Bayesian Robustness and Conflict Resolution.

Pericchi Guerra, Luis Raul(1)(2)
1) Department of Mathematics and Center for Biostatistics and Bioinformatics,
University of Puerto Rico, Rio Piedras, San Juan, PR
2) Center for Biostatistics and Bioinformatics Consortium
MDAnderson Cancer Center and Univ Puerto Rico CCC.

June 2015, Valencia, Objective Bayes, Course

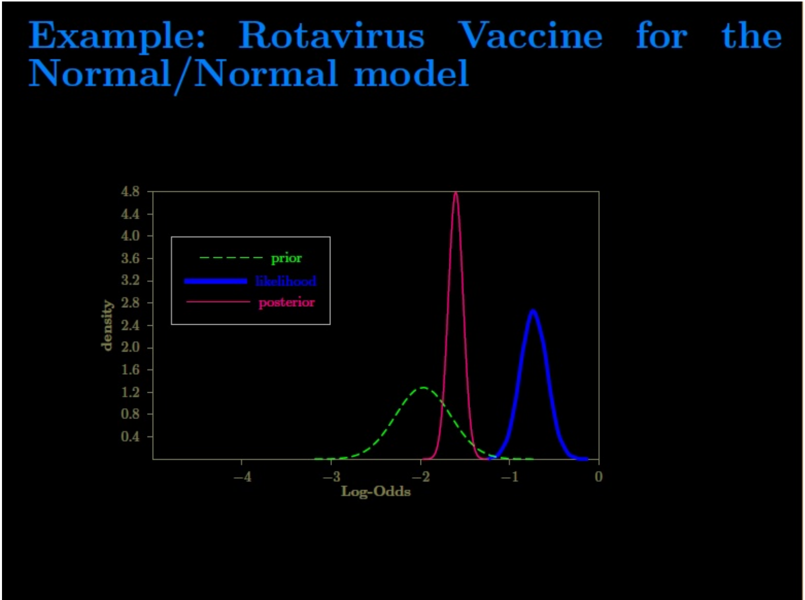
Abstract

-We are not going to address Bayesian Robustness from the point of view of Classes of Priors: Classical reference Berger (1994) An overview of robust Bayesian analysis [with Discussion]. 'Test' , 3, 5–124.
-Rather we focus here on a relationship between a prior and a likelihood, and between observations. "Who is the winner in a conflict???"
Bayes is not Conjugate Bayes....
Its mistaken perception: "What I dislike about Bayesian Statistics is that you can produce the results you wish to produce..."""
-This tutorial is an update of: O'Hagan A. and Pericchi L. (2012) "Bayesian heavy-tailed models and conflict resolution: a review".Brazilian Journal of Probability and Statistics, Vol. 26, No. 4, 372-401.

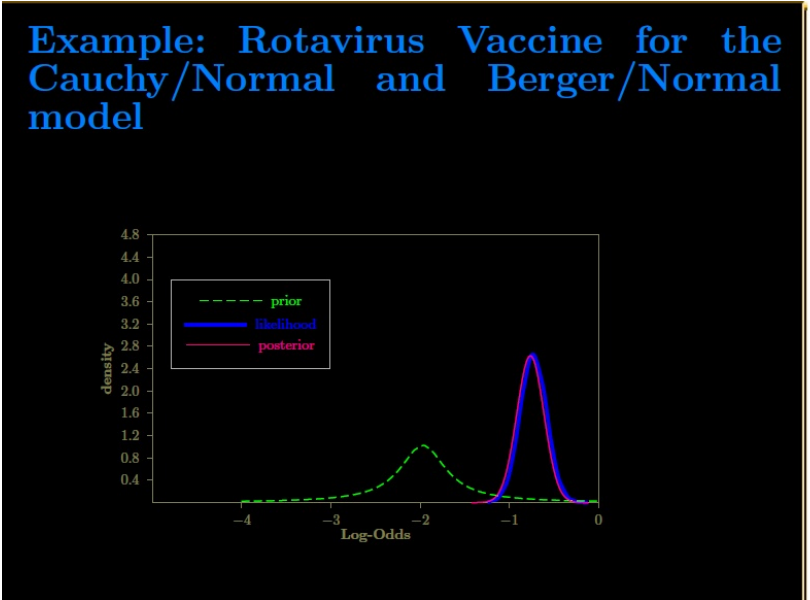
Is Robust Bayes-Conflict Resolution part of Objective Bayes?

-Objective Bayes VS Casual Objective Bayes
-Weakly Informative or Proper Objective Bayes (Objective Bayes with proper or partially proper priors)
-Conflicting information may also be among groups of observations
-The Theory of Conflict Resolution is necessary for Objective Bayes Conventional Priors.
-Many selections of $[Likelihood, Prior]$ can be done by Model Selection methodology.
-For heavy tail priors in the case of conflict the prior is discarded and an objective prior is recovered.

Motivation Cinical Trial Finland and Venezuela. Light tailed prior



Motivation: Cinical Trial Finland and Venezuela. Heavy tailed prior



History

- de Finetti (1961): if the observations had different, independent, unknown variances, then the outlier would be estimated as having a large error variance, and so would be given less weight. The greater the conflict, in the sense of the outlier being further from the remaining observations, the less weight it would get. In the limit, as the outlier became infinitely separated from the other observations it would be given zero weight, which de Finetti described as Bayesian outlier rejection.
- Lindley (1968), in response to a discussion contribution from E.M.L. Beale, gives an approximate analysis based on the leading term of an expansion for the posterior in the case of a t prior distribution
- Dawid (1973), Hill (1975) and O'Hagan (1979) gave first formal analyses.

Connection with Heavy Tails and Scale Mixtures

The connection between this kind of outlier rejection and heavy tails arises from the uncertainty about the error variance for each observation. Suppose that $x_i \sim N(\mu, \sigma_i^2)$, so that observation i has error variance σ_i^2, and now let σ_i^2 have distribution $F(\cdot)$. Assuming, as de Finetti did, that the σ_i^2s are independent, then the marginal density of x_i is of the form known as a scale mixture of normals. It has density

$$f(x_i \mid \mu) = \int \sigma^{-1} \phi((x_i - \mu)/\sigma) dF(\sigma^2) \ ,$$

where $\phi(\cdot)$ is the standard normal density function. It is clear that de Finetti's argument requires $F(\cdot)$ to give non-zero probability to arbitrarily large values of σ^2, so that the weight attached to an outlier can become arbitrarily small. The best known family of scale mixtures of normals is the t family, where $F(\cdot)$ is an inverse gamma distribution, and t distributions have heavier tails than normal distributions.

Tails and duality

We begin with the simplest case, in which we have a single observation x having density $f(x-\theta)$, so that θ is a location parameter. We let θ have prior density $g(\theta)$. We can think of x as composed of the location parameter θ plus 'observation error' $\phi = x - \theta$. Notice that $x = \theta + \phi$, and that θ and ϕ are independent random variables with densities g and f respectively. Our interest is in the limiting behaviour of the posterior distribution for θ when x becomes large. That limiting behaviour will depend on to what extent the posterior distribution attributes the large value of $x = \theta + \phi$ to θ being large or to ϕ being large, which in turn depends on the forms of their densities, g and f. An important feature of this model is a duality between θ and ϕ that was pointed out by Dawid (1973). Whatever results we can prove about the posterior distribution of θ in this system will apply instead to ϕ if we reverse the roles of f and g.

Normal Likelihood Normal prior as observation grows

378 A. O'Hagan and L. Pericchi

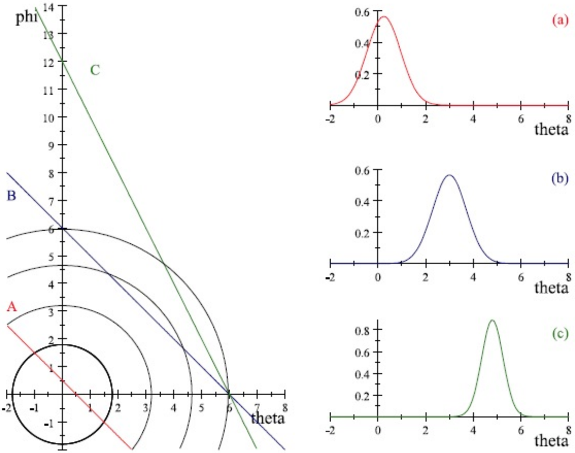


Figure 2 *Joint and posterior densities for normal f and g.*

Cauchy Likelihood Normal prior as observation grows

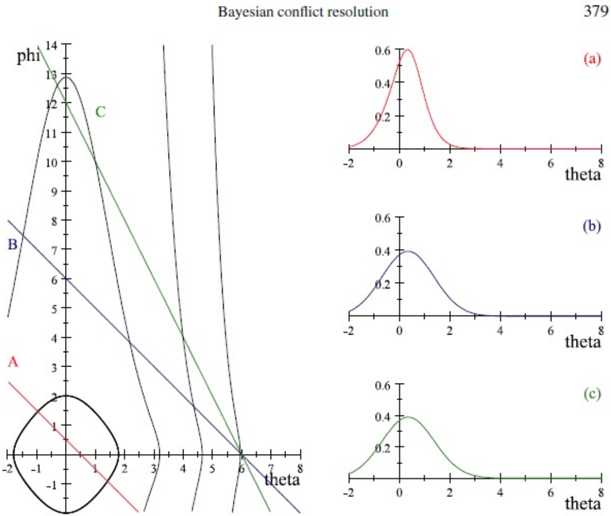


Figure 3 *Joint and posterior densities for Cauchy f and normal g.*

Cauchy Likelihood Cauchy prior as observation grows

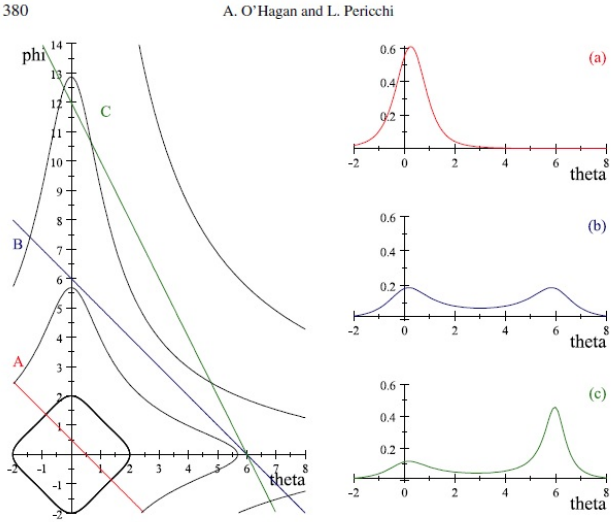


Figure 4 *Joint and posterior densities for Cauchy f and g.*

Dawid's Theorem: Rejection of Information

First, Dawid (1973) gave the following sufficient set of conditions for the posterior tending to the prior as $x \to \infty$.(Duality)

(A1) **"Heavy Tailed"**
Given $\varepsilon > 0$, $h > 0$, there exists A such that if $y > A$ then

$$\left| f(y') - f(y) \right| < \varepsilon f(y)$$

whenever $|y' - y| < h$.

(A2) **"Regularity"**
For some constants B, M,

$$0 < f(y') < Mf(y)$$

whenever $y' > y > B$.

(A3) **"Light Right Tail of g"**
$\int^{\infty} k(\theta) g(\theta) d\theta < \infty$, where $k(\theta) = \sup_x \{f(x - \theta)/f(x)\}$.

O'Hagan (1979) sought to simplify Dawid's condition (A3), which may be difficult to verify in practice because of the need to derive $k(\theta)$. He showed that if (A2) and (A3) were replaced by the following (slightly stronger) conditions then together with (A1) they would still be sufficient for the posterior density of θ to tend to $g(\theta)$ as $x \to \infty$.

(B2) **"Regularity"**

0.1 $f(y)$ is continuous and positive for all y.

0.2 There exists a B such that for all $y \geq B$

0.2.1 $f(y)$ is decreasing in y,

0.2.2 $d \log f(y)/dy$ exists and is increasing in y.

0.3 There exists a C such that, for all $y \leq C$, $f(y)$ is increasing in y.

(B3) **"Light Right Tail of g"**

$\int^{\infty} \{f(\theta)\}^{-1} g(\theta) d\theta < \infty$.

O'Hagan (1990) introduces the notion of *credence* according the following definition.

Definition

A density $f(y)$ has credence c if there exist $K \geq k > 0$ such that for all $y \in \mathbb{R}$,

$$k \leq (1+y^2)^{c/2} f(y) \leq K \ .$$

Thus f has credence c if it is bounded above and below by multiples of a t density with $c-1$ degrees of freedom (and in particular the t density itself has credence $d+1$). He then proves that if f has credence c and g has credence $c' > c$ then

(a) for any given $d > 0$, there exists an A such that for all $|d| > A$ the posterior density is bounded for all $|\theta| \leq d$ above and below by multiples of g, and

(b) for all $r : \mathbb{R}^+ \to \mathbb{R}^+$ such that $r(x) \to \infty$ as $x \to \infty$, the posterior probability that $P(|\theta| > r(|x|)$ tends to 0 as $|x| \to \infty$.

Andrade and O'Hagan (2006) consider distributions with *regularly varying* tails.

Definition

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

$$\frac{f(\lambda y)}{f(y)} \to \lambda^{\rho}$$

as $y \to \infty$ for all $\lambda > 0$.

The regular variation index for a proper density must be negative, so Andrade and O'Hagan say that f has *RV-credence c* if its right-hand tail is regularly varying with index $-c$. A t distribution with d degrees of freedom has credence $d+1$ and also RV-credence $d+1$, but the two definitions are not equivalent

Another similar result is the Generalised Polynomial Tails Theorem of Fúquene, Cook and Pericchi (2009). Their conditions are

(D1&2) There exist constants A_1, c, C_1, C_2 and C_3 such that for all $y > A_1$

$$C_1 y^{-c} \leq f(y) \leq C_2 y^{-c} \ ,$$
$$\frac{d}{dy} f(y) \leq C_3 y^{-c-1} \ .$$

"Polynomial right tail Does not wiggle too much"

(D3) There exist constants A_2, $c' > c$ and C_4 such that for all $m > A_2$

$$\int_{\theta > m} g(\theta) d\theta \leq C_4 m^{-c'} \ .$$

"Right tail of g thinner"

Then the observation will be rejected as $x \to \infty$.

Regular log-convex tails: Desgagné and Angers (2007)

a density f has a **regular log-convex** right-hand tail if it is positive, bounded above and satisfies both the condition (A1) and a new condition:

(E2) There exist constants $A_2 > 0$ and $M > 1$ and proper density functions f^* and f^+ such that for all $y > A_2$

$$\frac{f^2(y/2)}{f(y)f^+(y)} \leq M \,,$$
$$\frac{d^2}{dy^2} \log f^*(y) \geq \frac{d^2}{dy^2} \log f^+(y) \geq 0 \,,$$

where f^* must be such that there exist constants $B > 0$ and $0 < K_1 < K_2 < \infty$ for which $K_1 \leq f(y)/f^*(y) \leq K_2$ whenever $y > B$.

(E3) $\lim_{y\to\infty} g(y)/f(y) = 0$. (Straightforward)

Then, the observation is asymptotically rejected as $x \to \infty$.

Comparison of the Approaches

- The theory of credence requires both f and g to have finite credence, whereas other approaches allow g to be any kind of distribution subject only to it being lighter-tailed than f in the sense of condition (A3), (B3), (D3) or (E3). Andrade and O'Hagan allow g to be rapidly varying as well as regularly varying, which for instance includes the case of g being normal.
- Distributions with finite credence can have tails that 'wiggle' in ways that are not allowed by Dawid's conditions or regular variation.
- A density $f(y)$ whose tails are proportional to a t density with d degrees of freedom multiplied by a slowly-varying function like $\log y$ does not have any credence value but is covered by the regular variation approach with RV-credence $d + 1$. Desgagné and Angers prove that such a f is heavier-tailed than the t distribution with d degrees of freedom and the observation will be rejected in the limit if g has that t distribution.

Comparison

- There are other kinds of distribution, for instance those whose tails decay like $\exp(-y^b)$ for $0 < b < 1$, that are covered by Dawid's conditions and by Desgagné and Angers' log-convexity conditions, but these are lighter-tailed than the t distributions.

Asymptotic results for moments

(A4) $\int^{\infty} m(\theta)k(\theta)g(\theta)d\theta < \infty$ Dawid Conditions

the posterior expectation of $m(\theta)$ tends to its prior expectation. In the case of f having a t distribution with d degrees of freedom, this means that the posterior moments of order up to n converge to the corresponding prior moments if g is normal, or if g is a t density with degrees of freedom $d' > d + p + 1$. The same will be true under the alternative conditions of O'Hagan (1979) and Andrade and O'Hagan (2006). In fact, again we find empirically that this convergence holds if $d' > d + p$, but O'Hagan (1990) only obtains the same requirement, $d' > d + p + 1$, using the credence approach. The tighter condition $d' > d + p$ is proved specifically for the case of two t distributions by Fan and Berger (1992), at least for the mean and variance ($p = 1$ or 2), but this issue is now fully resolved by Desgagné and Angers (2007). Their condition for convergence of the posterior expectation of $m(\theta)$ to its prior expectation is simply

(E4) $\lim_{y\to\infty} m(y)g(y)/f(y) = 0$.

Related Research

- **Never Rejected** Normal Density, O'Hagan (1979)
- **Bounded Influence or Partial Rejection** Double Exponential Density, O'Hagan (1979), Pericchi and Smith (1992), Pericchi and Sansó (1995)
- **A Moment of Indecision** Normal and Heavy Tailed as $x \to \infty$ the posterior Variance first grows and then decrease to its asymptotic value. O'Hagan (1981), Fan and Berger (1992)

A Moment of Indecision through the Posterior Variance

PERICCHI AND SMITH

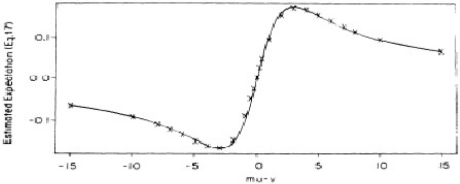


Fig. 3. Student prior: $n = 10$, $\alpha = 9$, $\sigma = \tau = 1$ (×, calculated value)

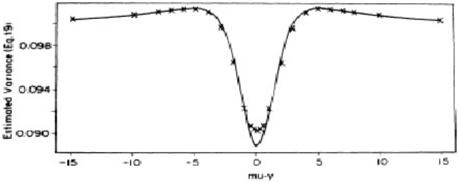


Fig. 4. Student prior: $n = 10$, $\alpha = 9$, $\sigma = \tau = 1$ (×, calculated value)

Generalized Exponential Power

Desagne and Angers (2007) and Desagné and Angers (2007), extend the definition of **credence**.
redefine p-credence: $f(y)$ is now said to have p-credence (b, a, c, d) in its right-hand tail if

$$\lim_{y \to \infty} \frac{f(y)}{\exp(-a(y^*)^b)(y^*)^{-c} \log^{-d}(y^*)} = K$$

for some K. With this modification they show that the distributions with p-credence (b, a, c, d) for $b < 1$ have regular log-convex tails, so that their basic results apply for any f and g in this wide class of densities.

Multiple Observations, Single Location Parameters

Both O'Hagan (1979) and Desgagné and Angers (2007) address outlier rejection in these general terms. Consider m groups of observations, identified by subsets S_j, $j = 1, 2, \ldots, m$, of the indices. Thus, $\cup_{j=1}^{m} S_j = \{0, 1, 2, \ldots, n\}$ and $S_j \cap S_{j'} = \varnothing$ when $j \neq j'$. We suppose that the observations in group 1 remain fixed while the other groups move increasing far apart from the first group and from each other.

We now seek conditions under which the posterior distribution of θ tends to the posterior

$$g^*(\theta) \propto \prod_{i \in S_1} f_i(x_i - \theta)$$

that would arise given only the information sources in group 1.

O'Hagan (1979) proves that if f_i has credence c_i, $i = 0, 1, 2, \ldots, n$ then the asymptotic rejection of all the other groups occurs provided $C_1 = \max_j C_j$, where $C_j = \sum_{i \in S_j} c_i$.

Multivariate Location

O'Hagan and Le (1994) introduce a family of bivariate heavy-tailed distributions, the bivariate T family, generalising the two forms above. The $T(c, c_1, c_2)$ distribution is defined to have density function

$$f(y_1, y_2) \propto (1 + y_1^2 + y_2^2)^{-c/2}(1 + y_1^2)^{-c_1/2}(1 + y_2^2)^{-c_2/2} \ . \qquad (1)$$

O'Hagan and Le provide several numerical examples to illustrate different asymptotics for a single bivariate observation when both f and g have bivariate T distributions.

Finegold and Drton (2014) Alternative T Mult Outliers

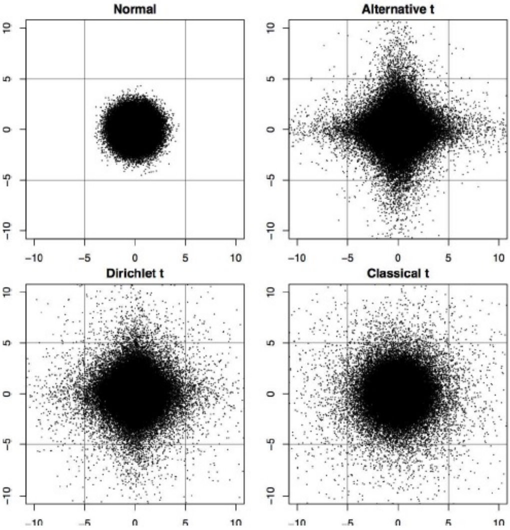


Figure 2: We simulate 100,000 draws from four bivariate distributions: the normal

Exchangeable Locations

1. Observations x_i have densities $f_i(x_i - \theta_i)$ for $i = 1, 2, \ldots, p$. Often we have replication and would suppose that observations x_{ij} have densities $f_i(x_{ij} - \theta_i)$ for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, n_i$.
2. Parameters θ_i have independent densities $g_i(\theta_i - \xi)$ given ξ, $i = 1, 2, \ldots, p$.
3. The hyperparameter ξ has density $h(\xi)$.

Exchangeable means No Conflict

Bayesian conflict resolution

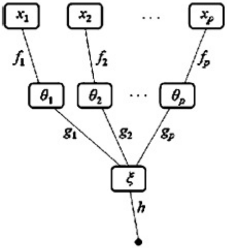


Figure 5 *Exchangeable means, no conflict.*

Exchangeable means Conflict

392 A. O'Hagan and L. Pericchi

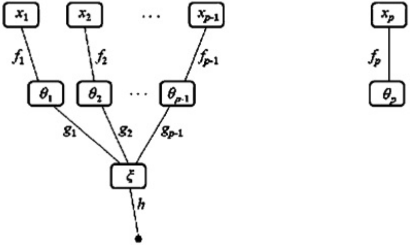
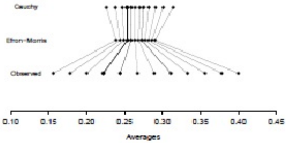
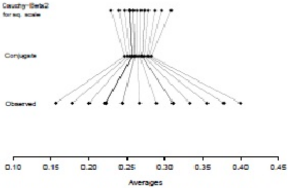


Figure 6 *Exchangeable means with an outlier.*

The Clemente Problem (Efron): How to protect exceptionals from too much shrinkage? Perez and Pericchi (2015)



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



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

Gaps in the Theory

We have quite complete theory for the case of many observations and a single location parameter.
Unfortunately, in models with two or more parameters we have only a few sparse results. We need more general theory of multivariate heavy-tailed distributions, addressing more of the potential complexity that is opened up by allowing different tail thicknesses in different directions.
We need some theory dealing with the interaction of heavy-tailed distributions at different levels of a hierarchical model.

Scale Mixture of Normals

One way to generate heavy-tailed distributions, and t distributions in particular, is as scale mixtures of normal distributions. Thus, if f_N is the density of the $N(0, \omega)$ distribution and we let ω have a density $p(\omega)$, then integrating out ω gives the density

$$f(y) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \omega^{-1/2} \exp(-\omega^{-1} y^2/2) p(\omega) \; d\omega \; .$$

If p is an inverse-gamma density, then f is a t density. Other choices of $p(\omega)$ can give exponential power distributions or stable distributions. If the tail of p is sufficiently heavy then f will be a heavy-tailed distribution. The representation as a scale mixture of normals can facilitate computation of the posterior distribution by MCMC; see for instance West (1981, 1984), Carlin and Polson (1991), Choy and Smith (1997). Essentially, rejection of an information source represented by a scale mixture arises through the posterior distribution of ω concentrating on larger values of ω.

Scale Mixture of Normals Cont.

With multiple observations, heavy-tailed distributions f_i can be modelled as scale mixtures, each with its own ω_i. Therefore individual observations can be rejected when they conflict with the remaining observations, and the posterior estimates of the ω_is provide an indication of which observations are being discounted as outliers

Scale mixtures for scales

Thus far, we have effectively assumed known variances.

Suppose x_i has the distribution $N(\theta, \sigma^2)$ with both θ and σ^2 unknown. If σ^2 has an inverse-gamma distribution, which is the standard conjugate prior distribution, then after integrating out σ^2 each x_i has a t distribution. But these are not independent t distributions. Instead, the x_is jointly have a multivariate t distribution. **The difference is substantial when conflict arises.** Instead of being able to reject individual observations as outliers, we can only reject the entire sample or none of it. In practice, it is very important to give each information source its own ω_i.

A recent proposal for scale mixtures of scales is the Scaled Beta 2, Perez, Pericchi and Ramirez (2015) which is a gamma scaled mixture of gammas:

$$g(\sigma^2) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)b} \frac{(\sigma^2/b)^{p-1}}{(1+\sigma^2/b)^{(p+q)}}, p > 0, q > 0,$$

which for small q is heavy tail and p control behaviour at the origin.

Single Scale Parameter

Andrade and O'Hagan (2006) when f has regularly varying right tail. If the RV-credence of f is c they have a single condition for g to be lighter-tailed than f.

(F) For some $\delta > 0$, $\int^{\infty} \theta^{c+\delta-1} g(\theta)\, d\theta < \infty$.

In particular, if g also has regularly varying right tail with RV-credence c', then condition (F) simply requires $c' > c$. Andrade and O'Hagan (2006) prove that then the limiting posterior density of θ is

$$\frac{\theta^{c-1} g(\theta)}{\int_0^{\infty} \theta^{c-1} g(\theta)\, d\theta}\ . \tag{2}$$

Notice that this is not the prior distribution,but a **partial rejection of the observation**.

Single Scale Parameter Cont.

In the location parameter models, rejection of the prior means a limiting posterior that is the same as would have been obtained from an improper uniform prior.the limiting posterior corresponds to an improper prior with density proportional to θ^{c-1}.
Andrade and O'Hagan (2006) also consider the case of multiple observations, and prove that when there are two or more outlying groups the group with largest total RV-credence dominates and all the other information sources are 'partially rejected'.

Location and Scale Parameters

x has density $\theta^{-1}f((x-\mu)/\theta)$.O'Hagan and Andrade (2011) deal with this model, but only for a specific form of prior distribution: μ given θ has the form $\theta^{-1}g(\mu/\theta)$, while the prior density for θ is $h(\theta)$. All three densities have regularly varying right tails, such that the RV-credences of f, g and h are c, c' and c'', respectively. They prove that as $x \to \infty$ the following three forms of resolution are possible.

1. If $c < \min(c', c'')$, then subject to additional regularity conditions on f the observation is 'partially rejected' and the posterior joint distribution of μ and θ is in the limit proportional to $\theta^{c-1}g(\mu/\theta)h(\theta)$.
2. If $c' < \min(c, c'')$, then subject to additional regularity conditions on g the prior information on μ is 'partially rejected' and the posterior joint distribution of μ and θ is in the limit proportional to $\theta^{c'-1}f((x-\mu)/\theta)h(\theta)$.
3. If $c'' < \min(c, c')$, then the prior information on θ is 'partially rejected' and the posterior joint distribution of μ and θ is in the limit proportional to $\theta^{-c''}f((x-\mu)/\theta)g(\mu/\theta)$.

Log-Regularly varying distributions for Location-Scale parameters

Desagné (2015) put forward the log-Pareto-tailed symmetric distributions that belongs to the log-regularly varying distributions. These have a center distribution that is for example Normal-g, but tails which are "super-heavy tailed"

$$f(z|\theta,\omega,\beta) \propto g(z|\theta) I_{[-\omega,\omega]} + g(\omega|\theta)\frac{\omega}{|z|}\left(\frac{log(\omega)}{log(|z|)}\right)^{\beta} I_{(\omega,\infty)}(|z|),$$

where $\omega > 1$, $\beta > 1$. Tail is governed by β.
Desagné assumes a reference prior and a center distribution with log-Pareto tails, and shows complete rejection to outliers.

Other kinds of parameters

Suppose that the observation x follows an exponential family distribution with canonical parameter θ with an arbitrary prior distribution for θ, and consider the posterior expectation of a function $m(\theta)$. Extending a result of Meeden and Isaacson (1977), Pericchi, Sansó and Smith (1993) show that if $m(\theta)$ is bounded for large θ by a power of θ then, subject to some regularity conditions, the posterior expectation of $m(\theta)$ tends to $m(\tilde{\theta})$ as $x \to \infty$, where $\tilde{\theta}$ is the posterior mode. They also show that $\tilde{\theta}$ may be found by solving a simple equation. It is now possible to explore cases under which the observation (or the prior distribution) is asymptotically rejected by studying the behaviour of $\tilde{\theta}$.

Expoential Familiy Parameters

As an example, Pericchi et al (1993) consider a Poisson likelihood, $f(x|\theta) \propto \exp(\theta x - e^{\theta})$. They establish that: (i) if the prior is normal, the posterior mean of the mean parameter e^{θ} diverges from the observation, and thus the prior has unbounded influence; (ii) if the prior is a t distribution the posterior mean approaches x, and so the prior is discarded; (iii) if the prior is logistic with σ^2 then $E[e^{\theta}|x]$ behaves like $x - \frac{\pi}{\sigma\sqrt{3}}$, reflecting a situation of bounded influence.

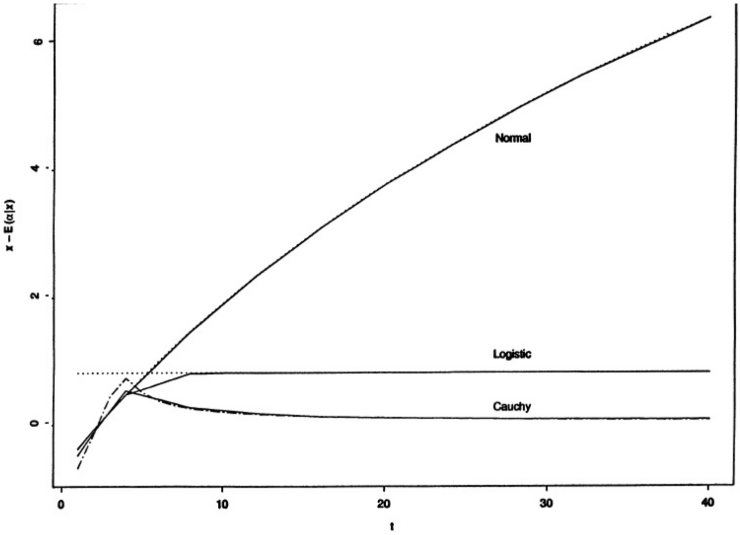


Figure 1. Exact (Solid Line) and Approximate (Dotted Line) Values of Deviations of Posterior From Sample Mean for Three Different Priors.

Figure: Difference between Observation and Posterior Expectation as the conflict grows in a Poisson Likelihood.

Discussion: What if?

This theory is one essential component in the specification of the prior and the likelihood. When eliciting such distributions, the practitioner's substantive knowledge and past experience may suggest center and spread of the distribution to be approximated, but it is much more difficult to elicit meaningful beliefs about tails. In this context, the theory of conflict resolution is a powerful tool for determining the relative weights of tails. It is quite natural to ask the practitioner, for example, "What if the next observation is in conflict with prior expectations, would you believe the data (the prior is wrong) or the prior (the data is an outlier) or both (I would not decide yet, but will wait until more information is gathered)?" The three different answers for the "What if" question immediately settle the question of tail characteristics of likelihood and prior. Heavy-tailed distributions can be used precisely to achieve whatever resolutions of conflicts are judged to be appropriate. For instance, if it is felt that when observations conflict with the prior information the prior should be rejected, then this can be achieved by a suitably heavy-tailed prior distribution.

What if? Cont.

Equally, if it is felt that extreme data should be discounted as outliers, then this can be achieved with appropriate heavy-tailed distributions for the data. And if the judgement is that a small number of observations conflicting with the prior might be discounted but that a larger number should lead to rejection of the prior then this, too, can be 'built-in' through careful choices of tail weights.

The existing theory gives us clear guidance on how to model prior and likelihood tails in order to obtain the desired behavior for a "what if" question as simple as the one above. For more complex and realistic models, the Theory of Conflict Resolution in Bayesian Statistics is still lagging behind.

Main Gaps in the Theory

- *Hierarchical models*. With three or more layers of hierarchy, and even with known variances throughout, we do not know how the posterior behaves when all layers have heavy-tailed models.
- *Unknown variances*. In models with a single location parameter and a single scale parameter, we do not know how the posterior behaves when there are multiple outlying observations. We do not even know what happens with a single observation when the joint prior distribution does not fit the structure assumed by O'Hagan and Andrade (2011), for instance when the parameters have independent priors. There is no theory at all for models with two or more scale parameters, such as routinely arise in hierarchical modelling.

Main Gaps Cont.

- *More general models*. An important class of models in which heavy-tailed models have been used but for which there is almost no theory is the linear and generalized linear models. Research is needed here and in time-series models. Although there has been some work on observations with exponential family distributions, the whole area of models that do not fit the structure of location and/or scale parameters is more or less unexplored.
- *Other questions*. The more abstract question of what constitutes a source of information demands careful study. Observations with independent t distributions behave like separate information sources but multivariate t distributions behave like a single source. It is not just a matter of independence, because in the bivariate priors of (1) the parameters are generally not independent and yet it seems that each of the three components represents a separate source. which can be separately rejected.

Enthusiastic VS Skeptical Priors in Clinical Trials: Spiegelhalter et al. Robustified Fuquene et. al. (2013)

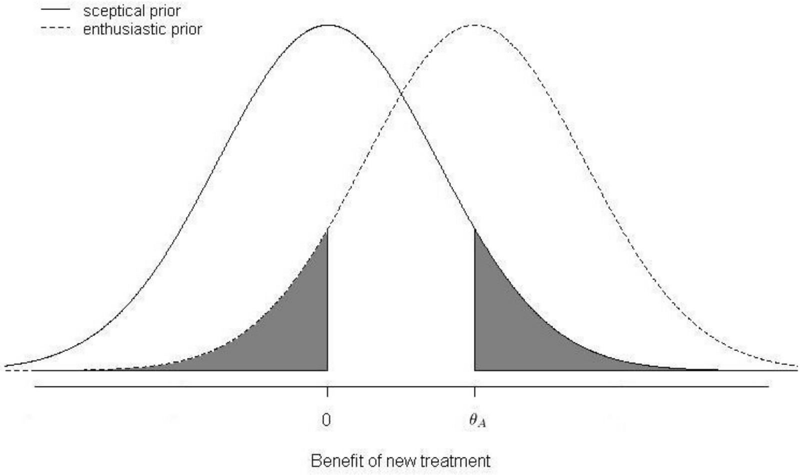


Figure: Robust Priors Non Dogmatic.

References

-Andrade, J.A.A. and O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Analysis* **1**, 169–188. Available electronically at: http://ba.stat.cmu.edu/vol01is01.php.
-Angers, J-F. (2000). P-credence and outliers. *Metron* **58**, 81–108.
-Angers, J-F. and Berger, J.O. (1991). Robust hierarchical Bayes estimation of exchangeable means. *Canadian Journal of Statistics* **19**, 39–56.
-Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd ed. Wiley: Chichester.
-Carlin, B.P. and Polson, N.G. (1991). Inference for nonconjugate Bayesian models using the Gibbs sampler. *Canadian Journal of Statistics* **19**, 399–405.

-Choy, S.T.B. and Smith, A.F.M. (1997). On robust analysis of a normal location parameter. *Journal of the Royal Statistical Society B* **59**, 463–474.
-Choy, S.T.B. and Walker, S.G. (2003). The extended exponential power distribution and Bayesian robustness. *Statistics and Probability Letters* **65**, 227–232.
-Dawid, A.P. (193). Posterior expectations for large observations. *Biometrika* **60**, 664–667.
-de Finetti, B. (1961). The Bayesian approach to the rejection of outliers. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I pp. 199-210. Univ. California Press: Berkeley, California.
-Desgagné, A. and Angers, J-F. (2007) Conflicting information and location parameter inference, *Metron* **65**, 67–97.

-Desgagné, A. (2013). Full Robustness in Bayesian Modelling of a Scale Parameter. *Bayesian Analysis*, **8**, 187-220. MR3036259
-Desgagné, A. (2015). Robustness to Outliers in Location-Scale Parameter Model Using Log-Regularly Varying Distributions. *Annals of Statistics*. (Accepted).
-Fan, T.H. and Berger, J.O. (1992). Behaviour of the posterior distribution and inferences for a normal mean with t prior distributions. *Statistics & Decisions* **10**, 99–120.
-Fernández, C., Osiewalski, J. and Steel, M.F.J. (1995). Modelling and inference with v-spherical distributions. *Journal of The American Statistical Association* **90**, 1331–1340.
-Finegold M. and Drton M. (2014) Robust Bayesian Graphical Modeling Using Dirichlet t-Distributions. *Bayesian Analysis* (2014) **9**, Number 3, 521-550.

-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**, 817–846. Available electronically at: http://ba.stat.cmu.edu/vol04is04.php.
-Goldstein, M. (1983). Outlier resistant distributions: Where does the probability go? *Journal of the Royal Statistical Society B* **45**, 355–357.
-Haro-López, R.A. and Smith, A.F.M. (1999). On robust Bayesian analysis for location and scale parameters. *Journal of Multivariate Analysis* **70**, 30–56.
-Hill, B.M. (1975). On coherence, inadmissibility and inference about many parameters in the theory of least squares. In *Studies in Bayesian Econometrics and Statistics: In Honor of Leonard J. Savage* (S.E. Fienberg and A. Zellner, eds.), 555–584. North-Holland.
-Le, H. and O'Hagan, A. (1998). A class of bivariate heavy-tailed distributions. *Sankhya* **B 60**, 82–100.

-Lindley, D.V. (1968). The choice of variables in multiple regression (with discussion). *Journal of the Royal Statistical Society B* **30**, 31–66.
-Meeden, G. and Isaacson, D. (1977). Approximate behavior of the posterior distribution for a large observation. *Annals of Statistics* **5**, 899–908.
-Meinhold, R.J. and Singpurwalla, N.D. (1989). Robustification of Kalman filter models. *Journal of The American Statistical Association* **84**, 479–486.
-Mitchell, A.F.S. (1994). A note on posterior moments for a normal mean with double-exponential prior. *Journal of the Royal Statistical Society B* **56**, 605–610.
-O'Hagan, A. (1979). On outlier rejection phenomena in Bayes inference. *Journal of the Royal Statistical Society B* **41**, 358–367.

-O'Hagan, A. (1981). A moment of indecision. *Biometrika* **68**, 329–330.
-O'Hagan, A., (1988). Modelling with heavy tails. In *Bayesian Statistics 3* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.), 345–359. Oxford: Clarendon Press.
-O'Hagan, A. (1990). Outliers and credence for location parameter inference. *Journal of the American Statistical Association* **85**, 172–176.
-O'Hagan, A. and Andrade, J.A.A. (2011). Bayesian robustness modelling of location and scale parameters. *Scandinavian Journal of Statistics* (in press).
-O'Hagan, A. and Le, H. (1994). Conflicting information and a class of bivariate heavy-tailed distributions. In *Aspects of Uncertainty: a Tribute to D. V. Lindley* (P.R. Freeman and A.F.M. Smith, eds.), 311–327. New York: Wiley.

-O'Hagan A. and Pericchi L. (2012) "Bayesian heavy-tailed models and conflict resolution: a review".Brazilian Journal of Probability and Statistics, Vol. 26, No. 4, 372-401.
-Pérez M.E., Pericchi L.R. and Ramirez I. (2015) The Scaled Beta2 Distribution as a Robust Prior for Scales. (submitted)
-Pérez M.E. and Pericchi L.R. (2015) Limiting the Shrinkage for the Exceptional by Objective Robust Bayesian Analysis. (in revision)
-Pericchi, L.R. and Sansó, B. (1995). A note on Bounded Influence in Bayesian Analysis. *Biometrika* **82**, 223–225
-Pericchi, L.R. and Smith, A.F.M. (1992). Exact and Approximate Posterior Moments for a Normal Location Parameter. *Journal of the Royal Statistical Society B* **54**, 793–804.

-Pericchi, L.R., Sansó, B. and Smith, A.F.M. (1993). Posterior cumulant relationships in Bayesian inference involving the exponential family. *Journal of the American Statistical Association* **88**, 1419–1426.
-Wakefield, J.C., Smith, A.F.M., Racine-Poon, A. and Gelfand, A.E. (1994). Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics* **43**, 201–221.
-West, M. (1981). Robust sequential approximate Bayesian estimation. *Journal of the Royal Statistical Society B* **43**, 157–166.
-West, M. (1984). Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society B* **46**, 431–439.