Bayesian Analysis (2017) **12**, Number 3, pp. 615–637

The Scaled Beta2 Distribution as a Robust Prior for Scales

María-Eglée Pérez[*], Luis Raúl Pericchi[†], and Isabel Cristina Ramírez[‡]

Abstract. We put forward the Scaled Beta2 (SBeta2) as a flexible and tractable family for modeling scales in both hierarchical and non-hierarchical settings. Various sensible alternatives to the overuse of vague Inverted Gamma priors have been proposed, mainly for hierarchical models. Several of these alternatives are particular cases of the SBeta2 or can be well approximated by it. This family of distributions can be obtained in closed form as a Gamma scale mixture of Gamma distributions, as the Student distribution can be obtained as a Gamma scale mixture of Normal variables. Members of the SBeta2 family arise as intrinsic priors and as divergence based priors in diverse situations, hierarchical and non-hierarchical.

The SBeta2 family unifies and generalizes different proposals in the Bayesian literature, and has numerous theoretical and practical advantages: it is flexible, its members can be lighter, as heavy or heavier tailed as the half-Cauchy, and different behaviors at the origin can be modeled. It has the *reciprocality* property, *i.e* if the variance parameter is in the family the precision also is. It is easy to simulate from, and can be embedded in a Gibbs sampling scheme. Short of not being conjugate, it is also amazingly tractable: when coupled with a conditional Cauchy prior for locations, the marginal prior for locations can be found explicitly as proportional to known transcendental functions, and for integer values of the hyperparameters an analytical closed form exists. Furthermore, for specific choices of the hyperparameters, the marginal is found to be an explicit "horseshoe prior", which are known to have excellent theoretical and practical properties. To our knowledge this is the first closed form horseshoe prior obtained. We also show that for certain values of the hyperparameters the mixture of a Normal and a Scaled Beta2 distribution also gives a closed form marginal.

Examples include robust normal and binomial hierarchical modeling and meta-analysis, with real and simulated data.

Keywords: Scaled Beta2 distribution, prior for scale parameters, horseshoe prior, intrinsic priors, divergence priors, reciprocality.

1 Introduction

The focus of this paper is to propose the Scaled Beta2 (SBeta2) family of distributions,

$$\text{SBeta2}(\psi|p,q,b) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)\cdot b}\cdot\frac{(\frac{\psi}{b})^{(p-1)}}{((\frac{\psi}{b})+1)^{(p+q)}},\ \text{for } \psi>0, b>0, p>0, q>0 \qquad (1)$$

[*] University of Puerto Rico, Río Piedras Campus, maria.perez34@upr.edu
[†] University of Puerto Rico, Río Piedras Campus, luis.pericchi@upr.edu
[‡] Universidad Nacional de Colombia Sede Medellín, iscramirezgu@unal.edu.co

DOI: 10.1214/16-BA1015

as a convenient family of prior distributions for scale parameters, both for informative and quasi-non-informative scenarios, and for hierarchical and non-hierarchical models. The intention is to provide an alternative to the use of the Inverted Gamma distribution, and to show that the SBeta2 has a natural motivation, is flexible and tractable.

The SBeta2 is a comprehensive family that encompasses and expands several previous proposals. Two of the most noteworthy, in the context of random effects models, are Gelman (2006) and Berger (2006). Gelman (2006) proposes the half-Cauchy and the Uniform distributions for the between groups standard deviations. Berger (2006) proposes using $1/\sqrt{\sigma^2}$ as prior for the between groups variance. We claim that the SBeta2 contains these proposals, exactly or approximately. Take the case of the half-Cauchy for the standard deviation, where an application of the change of variables formula leads to a SBeta2$(1/2, 1/2, b)$ for the variance. On the other hand, the SBeta2$(1/2, q, b)$, for small q, say $0 < q \leq 1/2$ and large b is a useful approximation to $1/\sqrt{\sigma^2}$. Other proposals are contained in Pericchi (2010) and Polson and Scott (2012), both putting forward a SBeta2 but without explicit mention of scale. However, a more flexible family is achieved including an adjustable scale, without which, for example, the distribution may not approximate sufficiently well the Uniform distribution.

Certainly, other alternative families to the Gamma/Inverted-Gamma have been proposed, for instance Griffin and Brown (2010) and Frühwirth-Schnatter and Wagner (2010), among others. Nevertheless, we argue that the SBeta2 is a flexible family that is able to model the advantages of the previous proposals, like heavy tails or boundedness/unboundedness at the origin, etc. Besides it has additional advantages that will be discussed in the sequel. It is no surprise that in Bayesian Statistics modeling and testing, scattered particular cases of the SBeta2 or of the Beta2 distribution have appeared (like Gelman's half-Cauchy). These include: Bradlow et al. (2002), Scott and Berger (2006), Liang et al. (2008), Maruyama and George (2011), Wang and Sun (2013) and Sparks et al. (2013). Noteworthy is the appearance of particular members of the SBeta2 family in Pericchi (2005) as intrinsic priors for testing the scale of an Exponential Law and in Giron et al. (2006) as intrinsic priors for scales in the Linear Model. In Supplementary Appendix 3 (Pérez et al., 2016) we show that in a normal model with known mean, the SBeta2 distribution is the intrinsic prior of the scale parameter.

We justify our proposal of using the SBeta2 family for modeling scales based on a combination of theoretical and practical considerations.

First of all, it has a natural motivation as a Gamma scaled mixture of a Gamma distribution as shown in Lemma 1. It has the property of reciprocality, *i.e.*, if $p(\psi)$ belongs to the SBeta2 family, so does $p(1/\psi)$, which is not a property of the Gamma/Inverted-Gamma family (the half-Cauchy distribution proposed by Gelman to model standard deviations also has this attractive property, and it is reassuring that this is a particular case of our proposal as mentioned before). It is flexible enough for modeling a variety of behaviors at the origin and at the tail, and for specific hyperparameters boundedness at the origin and heavy right tail is obtained, as heavy or even heavier than the Cauchy distribution.

Secondly, it is convenient practically: it can be simulated from in various ways (as a Gamma scaled mixture of Gammas or as the scaled odds of a Beta distribution), and thus it can be easily embedded in a Gibbs sampler scheme. Also the SBeta2 family is amenable for elicitation as one of its parameter controls the behavior at the origin, another the right tail behavior and the scale can be assessed in a variety of ways, as will be seen later in this work.

Thirdly a variety of analytical results spring from the SBeta2 family, that we summarize here:

a) With conditional Normal priors for location and SBeta2 for the precision, the marginal prior for location can be found in closed form for specific values of hyperparameters. It is bounded at zero and heavy tailed.

b) When a SBeta2 prior for the scale is coupled with a conditional Cauchy prior for locations, the marginal prior for locations can be found explicitly as proportional to known transcendental functions, and for integer values of the hyperparameters they are found in analytical closed forms. Furthermore, for specific choices of the hyperparameters, the marginal is found to be an explicit horseshoe prior (Carvalho et al., 2010) with a pole at the origin and heavy tail, leading to a sort of nearly optimal choice as a prior for sparse locations. This seems to be the first explicit horseshoe prior in the literature.

c) Again for a conditional Cauchy prior for location, if now the square of the scale is modeled as a SBeta2, the marginal is no longer a horseshoe prior, but a general closed form result is obtained. This strategy leads also to a very useful prior distribution, called the Student-SBeta2 distribution (Fúquene et al., 2014).

It is important to emphasize that analytical results in cases (b) and (c), are obtained for heavy tailed distributions for locations and in (a) for light tailed distributions.

This paper is organized as follows: in Section 2 we motivate the SBeta2 family showing that the SBeta2 distribution can be obtained as a scale mixture of Gamma distributions, we present some of its properties and we discuss how to use the SBeta2 distribution as a prior for variances and precisions. Section 3 deals with closed forms for mixtures of SBeta2 and Cauchy, Student and Normal distributions. In Section 4 we give examples to illustrate the advantages of the use of SBeta2 distributions as priors for scale parameters. Finally, we summarize some conclusions about those advantages.

2 The SBeta2 distribution

2.1 Motivation

A simple and natural motivation of the SBeta2 springs from the robustification of hierarchical models. As a simple example consider a balanced analysis of variance (ANOVA) model with k groups, n observations by group and possibly different variances:

$$X_{ij} = \mu_i + \varepsilon_{ij},\ \varepsilon_{ij} \sim N(0, h_i) \tag{2}$$

where μ_i and h_i refer to the mean and precision of group i, respectively, and N stands for the Normal distribution.

The usual second level in the hierarchy reads as

$$\begin{aligned} \mu_i &\sim N(\theta, h_0), && (3)\\ h_i &\sim \text{Gamma}(p, b). && (4) \end{aligned}$$

This model is known to be "non-robust" in the sense that the amount of shrinkage may be too heavy for an outlying observation. To reduce the excessive shrinkage, following a suggestion that goes back to De Finetti (1961), a level of hierarchy is added to have a scale mixture of Normals, replacing (3) by

$$\mu_i \sim N(\theta, h_0\rho_{1i}),\ \ \rho_{1i} \sim \text{Gamma}(\nu/2, \nu/2). \tag{5}$$

This effectively replaces the Normal by a Student distribution since

$$\text{Student}_\nu(\mu_i|\theta, 1/h_0) = \int \text{Normal}(\mu_i|\theta, h_0\rho_{1i}) \cdot \text{Gamma}(\rho_{1i}|\nu_2, \nu/2) d\rho_{1i}.$$

Similarly, replacing (4) as

$$h_i \sim \text{Gamma}(p, b/\rho_{2i}),\ \ \rho_{2i} \sim \text{Gamma}(q, 1),$$

yields the SBeta2 distribution as prior for the precisions h_i. This, we prove in the sequel, effectively replaces the Gamma by a scaled version of the Beta2 distribution, the *Scaled Beta2* distribution given in (1). This result, formalized in Lemma 1 below, describes an effective way to generate SBeta2 random variables.

Lemma 1. *The SBeta2 density is obtained as a Gamma mixture of Gamma densities or Inverted Gamma densities as follows:*

$$\text{SBeta2}(\psi|p, q, b) = \int_0^\infty \text{Gamma}(\psi|p, \frac{b}{\rho})\text{Gamma}(\rho|q, 1)d\rho.$$

Similarly,

$$\text{SBeta2}(\sigma^2|q, p, 1/b) = \int \text{Inverse-Gamma}(\sigma^2|p, \tau^2) \cdot \text{Gamma}(\tau^2|q, b^{-1})d\tau^2.$$

Proof. See Supplementary Appendix 1. □

In Section 4 we will present examples of the use of the SBeta2 in practical situations, where it will be seen that the use of this distribution as prior for scale parameters promotes robustness in hierarchical models and produces sensible analyses in diverse settings.

2.2 Properties of the SBeta2

We now explore the properties of SBeta2 distribution that can be helpful for assessment of the hyperparameters.

For the SBeta2 distribution defined in (1),

$$\begin{aligned} E[\psi] &= \frac{p}{q-1} b \text{ when } q > 1 \\ Var[\psi] &= \frac{p(p+q-1)}{(q-1)^2(q-2)} b^2 \text{ when } q > 2 \end{aligned}$$

It is also easy to see that if $\psi \sim \text{SBeta2}(p, q, b)$, then $\phi = \frac{1}{\psi} \sim \text{SBeta2}(q, p, \frac{1}{b})$ (the reciprocality property).

As we already discussed, the SBeta2 can be generated using Lemma 1, as a Gamma scale mixture of Gamma distributions. Another easy way to generate $\psi \sim \text{SBeta2}(p, q, b)$ random variables is as $\psi = \frac{\theta}{(1-\theta)} b$, where θ follows a Beta(p, q) distribution, that is, as a scaled odds of Beta random variables.

Consider a distribution in a scale or location-scale family with unknown scale parameter σ. We suggest to specify the prior as $\sigma^2 \sim \text{SBeta2}(p, q, b)$, or equivalently for the reciprocal (precision), $h = \frac{1}{\sigma^2} \sim \text{SBeta2}(q, p, \frac{1}{b})$.

For selecting values for the hyperparameters p, q and b, the following properties can be useful.

1. Behavior at zero:

$$f_{\sigma^2}(0) = \begin{cases} \infty & p < 1 \\ \frac{q}{b} & p = 1 \\ 0 & p > 1 \end{cases}, \qquad f_h(0) = \begin{cases} \infty & q < 1 \\ pb & q = 1 \\ 0 & q > 1 \end{cases}$$

2. When $q \leq 1$, $E(\sigma^2) = \infty$ Similarly, when $p \leq 1$, $E(h) = \infty$.

3. Location of the mode:

$$\text{mode}(\sigma^2) = \begin{cases} \frac{p-1}{q+1} b & p \geq 1 \\ 0 & \text{Otherwise} \end{cases}$$

$$\text{mode}(h) = \begin{cases} \frac{q-1}{p+1} \frac{1}{b} & q \geq 1 \\ 0 & \text{Otherwise} \end{cases}$$

4. When $p = q$, the median of the SBeta2 distribution is the scale parameter b. For $p = 1$, the median turns out to be $b \cdot (2^{1/q} - 1)$.

5. If σ is half-Cauchy with scale b, then a direct application of the change of variable formula shows that $\sigma^2 \sim \text{SBeta2}(1/2, 1/2, b^2)$.

2.3 Robustness of the SBeta2

One way to measure the thickness of the tails is to measure the index ρ of a regularly varying density (Andrade and O'Hagan, 2006).

Definition. The right-hand tail of a density $f(y)$ is regularly varying with index ρ if

$$\frac{f(\lambda y)}{f(y)} \to \lambda^{\rho}, \text{as } y \to \infty \text{ for all } \lambda > 0.$$

Note that a Student-t distribution with ν degrees of freedom has index $\rho = -(\nu + 1)$. Computation shows that the SBeta2(p, q, b) has $\rho = -(q + 1)$, so the tail behavior is totally defined by q. Thus for $q = 1$ we have a tail behavior of the same index as a Cauchy distribution. More generally the tail index of a SBeta2(p, q, b) is that of a Student-t with q degrees of freedom. On the other hand the behavior at the origin is commanded by the value of the parameter p. For $p > 1$ the density function is zero at the origin, and for $p < 1$ it is infinity at the origin. For $p = 1$ the density function is bounded at the origin.

2.4 Some thoughts about elicitation

In general, we want to have heavy tails for a robust inference, but we don't want to give high weight to $\sigma^2 = 0$. So, our suggestion for selecting the hyperparameters p, q and b is taking $1/2 \leq p \leq 1$ and $0 < q \leq 1$. Note however that other values of (p, q) may be necessary, like $q > 1$, based on stability considerations in complex Markov Chain Monte Carlo (MCMC) modeling (Pericchi et al., 2011). One way to assess b is to fix it empirically as the median (or somehow higher than the median) or based on subject matter knowledge. Another possibility is to assess probability statements like $P(\psi > a) = c$ for $\psi \sim SBeta2(p, q, b)$. This approach can be very useful since the SBeta2 distribution can be regarded as the distribution of the scaled odds of a Beta random variable. We then have $P(\psi > a) = P(\theta > \frac{a}{(a+b)})$ where $\theta \sim$ Beta(p, q), which can be easily solved using statistical software. Note that the $p = q = 1$ and $p = q = \frac{1}{2}$ cases have a special standing regarding objective priors, as the first corresponds to a Bayes–Laplace Uniform prior for θ and the second corresponds to the Jeffreys prior for θ. There are several other possibilities for elicitation, for example by empirical Bayes methods. We will return to the assessment in the examples.

3 Closed form results for mixtures with SBeta2 distribution

In this section, we show that the SBeta2 is amazingly tractable for Bayesian analysis, and produces some interesting heavy tail distributions for locations. Here the SBeta2 will be used as a prior distribution of the precision of a Normal and the scale and square scale of a Cauchy distribution (some results in this section were obtained using Wolfram Alpha LLC., 2014).

3.1 Normal-Scaled Beta2 distribution prior

Let $\theta \sim N(0, \tau)$, i.e.

$$f(\theta|\tau) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\tau\theta^2\right)$$

where the precision τ follows a SBeta2(p, q, b) distribution. We use the representation of the SBeta2 distribution as Gamma scale mixtures of Gamma distributions in Lemma 1 for calculating the marginal distribution of θ by changing the order of integration

$$\pi(\theta|p,q,b) \quad = \quad \int_0^\infty \int_0^\infty f(\theta|\tau) \cdot \text{Gamma}(\tau|p, b/\rho)\text{Gamma}(\rho|q,1) d\tau d\rho,$$

The integrand with respect to τ simplifies to:

$$\int_0^\infty \tau^{p-1/2} \exp\left(-\tau\left(\frac{\theta^2}{2}+\frac{\rho}{b}\right)\right) d\tau = \frac{\Gamma(p+1/2)}{\left(\frac{\theta^2}{2}+\frac{\rho}{b}\right)^{(p+1/2)}}.$$

and the integral becomes,

$$\pi(\theta|p,q,b) = \frac{\Gamma(p+0.5)}{\Gamma(p)\Gamma(q)\sqrt{2\pi}b^p} \int_0^\infty \exp(-\rho) \cdot \frac{\rho^{p+q-1}}{(\theta^2/2+\rho/b)^{(p+0.5)}} d\rho.$$

For the important particular case $p = q = 1$, the integral reduces to

$$\pi(\theta|1,1,b) = \frac{\Gamma(1.5)}{\sqrt{2\pi}\cdot b} \int_0^\infty \frac{\rho}{\exp(\rho)\cdot(\theta^2/2+\rho/b)^{1.5}} d\rho$$

This last integral can be explicitly calculated, and the result is

$$\pi(\theta|1,1,b) = \frac{1}{2}\sqrt{\frac{b}{2}}\left[2\sqrt{\pi}e^{\frac{b\theta^2}{2}}(1+b\theta^2)(1-\Phi(\sqrt{b}|\theta|)) - \sqrt{2b}|\theta|\right]$$

We will call this the *Normal-Scaled Beta2 distribution*. It is a scale family with scale $\frac{1}{\sqrt{b}}$. The density is shown in Figure 1 for different values of b.

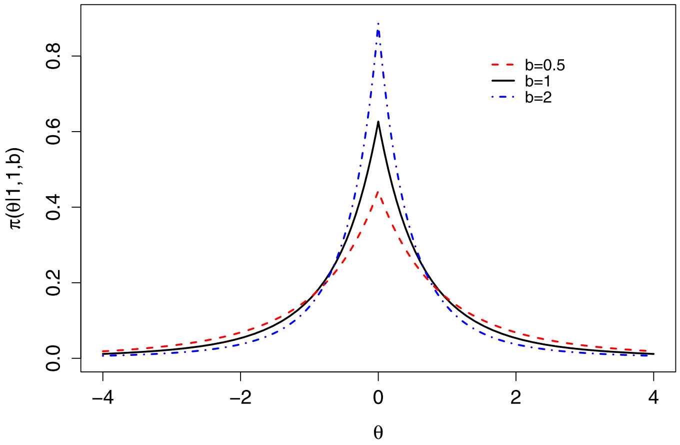

Figure 1: Normal-Scaled Beta2 density for different values of b $(p = q = 1)$.

It can be shown that tails of this distribution go to zero as $O(\theta^{-3})$. This implies that this distribution has a finite mean, but does not have a finite second moment. Its cumulative distribution function (CDF) can be calculated also in closed form,

$$\Pi(\theta|1,1,b) = 0.5 + \sqrt{\frac{\pi b}{2}}\theta e^{\frac{b\theta^2}{2}}(1-\Phi(\sqrt{b}\theta)), \theta > 0$$

For $\theta < 0$, symmetry can be used for finding the CDF.

3.2 Cauchy-Scaled Beta2 distribution: an explicit horseshoe distribution

Now, instead of a normal, let θ be a Cauchy random variable with location parameter 0 and scale parameter τ, and assume $\tau \sim \mathrm{SBeta2}(p,q,b)$. Then, the joint distribution of θ and τ is given by

$$\begin{aligned}
\pi(\theta,\tau) &= \frac{1}{\pi\tau\left(1+\left(\frac{\theta}{\tau}\right)^2\right)}\frac{1}{\mathrm{Beta}(p,q)}\frac{1}{b}\frac{\left(\frac{\tau}{b}\right)^{p-1}}{\left(\frac{\tau}{b}+1\right)^{p+q}} \\
&= \frac{(b)^q}{\pi\mathrm{Beta}(p,q)}\frac{\tau^p}{(\tau^2+\theta^2)(b+\tau)^{p+q}}
\end{aligned}$$

The marginal distribution of θ can be calculated as

$$\pi(\theta) = \int_0^\infty \pi(\theta,\tau)d\tau$$

Note that when p and q are integers, the integrand is a rational function. For example, if $p = q = 1$,

$$\begin{aligned}
\pi(\theta) &= \int_0^\infty \frac{b\tau}{\pi(b+\tau)^2\left(\theta^2+\tau^2\right)}d\tau \\
&= \frac{b}{\pi\left(b^2+\theta^2\right)^2}\left[-\left(b^2+\theta^2-\pi b|\theta|\right)+\left(b^2-\theta^2\right)\left(\log(b)-\log(|\theta|)\right)\right]
\end{aligned}$$

Note that this density function depends on θ only through $|\theta|$, and so it is clearly symmetric around 0.

In this case, the cumulative distribution function also has a closed form, given by

$$\Pi(\theta) = \frac{2\pi\theta^2 - \pi b^2 - \theta b \log\left[\left(\frac{\theta}{b}\right)^2\right]}{2\pi(\theta^2+b^2)},\ \theta > 0$$

For $\theta < 0$, symmetry can be used for finding the CDF.

The density functions of Cauchy-Scaled Beta2 variables for $p = q = 1$ and different values of b are shown in Figure 2. Larger values of b are associated to lower areas around the origin. This distribution has a pole at $\theta = 0$ and flat tails, which is an example of

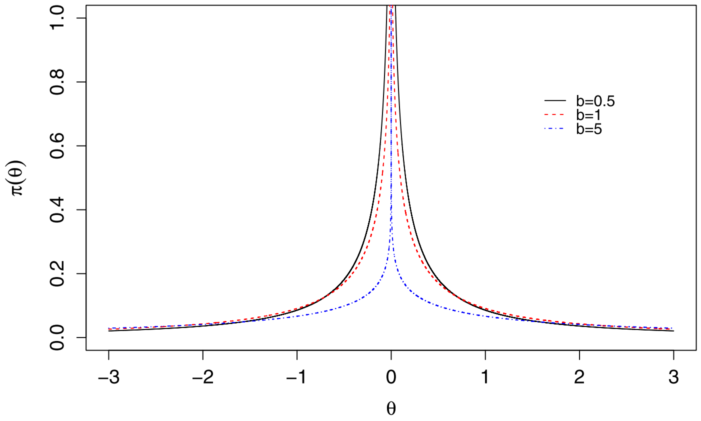

Figure 2: Cauchy-Scaled Beta2 distribution for different values of b ($p = q = 1$).

a horseshoe prior (Carvalho et al., 2010). To the best of our knowledge, this is the first horseshoe prior in explicit algebraic form.

Figure 3 compares densities for quartile matching ($|q_1| = q_3 = 1$) Normal, Cauchy, Normal-SBeta2($p = q = 1$) and Cauchy-Beta2 ($p = q = 1$) distributions. The heaviest tails correspond to the Cauchy-SBeta2(1,1,1), while the Normal-SBeta2 tails are lighter than those of the Cauchy.

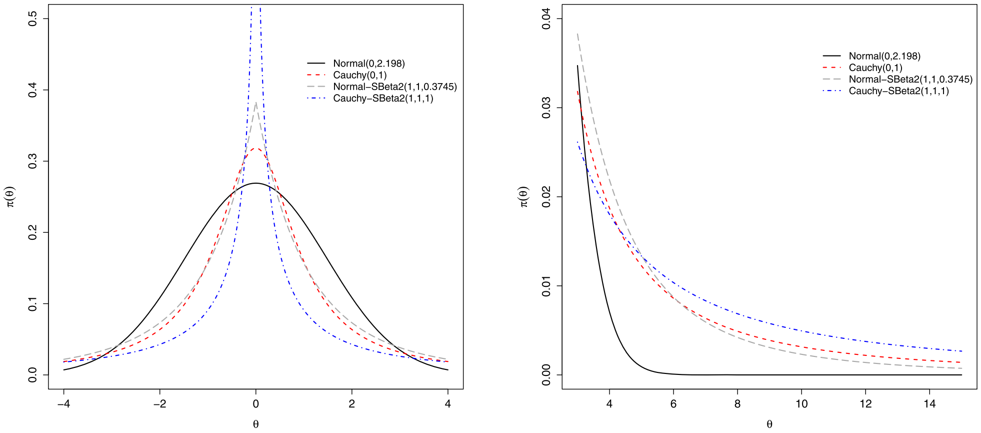

Figure 3: Comparison of quartile matching ($|q_1| = q_3 = 1$) Normal, Cauchy, Normal-SBeta2 ($p = q = 1$) and Cauchy-Beta2 ($p = q = 1$) distributions. The right plot shows the behavior of the tails.

Other choices of the hyperparameters may lead to a closed form marginal but not necessarily a horseshoe prior. For instance, for p = 2 and q = 1, we obtain

$$\begin{aligned}\pi(\theta) &= \int_0^\infty \frac{b}{\pi \text{Beta}(2,1)} \frac{\tau^2}{(\tau^2+\theta^2)(b+\tau)^3} d\tau \\ &= \frac{b\left((b^2-3\theta^2)(b^2+\theta^2-\pi b|\theta|)+(\theta^4-3b^2\theta^2)\log\left(\left(\frac{\theta}{b}\right)^2\right)\right)}{\pi(b^2+\theta^2)^3}.\end{aligned}$$

This density does not have a pole at $\theta = 0$. The corresponding CDF is

$$\Pi(\theta) = \frac{2\pi\theta^4 + 2b\theta^3 + \pi b^2\theta^2 + 2b^3\theta + \pi b^4 - b\theta^3 \log\left(\left(\frac{\theta}{b}\right)^4\right)}{2\pi\left(\theta^2+b^2\right)^2}, \ \theta > 0$$

Again, symmetry can be used for calculating the CDF value when $\theta < 0$.

3.3 Assigning a SBeta2 prior to the square of the scale: a general result

As before, suppose that Cauchy or, more generally, Student-t distributions are assumed for locations. What if instead of the scale, the square of the scale is assumed to be SBeta2? For example,

$$\pi(\theta|\mu,\tau,b) = \frac{1}{\pi\sqrt{b}\tau} \cdot [1 + \frac{(\theta-\mu)^2}{b\tau^2}]^{-1},$$

and

$$\tau^2 \sim \text{SBeta2}(\tau^2|1,1,1/b).$$

In this case, the marginal for θ is:

$$\pi(\theta) = \frac{1}{2\sqrt{b}\cdot(1+\frac{|\theta-\mu|}{\sqrt{b}})^2}$$

This is an interesting distribution, close to a Cauchy. It does not have a pole at zero, so it is not a horseshoe prior.

In Fúquene et al. (2014) this distribution, called Student-t-Beta2, is studied and applied in detail. There a general result for the marginal of the location for any p and q is obtained in terms of the Hypergeometric Function, as summarized in Supplementary Appendix 2.

4 Examples

Here we analyze three datasets found in literature and some simulated data. The first dataset is the "8-schools example" presented in Gelman (2006), where it is shown that the SBeta2 behaves sensibly in the sense that it does not promote very small variances and large shrinkages. In the second example, we use data from Normand and Shahian (2007) to illustrate that the SBeta2 promotes robustness in the hierarchical model. In the third example we revisit the famous baseball dataset in Efron and Morris (1972) and we robustly predict the batting averages of 18 baseball players, protecting the exceptional players from too much shrinkage to the mean (the so called "Clemente Problem", Efron, 2010) and at the same time improving the mean squared error (MSE). Finally, we use simulated data to compare the SBeta2 with the half-Cauchy proposed by Gelman for the schools example and find that the SBeta2$(1, 1, b)$ seems preferable in this settings.

4.1 A normal hierarchical model

In this section we will consider the normal hierarchical model described by Gelman (2006) in the so called "8-schools example". We wish to compare the changes in the posterior distribution of the precision when using either the Inverse-Gamma or the SBeta2 as prior distributions for the random effects variance.

Gelman works with a simple two-level normal model of data y_{ij} with group-level effects α_j:

$$\begin{aligned} y_{ij} &\sim N(\mu + \alpha_j, \sigma_y^2), \quad i = 1, \cdots, n_j, \quad j = 1, \cdots, J \\ \alpha_j &\sim N(0, \sigma_\alpha^2), \quad j = 1, \cdots, J \end{aligned} \tag{6}$$

where the parameters $\alpha_1, \cdots, \alpha_8$ represent the relative effects of Scholastic Aptitude Test coaching programs in eight different schools, and σ_α stands for the between-school standard deviations of these effects. The effects are measured in points within a range between 200 and 800. The approximate average and standard deviation are 500 and 100 respectively. The model has three hyperparameters μ, σ_y, and σ_α. Here we will only study the effect of the prior distribution for the variance of the random effects, σ_α^2.

Gelman proposes a half-Cauchy(25) as prior distribution for σ_α, which corresponds to a SBeta2$(0.5, 0.5, 625)$ prior distribution for σ_α^2. Therefore, for Bayesian estimation, the model is fitted with three different prior distributions for σ_α^2: SBeta2$(1, 1, 625)$, SBeta2$(0.5, 0.5, 625)$ and Inverse-Gamma$(0.001, 0.001)$.

Results in Figure 4 are based on 6000 iterations from a model fit using OpenBUGS (Thomas et al., 2006), and correspond to the posterior distribution for σ_α obtained with each of these three prior distributions .The left and middle histograms show the posterior distributions for σ_α using priors SBeta2(0.5,0.5,625) and SBeta2(1,1,625), respectively. We can observe that the range of values is mainly between 0 and 20 with a light tail after this last value.

The histogram on the right shows the posterior distribution for the same parameter using an Inverse-Gamma$(0.001, 0.001)$ prior distribution. We can see that the range of values for σ_α is concentrated in a short interval near 0 (0 to 5), and the posterior has a sharp peak near zero. This is the anomalous behavior highlighted by Gelman (2006) which is not present when the SBeta2 distributions are used as priors.

It can be seen that the SBeta2 distribution works properly when the analysis is performed for all 8 schools. Gelman (2006) comments that some problems could arise when the number of groups J is small because the data give little information about the variance between groups. In the analysis of the schools example Gelman only in-

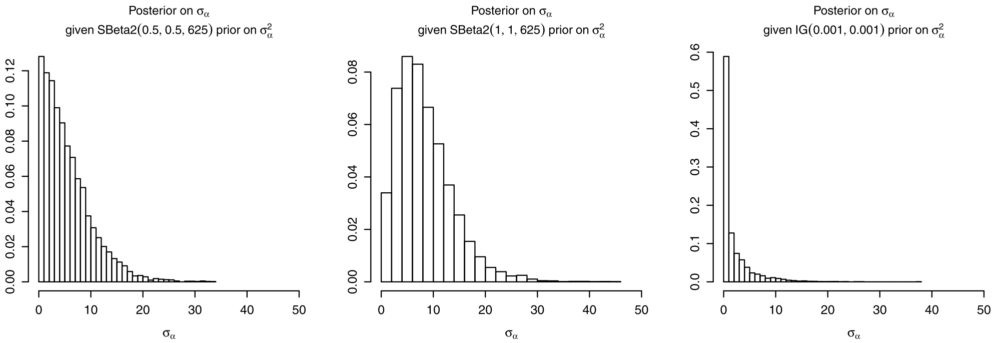

Figure 4: Histograms of posterior simulations of the between-school standard deviation, σ_α, from models with three different prior distribution: (i) $\text{SBeta2}(0.5, 0.5, 625)$, (ii) $\text{SBeta2}(1, 1, 625)$ and (ii) Inverse-Gamma(0.001,0.001).

cludes data for the first three schools, and uses the uniform and the half-Cauchy as prior distributions for σ_α. He concludes that the half-Cauchy gives good results in this example with plausible posterior values for $\sigma_\alpha < 50$. However, when a uniform prior distribution is used, the posterior distribution for σ_α presents an extremely long right tail with values for σ_α too high to be reasonable for this example, and therefore its use could result in an "under-shrinking" of the estimates for the effects α_j.

Figure 5 shows the histograms of the posterior distributions for σ_α with prior distributions SBeta2(0.5,0.5,625) and SBeta2(1,1,625) when only data from the first three schools are used. We see that they present a range of plausible values for σ_α between 0 and 50; after this last value, the posterior densities decrease rapidly. Like the half-Cauchy, the SBeta2 prior distribution with $p = q = 1$ has a good performance in this example because it shows plausible values for σ_α without the presence of a heavy right tail.

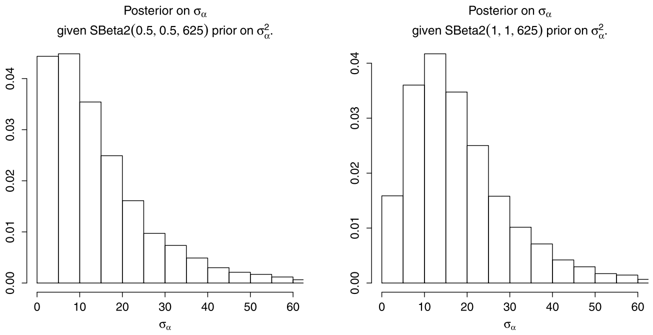

Figure 5: Histograms of posterior simulations of the between-school standard deviation, σ_α, from models with two different prior distribution: (i) SBeta2(0.5,0.5,625), (ii) SBeta2(1,1,50) and data for the first three schools.

In summary, both the half-Cauchy(25) (SBeta2(0.5,0.5,625)) and the SBeta2(1,1,625) produce sensible results for $J = 3$ schools and for $J = 8$ schools.

4.2 A tale of unwanted attraction, or how "naïve" hierarchical Bayes pulls up perfect hospitals

In an application of Bayesian hierarchical models to hospital profiling, Normand and Shahian (2007) analyze data for 30-day mortality following isolated coronary artery bypass grafting (CABG) surgery in 13 non-governmental hospitals in Massachusetts, USA. In their work, it can be seen that, unfortunately, random effects of hospitals with no deaths are subject to large shrinkage with non-robust hierarchical models. This is an instance of too much shrinkage: a perfect hospital (no deaths) is pulled strongly towards the general mean regardless of its exceptional quality. The assumption of a vague Inverse-Gamma produces a hierarchical model that does not predict outliers, very bad or very good hospitals, and thus implies large corrections for outlying values. In that sense the selected model is non-robust. Then we should change the assumptions if a robust behavior is desired. Notice also that a robust model is fairer with exceptional individuals: the amount of shrinkage is not constant, but depends on performance.

We revisit the data shown in Table 1 without using explanatory data (which is not available). Additionally, we enlarge the sample size of one of the hospitals with no deaths to the average size of all hospitals, as an alternative scenario to explore the differences between different models.

We will focus on the probability of death for each hospital, θ_i, and its corresponding log-odds, β_i. We also calculate the predictive probability of 0 deaths for the following 100 patients for each hospital.

Hospital	Patients	Deaths
1	508	11
2	454	11
3	381	15
4	623	11
5	26 (350)	0
6	393	7
7	718	18
8	149	1
9	80	0
10	296	5
11	191	3
12	365	4
13	419	15

Table 1: 30-day mortality in 13 non-governmental hospitals following isolated CABG surgery, Massachusetts, USA (Normand and Shahian, 2007). Hospital 5 was changed from 26 patients to the approximate mean number of patients 350.

We compare three models:

- **Model 1**: A non-robust Normal–Inverse-Gamma model

  $$\begin{aligned} y_i &\sim \text{Bin}(n_i, \theta_i) \\ \beta_i &\sim N(\mu, \sigma^2) \\ \mu &\sim N(0, 10^3) \\ \sigma^2 &\sim \text{Inverse-Gamma}(0.001, 0.001) \end{aligned}$$

  where $\beta_i = \log(\theta_i/(1-\theta_i))$.

- **Model 2**: We substitute the Inverse-Gamma distribution by a SBeta2$(1, 1, 1)$.

  $$\sigma^2 \sim \text{SBeta2}(1, 1, 1)$$

- **Model 3:** A Student-t distribution with 2 degrees of freedom is used as a prior for the log-odds instead of a Normal.

  $$\begin{aligned} y_i &\sim \text{Bin}(n_i, \theta_i) \\ \beta_i &\sim t_2(\mu, \sigma^2) \\ \mu &\sim N(0, 10^3) \\ \sigma^2 &\sim \text{SBeta2}(1, 1, 1) \end{aligned}$$

In Models 2 and 3, we choose a conventional SBeta2$(1, 1, 1)$ as a plausible nearly objective model, since it is symmetric in the information of the scale and its reciprocal. We may add that this choice also makes sense from an Empirical Bayes approach, since the observed variance of the log-odds (from the modified hospital data) is around 0.8, close to the assessed median of 1.

Figure 6 shows 95% posterior probability intervals for the log-odds and the probabilities of death for each hospital, and the posterior probabilities of 0 deaths for 100 patients for each hospital are shown in Figure 7. For the fully robust Model 3, the inference for the "perfect" hospital 5 is the most reasonable, followed by Model 2. The assumption of a flat tail location, as the Cauchy (which is widely accepted as a more robust model in a setting like this) is made even more robust by the assumption of a Scaled Beta2 for the scale.

4.3 The Clemente problem

Efron and Morris (1972) obtained a sample of batting averages for 18 baseball players for the 1970 season. They used the average obtained during the first 45 at bats for predicting the batting average of each player for the rest of the season. The initial assumption about the data is

$$Y_i \sim \frac{1}{45}\text{Bin}(45, p_i)$$

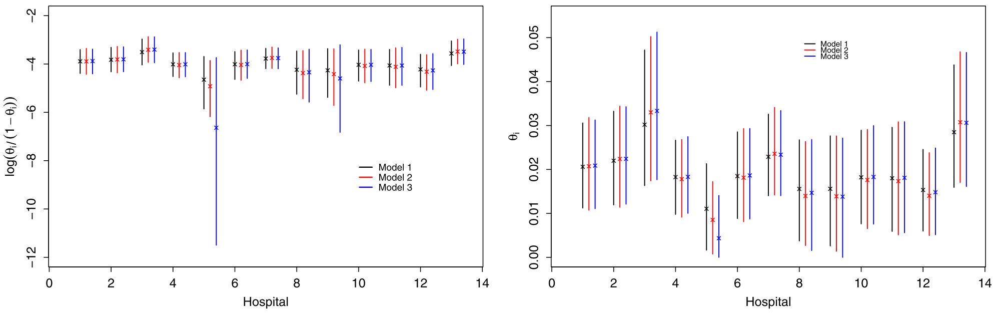

Figure 6: 95% posterior probability intervals for the log-odds and the probabilities of death for each hospital under the three models. Models 2 and 3 reduce the shrinkage of the "perfect" hospital 5 (with no observed deaths) towards the mean.

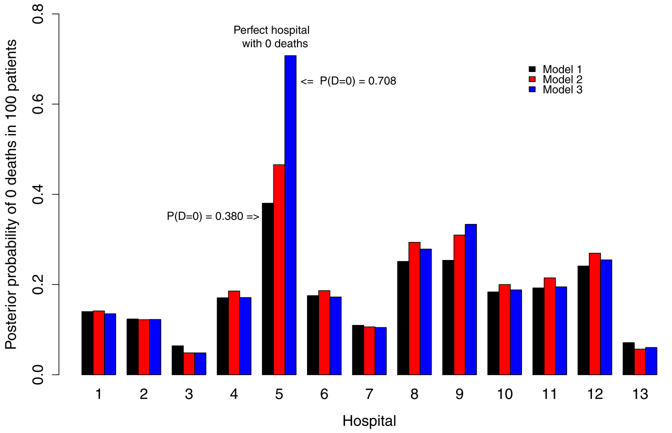

Figure 7: Posterior probability of no deaths for 100 new patients. Note the differences in the probabilities estimated for each model for hospital 5, in which no deaths were observed.

where Y_i is the batting average for the first 45 at bats, and p_i depends on each player's ability.

The batting average for the rest of the season, R_i, can be modeled as

$$R_i \sim \frac{1}{n_i}\text{Bin}(n_i, p_i)$$

where n_i is the number of at bats for player i during the remainder of the season.

Efron and Morris applied a variance stabilizing transformation to Y_i,

$$X_i = \sqrt{45}\arcsin(2Y_i - 1)$$

so that $X_i \sim N(\mu_i, 1)$, with μ_i approximately equal to the transformed value of p_i. In the sequel, we will use this transformed variable.

In his talk at the '09 Objective Bayes Conference June 2009, *"The Future of Indirect Evidence"* Professor Bradley Efron exposed a fundamental problem: *"The Clemente Problem: How to protect atypical cases from too much indirect evidence?"*. Professor Efron is referring to the Puerto Rican sportsman Roberto Clemente, an outstanding batter and human being, who had the highest batting average of the list of 18 players. After the first 45 turns, Clemente had a score of 400 (40% of hits). Even though shrinking to a general mean improves the overall prediction of the 18 batters, for Clemente under the conjugate prior a batting average of 282 is predicted (see Table 2). The atypical Clemente was not protected from "too much of a good thing", and his personal prediction was very poor: he finished with a score of 346, much higher than predicted. The problem lies in the fact that the usual method shrinks equally all players. This problem extends beyond location parameters: formula (2.10) in Diaconis and Ylvisaker (1979) shows that if the prior for an Exponential Family parameter is chosen in the usual conjugate family its posterior mean is a linear combination of prior expectation and arithmetic mean. This implies a serious lack of robustness, since the shrinkage rate is constant regardless the conflict between prior expectations and sample means.

As an illustration assume that h is the precision parameter of Normal data with known mean. Applying (2.3) in Diaconis and Ylvisaker (1979), the conjugate prior of the precision is a Gamma distribution with prior location W_0 and "prior sample size" n_0. It turns out that the posterior mean of the mean parameter W can be written as:

$$E(E(W)|\mathbf{Data}) = \bar{W} - \left(\frac{n_0}{n_0+n}\right)(\bar{W} - W_0).$$

It is clear then that the rate of shrinkage $\frac{n_0}{n_0+n}$ is constant, regardless the conflict $\bar{W}-W_0$. So any Exponential Family parameter shares the same behavior when conjugate priors are employed, regardless if the parameter is location, scale, etc. The goal, then, is a model that shrinks less the exceptional, without inflating the mean square error of prediction.

We fit the model

$$\begin{aligned} X_i &\sim N(\mu_i, 1),\ \ i = 1, \cdots, 18 \\ \mu_i &\sim \text{Cauchy}(M, \sigma) \\ M &\sim \text{Cauchy}(0, 10^3),\ \ \sigma^2 \sim \text{SBeta2}(p, q, b) \end{aligned}$$

for different hyperparameters (p, q, b). We used one of the assessment strategies discussed in Section 2, selecting b such that $P(\sigma^2 > 1.5) = 0.1$ (the value 1.5 was chosen using the empirical variance for the transformed data, $s^2 = 1.116$). Under this condition, the values $b = 0.17$ for $p = q = 1$ and $b = 0.038$ for $p = q = 1/2$ were elicited. The former prior leads to a reduction of MSE of 8.4% over the conjugate, while the SBeta2(1/2,1/2,0.038) prior leads to a lesser reduction of 5%, so here again the SBeta2(1,1,b) performed slightly superior than the half-Cauchy (though it can be argued that in both cases the shrinkage of Clemente is still excessive). However, assigning higher values to b, as in a SBeta2(1,1,1), arguably leads to a satisfactory reduction in mean square error of almost 7% simultaneously with less shrinkage of the extreme values, Clemente and Alvis (see Table 2). It should be noted that the SBeta2 is somewhat sensitive to the assessed values of the scale b. Also note that the improvement in the overall mean squared error does not imply that all the individual predictions are better.

		Predicted batting probabilities			
Player	Observed season	Conjugate model	Normal likelihood. Structural Cauchy. Cauchy prior for common location. SBeta2(1,1,0.17) prior for squared scale parameter	Normal likelihood. Structural Cauchy. Cauchy prior for common location. SBeta2(0.5,0.5,0.038) prior for squared scale parameter	Normal likelihood. Structural Cauchy. Cauchy prior for common location. SBeta2(1,1,1) prior for squared scale parameter
Clemente	0.346	0.2825	0.2974	0.2838	0.3066
Robinson	0.298	0.2797	0.2875	0.2775	0.2958
Howard	0.276	0.2765	0.2803	0.2731	0.2861
Johnstone	0.222	0.2733	0.2745	0.2697	0.2791
Berry	0.273	0.2701	0.2704	0.2675	0.2728
Spencer	0.270	0.2699	0.2701	0.2670	0.2733
Kessinger	0.263	0.2668	0.2665	0.2650	0.2683
Alvarado	0.210	0.2645	0.2641	0.2639	0.2643
Santo	0.269	0.2603	0.2599	0.2612	0.2587
Swoboda	0.230	0.2606	0.2602	0.2616	0.2593
Unser	0.264	0.2568	0.2567	0.2595	0.2546
Williams	0.256	0.2573	0.2570	0.2593	0.2548
Scott	0.303	0.2568	0.2566	0.2590	0.2541
Petrocelli	0.264	0.2566	0.2562	0.2590	0.2539
Rodriguez	0.226	0.2572	0.2569	0.2595	0.2545
Campaneris	0.285	0.2532	0.2520	0.2567	0.2488
Munson	0.316	0.2502	0.2465	0.2535	0.2402
Alvis	0.200	0.2476	0.2408	0.2483	0.2321
MSE(x10^3)		1.2126	1.1108	1.1519	1.1307

Table 2: Predicted values and mean squared error of prediction (MSE) for the batting averages data in Efron and Morris (1972) using a conjugate model and three robust models with SBeta2 priors for the squared scale parameter.

4.4 Simulation study

A simulation study was performed in order to analyze the performance of the SBeta2 as a prior distribution for scales in hierarchical models in different scenarios. Back in the setting of the 8-schools example, we generated data according to the model in equation (6). Values for the effects α_j were fixed according with three scenarios: all similar in magnitude, a few medium outliers and one large outlier. We simulated data for $J = 3, 4, 5, 6, 7, 8, 9$ and 10. The value for σ_y^2 is known and we want to determine the effect of the prior distribution of σ_α^2 on the estimation error of the effects α_j. For each fixed value of α_j and σ_y^2 we generated 1000 samples from a normal distribution for the error terms.

We fitted the model using five different prior distributions for σ_α^2: SBeta2(1,1,625), SBeta2(0.5,0.5,625), Inverse-Gamma(0.001,0.001), SBeta2(1,1,b) and SBeta2(0.5,0.5,b), where values for b were assigned such that p(σ_α^2 >Var(α_j)) = 0.5 (and therefore b =Var(α_j)). We carried out 10000 MCMC simulations with a burn in of 2000 for each case.

In order to study the estimation error we computed the global estimation error

$$G = \sqrt{\frac{\sum_{j=1}^{J}\sum_{k=1}^{1000}(\alpha_j - \hat{\alpha}_{kj})^2}{1000}},$$

where $\hat{\alpha}_{kj}$ is the posterior mean of α_j calculated for simulated dataset k. Table 3 shows the results for the first scenario, where all α_j's are similar in magnitude. The values for these effects were set within a range of ±0.5 with $\text{Var}(\alpha_i) = 0.093$, 0.067, 0.150, 0.135, 0.130, 0.156, 0.138, 0.178 for J from 3 to 10, respectively. We observe that for each J the global estimation error is smaller when the model is fitted using SBeta2(1,1,Var(α_j)) as prior distribution for σ^2_α, followed by the fitting with SBeta2(0.5,0.5,Var(α_j)). The distributions that exhibit the largest errors are the SBeta2 with $b = 625$; this seems reasonable since we are assigning a prior distribution with big scale parameter in a situation where the effects are similar in magnitude, and therefore their variance is small.

	Prior distribution on σ^2_α				
J	Inverse-Gamma (0.001,0.001)	SBeta2 (0.5,0.5,625)	SBeta2 (1,1,625)	SBeta2 (0.5,0.5,Var(α_j))	SBeta2 (1,1,Var(α_j))
3	0.6521	1.1304	1.4502	0.5060	0.4574
4	0.6947	1.2017	1.6316	0.5266	0.4741
5	0.8856	1.2833	1.6934	0.8156	0.7855
6	0.9519	1.3288	1.7373	0.8818	0.8514
7	2.8659	3.4144	3.8001	0.9765	0.9453
8	3.4815	3.9480	4.2805	1.0964	1.0640
9	3.9462	4.4790	4.8496	1.1066	1.0797
10	3.1022	3.6381	4.0216	1.1333	1.1025

Table 3: Simulation study: Global estimation error G for the scenario in which all α_j have similar magnitude.

Table 4 shows the results for the second scenario: a few medium outliers. The majority of values for the effects were set within the range ±0.5, and a few in the range ±2 and ±3. One medium outlier was assigned for $J = 3, 4, 5$, two medium outliers for $J = 6, 7$ and three medium outliers for $J \geq 8$, with $\text{Var}(\alpha_j) = 1.803$, 1.200, 1.332, 2.647, 2.246, 2.876, 2.520, 2.299 for J from 3 to 10, respectively. The global estimation error is smaller using SBeta2(1,1,Var(α_j)) prior except in the case $J = 10$. For $J \geq 6$ (with more than one medium outlier) using the Inverse-Gamma prior leads to larger global estimation errors.

Table 5 shows the results for the third scenario: one large outlier. Again the values for all effects were set within a range ±0.5 except one, which was assigned an absolute value greater than 5. The variance for the random effects is $\text{Var}(\alpha_i) = 9.403$, 6.650, 6.400, 5.180, 3.878, 4.039, 3.436, 3.290 for J from 3 to 10, respectively. The largest global error corresponds to the Inverse-Gamma(0.001, 0.001). The values fitted with SBeta2 priors exhibit similar global errors. However, when $b = 625$ the individual estimation errors for those α's within the range ±0.5 are large compared to the individual estimation

	Prior distribution on σ^2_α				
J	Inverse-Gamma	SBeta2	SBeta2	SBeta2	SBeta2
	(0.001,0.001)	(0.5,0.5,625)	(1,1,625)	(0.5,0.5,Var(α_j))	(1,1,Var(α_j))
3	1.9008	1.8538	1.9158	1.8347	1.7986
4	1.8966	1.9197	2.0588	1.8489	1.8202
5	2.1835	2.1581	2.2710	2.1134	2.0755
6	2.5672	2.3760	2.3671	2.4033	2.3188
7	2.6535	2.4811	2.4775	2.4979	2.4238
8	3.1622	2.9995	2.9761	3.0213	2.9523
9	3.3462	3.1590	3.1204	3.1930	3.1186
10	3.4352	3.2333	3.1895	3.2758	3.1949

Table 4: Simulation study: Global estimation error G for the scenario in which there are few medium outliers.

error for the α with absolute value greater than 5. For instance, for $J = 3$ we fixed $\alpha_j = -0.3, 0.1, 5.2$. When the model was fitted with SBeta2(1,1,625), the individual estimation errors were 3.3086, 3.1203, 4.5850 for $j = 1, 2$ and 3 respectively, whereas with the SBeta2(1,1,Var(α_j)) these values were 2.4730, 2.3937, 6.5967. Therefore the simulation study shows smaller individual estimation errors for the outliers when b is large.

	Prior distribution on σ^2_α				
J	Inverse-Gamma	SBeta2	SBeta2	SBeta2	SBeta2
	(0.001,0.001)	(0.5,0.5,625)	(1,1,625)	(0.5,0.5,Var(α_j))	(1,1,Var(α_j))
3	3.6400	3.3783	3.3187	3.4457	3.3858
4	3.5496	3.3040	3.2371	3.3694	3.2948
5	3.3458	3.2175	3.1934	3.2409	3.1962
6	3.5963	3.4181	3.3693	3.4566	3.3953
7	3.4577	3.3549	3.3463	3.3593	3.3154
8	3.8349	3.6049	3.5295	3.6620	3.5857
9	3.6869	3.5509	3.5244	3.5707	3.5177
10	3.9907	3.7542	3.6676	3.8117	3.7275

Table 5: Simulation study: Global estimation error G for the scenario with one large outlier.

We calculated 95% highest posterior density intervals for the effects in each of the simulations corresponding to the three scenarios. With these intervals we calculated the coverage rate. When all effects are similar in magnitude the coverage rates are almost equal regardless the prior employed.

For the second scenario (a few medium outliers) the coverage rates for medium outliers change with the prior, and they are smallest when the Inverse-Gamma(0.001,0.001) is used. For example, Table 6 shows the coverage rates when $J = 8$, where $\alpha_j = -0.3, 0.1, 0.3, -0.2, 0.5, 2.5, -2.6, 2.8$. It can be seen that when the effects are similar in magnitude the coverage rates are almost equal but for α_6, α_7 and α_8 the coverage rates are smaller. The biggest coverage rate for these three medium outliers is obtained with the SBeta2(1,1,625) prior.

	Prior distribution on σ^2_α				
j	Inverse-Gamma	SBeta2	SBeta2	SBeta2	SBeta2
	(0.001,0.001)	(0.5,0.5,625)	(1,1,625)	(0.5,0.5,V(α))	(1,1,V(α))
1	99.9	99.9	99.9	99.9	99.7
2	100	100	100	99.9	99.8
3	99.8	99.9	99.9	99.8	99.7
4	99.9	99.9	99.9	100	100
5	99.9	99.9	100	99.2	99.7
6	78.5	91.3	95.9	82.0	86.0
7	84.8	94.5	98.4	89.9	92.9
8	77.2	88.3	94.8	80.6	82.9

Table 6: Simulation study: Coverage rates for $J = 8$ in the scenario with a few medium outliers.

Consider the case when $J = 3$ and the scenario is one large outlier. In this situation we selected α_j = -0.3, 0.1, and 5.2, as commented before. Table 7 shows that the coverage rate for the large outlier obtained with any of the SBeta2 distributions as prior on σ^2_α is greater than the one obtained using the Inverse-Gamma(0.001,0.001). Similar results are obtained for other values of J: the coverage rates are almost equal when the effects are similar in magnitude and smaller when the model is fitted using the Inverse-Gamma(0.001,0.001) prior. Even though the main intention in this subsection is to compare SBeta2 with different parameters and Inverse-Gamma, it is to be expected that assuming t-priors for the random effects would improve the robustness of the methodology, as in the previous subsections.

		Prior distribution on σ^2_α			
j	Inverse-Gamma	SBeta2	SBeta2	SBeta2	SBeta2
(0.001,0.001)	(0.5,0.5,625)	(1,1,625)	(0.5,0.5,V(α))	(1,1,V(α))	
1	100	100	100	100	100
2	100	100	100	100	100
3	87.0	99.4	100	92.5	94.9

Table 7: Simulation study: Coverage rate for $J = 3$ in the scenario with one large outlier.

5 Final remarks

In this article we put forward the idea of the Scaled Beta2 as a standard family for prior distributions of scales for both hierarchical and non-hierarchical models. The Scaled Beta2 distribution is naturally motivated, flexible and amazingly tractable. For specific values of hyperparameters, it leads to closed form results, and generalizes previous proposals in the literature like a half-Cauchy for standard deviations. For ranges of hyperparameters, both flat tails and low probabilities of small scale values can simultaneously be achieved. In this manner, undesirable features like excessive shrinkage and very low scale values can be avoided.

Supplementary Material

Supplementary Material of "The Scaled Beta2 Distribution as a Robust Prior for Scales" (DOI: 10.1214/16-BA1015SUPP; .pdf).

References

Andrade, J. A. A. and O'Hagan, A. (2006). "Bayesian robustness modelling using regularly varying distributions." *Bayesian Analysis*, 1: 169–188. MR2227369. doi: `http://dx.doi.org/10.1214/06-BA106`. 619

Berger, J. (2006). "The case for objective Bayesian analysis." *Bayesian Analysis*, 1(3): 385–402. MR2221271. 616

Bradlow, E., Hardie, B., and Fader, P. (2002). "Closed-Form Bayesian Inference for the Negative Binomial Distribution." *Journal of Computational and Graphical Statistics*, 11: 189–202. MR1937285. doi: `http://dx.doi.org/10.1198/106186002317375677`. 616

Carvalho, C., Polson, N., and Scott, J. (2010). "The horseshoe estimator for sparse signals." *Biometrika*, 87(2): 465–480. MR2650751. doi: `http://dx.doi.org/10.1093/biomet/asq017`. 617, 623

De Finetti, B. (1961). "The Bayesian Approach to the Rejection of Outliers." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 199–210. Berkeley, Calif.: University of California Press. `http://projecteuclid.org/euclid.bsmsp/1200512167`. MR0133935. 618

Diaconis, P. and Ylvisaker, D. (1979). "Conjugate priors for exponential families." *The Annals of Statistics*, 7: 269–281. MR0520238. 630

Efron, B. (2010). "The future of indirect evidence." *Statistical Science*, 25(2): 145–157. MR2789983. doi: `http://dx.doi.org/10.1214/09-STS308`. 625

Efron, B. and Morris, C. (1972). "Limiting the risk of Bayes and empirical Bayes estimators-part II: The empirical Bayes case." *Journal of the American Statistical Association*, 67: 130–139. MR0323015. 624, 628, 631

Frühwirth-Schnatter, S. and Wagner, H. (2010). "Bayesian Variable Selection for Random Intercept Modeling of Gaussian and non-Gaussian Data." In *Bayesian Statistics 9*. J. Bernardo and M. Bayarri and J. O. Berger and A. P. Dawid and D. Heckerman and A. F. M. Smith and M West (eds.). MR3204006. doi: `http://dx.doi.org/10.1093/acprof:oso/9780199694587.003.0006`. 616

Fúquene, J., Pérez, M., and Pericchi, L. (2014). "An alternative to the Inverted Gamma for the variances to modelling outliers and structural breaks in dynamic models." *Brazilian Journal of Probability and Statistics*, 28(2): 288–299. MR3189499. doi: http://dx.doi.org/10.1214/12-BJPS207. 617, 624

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1(3): 515–533. MR2221284. 616, 624, 625

Giron, J., Martinez, L., Moreno, E., and Torres, F. (2006). "Objective testing procedures in linear models: Calibration of the p-values." *Scandinavian Journal of Statistics*, 33(4): 765–784. MR2300915. doi: http://dx.doi.org/10.1111/j.1467-9469.2006.00514.x. 616

Griffin, J. E. and Brown, P. J. (2010). "Inference with normal-gamma prior distributions in regression problems." *Bayesian Analysis*, 5(1): 171–188. MR2596440. doi: http://dx.doi.org/10.1214/10-BA507. 616

Liang, F., Paulo, R., Molina, G., Clyde, M., and Berger, J. (2008). "Mixtures of g priors for Bayesian variable selection." *Journal of the American Statistical Association*, 103(481): 410–423. MR2420243. doi: http://dx.doi.org/10.1198/016214507000001337. 616

Maruyama, Y. and George, E. (2011). "Fully Bayes factors with a generalized g-prior." *The Annals of Statistics*, 39: 2740–2765. MR2906885. doi: http://dx.doi.org/10.1214/11-AOS917. 616

Normand, S. T. and Shahian, D. M. (2007). "Statistical and clinical aspects of hospital outcomes profiling." *Statistical Science*, 22(2): 206–226. MR2408959. doi: http://dx.doi.org/10.1214/088342307000000096. 624, 627

Pérez, M., Pericchi, L., and Ramírez, I. (2016). "Supplementary Material of "The Scaled Beta2 Distribution as a Robust Prior for Scales"." *Bayesian Analysis*. doi: http://dx.doi.org/10.1214/16-BA1015SUPP. 616

Pericchi, L. (2005). "Model selection and hypothesis testing based on objective probabilities and Bayes factors." In *Handbook of Statistics, 25*, 115–149. MR2490524. doi: http://dx.doi.org/10.1016/S0169-7161(05)25004-6. 616

Pericchi, L. (2010). "Discussion of Polson, N, and Scott, J." In *Bayesian Statistics 9*, 531. Oxford University Press. 616

Pericchi, L., Figueroa, N., Perez, J., and Torres, D. (2011). "Age-Period-Cohort Robust Bayesian Models for Projecting Cancer Incidence and Mortality in Puerto Rico." *Working Paper*, Center for Biostatistics and Bioinformatics, UPR-RRP. http://hdl.handle.net/10586/145 620

Polson, N. and Scott, J. (2012). "On the half-Cauchy prior for a global scale parameter." *Bayesian Analysis*, 4: 771–1052. MR3000018. doi: http://dx.doi.org/10.1214/12-BA730. 616

Scott, J. G. and Berger, J. O. (2006). "An exploration of aspects of Bayesian multiple testing." *Journal of Statistical Planning and Inference*, 136: 2144–2162. MR2235051. doi: http://dx.doi.org/10.1016/j.jspi.2005.08.031. 616

Sparks, D., Khare, K., Ghosh, M., and Xiang, R. (2013). "Necessary and sufficient conditions for posterior consistency under g prior." OBayes 2013, Duke University, Raleigh, USA. http://bayesian.org/sites/default/files/MGhosh.pdf. 616

Thomas, A., Hara, B. O., Ligges, U., and Sturtz, S. (2006). "Making BUGS open." *R News*, 6: 12–17. 625

Wang, M. and Sun, X. (2013). "Bayes factor consistency for one-way random effects model." *Statistics: A Journal of Theoretical and Applied Statistics*, 47(5): 1104–1115. MR3175737. doi: http://dx.doi.org/10.1080/02331888.2012.694445. 616

Wolfram Alpha LLC. (2014). "Wolfram—Alpha." Access August 20, 2014. http://www.wolframalpha.com. 620

Acknowledgments

M. E. Pérez and L. R. Pericchi's research funded by NIH/NCRR project No. 5P20RRQ16470-09 and NIH/NCI project No. 3U54CA096297-11S1 UPR/MDACC Partnership for Excellence in Cancer Research 2013–2018. M. E. Pérez research was also partially funded by NSF/CREST project No. 0206200. L.R. Pericchi benefited from discussions with J. Mulder about the subject of this paper. The authors are grateful to the anonymous associated editor and referees for their thoughtful and detailed comments, which led to a substantial improvement of the manuscript.