LIMITING THE SHRINKAGE FOR THE EXCEPTIONAL BY OBJECTIVE ROBUST BAYESIAN ANALYSIS *

By Luis R. Pericchi† and Maria-Eglée Pérez†

University of Puerto Rico, Rio Piedras Campus

Modern Statistics is made of the sensible combination of direct evidence (the data directly relevant or the "individual data") and indirect evidence (the data and knowledge indirectly relevant or the "group data"). The admissible procedures are a combination of the two sources of information, and the advance of technology is making indirect evidence more substantial and ubiquitous. It has been pointed out however, that in "borrowing strength" an important problem of Statistics is to treat in a fundamentally different way exceptional cases, cases that do not adapt to the central "aurea mediocritas". This is what has been recently coined as "the Clemente problem" (Efron, 2009). In this article we put forward that the problem is caused by the simultaneous use of square loss function and conjugate (light tailed) priors which is the usual procedure. We propose in their place to use robust penalties, in the form of losses that penalize more severely huge errors, or (equivalently) priors of heavy tails which make more probable the exceptional. Using heavy tailed prior we can reproduce in a Bayesian way, Efron and Morris' "limited translated estimators" (with Double Exponential Priors) and "discarding priors estimators" (with Cauchy-like priors) which discard the prior in the presence of conflict. Both Empirical Bayes and Full Bayes approaches are able to alleviate the Clemente Problem and furthermore beat the James-Stein estimator in terms of smaller square errors, for sensible Robust Bayes priors. We model in parallel Empirical Bayes and Fully Bayesian hierarchical models, illustrating that the differences among sensible versions of both are minute, as compared with the effect due to the robust assumptions. We propose a heavy tailed Beta2 distribution for variances that arises naturally as an alternative to the usual Inverted-Gamma distribution. The combination of a Cauchy Prior for location and Beta2 for scale, yields a novel closed form prior for location that we call Beta2-Cauchy, extremely suitable for Objective Robust Bayesian Analysis (ORBA).

1. Shrinkage estimators that borrow strength from indirect Evidence: Too much of a good thing?.

*To Roberto Clemente Walker, first latin-american to reach the Baseball Hall of Fame.

†This work was sponsored in part by NIH Grant: P20-RR016470. M.E. Pérez research was also sponsored by NSF Grant: HRD 0734826.

1.1. *An Historical Example.* Efron and Morris (1975) obtained a sample of batting averages for 18 baseball player during the 1970 season. They used the average obtained during the first 45 at-bats for predicting the batting average for the rest of the season for each player.

Player	Batting average for first 45 at bats	Batting average for remainder of season	At bats for remainder of season
Clemente (Pitts, NL)	0.400	0.346	367
F. Robinson (Balt, AL)	0.378	0.298	426
F. Howard (Wash,AL)	0.356	0.276	521
Johnstone (Cal, AL)	0.333	0.222	275
Berry (Chi, AL)	0.311	0.273	418
Spencer (Cal, AL)	0.311	0.270	466
Kessinger (Chi, NL)	0.289	0.263	586
Alvarado (Bos, AL)	0.267	0.210	138
Santo (Chi, NL)	0.244	0.269	510
Swoboda (NY, NL)	0.244	0.230	200
Unser (Wash, AL)	0.222	0.264	277
Williams (Chi, AL)	0.222	0.256	270
Scott (Bos, AL)	0.222	0.303	435
Petrocelli (Bos, AL)	0.222	0.264	538
E. Rodriguez (KC, AL)	0.222	0.226	186
Campaneris (Oak, AL)	0.200	0.285	558
Munson (NY, AL)	0.178	0.316	408
Alvis (Mil, NL)	0.156	0.200	70

TABLE 1
Original data: 1970 *batting averages for* 18 *MBL players*

In Table 1, the relevant data are presented. Direct evidence is the observed individual average, thus the temptation to predict by the observed individual average, although it is known that this estimator is inadmissable. This is a very bad practical predictor in this case, and this has been corroborated in other cases, see Brown (2008). On the other end, the indirect evidence comes in the form of the overall mean $M = 0.265$ of all batters.
This "pure indirect evidence" estimator is surprisingly good in this case, and far better than the "pure direct evidence", MLE or naive estimator as Brown calls it. However, both intuition and theory point to a sensible combination of the two sources of evidence to improve overall predictions. The problem is then: How much to weight direct and indirect evidence in each individual case? Wouldn't it be reasonable to weight less the common indirect evidence, when there is reason to believe that the individual is exceptional? This was

the original motivation of Efron and Morris for their clever (and "*ad-hoc*") "limited translation estimators" (Efron and Morris , 1972).

1.2. *Efron and Morris set up.* The initial assumption about the data in Efron and Morris (1972) is:

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

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

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

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

They applied a variance stabilizing transformation to Y_i,

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

In the sequel, we will use this transformed variable.

The analysis of these baseball data by Efron and Morris (1972) was widely cited and remains one of the clearest expositions in favor of combining *"indirect evidence"*, with *"direct evidence"*, a practice often termed "borrowing strength" and "shrinkage estimation".

1.3. *The Clemente Problem.* In his conference in 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 0.400, or 40% of hits. Even though shrinking to a general mean improves the overall prediction of the 18 batters, for Clemente his score was predicted as 0.290. 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 0.346, much higher than the predicted. The problem lies in the fact that the usual method shrinks a fixed proportion to all players, see equation (4) below. It does not make any exception, for the too good or too bad players. This is a logical consequence of the assumptions made, since it corresponds to an optimal decision in Decision Theory. So, in *"What If?"* mode of thought if the logical consequences are not *"pleasant to the mind"*, *a fortiori* assumptions have to be changed. Fixed proportion estimators are not robust in the sense that the amount of shrinkage is not limited, that is the potential influence of indirect evidence is unbounded, or using a metaphoric expression the procedure is "*myopic*" to the conflict between the bulk of the data and the individual.

1.4. *Robust Penalties.* Our starting point is re-analyzing the Loss (or minus Utility) function. In decision analysis the Square Loss is by far the most used (or over-used?) and the Clemente problem is (in part) a coherent consequence of its assumption. To see this, we recall the following result in Decision Theory.

RESULT 1. Suppose a function of the parameter θ, $g(\theta)$, is being estimated by $\delta(X_1, \ldots, X_m)$. Assume the weighted square loss function:

$$L(g(\theta), \delta) = \sum_{i=1}^{m} w(\theta_i) \cdot (\delta_i(X_1, \ldots, X_m) - g(\theta_i))^2).$$

Then the optimal Bayes estimator is:

$$\delta(g(\theta_i)) = \frac{E[w(\theta_i) \cdot g(\theta_i)|\text{data}]}{E[w(\theta_i)|\text{data}]}. \qquad (1)$$

Proof: Ferguson (1967, p. 47).

Although this result may be termed as classical, its statistical consequences have not being fully appreciated. In fact (1) invites two strategies: the first is to weight the square loss and use (1) as the individual estimator and the second is to change the prior in a way suggested by (1), and keep the square loss function. These are two different view points that shed different lights and possibilities. In this paper we highlight assumptions to get robust solutions both ways.

2. Heavier than Quadratic Losses. For simplicity we will assume we are estimating θ, that is, $g(\theta) = \theta$.

To diminish the shrinkage on the extremes, the loss function (centered around the overall group location) has to penalize large errors more heavily than square loss. We also need an origin where to anchor the weighting function $w(\theta)$. The natural origin in this kind of examples is the measured sample M of the common mean after the first 45 turns at bat.

Example 1 (Exponentially weighted loss:).

$$L^{exp}(\theta_i, \delta) = \exp[r \cdot |\theta_i - M|] \cdot (\theta_i - \delta_i)^2, i = 1, \ldots, m, r > 0$$

Calculation yields as the optimal estimator in (1)

$$\delta^{exp}(\theta_i) = \frac{a[b(\mu_1 - rv) - c] + a'[c' + (1 - b')(\mu_1 + rv)]}{ab + a'(1 - b')},$$

where $a = \exp(r(M-\mu_1)+vr^2/2))$, $b = \Phi(\frac{(M-(\mu_1-rv)}{\sqrt{v}})$, $c = \sqrt{v}\phi(\frac{M-(\mu_1-rv)}{\sqrt{v}})$, $a' = \exp(r(\mu_1 - M) + vr^2/2))$,$b' = \Phi(\frac{M-(\mu_1+vr)}{\sqrt{v}})$,$c' = \sqrt{v}\phi(\frac{M-(\mu_1+rv)}{\sqrt{v}})$, and μ_1, v are the posterior mean and variance respectively of a Normal likelihood with a Normal prior.

This and other loss functions can lead to tractable results, but it is more convenient for the purposes of the present paper to work with losses that have also a direct interpretation in terms of heavy tailed (robust) priors, which penalizes more heavily discrepancies and that are naturally scaled.

Example 2 (Cauchy over Gaussian Loss:). Here we place a Cauchy density over a Gaussian both centered at M with matched interquartile range.

$$L^{CG}(\theta_i, \delta_i) = \frac{\text{Cauchy}(\theta_i|M, 1)}{N(\theta_i|M, 2.19)} \cdot (\theta_i - \delta_i)^2.$$

Figure 2 shows the Cauchy over Gaussian loss, being quite close to square loss around zero but growing fast without bound for "exceptional" values.

Result 2. In fact the optimal estimator (1) under the $L^{CG}(\theta, \delta)$ is the posterior expectation under a Cauchy prior, since, using the Gaussian with mean M and variance 2.19 as a prior

$$\begin{aligned}\delta(\theta_i) = \frac{E[w(\theta_i) \cdot \theta_i|\text{data}]}{E[w(\theta_i)|\text{data}]} &= \frac{\int \frac{\text{Cauchy}(\theta_i|M,1)}{N(\theta_i|M,2.19)}\theta_i \cdot \pi(\theta_i|\text{data})}{\int \frac{\text{Cauchy}(\theta_i|M,1)}{N(\theta_i|M,2.19)}\pi(\theta_i|data)} \\ &= E[\theta_i|\text{data, Cauchy Prior}]\end{aligned}$$

In words, the optimal estimator under the Robust Cauchy over Gaussian penalty (and Gaussian prior) is the posterior expectation under a Cauchy Prior, since with square loss the optimal estimator is the posterior expectation.

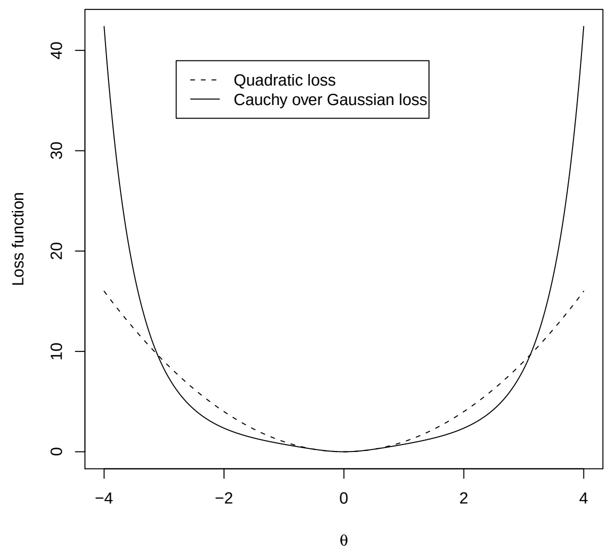


FIG 1. *Quadratic loss and Cauchy over Gaussian loss ($M = 0$).*

So a bridge has been established between Robust Losses and Robust Priors trough equation (1). We call it comprehensive Robustness, the use of Robust Loss Functions or the use of Robust (heavy tailed) Priors. For convenience, now we go to the "Robust Prior Route".

In what follows we try different models, in two methods: Empirical Bayes and Fully Robust Bayes, trying to solve, or at least alleviate the Clemente problem, that is, the lack of robustness of exponential family models with conjugate (light tailed) priors and squared loss function.

3. Robust Consequences of Robust (Heavily Tailed) Priors. In this section we motivate briefly the use of heavy tailed priors as a tool for robustness. Let us make the usual assumption of a Normal Likelihood and Prior:

$$Y \sim \text{Normal}(\mu, \tau^2), \tag{2}$$

$$\mu \sim \text{Normal}(\text{M}, \sigma_0^2), \tag{3}$$

where (M, σ_0^2) has been assigned, for instance via an Empirical Bayesian method. These assumptions coupled with square error loss, lead to the posterior conditional expectation as the optimal estimator, which can be written as:

$$E(\mu|y) = y + \frac{\tau^2}{\tau^2 + \sigma_0^2}(M - y). \tag{4}$$

There are several ways to analyze the lack of robustness of (4), but the one that is more relevant here: *all batters either exceptional or average are shifted to the general mean M a* **fixed** proportion $\tau^2/(\tau^2 + \sigma_0^2)$. That is the "Clemente Problem". Notice that increasing the prior variance σ_0^2 is **not** a fix to the problem: it would reduce the shrinkage certainly, but to all batters in equal proportion, even to those who are not exceptional. The posterior mean (4) is *myopic* to the exception. Now (4) is the logical consequence of the assumptions. Thus the only logical way to change it is to change the assumptions. We could change the loss, but equivalently we change the tail behavior of the priors: Using flatter tails, is giving to Bayes Theorem the input that the exceptional is more likely.

Our first alternative is a Double Exponential prior:

$$\mu \quad \sim \quad \mathrm{DE}(\mathrm{M}, \nu_0) = \frac{1}{\nu_0\sqrt{2}} \exp(-\frac{\sqrt{(2)}}{\nu_0}|\mu - \mathrm{M}|), \tag{5}$$

where $\nu_0 = \frac{\sqrt{(2)}\sigma_0}{\log(2)}\Phi^{-1}(0.75)$, to match the quartiles of the Normal prior. Finally for even heavier tails we explore with a Cauchy prior.

$$\mu \quad \sim \quad \mathrm{C}(\mathrm{M}, \gamma_0) = \frac{1}{\pi\gamma_0} \frac{1}{1 + (\mu - \mathrm{M})^2/\gamma_0^2}, \tag{6}$$

where $\gamma_0 = \sigma_0\Phi^{-1}(0.75)$, again to match Normal quartiles.

For exact and approximate results with these, and other priors, see for example Pericchi and Smith (1992), but Figures 2 and 3 tell the story. In Figure 2 the data are kept fixed at zero and the prior location is moved as to create a conflict between one data point and a prior. With a Normal prior the shrinkage grows linearly without bound. The other two priors yield robust estimators. For a Double Exponential prior, the posterior expectation become essentially a"*limited translation estimator*" in Efron and Morris' terminology. The influence of the overall mean M is bounded and monotonic. Finally for the Cauchy prior, the posterior expectation is **not** monotonic in the conflict between the MLE and the General Mean. Furthermore, the prior is progressively discarded in favor of the MLE, as the general mean and MLE diverge. It is also quite interesting that the shrinkage of the two robust priors is almost the same close to the center, that is around the general average M, and actually they give more shrinkage near the center than the Normal.

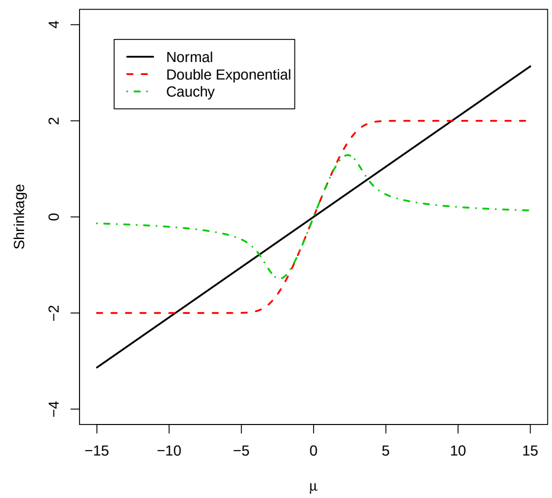


FIG 2. *Observation fixed at zero prior location varying. Normal linear unbounded influence of prior location, monotone limited translation in Double Exponential and discarding influence in the Cauchy.*

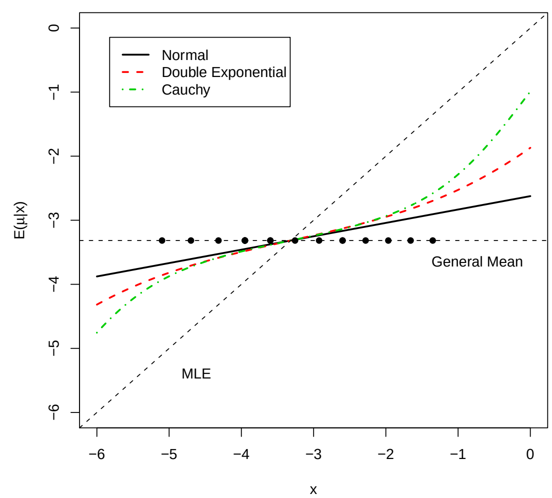


FIG 3. *Prior location fixed at the general average. Observations displayed in the X-Axis. Different Shrinkage behavior displayed of Normal, Double Exponential and Cauchy priors*

Coming closer to the current data and models, in Figure 3 the prior location M is fixed, and the different data points are presented in the X-Axis. Around the middle-ground the three estimators are very close together. But at some point, the robust estimators separate from the Normal, in the direction of the MLE, being the adjustment towards the MLE of the Cauchy somewhat stronger than that of the Double-Exponential.

4. Models. We will consider two types of models, Empirical Bayes and Full Bayesian strategies.

4.1. *Empirical Bayes Strategy.* Following Efron and Morris (1975), general location and scale parameters are calculated from the sample: $\mathrm{M} = \bar{X} = -3.3166$ and $\tilde{\sigma}^2$ such that $\frac{1}{(1+\tilde{\sigma}^2)} = \frac{k-3}{\sum_{i=1}^{k}(X_i-\bar{X})^2}$, so $\tau = (\sigma^2)^{-1} = 3.7853$.

Models 1, 2 and 3, are defined with the likelihood (2) and respectively priors (3), (5) and (6). The justification is as follows: the first prior is the original analysis by Efron and Morris, Double Exponential and Cauchy priors are two heavy tailed (as compared with Normal) but that promote a qualitative different behavior of the estimators. The three priors have the same origin given by an Empirical Bayes analysis, and the scales have been matched by equating the interquartile ranges, to ease the comparison.

4.2. *Full Bayesian Strategy.*

4.2.1. *Model 4: Full Bayesian conjugate model.* This model assigns vague conjugate priors to the common mean M and the common variance σ^2. This is the fully Bayes version of Efron and Morrris Empirical Bayes model.

$$\begin{aligned} X_i &\sim \text{Normal}(\mu_i, 1), i = 1, \ldots, 18 \\ \mu_i &\sim \text{Normal}(\text{M}, \sigma^2) \\ M &\sim N(0, 10^5), \quad \sigma^2 \sim \text{Inv-Gamma}(0.01, 0.01) \end{aligned}$$

4.2.2. *Model 5: Normal likelihood, Double Exponential prior for μ_i, vague Double Exponential prior for the general mean M, Beta2$(1,1)$ prior for the squared location parameter $\sigma = \frac{\nu}{\sqrt{2}}$.*

$$\begin{aligned} X_i &\sim \text{Normal}(\mu_i, 1), i = 1, \ldots, 18 \\ \mu_i &\sim \text{DE}(\text{M}, \sqrt{2}\sigma) \\ M &\sim \text{DE}(0, \sqrt{2} \times 10^{-3}), \quad \sigma^2 \sim \text{Beta2}(1, 1) \end{aligned}$$

4.2.3. *Model 6: Normal likelihood, Cauchy prior for μ_i, vague Cauchy prior for the general mean M, Beta2$(1,1)$ prior for the squared location parameter.*

$$\begin{aligned} X_i &\sim \text{Normal}(\mu_i, 1), i = 1, \ldots, 18 \\ \mu_i &\sim \text{Cauchy}(\text{M}, \sigma) \\ M &\sim \text{Cauchy}(0, 10^{-3}), \;\; \sigma^2 \sim \text{Beta2}(1,1) \end{aligned}$$

In this two last models, robust priors (Double Exponential and Cauchy) have been assigned for the locations. On the other hand we propose the use of the **Beta distribution of the second kind** with parameters p and q family (Beta2(p,q)) as priors for the squared scale parameters. Let Y be a random variable such that $Y \sim$ Beta2(p,q); its density function (see Johnson, Kotz, and Balakrishnan, 1995) is given by:

$$p(y|p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \cdot \frac{y^{p-1}}{(1+y)^{(p+q)}}, y > 0,$$

This family has a very natural justification as a prior for variances in hierarchical models, as it is obtained as a scale mixture of Gamma distributions, through a Gamma mixing distributions, in much the same way that the Student-t is obtained as a scale mixture of Normal distributions, see Pérez and Pericchi (2009). The usual prior assumed for variances is the Inverted Gamma family with very large prior variance. This practice has come under criticism by Gelman (2006), who among other alternatives propose a half-Cauchy prior. The Beta2 prior has flexible tail behavior which makes it particularly suitable for modeling. When p = q = 1, the Beta2(1,1) prior is very close to the half-Cauchy, so we use it here. We also tried a very small q, which leads to a flatter prior than that with a Cauchy tail, with very similar results.

The big question now is: what is the assumed prior for the location parameter in Model 6 after integrating out the variance? This is a novel distribution, to the best of our knowledge, that we call Cauchy-Beta Type 2, which deserves on its own a special note.

Definition (Beta2-Cauchy Prior). Assume that $\theta|\sigma \sim$ Cauchy$(0,\sigma)$, and $\sigma^2 \sim$ Beta2(p,q), then θ is distributed as a Cauchy-Beta2$(0,p,q)$.

It is remarkable that when $p = q = 1$ (assignment that makes the Beta2 to have Cauchy tails) the marginal density for θ has an explicit formulae (after a long integration by simple fractions):

Result 3. The marginal density for the location θ of a Beta2-Cauchy prior is:

$$\pi(\theta) = \frac{\pi \cdot |\theta| - 1 - \theta^2 - (1-\theta^2) \cdot \log(|\theta|)}{\pi \cdot (1+\theta^2)^2}. \tag{7}$$

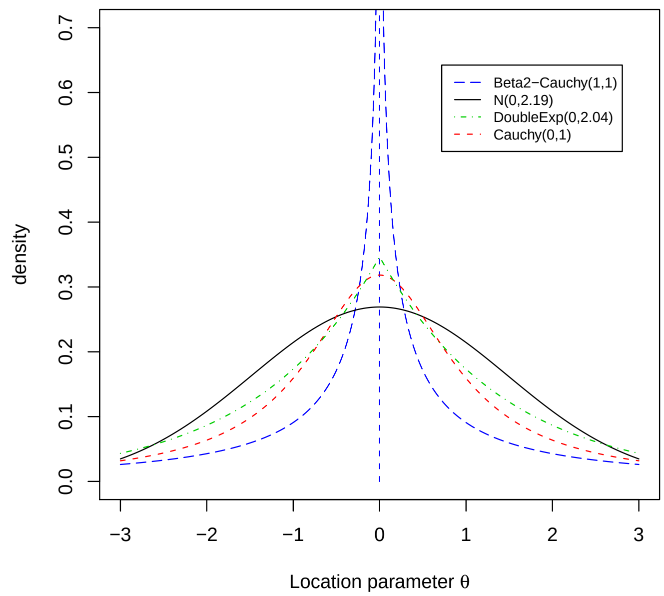
In Figure 4, the new prior (7) is displayed along with the Cauchy, Double-Exponential and Normal. This prior enjoys a number of features that explain why it is optimal in predicting the batters averages. This prior is unbounded as $\theta \to 0$ and has tails even heavier than Cauchy. In a recent paper Carvalho, Polson and Scott (2010) (see also Polson and Scott , 2010) find such characteristics for a prior to be both robust and leading to efficient estimation (they propose particular prior which does not have an explicit form and that they call "horseshoe" prior). Furthermore, the Beta2-Cauchy besides obeying such desiderata, has an explicit form (at least for p=q=1) which makes it amenable for mathematical analysis. We do not know of any other explicit "horseshoe" prior.

5. Results. In Figure 5 we display the milder shrinkage of the extremes using robust priors, particularly Cauchy priors, when compared to that of Model 1 and Model 4.

Robustifying the priors pays dividends twice: the relative shrinkage of the extremes is lower and at the same time the error of prediction is diminished up to 7%, for Model 6 which uses the Beta2-Cauchy prior. Model 3 for example, which incorporates the Cauchy prior has decreased 5% the square prediction error than Model 1, and predicts for Clemente a more respectful 0.314 average, much higher than the 0.290 from Model 1. Something similar may be said for Model 6. On the other hand, the price paid seems less than modest: computationally now there exist approximations and MCMC algorithms that make the computations routine. Thus by a very modest cost in computation, the robustified model has achieved both goals, decreasing the MSE and solving or at least alleviating the Clemente problem.

In terms of alternative approaches of Statistics that merge direct and indirect evidence, the difference between (sensible and objective versions) of Empirical and Fully Bayes Hierarchical Modeling is tiny as compared with the difference between heavy and light tail priors.

In general the Clemente problem is closely related to the implicit dogmatism inherent, not in Bayes in general, but in "Conjugate Bayes with Square Loss". The way out seems to be: either Empirical or Fully Bayesian Hierarchical Model, but making emphasis on Robustness.



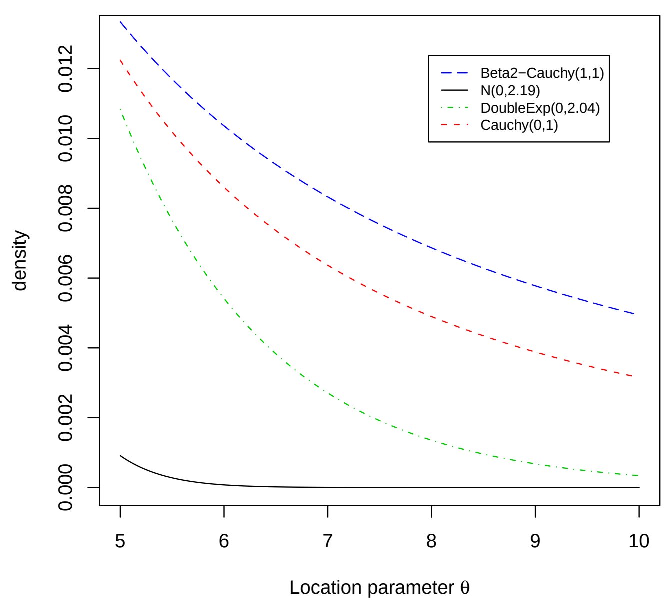
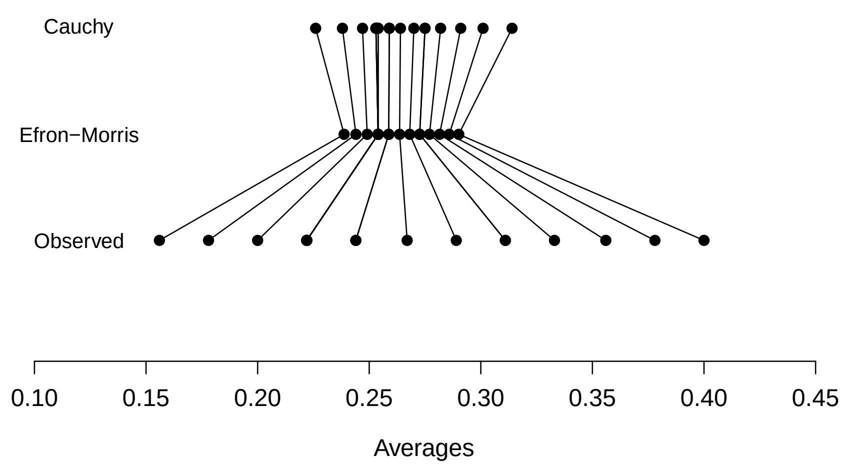


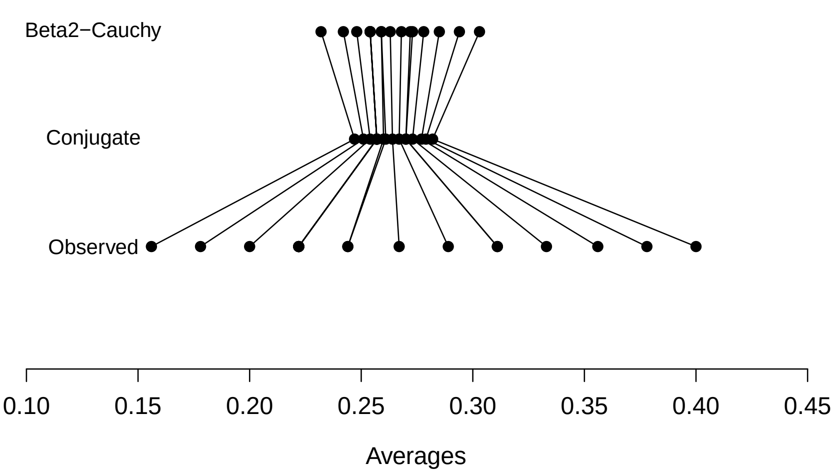
FIG 4. *Comparison of Beta2-Cauchy, Normal, Double Exponential and Cauchy distributions.*

Player	Observed season	First 45 (MLE)	General mean	Model 1	Model 2	Model3	Model 4	Model 5	Model 6
Clemente	0.346	0.400	0.265	0.290	0.304	0.314	0.282	0.2978	0.303
Robinson	0.298	0.378	0.265	0.286	0.296	0.301	0.279	0.2912	0.294
Howard	0.276	0.356	0.265	0.282	0.288	0.291	0.277	0.2848	0.285
Johnstone	0.222	0.333	0.265	0.277	0.281	0.282	0.273	0.2788	0.278
Berry	0.273	0.311	0.265	0.273	0.275	0.275	0.270	0.2732	0.272
Spencer	0.270	0.311	0.265	0.273	0.275	0.275	0.270	0.2729	0.273
Kessinger	0.263	0.289	0.265	0.269	0.269	0.270	0.267	0.2683	0.268
Alvarado	0.210	0.267	0.265	0.265	0.264	0.264	0.264	0.2637	0.263
Santo	0.269	0.244	0.265	0.259	0.258	0.259	0.261	0.2585	0.259
Swoboda	0.230	0.244	0.265	0.259	0.258	0.259	0.260	0.2593	0.259
Unser	0.264	0.222	0.265	0.255	0.252	0.254	0.257	0.2539	0.254
Williams	0.256	0.222	0.265	0.255	0.252	0.254	0.257	0.254	0.254
Scott	0.303	0.222	0.265	0.255	0.252	0.253	0.257	0.2538	0.254
Petrocelli	0.264	0.222	0.265	0.255	0.252	0.253	0.257	0.2535	0.254
Rodriguez	0.226	0.222	0.265	0.255	0.252	0.253	0.257	0.2538	0.254
Campaneris	0.285	0.200	0.265	0.250	0.245	0.247	0.254	0.248	0.248
Munson	0.316	0.178	0.265	0.245	0.237	0.238	0.251	0.2416	0.242
Alvis	0.200	0.156	0.265	0.240	0.228	0.226	0.247	0.2339	0.232
MSE ($x10^3$)		**4.184**	**1.348**	**1.196**	**1.187**	**1.137**	**1.1975**	**1.1681**	**1.1159**
$\frac{MSE(\text{Model})}{MLE(\text{Model 1})}$		**350%**	**113%**	**100%**	**99%**	**95%**	**100%**	**98%**	**93%**

Table 2. *Estimators and mean square error of prediction for MLE, general mean, empirical Bayes models and full Bayesian models*



(a)MLE, Model 1, and Robust Empirical Bayes Model 2.



(b)MLE, Model 1, and Robust Full Bayes Hierarchical Model 6.

FIG 5. *MLE, Model 1, and Robust Bayes Estimators.*

6. The resurgence of "Objective Robust Bayesian Analysis". It can be argued that there is a resurgence of Objective Robust Bayesian Analisis (ORBA). To put this article in perspective we should mention some recent contributions. The first is Andrade and O'Hagan (2006), on which the theory of Regularly Varying (RV) functions is used to check robustness, for general location and scale parameters. The second is Fúquene, Cook and Pericchi (2009) on which it is proved the Generalized Polynomial Tails Comparison (GPTC) Theorem, and the properties of Berger's prior are analyzed (prior mentioned in the Polson and Scott's paper). In Pérez and Pericchi (2009), the relationship between RV and GPTC is established, and naturally the Beta of the Second Kind appears in a Meta-analysis of different hospitals. On the other hand, Carvalho, Polson and Scott (2010) introduced the "Horseshoe Estimator" which obeys certain desiderata convenient for robust analysis in hierarchical models.

In the present article we add the following insights: 1) We established a bridge between robust losses and robust priors. The theoretical duality between losses and priors has been mentioned before, for example by Berger (1985, p. 161) but to the best of our knowledge the consequences of looking at robust procedures with the glass of robust penalties is new. Furthermore note that the way on which the bridge between priors and Losses is done, is through an empirical Bayes component, since the loss is centered and scaled through an EB reasoning 2) We introduce here a robust conventional prior for variances, the Beta type 2, that generalizes previous proposals. Furthermore we show the novel Beta2-Cauchy prior which is an explicit objective prior that obeys the desiderata of a "Horseshoe" prior. 3) We illustrate, using a classical data set, that it is possible to alleviate the "Clemente problem" and at the same time reduce mean square error of prediction as compared with non-robust conjugate approaches and with the James-Stein estimator. 4) We illustrate that the differences between sensible versions of Empirical Bayes and Objective Bayes are minute in practice. Much more important is the difference between robust Bayes and conjugate Bayes. This coupled with the fact that Robust Bayes is in general less dogmatic, in the sense that in conflict prior information is discarded, makes Objective Robust Bayes a strong candidate for statistical synthesis.

The scope of applications of Robust Objective Procedures is wide ranging, and includes any problem calling for the use of hierarchical models. Just to mention a few, in the analysis of population dynamics for the orchid genus *Caladenia* presented in Tremblay *et al* (2009), hierarchical models based on vague conjugate priors were used, but even though model selection procedures clearly pointed towards a hierarchical model as the selected one, it gave results that were not biologically sound, probably due to excessive shrinkage towards the species with more data. Another example is the unfair "pulling down" of hitherto perfect score hospitals in hospital profiling, just because there are a couple of hospitals with bad performance (see Normand and Shahian , 2007). Shrinkage is a good thing, but non-robust methods yields too much of a good thing.

References.

Andrade J.A.A. and O'Hagan, A.(2006). Bayesian Robustness Modeling Using Regularly Varying Distribution. *Bayesian Analysis* **1**, pp. 169-188.

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

Brown, L.D (2008). In-season prediction of batting averages: a field test of empirical Bayes and Bayes methodologies. *The Annals of Applied Sttistics* **2** 113–214.

Carvalho C.M., Polson N.G. and Scott J.G. (2010). The Horseshoe Estimator for Sparse Signals. *BIOMETRIKA.* (Accepted).

Efron B. (2009). The future of indirect evidence. *The 2009 International Workshop on Objective Bayesian Methodology*, Wharton School of Business. `http://stat.wharton.upenn.edu/statweb/Conference/OBayes09/lectures.htm`

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.

Efron B. and Morris C. (1975). Data analysis using Stein's estimators and its generalizations. *J. Amer. Stat. Assoc.* **70**, 311–319.

Ferguson, T. (1967). *Mathematical Statistics. A Decision Theoretic Approach.* Academic Press.

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

Gelman, A.(2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1(3)** 515-533

Jonhnson, N.L., Kotz, S., Balakrishnan, N. (1995). *Continuous Univariate Distributions.* Vol. 2, 2nd ed., Wiley, New York.

Normand, S.T. and Shahian, D.M.(2007). Statistical and clinical sspects of hospital outcomes profiling. *Statistical Science*, **22(2)**, 206–226.

Pérez M.E. and Pericchi L.R. (2009). The case for a fully Robust Hierarchical Bayesian analysis of clinical trials. *The 2009 International Workshop on Objective Bayesian Methodology*,Wharton School of Business.

Pericchi L.R. and Smith A.F.M. (1992.) Exact and approximate posterior moments for a Normal location parameter. *J. R. Statist. Soc. B*, **54(3)**,793–804.

Polson N.G. and Scott J.G. (2010). Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction (with discussion). In, *Bayesian Statistics 9, Oxford University Press.* (To appear).

Tremblay, R.L., Pérez, M.E., Larcombe, M., Brown, D., Quarmby, J., Bickerton, D., French, G. and Bould, A.(2009). Population dynamics of *Caladenia*: Bayesian estimates of transition and extinction probabilities. *Australian Journal of Botany*, **57**, 351-360.

Department of Mathematics
University of Puerto Rico, Rio Piedras Campus
POBox 70377
San Juan, PR 00936-8377
E-mail: luarpr@gmail.com
meglee@uprrp.edu