Skeptical and Optimistic Robust Priors for Clinical Trials

John Cook.*
Jairo Fúquene. †
Luis Pericchi Guerra.‡

January 19, 2011

Abstract

A useful technique from the subjective Bayesian viewpoint, suggested by Spiegelhalter et al. (1994), is to ask the subject matter researchers and other parties involved, such as pharmaceutical companies and regulatory bodies, for reasonable optimistic and pessimistic priors regarding the effectiveness of a new treatment. Up to now, the proposed skeptical and optimistic priors have been limited to conjugate priors, though there is no need for this limitation. The same reasonably adversarial points of view can be taken with robust priors. A recent reference with robust priors usefully applied to clinical trials is in Fuquene, Cook, and Pericchi (2009). Our proposal in this paper is to use Cauchy and intrinsic robust priors for both skeptical and optimistic priors leading to results more closely related with the sampling data when prior and data are in conflict. In other words, the use of robust priors removes the dogmatism implicit in conjugate priors.

Keywords: *Clinical Trials, Skeptical and Optimistic Priors, Robust Priors.*

1 Introduction

Clinical trials are contentious. Pharmaceutical companies are eager to show that their new drug, on which they may have been invested millions of dollars, is a substantial improvement over the current treatment. On the other hand, government regulatory agencies take the opposite view and ask for substantial evidence that the new drug is not less effective than the current standard.

Bayesian statistics permits a useful technique for modeling adversarial positions by using *two* prior distributions on the parameters of interest. One, the optimistic prior, corresponds to positive expectations. The other, the pessimistic prior, corresponds to a more skeptical position.

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

†Institute of Statistics, School of Business Administration, University of Puerto Rico, Río Piedras Campus. PO Box 23332. San Juan, PR 00931-3332, USA. jairo.a.fuquene@uprrp.edu, supported by Institute of Statistics, School of Business Administration, UPR-RRP.

‡University of Puerto Rico, Rio Piedras. Department of Mathematics. luarpr@gmail.com, sponsored in part by NIH P20-RR016470, The Comprehensive Cancer Center of the UPR and by Merck

To be more specific, let us consider the following one-sided hypothesis test. Suppose that we have two medical treatments with corresponding probabilities of events p_1 and p_2, the events may be for instance "disease recurrence" or "deaths". The Odds Ratio is defined as

$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)},$$

and to make the assumption of Normality more realistic, the focus is on the parameter $\theta = \log(OR)$. Often the test of interest is the following:

$$H_0 : \theta \geq \theta_H, \ vs \ H_1 : \theta < \theta_H. \tag{1}$$

Here, H_1 is the region of substantial improvement by the new treatment over the standard one. Here, $\theta_H < 0$, corresponding to $OR < 1$, for example $OR = 0.5$, the point of 50% of improvement.
Part of the clever proposal by Spiegelhalter *et al* (2004) is to assume two priors, π_S and π_O. The skeptical position (say the government regulatory body attitude) is to first center the prior on the point of no difference, i.e. $\theta = 0$, and secondly to give only a small probability, denoted by ξ, say $\xi = 0.05$ or $\xi = 0.025$ to the improvement (i.e H_1), that is

$$P_S(H_1) = \int_{-\infty}^{\theta_H} \pi_S(\theta) d\theta = \xi. \tag{2}$$

On the other hand, the optimistic position is centered on θ_H and it gives only probability ξ that there is no benefit with the new treatment (the investigator position), so

$$P_O(\theta > 0) = \int_{0}^{\infty} \pi_0(\theta) d\theta = \xi. \tag{3}$$

If under the skeptical prior, the posterior probability of H_1 is bigger than $1-\gamma$, say $1-\gamma = 0.95$, that is if under the skeptical prior $P(H_1|\text{ Data }, \pi_S) > 1-\gamma$, then it is safe to decide in favor of H_1. On the other hand, if with the optimistic prior $P_O(\theta > 0|\text{ Data }) > \gamma$, it is safe to decide that there is not benefit with the new treatment. If there is disagreement between the procedures then the trial is inconclusive, given the information so far collected.

This approach is intuitively satisfying. However, this framework may be overly cautious and lead to an enormous delay in the decisions. The implementation proposed in Spiegelhalter *et al* (2004) uses conjugate priors, which lead to simple computations. However, we show in this article that we may preserve the useful framework of skeptical and optimistic priors without the dogmatism inherent in conjugate priors.

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 proposed by J. O. Berger (Berger 1985) that leads to a closed form posterior and moments, a sort of the "best of both worlds." On the other hand, the "intrinsic prior" was obtained in (Berger and Pericchi 1996) as the implicit prior to which the arithmetic intrinsic Bayes factor converges, and in fact it turns out that it is a limiting case of Berger's prior. In Fúquene *et al* (2009), robust

priors are proposed for clinical trials. But the priors studied there are clinical priors based on previous trials. Here, we propose a class of robust priors, a novel proposal to the best of our knowledge, that improves the class of Normal skeptical and optimistic priors originally proposed by Spiegelhalter et al (2004)(Note that the original proposal by Spiegelhalter et al propose a class of priors of two elements, the skeptical and the optimistic normal priors. We propose a class of robust priors, again of two elements, the skeptical and the optimistic robust prior. Effectively however, both these classes allow a substantial variation on prior assessments). We illustrate the improvement in terms of an example and also provide mathematical results that lead us to expect that the speed of convergence of the robust class is substantially higher than the normal class.

2 Illustration: Skeptical and Optimistic Robust Priors

A useful suggestion under the *subjective Bayesian viewpoint*, taken by Spiegelhalter et al. (1994), is to ask the subject matter researchers, for reasonably optimistic and pessimistic priors (regarding the effectiveness of a new treatment). On the log-odds scale, a skeptical prior on the amount of improvement has mean zero (i.e. no difference between treatments, $\theta = 0$) and a substantial probability that the new treatment is not better. The prior scale is assessed in reference to an optimistic hypothesis θ_H (say $\theta_H = -0.69$, which corresponds to a 50% improvement, see below). Then a small (skeptical) probability ξ, is assessed, for example $\xi = 0.025$ that the effect of the treatment is equal or better than θ_H. For the Cauchy prior, the skeptical parameters are very easy to assess. The location μ is zero and the scale $\beta = \theta_H / \tan(\pi(\xi - 1/2))$. In Fuquene, Cook and Pericchi (2009) other robust heavy tailed priors are considered, apart from the Cauchy, and several mathematical results are presented. Two such priors are the Intrinsic Prior, calculated by Berger and Pericchi (1996) and the so called Berger's prior, Berger (1985). There is a close relationship between the Intrinsic and Berger's prior, and for the most part we present the results for the Intrinsic prior, for which the location (μ) is zero and the scale, τ, is found by solving the following equation

$$\int_{-\infty}^{\theta_H} \frac{1}{2\sqrt{\pi}} \frac{1 - \exp\left\{-\dfrac{(\theta-\mu)^2}{\tau^2}\right\}}{(\theta-\mu)^2/\tau} d\theta = \xi. \tag{4}$$

On the other hand for the optimistic priors, the assessments are similar, except that the location is changed to $\mu = \theta_H = -0.69$, that is the location is now placed at the point of 50% improvement and the scales are obtained by placing $1 - \xi$ probability on H_1. For example, in equation (3) the right hand side is changed from ξ to $1 - \xi$.

Illustrative Example: Suppose that a new drug is compared with a conventional treatment. With the conventional treatment it is obtained that 26 out of 97 patients died (26.8%). On the other hand, only 13 out of 193 patients died with the new drug (6.7%). We wish to test the hypothesis that the improvement with the new treatment is at least 50% of reduction of the odds of death. Is there enough evidence to make a skeptical prior to give over 90% of posterior probability?

We now make the comparison of two analysis: The first combining the normal log odds

coupled with a normal sceptical prior (with the current data). Alternatively, we assume heavy-tailed Cauchy and intrinsic priors, with the same location and the same probabilities of the region of interest as the normal prior. Thus we assume that the normal, Cauchy, and intrinsic sceptical priors have all mean log(OR)=0 and with a 95% interval running from 50% reduction in odds of death (log(OR) = -0.69) to a 50% increase (log(OR) = 0.69). In the normal prior case we assume $\pi_n(\theta) = N[0, \sigma^2/n_0]$, and n_0 is referred to as "the number of prior observations". On a log(OR) scale, this prior has a 95% interval from -0.69 to 0.69 and so has a standard deviation $0.69/1.96 = 0.35$ (for $\xi = 0.025$) and hence the number of prior observations is $n_0 = 4/0.35^2 = 32.3$. For Cauchy and intrinsic priors the scale is 0.05 and 0.06 respectively, so that the prior probability of the region on which there is a reduction of risk of 50% or more is 0.025. We use the normal approximation for binary data for the log-odds with the approximate standard error recommended in Spiegelhalter et al. (2004, page 26) for 2×2 tables, following their suggestion of an standard error of the likelihood normal and normal/normal posterior model equal to $\sigma = 2$. As we anticipated above, assume the evidence arising from the study about 30-day mortality was 26/97 on control and 13/193 on new treatment. If the ratio of the odds of death following the new treatment to the odds of death on the conventional is OR< 1 then the data favors the new treatment. We have that the estimated log(OR) is $\bar{y}_n = -1.6$ (OR=0.2 or 80% risk reduction) with estimated standard error 0.36 and $n = 4/0.36^2 = 30.5$, which in this case is approximately the same weight of the prior $n_0 = 32.3$, as calculated above. We use the R (R Development Core Team 2009) package named ClinicalRobustPriors, available from the Comprehensive R Archive Network at http://CRAN.R-project.org/package=ClinicalRobustPriors, which can be used to compute probabilities and figures for the prior, likelihood and posterior models. The posterior mean in the normal/normal model is $(n_0\mu + n\bar{X}_n)/(n_0 + n) = -0.77$ with standard deviation $\sigma/\sqrt{n_0 + n} = 0.25$, the estimated odds ratio is $e^{-0.77} = 0.46$ or 54% risk reduction. In the Cauchy/normal and intrinsic/normal models the posterior mean is -1.48 ($e^{-1.48} = 0.22$ or 78% risk reduction) with standard deviation 0.32. In the normal/normal an 95% credible interval on the log(OR) scale is between -1.27 and -0.28 that corresponds to odds ratios from 0.28 to 0.75, or a 95% probability that the true risk reduction lies between 25% and 72%. For the Cauchy/normal and intrinsic/normal posterior models the 95% credible interval show that the true risk reduction lies between 57% and 88%. On the other hand, the likelihood shows a risk reduction between 60% and 90%. In Figure 1 we can see that the normal skeptical prior is more dogmatic than both the Cauchy and intrinsic skeptical priors. When there is a discrepancy between prior and data, the robust priors are discarded to some extend in favor of the likelihood, but the normal prior is not.

As a counterbalance to the skeptical priors, Spiegelhalter et al. (1994) suggest an "enthusiastic" (or "optimistic" as we call it here) prior centered on the alternative hypothesis and with a high chance that the true treatment benefit is over 50%. In this example the alternative hypothesis is $\theta_H = -0.69$ (50% risk reduction) and the high chance is $1 - \xi = 0.975$. The scale and prior sample size are the same as with the skeptical priors. Figure 2 display the results. The posterior mean in the normal/normal conjugate model is -1.13 ($e^{-1.13} = 0.32$ or 68% risk reduction), much closer than the -1.39 ($e^{-1.39} = 0.24$ or 76% risk reduction) of the Cauchy/normal or intrinsic/normal models. The scale for the posterior models are 0.25 and 0.34 respectively with normal and Cauchy (or intrinsic) priors. For the Optimistic prior, there is no strong conflict with the likelihood and the inference with normal and robust priors is similar.

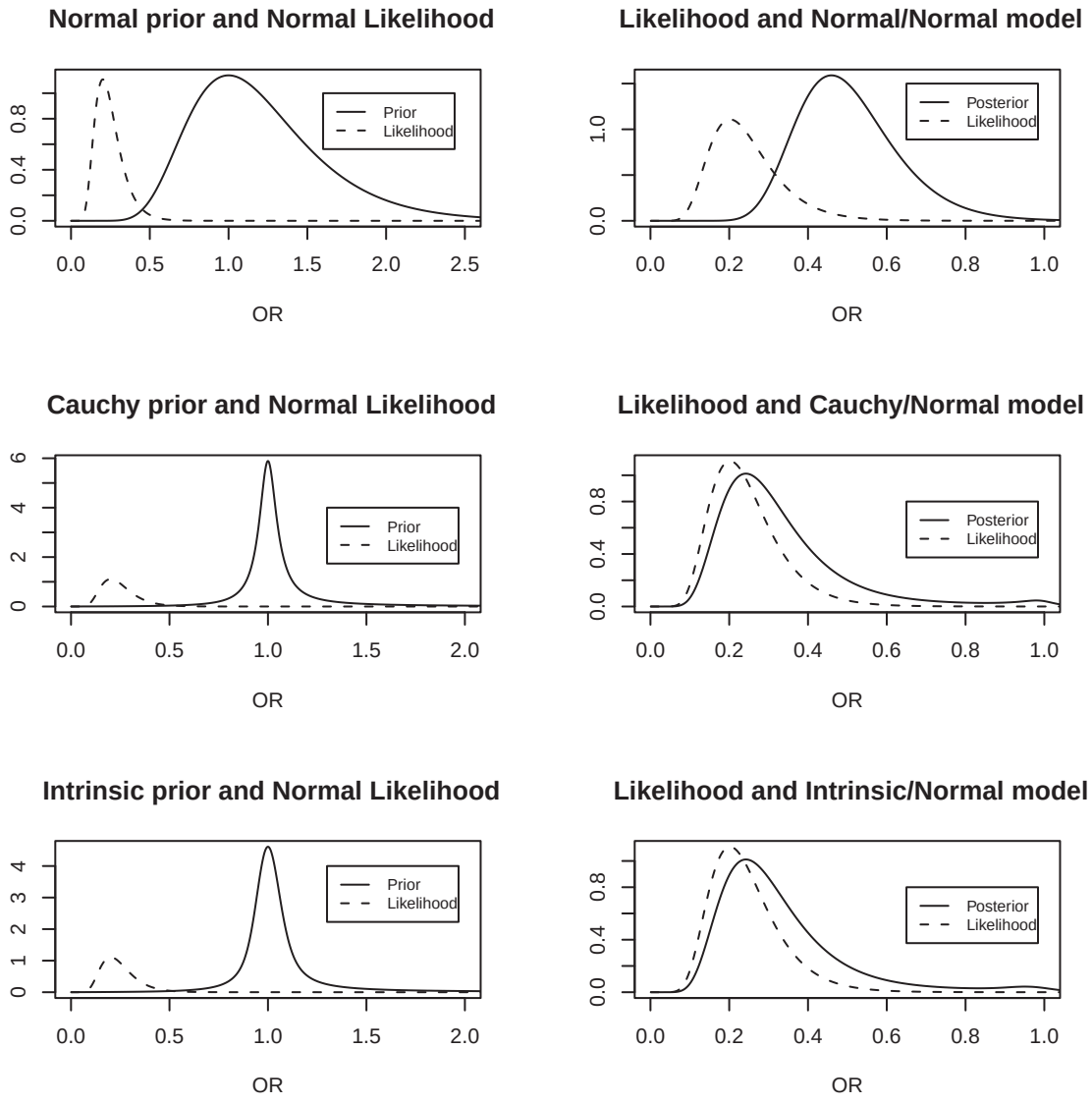

Figure 1: Skeptical priors, likelihoods and posterior models: normal/normal, cauchy/normal and intrinsic/normal.

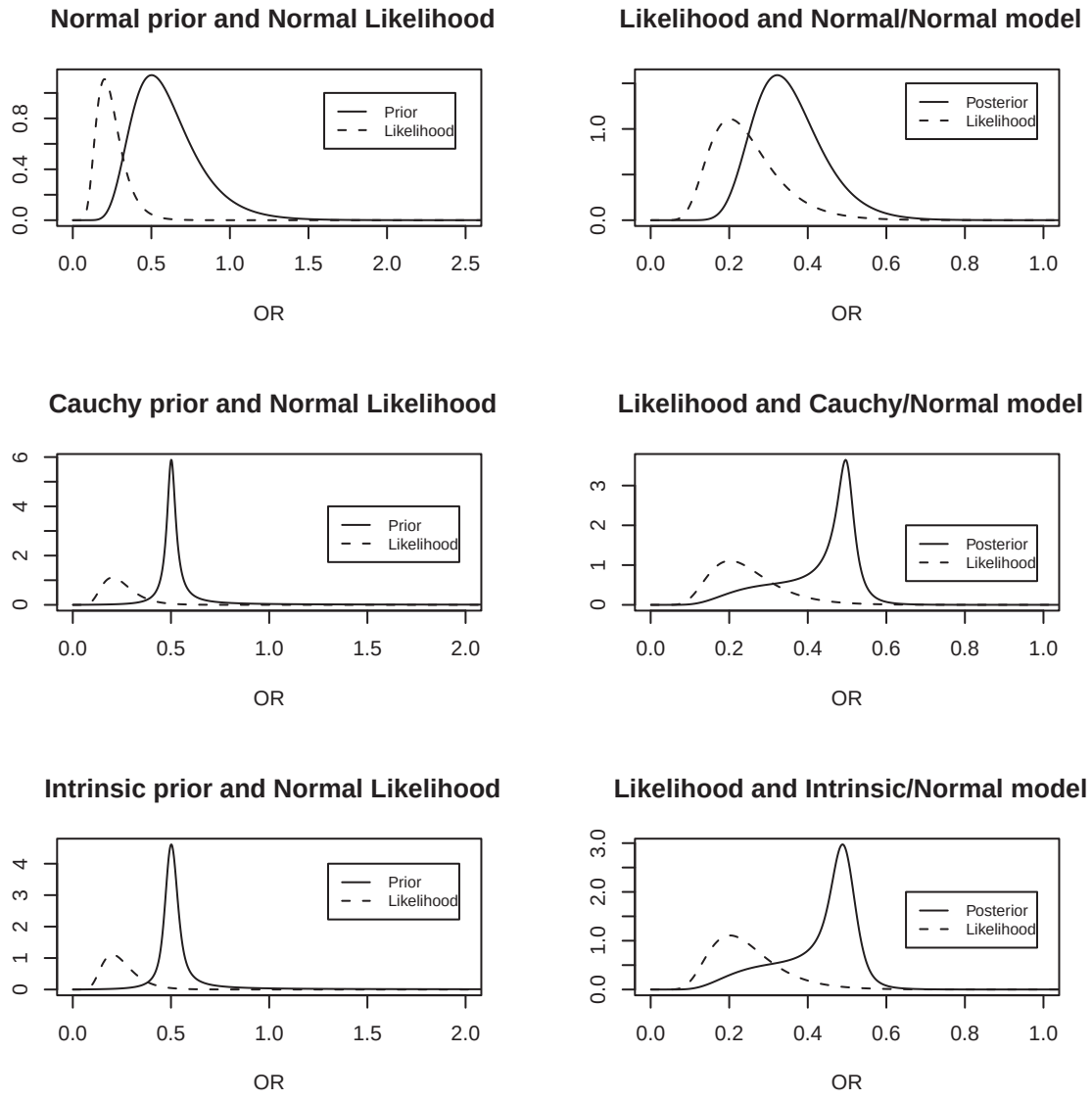

Figure 2: Optimistic priors, likelihoods and posterior models: normal/normal, cauchy/normal and intrinsic/normal.

Conclusion for the Example The quantity of interest is the set of a reduction of risk of at least of 50%, that is H_1. For the normal skeptical prior, the posterior probability can be calculated in the log-odds scale as $\Phi((-0.69 + 0.77)/0.25)) = \Phi(0.32) = 0.62$. This contrast with both the Cauchy and intrinsic skeptical robust priors, for which this probability is 0.95. In other words the hypothesis H_1 is supported by the data, even assuming the Cauchy and intrinsic sceptical robust priors. This is not the case with the normal skeptical prior. Next consider the optimistic priors case. This probability with the normal prior is $\Phi((0 + 1.13)/0.25) = \Phi(4.52) = 0.99$. The same probability is obtained with both the Cauchy and the intrinsic. Thus, in the example, the skeptical and optimistic robust priors have converged in over 95% of probability of the region of substantial improvement. Unfortunately, using the normal priors, the procedure has not yet converged, and much more evidence is still required for a definitive conclusion, and this is due entirely to the particular functional form chosen for the prior, and not in its condition of skepticism. In our experience, the results obtained using any of the considered robust priors, Cauchy or intrinsic or Berger's prior, yield essentially the same results.
That the behavior observed in the illustration is a general feature of robust and normal priors, rather than specific phenomenon of this example is demonstrated in the next section.

3 Asymptotic Results

In this section we show that robust priors can "change their mind" more readily than conjugate priors by looking at asymptotic properties of posterior means under each type of prior. When both the skeptical and optimistic prior are robust, they may reach agreement more quickly than if the two priors were not robust.

Consider a single sample y from a normal(θ, σ^2) distribution where θ has a conjugate normal(0, τ^2) prior. It is well-known that the posterior distribution on θ has mean

$$\frac{\tau^2}{\tau^2 + \sigma^2} y.$$

If we use a Cauchy(0, 1) prior rather than a normal(0, τ^2) prior on θ above, the posterior mean of θ is

$$y - \mathcal{O}\left(\frac{1}{y}\right)$$

as $y \to \infty$. See the technical report Cook (2010) for the derivation of this asymptotic result.

Under the normal prior, the posterior mean of θ under-estimates the data y by a constant ratio which depends on the relative scales of the sampling distribution and the prior distribution. Under the Cauchy prior, however, the posterior mean of θ asymptotically approaches the value of y, independent of the scales of the sampling and prior distributions. This is illustrated by Figure 3.

Next we consider the case of n samples y_i with mean $\overline{y}$ and consider the behavior of the posterior mean as $n \to \infty$. Under a normal(0, τ^2) prior, the posterior mean is given by

$$\frac{1}{1 + \frac{\sigma^2}{n\tau^2}} \overline{y} = \left(1 - \frac{\sigma^2}{n\tau^2} + \mathcal{O}\left(\frac{1}{n^2}\right)\right) \overline{y}.$$

Under a Cauchy prior, however, the posterior mean is

$$\overline{y} + \frac{(\overline{y}^2 - 3)\overline{y}}{(1+\overline{y}^2)^2}\frac{\sigma^2}{n} + \mathcal{O}\left(\frac{1}{n^2}\right).$$

The rate at which the posterior mean converges to $\overline{y}$ depends on τ in the case of the normal prior and on $\overline{y}$ in the case of the Cauchy prior. For any value of τ, the convergence is faster under the Cauchy prior for sufficiently large values of $\overline{y}$.

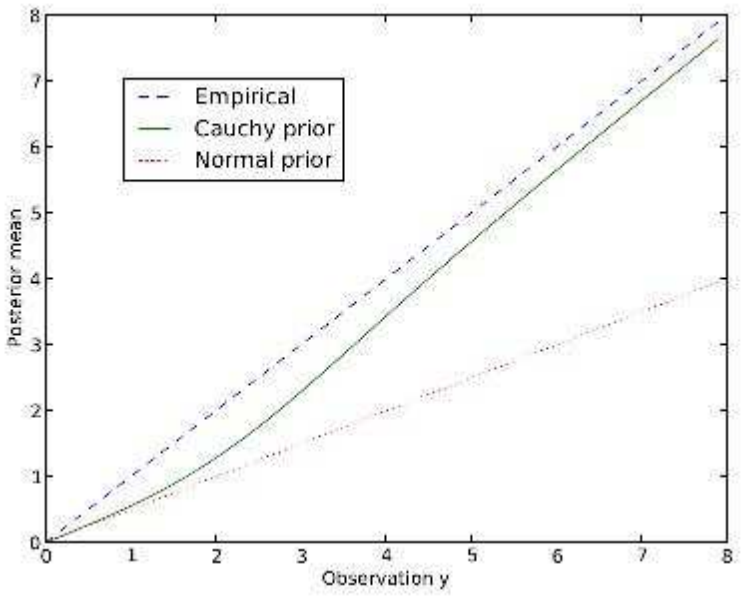

Figure 3: Posterior mean of θ under robust and conjugate priors compared to empirical value

We now obtain results for Berger's prior that are quite similar to those above for the Cauchy prior. If we observe n samples from a normal(θ, σ^2) distribution and θ has Berger's prior with location 0, the posterior mean of θ is

$$\frac{2\sigma^2\overline{y}}{(\sigma^2 + n\beta^2)\left(\exp\left(\frac{n\overline{y}}{\sigma^2+n\beta^2}\right) - 1\right)} - \frac{2\sigma^2}{n\overline{y}}$$

where $\overline{y}$ is the sample mean of the observations as shown in Fúquene, Cook, and Pericchi (2009). Therefore as a single sample $y \to \infty$, the posterior mean of θ is

$$y - \frac{2\sigma^2}{y} + \mathcal{O}(\exp(-y^2)).$$

Also, as $n \to \infty$, the posterior mean of θ is

$$\overline{y} - \frac{\sigma^2}{n}\frac{2}{\overline{y}} + \mathcal{O}(\exp(-n)).$$

These results show that for Cauchy and Berger priors, the influence of the prior diminishes more quickly when the data are far from the location parameter of the prior. When skeptical and optimistic priors using either of these robust distributions, the priors will reach consensus more quickly than corresponding conjugate priors when the data are in conflict with one or other of the priors.

4 Conclusions

The idea of skeptical and optimistic priors is an important one, and useful to increase the impact of Bayesian statistics in the whole of medical statistics. As introduced by Spiegelhalter et al. (2004) this idea has two components. The first is that the prior probability of the region where the new treatment is substantially better than the current treatment is high for the optimistic prior and low for the skeptical prior. The second component is the specific shape of the priors (effectively a class of two priors), that was assumed by Spiegelhalter et al. as normal priors. We show here, to the best of our knowledge for the first time, that the behavior of the procedure, even complying with the first component, crucially depends on the shape of the priors, particularly the tail size. We illustrate that with an example, similar as those in Spiegelhalter et al. (except that the equivalent sample sizes of the likelihood and priors are about the same in our example) and show it mathematically in Section 3. Furthermore, we put forward that the idea of skeptical and optimistic prior is actually improved by the use of heavy tailed priors, because the procedure is "less stubborn" and even though it places a handicap on both hypothesis, it is more willing to change its mind when there is clear information in favor of one of the hypothesis. As a consequence, the robust procedure put forward here is able to embrace better treatments or to discard bad ones in a more efficient manner: it is more likely that the optimistic and skeptical procedure reach a consensus, under robust heavy tailed priors than under normal priors. Finally, the use of different heavy tailed priors presented here, the Cauchy, Berger's or Intrinsic priors lead to essentially equivalent statistical behavior, and thus the choice among them is more a matter of convenience and taste. But the difference between normal and robust skeptical/optimistic priors is substantial, and the robust version of the idea seems to us much more acceptable for the players involved.

References

[1] Berger, J. O. (1985), Statistical Decision Theory and Bayesian Analysis, second edition, Springer-Verlag.

[2] Berger, J. O. and Pericchi, L. R. (1996). "The Intrinsic Bayes Factor for Model Selection and Prediction." JASA, 91: 112-115.

[3] Cook, J. D. (2010) "Asymptotic results for Normal-Cauchy model." UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series. Working Paper 61. http://www.bepress.com/mdandersonbiostat/paper61

[4] Fúquene, J., Cook J. and Pericchi L.R. (2009) "A Case for Robust Bayesian Priors with Applications to Clinical Trials", *Bayesian Analysis*, 4, Number 4, pp. 817-846.

[5] Pericchi, L. R. and Smith, A. F. M. (1992). "Exact and Approximate Posterior Moments for a Normal Location Parameter." JRSSB, 54: 793-804.

[6] Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). Bayesian Approaches to Clinical Trials and Health-Care Evaluation. London: Wiley.

[7] Spiegelhalter, D. J., Freedman, L. S. and Parmar, M. K. B. (1994), "Bayesian approaches to randomized trials (with discussion)", Journal of the Royal Statistical Society, 157: 357-416.