation, $E_B(\lambda|X_+)$, in the B/B model for values $n = 10$, $\hat{\lambda} = 0$ and $a + b = 50$.

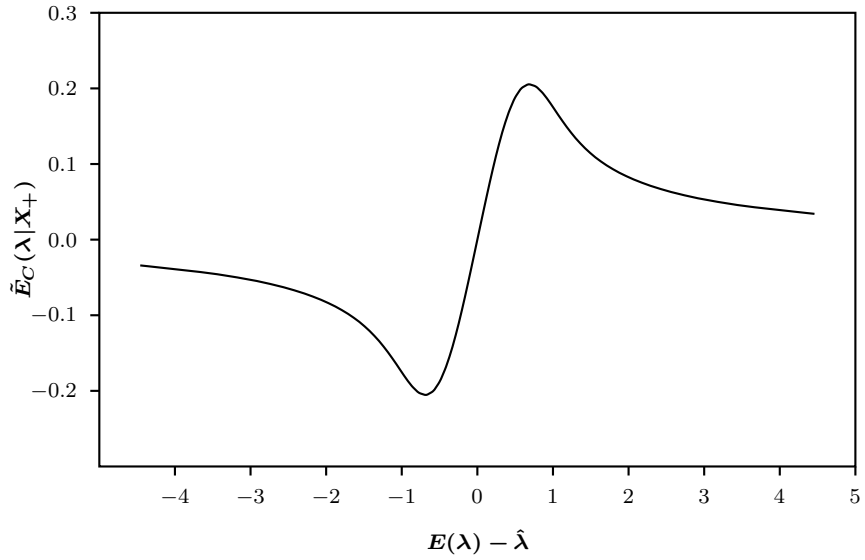


Figure 4: Behavior of the posterior expectation, $\tilde{E}_C(\lambda|X_+)$, in the C/B model for values $n = 50$, $\hat{\lambda} = 0$ and $a + b = 50$.

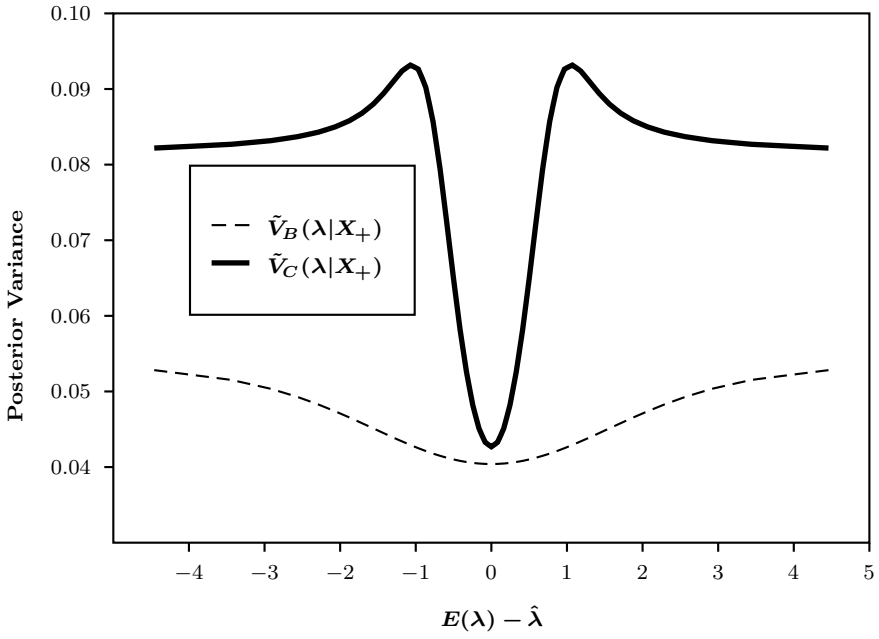


Figure 5: Behavior of the posterior variance, $V_B(\lambda|X_+)$ in the B/B and $\tilde{V}_C(\lambda|X_+)$ in the C/B for values $n = 50$, $\hat{\lambda} = 0$ and $a + b = 50$.

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

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

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

In Table 1 there is a big 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 $\theta$ that is the true response rate for a set Bernoulli trials, we know that the predictive mean of the total number of successes in $m$ trials is $E(X_m) = mE(\theta|X_+)$. 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 result of the C/B model is more closely related with to sampling data (because the maximum likelihood estimator of $\theta$ is 0.8 that is closer to 0.77) in contrast to the B/B model. The Beta prior is more "dogmatic" than the Cauchy prior leading to non robust results. Bayes is not dogmatic in general, but *conjugate* Bayes can be. This is a big 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. [7]. Pericchi and Smith [8] 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 and will then 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 the "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 [9] proposes for comparison of several means a robust prior (called the "Berger's Prior" in this work) that gives closed form results coupled Normal means. Berger's Prior (BP) is similar to a Cauchy prior in the tails but (for the same scale) it is less peaked at the mode. 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 $\mu$. This prior has the following form.

$$\pi_{BP}(\lambda) = \int_0^1 N(\lambda|\mu; \frac{d+b}{2\nu} - d) \cdot \frac{1}{2\sqrt{\nu}}\, d\nu \tag{5.1}$$

Here $N(\lambda|\mu, \tau^2)$ denotes a Normal density on the parameter $\lambda$ with mean and variance $\mu, \tau^2$ 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 $d = \sigma^2/n$. The posterior is proper whenever $b \geq d$ (see Berger [9]). Berger's prior then becomes

$$\pi_{BP}(\lambda) = \int_0^1 K \times \exp\left\{-\frac{n}{2}\left[\frac{2\nu(\lambda-\mu)^2}{\sigma^2(1-2\nu)+n\beta^2}\right]\right\} d\nu \tag{5.2}$$

where

$$K = \frac{\sqrt{n}}{\sqrt{4\pi(\sigma^2(1-2\nu)+n\beta^2)}}. \tag{5.3}$$

**Result 5.1.** *(see Berger [9] or Fúquene et al. [10]) Suppose that $X_1, \ldots, X_n \sim$ Normal$(\lambda, \sigma^2)$ where $\sigma^2$ is assumed known and $\lambda$ 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*

$$\pi_{BP}(\lambda|\bar{X}_n) = \frac{\pi_{BP}(\lambda)\exp\left\{-\frac{n(\bar{X}_n-\lambda)^2}{2\sigma^2}\right\}}{\frac{\sigma\sqrt{\sigma^2+n\beta^2}}{\sqrt{2}n(\bar{X}_n-\mu)^2}\left[1-\exp\left\{-\frac{n(\bar{X}_n-\mu)^2}{\sigma^2+n\beta^2}\right\}\right]}. \tag{5.4}$$

*The posterior expectation of the BP/N model*

$$E_{BP}(\lambda|\bar{X}_n) = \bar{X}_n + \frac{2\sigma^2 n(\bar{X}_n-\mu)^2 - 2\sigma^2(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)}{n(\bar{X}_n-\mu)(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)}, \tag{5.5}$$

*and the posterior variance of the BP/N model*

$$V_{BP}(\lambda|\bar{X}_n) = \frac{\sigma^2}{n} - \frac{\sigma^4}{n^2}\left\{\frac{4n^2(\bar{X}_n-\mu)^2 f(\bar{X}_n)}{(\sigma^2+n\beta^2)^2(f(\bar{X}_n)-1)^2}\right\} \tag{5.6}$$
$$+\frac{\sigma^4}{n^2}\left\{\frac{2(\sigma^2+n\beta^2)(f(\bar{X}_n)-1)((\sigma^2+n\beta^2)(f(\bar{X}_n)-1)-n)}{(\sigma^2+n\beta^2)^2(f(\bar{X}_n)-1)^2(\bar{X}_n-\mu)^2}\right\},$$

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

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 $\mu$ 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.

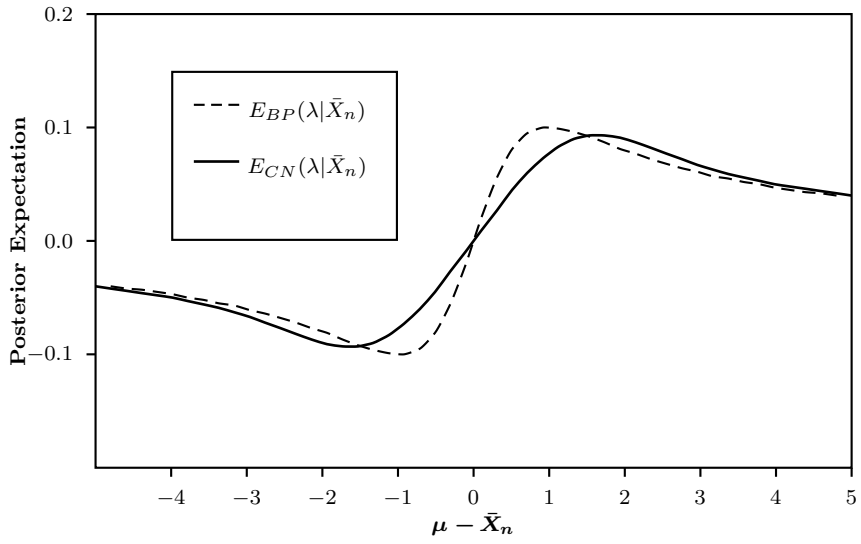


Figure 6: Behavior of the posterior expectation: $E_{BP}(\lambda|\bar{X}_n)$ in the BP/N and $E_{CN}(\lambda|\bar{X}_n)$ in the C/N. For values $n = 10$, $\bar{X}_n = 0$ and $\beta = \sigma = 1$.

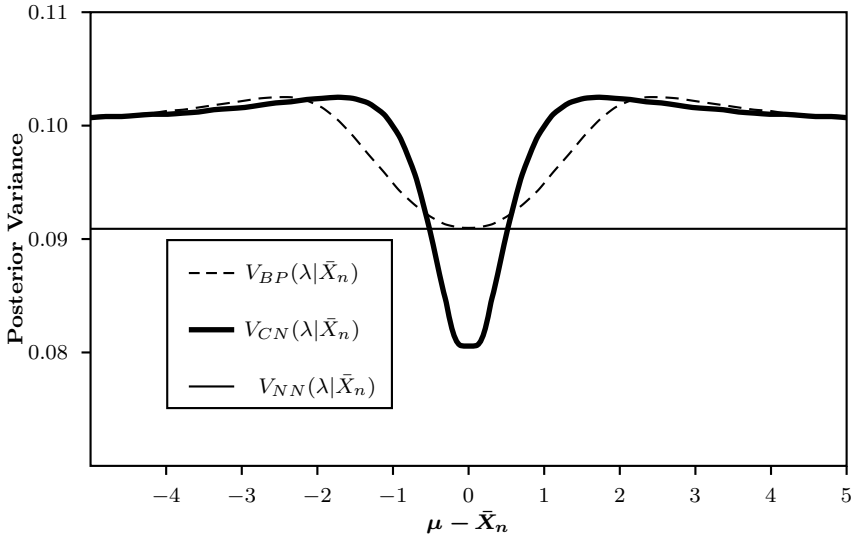


Figure 7: Behavior of the posterior variance: $V_{BP}(\lambda|\bar{X}_n)$ in the BP/N, $V_{CN}(\lambda|\bar{X}_n)$ in the C/N and $V_{NN}(\lambda|\bar{X}_n)$ in the N/N model. For values $n = 10$, $\bar{X}_n = 0$ and $\beta = \sigma = 1$.

## 5.1 Assessments of hyper-parameters

**"Sceptical" and "Enthusiastic" priors.** A useful suggestion under the *subjective Bayesian viewpoint*, taken by Spiegelhalter et al. [18], is to ask the the subject matter researchers, for reasonably optimistic and pessimistic priors (regarding the effectiveness of a new treatment). They do it for conjugate priors, but there is no need to do so. The same reasonably adversarial points of view can be taken with robust priors.

On the log-odds scale, an sceptical prior has mean zero (i.e. no difference between treatments or $\lambda = 0$) and a substantial probability that the new treatment is not better. The prior scale is assessed in reference to an optimistic hypothesis $\lambda_H$. Then a small probability $\xi$, is assessed, for example $\xi = 0.05$ that the effect of the treatment is equal or better than $\lambda_H$. For the Cauchy prior, the sceptical parameters are very easy to assess. The location $\alpha$ is zero and the scale $\beta = \lambda_H / \tan(\pi(\xi - 1/2))$. On the other hand for Berger's prior, again the location is zero and the scale is found from:

$$\int_0^1 \int_{-\infty}^{\lambda_H} N(\lambda|\mu; \frac{d+b}{2\nu} - d) \cdot \frac{1}{2\sqrt{\nu}}\, d\lambda\, d\nu = \xi. \tag{5.7}$$

It only remains to find $b$ from

$$\int_0^1 \frac{1}{2\sqrt{\nu}}\,\Phi\left(\frac{\lambda_H\sqrt{2\nu n}}{\sqrt{\sigma^2(1-2\nu)+bn}}\right)d\nu = \xi. \tag{5.8}$$

To solve $b$, it is useful to graph the integral as a function of $b$, computing the integral by the trapezoidal rule, or any other more accurate method, and read off the graph the intersection with the chosen value of $\xi$.

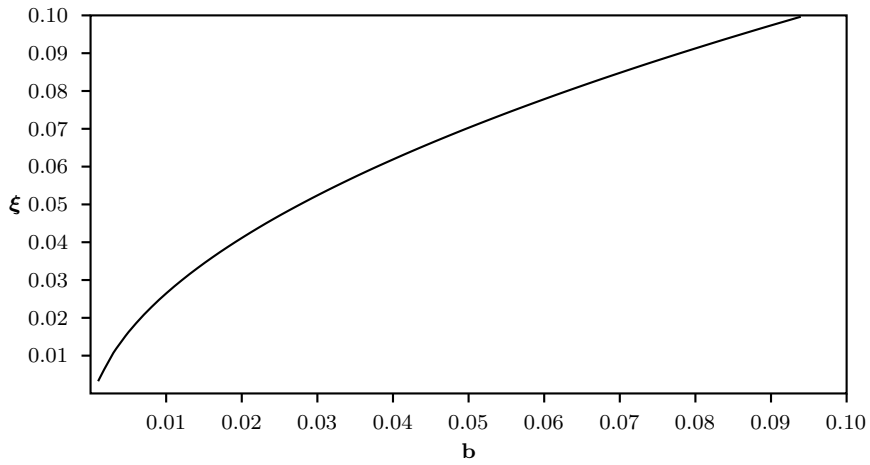


Figure 8: Assessments of the parameter b for $n = 25$ and $\lambda_H = 1$.

The Optimistic prior is assessed in an analogous way. For a more empirical point of view, the assessments can be based on a previous related experiment as it is in the next application.

**Example 5.1. Application B/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 (3.1), the Fisher information is $I_n(\lambda) = (ne^{\lambda}/(1+e^{\lambda})^2)$. So that 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 $\lambda$ 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.

## 5.2 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 5.2. Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine.**
**Reference:** Pérez-Schael et al. [19].
**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 localization parameters are the same with respect to Normal prior. The scale is the same in the Cauchy and Normal prior.
**Prior Distribution:** Was based in a published trial: Vesikari [20], where it is shown that in Finland the vaccine has high success rate in preventing severe rotavirus diarrhea.
**Loss function or demands:** None specified.
**Computation/software:** Conjugate Normal analysis and C/N and BP/N models.
**Evidence from study:** The following data show the episodes of gastroenteritis in Venezuela.

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

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

**Results:** We show the Normal approximation for binary data for the log-odds with the approximate standard error recommended in Spiegelhalter et al. [7] for $2 \times 2$ tables, following their suggestion of an standard error of the likelihood Normal and N/N posterior model equal to $\sigma = 2$. The posterior mean for the posterior model N/N is equal to $(n_0\mu + n\bar{X}_n)/(n_0 + n) = -1.60$.

Table 3: Exact and approximate moments of the N/N, C/N and BP/N models in the scale of log-odds.

| | Location | | | Scale | | |
|---|---|---|---|---|---|---|
| | Prior | Like | Post | Prior | Like | Post |
| N/N | -1.97 | -0.73 | -1.60 | 0.31 | 0.15 | 0.08 |
| C/N | -1.97 | -0.73 | -0.76 | 0.31 | 0.15 | 0.15 |
| BP/N | -1.97 | -0.73 | -0.76 | 0.31 | 0.15 | 0.15 |

In the Table 3. 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 ($n_0 + n = 406 + 170 = 576$) over the standard error ($\sigma/\sqrt{n_0 + n}$) is very high in the N/N model.

Table 4: Odds ratio and Credible Interval of the Posterior Model.

| | OR | 95% Credible Interval (Scale OR)) |
|---|---|---|
| N/N | 0.20 | [0.17; 0.23] |
| C/N | 0.46 | [0.35; 0.63] |
| BP/N | 0.46 | [0.35; 0.63] |
| Like | 0.48 | [0.36; 0.65] |

The data of the current experiment (data in the Venezuela experiment) dominated the C/N and BP/N models, resulting in a posterior expectation much closer to the MLE. The expectations of the BP/N model and the MLE are equal. We can see in the Table 3 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 80% and in the normalized likelihood and BP/N is around 52% (in the C/N model is 54%). The credible interval of the C/N and BP/N posterior model is closely related to the data in the trial.

Finally, the discrepancies between the expectation of the posterior models and the MLE are approximately 11 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 9 and 10 show the posterior distributions obtained in the conjugate analysis and non-conjugate analysis. We see that the prior distribution receive more weight in the N/N model. The posterior model C/N is very similar to the normalized likelihood. For the Figure 10 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 and the posterior precision is unacceptably high 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 conclusion the robust analyzes are giving the sensible answer, and 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 analyzes 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.

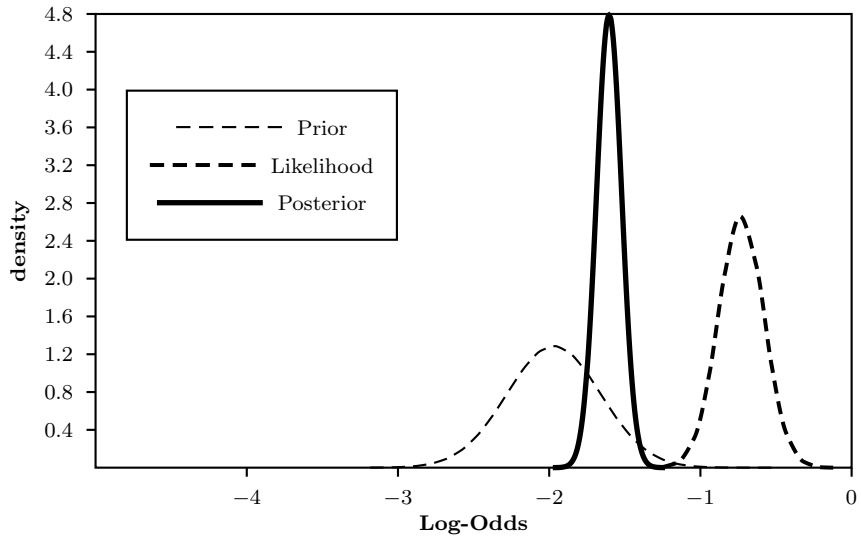


Figure 9: Prior(Finland), normalized likelihood(Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the N/N model.

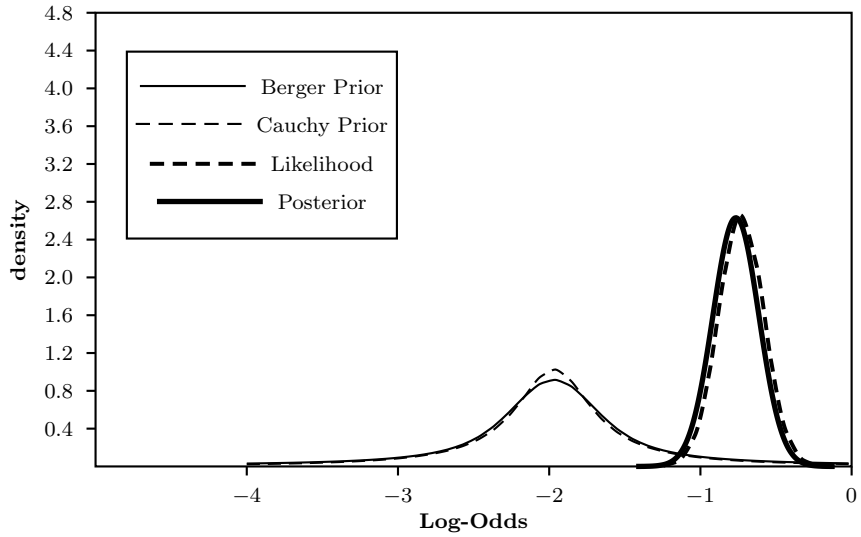


Figure 10: Prior(Finland), normalized likelihood(Venezuela) and posterior distributions in the Bayesian analysis of a trial of the Rhesus Rotavirus-Based Quadrivalent Vaccine for the C/N model.

# 6 Conclusions

The issues discussed in this paper have led us 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 the inference of 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 a 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 anti- "Bayesian" criticism: "With Bayes, you can get the results you want, by changing your prior!", should read instead: "With *conjugate* Bayes, you can get the results you want, by changing your conjugate prior!"

# References


1. Greenhouse JB, Wasserman LA. Robust bayesian methods for monitoring clinical trials. *Statistics in Medicine* 1995; **14**: 1379–1391.

2. Carlin BP, Louis TA. Identifying prior distribution that produce specific decisions, with application to monitoring clinical trials. *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner.* Wiley, New York. 1996; 493–503.

3. Carlin BP, Sargent DJ. Robust bayesian approaches for clinical trial monitoring. *Statistics in Medicine* 1996; **15**: 1093–1106.

4. Dawid AP. Posterior expectations for large observations. *Biometrika* 1973; **60**: 664–667.

5. O'Hagan A. On outlier rejection phenomena in bayes inference. *Journal of the Royal Statistics Society B* 1979; **41**: 358-367.

6. Pericchi LR, Sansó B. A note on bounded influence in Bayesian analysis. *Biometrika B* 1995; **82(1)**: 223–225.

7. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation.* Wiley, London. 2004.

8. Pericchi LR, Smith AFM. Exact and approximate posterior moments for a normal localization parameter. *Journal of the Royal Statistics Society* 1992; **54**: 793–804.

9. Berger JO. *Statistical Decision Theory and Bayesian Analysis.* second edn, Springer-Verlag. 1985.

10. Fúquene JA, Cook JD, Pericchi LR. A Case for Robust Bayesian priors with Applications to Binary Clinical Trials. *UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series.* Working Paper 44. 2008.

11. Abramowitz M, Stegun I. Handbook of Mathematical Functions. National Bureau of Standards, *Applied Mathematics Series* 1970; **46**.

12. Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 1986; **81**: 82–86.

13. Pericchi LR, Sansó B, Smith AFM. Posterior cumulant relationships in Bayesian inference involving the exponential family. *Journal of the American Statistical Association* 1993; **88**: 1419–1426.

14. Gutierrez-Peña E. Moments for the canonical parameter of an exponential family under a conjugate distribution. *Biometrika* 1997; **84**: 727–732.

15. Tierney L, Kass RE, Kadane JB. Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association* 1989; **84**, 710–716.

16. Smith AFM, Gelfand AE. Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician* 1992; **46**: 84–88.

17. Berger JO. Generalization of BIC. *Technical report, SAMSI.* 2005.

18. Spiegelhalter DJ, Freedman LS, Parmar MKB. Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society* 1994; **157**: 357–416.

19. Pérez-Schael I, Guntiñas MJ, Pérez M, Pagone V, Rojas AM, González R, Cunto W, Hoshino Y, Kapikian AZ. Efficacy of the rhesus rotavirus-based quadrivalent vaccine in infants and young children in Venezuela. *The New England Journal of Medicine* 1997; **337**: 1181–1187.

20. Vesikari T. Clinical experience with rotavirus vaccine in Finland. *2nd Satellite Symposium of the 14th Annual Meeting of the European Society for Paediatric Infectious Diseases (ESPID).* Denmark. 1996.

21. De Finetti B. The bayesian approach to the rejection of outliers. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Neyman, J. Editor, Berkeley, 1961; **1**: 199–210.

**Funding** National Science Foundation (DMS-0604896 to LRP)
**Acknowledgements**

We thank Dr. María Egleé Pérez for helpful comments and several suggestions. JF was supported by the Dean of graduate students, UPR-RRP and by MDAnderson Cancer Center. LRP was in sabbatical leave by The University of Puerto Rico-RRP.