Bayesian Analysis (2016) **11**, Number 4, pp. 1107–1136

New Classes of Priors Based on Stochastic Orders and Distortion Functions

J. Pablo Arias-Nicolás*, Fabrizio Ruggeri†, and Alfonso Suárez-Llorens‡

Abstract. In the context of robust Bayesian analysis, we introduce a new class of prior distributions based on stochastic orders and distortion functions. We provide the new definition, its interpretation and the main properties and we also study the relationship with other classical classes of prior beliefs. We also consider Kolmogorov and Kantorovich metrics to measure the uncertainty induced by such a class, as well as its effect on the set of corresponding Bayes actions. Finally, we conclude the paper with some numerical examples.

MSC 2010 subject classifications: Primary 62F15; secondary 60E15.

Keywords: robust Bayesian analysis, Bayesian sensitivity, class of priors, stochastic orders, distortion functions.

1 Introduction

Robust Bayesian analysis, also called Bayesian sensitivity analysis, aims to quantify and interpret the uncertainty induced by the partial knowledge of one (or more) of the three elements in Bayesian analysis (prior, likelihood and loss). Studies mainly focus on computing the range of some quantities of interest when the prior distribution π varies in a class Γ. The use of a class of priors somehow overcomes the common criticism about the choice of a unique prior on the grounds of arbitrariness and bias. On the other hand, changes in the likelihood have been rarely addressed for several reasons, including the mathematical complexity and the fact that the choice of the statistical model is somehow considered more *objective*. The limited interest for classes of loss functions relies on the fact that in inference problems just the quadratic or absolute losses are, in general, considered and no elicitation method about preferences is used. Nonetheless, not long ago, researchers' attention turned to classes of loss functions (i.e., $L \in \mathcal{L}$), being mostly interested in the changes of posterior expected losses and optimal actions. A thorough review of the robust Bayesian approach can be found in Ríos Insua and Ruggeri (2000).

Regarding prior distribution, Basu (1994) nicely summarized that any elicitation process, leading to a prior π, is to some extent arbitrary and, therefore, any prior in a "neighborhood" of the elicited prior would also be a reasonable representation of the prior beliefs. Consequently, neighborhoods have been specified in literature in many different ways depending on their use. For instance, neighborhoods have been defined as parametric families, contamination classes, density bands, densities with a few

*Universidad de Extremadura, Caceres, Spain, jparias@unex.es
†CNR IMATI, Milano, Italy, fabrizio@mi.imati.cnr.it
‡Universidad de Cádiz, Spain, alfonso.suarez@uca.es

 DOI: 10.1214/15-BA984

specified percentiles, distributions bands, etc. The "neighborhood" classes entertained in literature are not always topological neighborhoods of a prior distribution π_0, i.e., neighborhoods determining a topology on the space of all the probability measures. The most widely used example is provided by the ϵ-contamination class, defined as:

$$\Gamma_\epsilon = \{\pi : \pi = (1 - \epsilon)\pi_0 + \epsilon Q,\, Q \in \mathcal{Q}\},$$

where π is given by a mixture and $\mathcal{Q}$ is called the class of contaminations; see, e.g., Berger and Berliner (1986), Moreno and Cano (1991), Betrò et al. (1994), among others. A variety of proper topological neighborhoods has been considered, such as the class based on the concentration function; see Fortini and Ruggeri (2000) and references therein. Another relevant class, important also for the current paper, is the one based on distribution bands proposed by Basu and DasGupta (1995). The classes based on concentration functions and distribution bands have connections with the proposed class of distortion functions, as thoroughly discussed in the next sections. For a detailed illustration of classes of priors, we refer to Berger (1985), Berger (1994) and Moreno (2000).

With respect to the loss functions, extensive surveys of sensitivity analyses have been presented in Martín et al. (1998), Dey et al. (1998), Makov (1994) and more recently in Arias-Nicolás et al. (2006) and Arias-Nicolás et al. (2009), among others.

The major contribution of the current paper consists in the introduction of new classes of priors (especially) and loss functions, based on notions from other fields in Statistics (e.g., stochastic ordering). In particular, we consider the distorted band class of priors (stemming from stochastic ordering), and a subclass of convex loss functions, the submodular ones, which contains the most widely used loss functions: quadratic, absolute value, quantile and LINEX loss functions, among others. The proposed class fulfills all the requirements that Berger (1994) discussed about the choice of a class. First of all, elicitation of the class should be easy, as well as its interpretation, and its size should reflect the prior uncertainty, with no exclusion of reasonable priors and inclusion of unreasonable ones (e.g., discrete ones in many problems). Furthermore, the class should be extendable to high dimensions and allow for incorporation of features like symmetry, unimodality, independence, etc. Finally, computations of sensitivity measures should be as easy as possible, possibly looking for the extremal distributions generating the class since such measures are usually computed over such set of distributions.

The concept of distorted distributions has been used in many applied fields to represent a change in the probability measure, see, e.g., Quiggin (1982), Yaari (1987) and Schmeidler (1989) in the Rank Dependent Expected Utility model, Denneberg (1990), Wang (1995) and Wang (1996) in actuarial science and Goovaerts and Laeven (2008) about insurance premiums and risk measures. In this context, it is assumed that each decision-maker replaces the probability $P[\pi \leq t]$, which models a particular event of interest, by a distorted probability $h(P[\pi \leq t])$, for a particular distortion function h. The choice of different distortion functions leads to different ways of measuring the uncertainty. For example, a convex (concave) distortion function will give more (less) weight to larger events.

We also find in the literature many papers connecting distorted distributions and stochastic orderings; see Shaked and Shanthikumar (2007) for a general survey on stochastic orderings. In a formal way, the theory of stochastic orders deals with the comparison of two random behaviors. Let X and Y be two random variables, then it is said that X is smaller than Y, denoted by $X \leq_* Y$, if some specific conditions are satisfied, where the meaning of $\leq_*$ depends on what we try to compare. For example, it is a very common practice to compare X and Y in terms of variability or magnitude of their values, e.g., the magnitude of the expectations. The connection between distorted distributions and stochastic orderings arises since some orderings are consistent with some measures associated with the distorted distributions of the underlying variables. Important contributions in this field are Chew et al. (1987) and Wang and Young (1998) and, for recent works on this topic, see Shaked et al. (2010) and Sordo and Suárez-Llorens (2011).

The major reasons for the proposed class of priors are the relative easiness in specifying it through stochastic ordering and the class of posterior distributions being again a class of distorted distributions. The latter property is very uncommon among the classes proposed in literature. In the paper we will be interested in the distance among the posterior probability measures, using different metrics. This is a different approach with respect to (w.r.t.) most of the works in literature, where the most common measures of interest are the posterior mean, the posterior variance and the posterior expected loss, see, e.g., Berger and Berliner (1986), Sivaganesan (1989), Sivaganesan and Berger (1989), Moreno and Cano (1991), and Ríos Insua and Martín (1994), among others. If $\psi(\pi)$ is the posterior functional of interest (e.g., the posterior mean), global robustness studies are concerned about finding $(\underline{\psi}, \overline{\psi})$ where

$$\underline{\psi} = \inf_{\pi \in \Gamma} \psi(\pi), \text{ and } \overline{\psi} = \sup_{\pi \in \Gamma} \psi(\pi).$$

The difference $\overline{\psi} - \underline{\psi}$, called *range*, is the most important sensitivity measure in global robustness. The range has a simple interpretation: when Γ reasonably reflects the uncertainty in the prior, a "small" range indicates robustness w.r.t. the changes of the latter. On the other hand, a "large" range is an indication that there is lack of robustness w.r.t. the prior and further elicitation, data collection, or analysis is necessary. As an example, when considering a neighborhood Γ of a prior π_0, then the sensitivity is related to deviations from $\psi_0 = \psi(\pi_0)$. If the range is "small", then one may use, for example, ψ_0, as the Bayesian answer being aware that it is robust w.r.t. possible misspecifications of the prior. An analogous measure will be defined in our context.

The class of distorted bands will be introduced in Section 2, along with its properties, whereas its connection with the class of concentration functions will be discussed in Section 3. In Section 4, the Kolmogorov and Kantorovich metrics are introduced. The class of submodular loss functions is introduced in Section 5. Sampling from the class of posterior distributions is presented in Section 6 and applied to numerical examples in Section 7. Concluding remarks and pointers for future researches are finally presented in Section 8.

2 A new distribution prior class: the distorted band

The following notation will be used in the paper. Given a random variable X with distribution function F, we define the quantile function as $F_X^{-1}(p) = \inf\{x : F_X(x) \geq p\}$, for all real values $p \in (0,1)$. The symbol $=_{st}$ means equality in law. For every random variable Z and an event A, let $[Z|A]$ denote a random variable whose distribution is the conditional distribution of Z given A.

We start by recalling the definition of the stochastic orders that we will consider in this paper. Let X and Y be two random variables with distribution functions F_X and F_Y, respectively, such that

$$F_X(t) \geq F_Y(t), \quad \forall t \in \mathbf{R}. \tag{1}$$

Then X is said to be *smaller than* Y *in the usual stochastic order* (denoted by $X \leq_{\text{st}} Y$). Roughly speaking, (1) says that X is less likely than Y to take on large values. We would like to point out that the usual stochastic order has an interesting characterization in terms of the expectations of increasing transformations, i.e., $X \leq_{\text{st}} Y$ if and only if

$$E[g(X)] \leq E[g(Y)] \tag{2}$$

holds for all increasing functions g for which the expectations exist.

Let X and Y be absolutely continuous [discrete] random variables with distribution functions F_X and F_Y and densities [discrete densities] f_X and f_Y, respectively, such that

$$\frac{f_Y(t)}{f_X(t)} \quad \text{increases over the union of the supports of } X \text{ and } Y, \tag{3}$$

(here $a/0$ is taken to be equal to ∞ whenever $a > 0$). Then X is said to be *smaller than* Y *in the likelihood ratio order* (denoted by $X \leq_{\text{lr}} Y$). It is well known that

$$X \leq_{lr} Y \Rightarrow X \leq_{st} Y. \tag{4}$$

For more details about stochastic and likelihood orders, see Müller and Stoyan (2002) and Shaked and Shanthikumar (2007).

After introducing stochastic orders, we also recall the standard Bayesian decision theoretic framework for statistical problems, see French and Ríos Insua (2000) among others. Let X be an observation from a distribution P_θ with density $p_\theta(x)$, π a prior belief, over the set of states Θ, belonging to a class of distributions Γ, L a loss function belonging to a class $\mathcal{L}$ and $\mathcal{A}$ a set of alternatives. We will denote by $\pi(\theta)$ the density of the prior distribution.

Let π_x denote the posterior distribution and let $\pi_x(\theta)$ be its density function, when x is observed, belonging to Γ_x, the class of all posterior distributions; $m_\pi(x)$ denotes the marginal density, $l(\theta)$ the likelihood function for a fixed experiment x and $\rho(\pi, L, a)$ the posterior expected loss of a, i.e.,

$$\rho(\pi, L, a) = \frac{\int L(a,\theta) l(\theta) \pi(\theta) d\theta}{m_\pi(x)} = E^{\pi_x}[L(a,\theta)].$$

On the other hand, for any $(L, \pi) \in \mathcal{L} \times \Gamma$, a Bayes action corresponding to (L, π), i.e., an action minimizing $\rho(\pi, L, a)$ in $\mathcal{A}$, will be denoted by $a^*_{(L,\pi)}$

$$\rho(\pi, L, a^*_{(L,\pi)}) = \min_{a \in \mathcal{A}} \rho(\pi, L, a).$$

Finally, we will denote by $B(L, \pi)$ the set of all Bayes actions associated with the pair (L, π) and by $\underline{a}^*_{(L,\pi)}$ and $\overline{a}^*_{(L,\pi)}$ the infimum and supremum, respectively, of all Bayes actions.

For our purpose of connecting stochastic orders with the Bayesian decision theoretic framework, we will finally recall the concept of distortion function. Let X be a random variable with distribution function F_X and let h be a distortion function, i.e., a non-decreasing continuous function $h : [0, 1] \to [0, 1]$ such that $h(0) = 0$ and $h(1) = 1$. For each distortion h, the transformation of the distribution function of X given by

$$F_h(x) = h \circ F(x) = h\left[F(x)\right], \tag{5}$$

represents a perturbation of the accumulated probability in order to measure the uncertainty about the underlying distribution function F. It is worth mentioning that, under such assumptions, $F_h(x)$ is also a distribution function for a particular random variable, denoted by X_h, that we will call the distorted random variable. The choice of h depends on the particular "flavor" of the distortion. For example, in insurance pricing and in financial risk management, a distortion typically represents a change in the probability measure. Under some particular hypothesis, it is assumed that each decision-maker has a distortion function h and that he values some characteristics of X as its distorted characteristics, see Section 2.6 in Denuit et al. (2005), for a review.

If we focus on the perspective of robust Bayesian analysis, it seems natural to incorporate the uncertainty about specification of a prior belief by considering a distorted one. It is then assumed that each decision-maker has a distortion function h and he is able to represent the changes of prior beliefs by reasonable distorted prior beliefs. As we have mentioned, a distribution function can be distorted according to different criteria. In this paper we are just going to consider convex and concave distortion functions because they represent satisfactorily a change in magnitude and variability of the underlying prior belief and have desirable properties when we compare the original variable with the distorted one.

At this point, it is natural to wonder about the relationship between a prior distribution belief and its distorted version. Given a prior belief π and a distortion function h we will denote by π_h its distorted version, i.e., a "distorted prior belief" having distribution function $F_{\pi_h}(\theta) = h \circ F_\pi(\theta)$. Next we present a lemma which plays an important role in the definition of a new neighborhood band.

Lemma 1. *Let π be a specific prior belief with distribution function F_π (absolutely continuous or discrete) and let h be a convex (concave) distortion function in $[0, 1]$. Then $\pi \leq_{lr} (\geq_{lr})\pi_h$.*

Proof. Let h be a convex distortion function. First, we will assume that F_π is absolutely continuous. Due to (5), the distorted distribution function associated with F_π is given by $F_{\pi_h}(\theta) = h(F_\pi(\theta))$ which is, trivially, a continuous function. It is well known that a convex function has left and right derivatives at all points, and these are monotonically non-decreasing, and there are only countably many points where they do not coincide. Therefore, the distribution function $F_{\pi_h}(\theta)$ is also absolutely continuous and its density function is given by $\pi_h(\theta) = h'(F_\pi(\theta))\pi(\theta)$ almost everywhere, where h' represents the derivative of h. The proof follows easily just considering the ratio defined in (3) and taking in account that $h'(F_\pi(\theta))$ is increasing almost everywhere. If we assume now that F_π is discrete, the result follows from the fact that the ratio of discrete probabilities can be expressed as

$$\frac{\pi_h(\theta_i)}{\pi(\theta_i)} = \frac{h(F_\pi(\theta_i)) - h(F_\pi(\theta_{i-1}))}{F_\pi(\theta_i) - F_\pi(\theta_{i-1})},$$

and taking in account that the function

$$R(a,b) = \frac{h(b) - h(a)}{b - a}$$

is symmetric in a and b and monotonically non-decreasing in b, for a fixed (or vice versa). For a concave distortion function the proof holds using a similar argument. □

Remark 2. *Note that given two random variables X and Y and assuming continuous and strictly increasing distributions functions F_X and F_Y, respectively, it is evident that the function $h_{XY}(t) = F_Y(F_X^{-1}(t))$ is a distortion function that maps the distribution function of X to the corresponding of Y. We would like to emphasize that Lehmann and Rojo (1992) proved that, under the mentioned regularity conditions, the likelihood ratio order between X and Y is equivalent to checking if the function $h_{XY}(t)$ is convex or, analogously, if $h_{YX}(t)$ is concave. Therefore, assuming continuous and strictly increasing distributions functions, the existence of a convex or a concave distortion that maps a distribution function to another one is a necessary and sufficient condition for the likelihood ratio order. In this paper, we are mainly interested in using some particular family of distortion functions in order to measure the uncertainty.*

Let us assume now that the decision-maker is able to represent the changes of a prior belief, π, by a concave distortion function, h_1, and a convex distortion function, h_2. This fact, jointly with Lemma 1, leads him to two distorted distributions π_{h_1} and π_{h_2} such that $\pi_{h_1} \leq_{lr} \pi \leq_{lr} \pi_{h_2}$. Clearly inspired by this fact, we present the following neighborhood band for π.

Definition 3. *Let π be a specific prior belief. We will define the distorted band $\Gamma_{h_1,h_2,\pi}$ associated with π based on h_1 and h_2, a concave distortion function and a convex distortion function, respectively (distorted band, for short), as*

$$\Gamma_{h_1,h_2,\pi} = \{\pi' : \pi_{h_1} \leq_{lr} \pi' \leq_{lr} \pi_{h_2}\}.$$

As a consequence of Lemma 1, it is evident that $\pi \in \Gamma_{h_1,h_2,\pi}$. Therefore, the distorted band can be seen as a particular "neighborhood" band of π, where the lower and upper bound distributions are given by the distorted distributions. It is also clear from the definition that uncertainty could be introduced just through an upper (lower) bound by considering h_1 (h_2) the identity function. In order to give a meaningful understanding of the distorted band, we provide the following interpretations.

First, from (4) the distorted distribution is a subclass of the well known band class:

$$\begin{aligned}\Gamma_{h_1,h_2,\pi} &\subseteq \{\pi' : \pi_{h_1} \leq_{st} \pi' \leq_{st} \pi_{h_2}\}, \\ &= \left\{\pi' : F_{\pi_{h_1}}(\theta) \geq F_{\pi'}(\theta) \geq F_{\pi_{h_2}}(\theta), \forall \theta \in \Theta\right\}. \end{aligned} \tag{6}$$

It is worth mentioning that, in the classical notation for the band class, the distribution function of π_{h_1} is called "the upper bound", F_U, and the distribution function of π_{h_2} is called "the lower bound", F_L.

On the other hand, using expression (1.C.6) in Shaked and Shanthikumar (2007), it is also remarkable that the distorted band has a nice interpretation in terms of prior probability sets:

$$\Gamma_{h_1,h_2,\pi} = \{\pi' : \pi_{h_1}(\cdot|A) \leq_{st} \pi'(\cdot|A) \leq_{st} \pi_{h_2}(\cdot|A)\},$$

for all measurable $A \subseteq \Theta$.

It is worth mentioning that the likelihood ratio order does not apply, in general, when comparing two priors π'_1 and π'_2 in $\Gamma_{h_1,h_2,\pi}$. Each of them is just ordered w.r.t. π_{h_1} and π_{h_2}.

Remark 4. *The class $\Gamma_{h_1,h_2,\pi}$ contains all the priors $\pi_\epsilon = (1-\epsilon)\pi_1 + \epsilon\pi_2$, for any pair of priors π_1 and π_2 in $\Gamma_{h_1,h_2,\pi}$, obtained as a mixture of π_1 and π_2, for any $0 \leq \epsilon \leq 1$. In particular, it contains the mixture between the underlying prior π and any other prior in the band. The proof that $\pi_{h_1} \leq_{lr} \pi_\epsilon$ follows from*

$$\frac{\pi_\epsilon(\theta)}{\pi_{h_1}(\theta)} = (1-\epsilon)\frac{\pi_2(\theta)}{\pi_{h_1}(\theta)} + \epsilon\frac{\pi_1(\theta)}{\pi_{h_1}(\theta)},$$

which is an increasing function of θ. Similarly, it is possible to prove that $\pi_\epsilon \leq_{lr} \pi_{h_2}$.

Since Definition 3 is based on h_1 and h_2, we can provide many possible bands just by considering different concave and convex distortion functions. Of course, the choice of those functions cannot be arbitrary and should represent the uncertainty about the prior belief in each problem. Next, we present some examples. A classical way of inducing a distortion is given by the power functions

$$h_1(x) = 1-(1-x)^\alpha \text{ and } h_2(x) = x^\alpha, \quad \forall \alpha > 1. \tag{7}$$

Note that if we take $\alpha = n \in \mathbf{N}$ in (7), then $F_{\pi_{h_1}}(\theta) = 1-(1-F_\pi(\theta))^n$ and $F_{\pi_{h_2}}(\theta) = (F_\pi(\theta))^n$ which correspond to the distribution functions of the minimum and the maximum, respectively, of an i.i.d. random sample of size n from the baseline prior distribution π.

Another classical family of concave and convex distortion functions is given by

$$h_1(x) = \min\{\frac{x}{\alpha}, 1\} \text{ and } h_2(x) = \max\{\frac{x-\alpha}{1-\alpha}, 0\}, \quad 0 < \alpha < 1. \tag{8}$$

In this case, h_1 and h_2 represent the truncated variables $\pi_{h_1} =_{st} \pi(\cdot|A_1)$ and $\pi_{h_2} =_{st} \pi(\cdot|A_2)$ where $A_1 = (-\infty, F_\pi^{-1}(\alpha)]$ and $A_2 = (F_\pi^{-1}(\alpha), \infty)$.

Finally, an interesting family of distortion functions is given by the concept of skewed distributions. If π represents a prior belief with an absolutely continuous symmetric distribution around 0, the skewed distribution associated with π, denoted by skew-π, is a continuous probability distribution that generalizes π to allow for non-zero skewness. Given π, the probability density function of the skew-π with parameter α is given by

$$\pi_\alpha(\theta) = 2\pi(\theta)F_\pi(\alpha\theta).$$

The distribution is right skewed if $\alpha > 0$ and left skewed if $\alpha < 0$. A simple computation shows that $\pi \leq_{lr} \pi_\alpha$ for all $\alpha > 0$ and $\pi_\alpha \leq_{lr} \pi$ for all $\alpha < 0$; see Azzalini (1985) for a detailed explanation of this class of distributions and Ferreira and Steel (2006) for more details about representation of skewed distributions. From Remark 2, a straightforward computation shows that the distortion function given by

$$h_{\pi\pi_\alpha}(x) = \int_{-\infty}^{F_\pi^{-1}(x)} 2\pi(\theta)F_\pi(\alpha\theta)d\theta,$$

maps the distribution function of the prior π to the corresponding one of the skewed version π_α and satisfies

$$h'_{\pi\pi_\alpha}(x) = 2F_\pi(\alpha F_\pi^{-1}(x)). \tag{9}$$

Since both distributions F_π and F_π^{-1} are increasing and differentiable, it easy to check that $h'_{\pi\pi_\alpha}(x)$ is increasing for all $\alpha > 0$ and decreasing for all $\alpha < 0$, which implies that $h_{\pi\pi_\alpha}(x)$ is convex or concave, respectively. In order to distinguish between concave and convex functions, we consider the family of functions given by

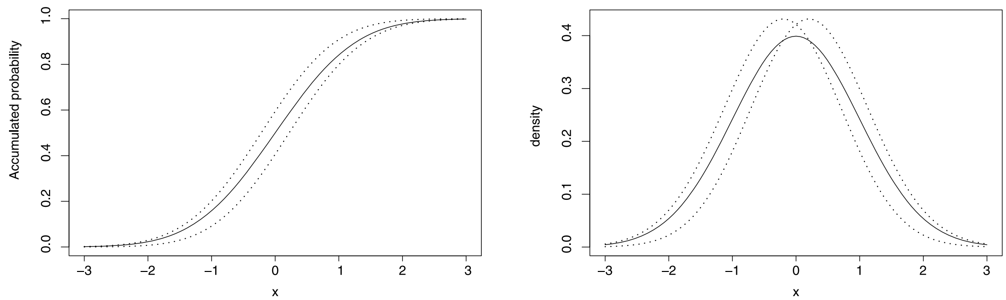
$$h_1(x) = \int_{-\infty}^{F_\pi^{-1}(x)} 2\pi(\theta)F_\pi(-\beta\theta)d\theta \text{ and } h_2(x) = \int_{-\infty}^{F_\pi^{-1}(x)} 2\pi(\theta)F_\pi(\beta\theta)d\theta, \tag{10}$$

for all $\beta \geq 0$.

Now we present the following example to clarify the previous ideas.

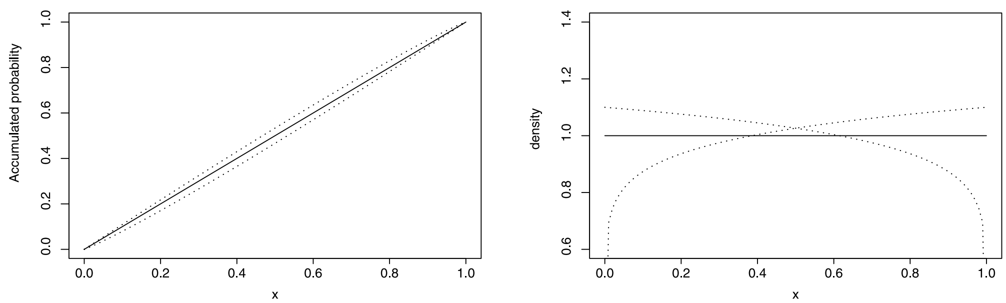
Example 5. *Suppose that the prior belief is given by $\pi \sim N(0,1)$, and consider the distortion functions defined in (7) with $\alpha = 1.3$. Then the distorted distributions are given by $F_{\pi_{h_1}}(\theta) = 1-(1-\Phi_Z(\theta))^{1.3}$ and $F_{\pi_{h_2}}(\theta) = (\Phi_Z(\theta))^{1.3}$, where Φ_Z is the standard normal distribution function. $F_{\pi_{h_1}}$ and $F_{\pi_{h_2}}$ have been represented with a dotted line in Figure 1(a), where we can see how they differ from the baseline prior distribution function. We also represent in Figure 1(b) the densities of the distorted distributions compared with the baseline prior density. With a similar argument, if the prior belief is given by $\pi \sim U(0,1)$ and we take $\alpha = 1.1$, we can see in Figure 2 (a)–(b) the effect of the distortion functions in both distribution and density functions, respectively.*

We conclude this section showing that the distorted band for a prior belief leads to another distorted band for posterior distributions. As a direct consequence of the likelihood ratio order definition, see Spizzichino (2001), if two prior distributions are ordered in the $\leq_{lr}$ order sense, then the corresponding posterior distributions are also ordered in the same sense, i.e., given two prior distribution π_1 and π_2 such that $\pi_1 \leq_{lr} \pi_2$ then $\pi_{1x} \leq_{lr} \pi_{2x}$. Therefore, since posterior distributions inherit the likelihood ratio order, we present the following proposition whose proof is omitted.



(a) Distorted distributions. (b) Densities of the distorted distributions.

Figure 1: Distorted N(0,1), $\alpha = 1.3$.



(a) Distorted distributions. (b) Densities of the distorted distributions.

Figure 2: Distorted U(0,1), $\alpha = 1.1$.

Proposition 6. *Let π be a specific prior belief and let $\Gamma_{h_1,h_2,\pi}$ be a distorted band associated with π based on h_1 and h_2. Then for all $\pi' \in \Gamma_{h_1,h_2,\pi}$ we obtain that $\pi_{h_1,x} \leq_{lr} \pi'_x \leq_{lr} \pi_{h_2,x}$.*

Proposition 6 says that the posterior distributions of the lower and upper bound of the prior distortion band are also lower and upper bounds for the family of all posterior distributions, Γ_x, in the $\leq_{lr}$ order sense. Assuming the regularity conditions in Remark 2, Γ_x can be also interpreted as a distortion band of the posterior belief for some particular concave and convex functions.

3 Relation with concentration functions

In this section we show that the distorted band class is, under very general conditions, contained in the concentration function class proposed by Fortini and Ruggeri (1995). The result is somehow surprising since it relates classes which, apparently, are obtained in very different ways. Exploiting the properties of the concentration function class, the new class can be embedded in a neighborhood of a prior in a proper topological sense and, furthermore, upper (lower) bound on the supremum (infimum) of the posterior quantity of interest can be computed. Furthermore, all the priors in the distorted band class are obtained as a mixture of the extremal distributions in the concentration function class. As a relevant difference, the prior distorted band class leads to another distorted band class a posteriori, as shown in Proposition 6; the concentration function class does not, in general, lead to a similar class.

Cifarelli and Regazzini (1987) introduced their notion of concentration function (c.f.) as an extension of the Lorenz curve to compare two probability measures Π and Π_0 defined on the same measurable space $(\Theta, \mathcal{F})$. According to the Radon–Nikodym theorem, there is a unique partition $\{N, N^C\} \subset \mathcal{F}$ of Θ and a nonnegative function m on N^C such that

$$\Pi(E) = \int_{E \cap N^C} m(\theta)\Pi_0(d\theta) + \Pi_s(E \cap N),$$

$\forall E \in \mathcal{F}$, $\Pi_0(N) = 0$, $\Pi_s(N) = \Pi_s(\Theta)$, where Π_a and Π_s denote the absolutely continuous and the singular part of Π w.r.t. Π_0, respectively. Set $m(\theta) = \infty$ all over N and define $H(y) = \Pi_0(\{\theta \in \Theta : m(\theta) \leq y\})$, $c(x) = \inf\{y \in \Re : H(y) \geq x\}$. Finally, let $L(x) = \{\theta \in \Theta : m(\theta) \leq c(x)\}$ and $L^-(x) = \{\theta \in \Theta : m(\theta) < c(x)\}$.

Definition 7. *The function $\varphi : [0,1] \to [0,1]$ is said to be the concentration function of Π w.r.t. Π_0 if $\varphi(x) = \Pi(L^-(x)) + c(x)\{x - H(c(x)^-)\}$ for $x \in (0,1)$, $\varphi(0) = 0$ and $\varphi(1) = \Pi_a(\Theta)$.*

Here we assume all the measures are absolutely continuous w.r.t. the Lebesgue measure. As a consequence, $m(\theta)$ is the likelihood ratio given by the densities of the probability measures and $\varphi(x) = \Pi(L(x))$, having got rid of the singular component.

To favor a better understanding of the c.f., we provide a constructive way to obtain it on the real space. Consider the probability measures Π and Π_0 on the real space, with densities $\pi(\theta)$ and $\pi_0(\theta)$, respectively. In Definition 7, likelihood subsets were determined by x via $c(x)$ but now we can fix their levels and compute, at the same time, their probabilities under Π_0 and Π, which are x and $\varphi(x)$, respectively.

Let $m(\theta) = \frac{\pi(\theta)}{\pi_0(\theta)}$ the likelihood ratio and consider the likelihood subsets $L_y = \{\theta : m(\theta) \leq y\}$, for all $y > 0$. Consider the pairs $(x_y, \varphi(x_y))$ where $x_y = \Pi_0(L_y)$ and $\varphi(x_y) = \Pi(L_y)$ for all $y > 0$. Removing from now on the dependence on y, it follows from Definition 7 that the function $\varphi : [0,1] \to [0,1]$ is the c.f. of Π w.r.t. Π_0.

Observe that $\varphi(x)$ is a nondecreasing, continuous and convex function, such that $\varphi(x) \equiv 0 \Longrightarrow \Pi \perp \Pi_0$, and $\varphi(x) = x, \forall x \in [0,1] \Longleftrightarrow \Pi = \Pi_0$.

We conclude the illustration of the properties of the c.f. mentioning that, regarding the simplifying assumptions made earlier, singularity of Π w.r.t. Π_0 implies $\varphi(1) < 1$ whereas singularity of both measures w.r.t. Lebesgue measure (e.g., a mixture with a Dirac measure in the same point and, possibly, different weights) implies an interpolating term in the definition of c.f.

Consider a nondecreasing, continuous and convex distortion function $h(x)$, a baseline random variable X with distribution function $F(x)$ and density $f(x)$ and a distorted random variable X_h with distribution function $F_h(x) = h[F(x)]$ and density $f_h(x) = h'[F(x)]f(x)$. The likelihood ratio $m(x) = \frac{f_h(x)}{f(x)} = h'[F(x)]$ is increasing since $h'' > 0$ because of the convexity of the function h (and this is also a consequence of the $\leq_{lr}$ order). Therefore it is possible to consider likelihood subsets $L_z = (-\infty, z]$ with probability $F(z)$ and $F_h(z)$ under the two probability measures.

Take $x_z = F(z)$ and assume F invertible so that $z = F^{-1}(x_z)$. Consider $\varphi_h(x_z) = F_h(z) = h[F(z)] = h[F(F^{-1}(x_z))] = h(x_z)$. Dropping the dependence on z it follows that $\varphi_h(x) = h(x)$ is the c.f. of the probability measure Π (corresponding to the random variable X_h) w.r.t. the probability measure Π_0 (corresponding to the random variable X). Therefore, given a distorted measure, its distortion function can be interpreted as the c.f. of the distorted measure w.r.t. the baseline. Given a nondecreasing, continuous and convex function $h(x)$, there exists an infinity of probability measures (including the corresponding distorted measure) whose c.f. w.r.t. the baseline measure is given by $h(x)$. Further details on this fact can be found in Fortini and Ruggeri (1995) and the example at the end of the proof of the next theorem.

Fortini and Ruggeri (1995) proved that it is possible to obtain topological neighborhoods of a baseline measure Π_0 when considering all the probability measures whose c.f. is above a nondecreasing, continuous and convex function $h(x)$, which is a c.f. as well (for an infinity of probability measures). We will denote such set of probability measures by $\Psi_{\pi_0,h}$, and we consider the distorted band $\Gamma_{\pi_0,h} = \{\pi' : \pi_0 \leq_{lr} \pi' \leq_{lr} \pi_h\}$. The latter class is properly included in the former as shown by the following

Theorem 8. *The distorted band class $\Gamma_{\pi_0,h}$ is properly included in the concentration function class $\Psi_{\pi_0,h}$.*

Proof. Consider a prior $\pi' \in \Gamma_{\pi_0,h}$ with distribution function $F_{\pi'}(x)$, and let $F(x)$ be the distribution function corresponding to the prior π_0, then the likelihood ratio $m(x) = \frac{\pi'(x)}{\pi_0(x)}$ is increasing because of the likelihood order and the computation of the c.f. $\tilde{\varphi}$ is as before:

$$\tilde{\varphi}(x) = F_{\pi'}(F^{-1}(x)),$$

since we consider likelihood subsets $L_z = (-\infty, z]$. Assuming $F_{\pi'}(x)$ and $F(x)$ continuous and strictly increasing, then Remark 2 implies that $\tilde{\varphi}(x)$ is a distortion function since it is obtained by the combination of the distribution function $F_{\pi'}(x)$ with the inverse of the distribution function $F(x)$. We now prove that $\tilde{\varphi}(x) \geq \varphi_h(x)$ for $x \in [0,1]$. We have that

$$\tilde{\varphi}(x) \geq \varphi_h(x),\ \forall x \in [0,1] \Leftrightarrow \int_{-\infty}^{z} \left(1 - \frac{\pi_h(t)}{\pi'(t)}\right) \pi'(t)dt \geq 0,\ \forall z.$$

The previous condition follows from observing that both likelihood ratios are increasing and $x = F(z)$, $\tilde{\varphi}(x) = F_{\pi'}(z)$ and $\varphi_h(x) = F_h(z)$.

Suppose there exists z^* such that

$$\int_{-\infty}^{z^*} \left(1 - \frac{\pi_h(t)}{\pi'(t)}\right) \pi'(t)dt < 0. \tag{11}$$

Since $\frac{\pi_h(t)}{\pi'(t)}$ is an increasing function (because of the likelihood order in $\Gamma_{\pi_0,h}$) then it is not possible that $\frac{\pi_h(z^*)}{\pi'(z^*)} < 1$ since the integral in (11) would be positive. As a consequence, the same integral over the interval (z^*, ∞) would be negative as well, which is impossible since the integral over the real line should be equal to 0.

We have proved that each measure in the distorted band has a corresponding c.f. lying above $h(x)$, so that it belongs as well to the family $\Psi_{\pi_0,h}$, as proved in Fortini and Ruggeri (1995). The class $\Gamma_{\pi_0,h}$ is properly included in $\Psi_{\pi_0,h}$ as shown by the following example.

Consider a uniform distribution on $[0, 1]$ as a baseline prior π_0 and the function $h(x) = x^2$. The corresponding distorted distribution has density $\pi_h(x) = 2x$, whose ratio w.r.t. the uniform density is increasing. The distribution with density $\pi^*(x) = 2(1 - x)$ has the same c.f. $\varphi^*(x) = x^2$ as the distorted distribution (w.r.t. π_0) but the ratio of its density w.r.t. the uniform one is decreasing so that it does not belong to $\Gamma_{\pi_0,h}$. □

An interesting mathematical result is obtained as a consequence of Theorem 8: based on Proposition 2 of Fortini and Ruggeri (1995), then the distorted band class is embedded in a topological neighborhood, $\Psi_{\pi_0,h}$, of the prior π_0. Furthermore, Theorem 3 in Fortini and Ruggeri (1995) applies to all the priors in the concentration function class, including those in the distorted band class, whose elements can be represented as mixture of extremal distributions in $\Psi_{\pi_0,h}$.

As a consequence of Theorem 8 and Theorem 4 in Fortini and Ruggeri (1995), it is possible to provide an upper bound on the supremum of the expectation of an integrable function $g(x)$ w.r.t. the class of priors $\Gamma_{\pi_0,h}$, since

$$\sup_{\pi \in \Gamma_{\pi_0,h}} E^\pi(g(X)) \le \sup_{\pi \in \Psi_{\pi_0,h}} E^\pi(g(X)).$$

The supremum of the expectation of $g(x)$ over the class $\Psi_{\pi_0,h}$ is obtained for a distribution with c.f. $h(x)$ w.r.t. π_0, as proved in Fortini and Ruggeri (1995). A lower bound on the infimum is obtained similarly. The finding can be useful, especially when the difference between upper and lower bounds is small, when performing a sensitivity analysis about a posterior expected value (e.g., of the function $g(x)$) aimed to measure the influence of the choice of a prior in a class.

When considering a distorted band with lower band equal to the identity, then a notion similar to the c.f. could be used: the only difference would be about the likelihood sets defined now for values of the likelihood ratio above some quantities and not below as before.

4 Using metrics to measure uncertainty

In robust Bayesian analysis, it is natural using probability metrics in order to incorporate the uncertainty in the elicitation process allowing for an error in the specification, see, e.g., Basu and DasGupta (1995). A nice summary of the most common probability metrics and the relationships among them is described in Gibbs and Su (2002). In our context, it is natural to use probability metrics to evaluate how a prior belief differs from its distorted version and how the corresponding posterior distributions differ, see López-Díaz et al. (2012) for a recent paper that deals with distances between probability distributions and their distortions. Due to the mathematical tractability when we compute probability metrics between a distribution function and its distorted version we are mainly interested in the Kolmogorov and Kantorovich metrics. We start by recalling the definition of the well known Kolmogorov (or uniform) metric. Given two random variables X and Y with distribution functions F_X and F_Y, respectively, the Kolmogorov (or uniform) metric is defined by

$$K(X,Y) = \sup_{x \in \mathbf{R}} |F_X(x) - F_Y(x)|, \tag{12}$$

which represents the largest absolute difference between F_X and F_Y. On the other hand, the Kantorovich metric (or Wasserstein metric) is defined by

$$KW(X,Y) = \int_{-\infty}^{\infty} |F_X(x) - F_Y(x)| dx. \tag{13}$$

The metrics above have been widely used in the literature. In particular, the Kolmogorov metric was used in Basu and DasGupta (1995) to measure the uncertainty in the distribution band and we can also find applications of that metric in Chapter 9 of the book by Denuit et al. (2005) in the context of distortion functions. It is worth mentioning that Kolmogorov metric suffers from the shortcoming of being completely insensitive to the losses in the tail of the distributions (this is because the difference $|F_X(x) - F_Y(x)|$ converges to zero as x increases or decreases). However, the Kantorovich metric provides aggregate information about the deviations between the two probabilities and, in contrast to the Kolmogorov metric, it is sensitive to the differences between the probabilities in the tails. Note that different decision makers may accentuate the differences in the body or in the tails of the baseline prior distribution. In our context, we can make use of both metrics, Kolmogorov and Kantorovich, in order to measure uncertainty in the elicitation process. Next we will show how those metrics work in our particular context.

4.1 The Kolmogorov metric

In practice, we can make use of Lemma 9 in order to compute the Kolmogorov distance between a prior belief π and its distortion function π_h, where the distortion function h is differentiable.

Lemma 9. *Let π be a prior belief with an absolutely continuous distribution function and let h be a differentiable (concave or convex) distortion function. The Kolmogorov distance between π and π_h is given by*

$$K(\pi, \pi_h) = \sup_{x \in \mathbf{R}} |F_\pi(x) - F_{\pi_h}(x)|,$$

$$= \begin{cases} p_0 - h(p_0) & \text{if } h \text{ is convex,} \\ h(p_0) - p_0 & \text{if } h \text{ is concave.} \end{cases}$$

where p_0 satisfies $h'(p_0) = 1$ and the argument of the maximum is achieved at $\theta_0 = F_\pi^{-1}(p_0)$.

Proof. Let h be a differentiable convex function. From Lemma 1 $\pi \leq_{lr} \pi_h$ holds and, using (4), then $\pi \leq_{st} \pi_h$ also holds and it trivially implies that

$$\begin{aligned} K(\pi, \pi_h) &= \sup_{\theta \in \mathbf{R}} |F_\pi(\theta) - F_{\pi_h}(\theta)|, \\ &= \sup_{\theta \in \mathbf{R}} \; F_\pi(x) - F_{\pi_h}(x), \\ &= \sup_{\theta \in \mathbf{R}} \; F_\pi(\theta) - h(F_\pi(\theta)). \end{aligned}$$

Just taking derivatives we obtain that $(F_\pi(\theta) - h(F_\pi(\theta)))' = 0$ if and only if $\pi(\theta) - \pi(\theta)h'(F_\pi(\theta)) = 0$, i.e., if and only if $1 - h'(F_\pi(\theta)) = 0$. As a direct application of the well-known mean value theorem applied to h in the interval $[0, 1]$, there exists p_0 such that $h'(p_0) = 1$. Since both h' and $F_\pi(\theta)$ are increasing, it follows directly that $\theta_0 = F_\pi^{-1}(p_0)$ is a maximum. □

We would like to emphasize that the distance in Lemma 9 depends only on the distortion function. Therefore, it seems intuitive that the Kolmogorov distance is useful to measure the uncertainty in the distorted band. Here we present two examples.

Example 10. *Let us consider the distortion functions defined in (7) and a baseline prior distribution π. Assuming that π is absolutely continuous, from Lemma 9, the Kolmogorov metric is given by the following expression:*

$$K(\pi, \pi_{h_1}) = K(\pi, \pi_{h_2}) = \frac{\alpha - 1}{\sqrt[\alpha-1]{\alpha^\alpha}}. \tag{14}$$

Just computing the derivative of the logarithm of the distance, we observe that expression (14) increases when α increases. Therefore, it can be useful to determine a fixed uncertainty ϵ. For example, $\alpha = 1.2$ will provide a distance equal to $\epsilon = K(\pi, \pi_{h_1}) = K(\pi, \pi_{h_2}) = 0.0669796$ which can be interpreted in terms of probability, i.e., the largest absolute uncertainty between the accumulated probabilities of the underlying prior belief and its distorted versions is about 6.7%. Note that we can also consider different values for α when we compute $K(\pi, \pi_{h_1})$ or $K(\pi, \pi_{h_2})$.

Example 11. *Let us consider the distortion functions defined in (10) and a baseline prior distribution π with an absolutely continuous symmetric distribution around 0. From Lemma 9 and using (9), the Kolmogorov metric is given by the following expression:*

$$K(\pi, \pi_{h_1}) = K(\pi, \pi_{h_2}) = \int_0^\infty 2\pi(\theta) F_\pi(\beta\theta) d\theta - \frac{1}{2}. \tag{15}$$

From the property of $F_\pi(\beta\theta)$, expression (15) increases when β increases. It is also apparent that the distance cannot be greater than 0.5.

Finally, Lemma 9 is not valid for non-differentiable distortion functions. For example, if we consider the distortion families defined in (8), it is easy to compute $K(\pi, \pi_{h_2}) = \alpha$ and $K(\pi, \pi_{h_1}) = 1 - \alpha$.

From the well-known fact that there is no closed-form expression of the posterior distributions, computing the Kolmogorov distance to evaluate the posterior distorted band could be difficult in practice. However, as we will see in Section 6, we will provide a way of estimating that distance based on the empirical posterior distributions.

4.2 The Kantorovich metric

Considering the Kantorovich metric, first observe that given any two random variables, X and Y, with distribution functions F_X and F_Y, respectively, such that $X \leq_{st} Y$ we have

$$\begin{aligned} KW(X,Y) &= \int_{-\infty}^{\infty} |F_X(x) - F_Y(x)| dx, \\ &= \int_{-\infty}^{\infty} (F_X(x) - F_Y(x)) dx, \\ &= E(Y) - E(X), \end{aligned} \tag{16}$$

where it is assumed that expectations exist. Note that the second equality in (16) follows from the fact that $F_X(x) \geq F_Y(x)$, for all x, and the last one from the well known expression of the expectation of a random variable in terms of its distribution function. It is also remarkable, see Theorem 1.A.8. in Shaked and Shanthikumar (2007), that if $X \leq_{st} Y$ and $E(X) = E(Y)$, i.e., $KW(X,Y) = 0$, then necessarily X and Y are equal in distributions, i.e., $X =_{st} Y$.

From expression (16) and using the implication given in (4), it follows directly that we can compute uncertainty, in the sense of the Kantorovich metric, in both prior and posterior distorted bands. In particular, we obtain that

$$\begin{aligned} KW(\pi_{h_1}, \pi_{h_2}) &= E^{\pi_{h_2}}(\theta) - E^{\pi_{h_1}}(\theta), \\ KW(\pi_{h_1,x}, \pi_{h_2,x}) &= E^{\pi_{h_2,x}}(\theta) - E^{\pi_{h_1,x}}(\theta), \\ KW(\pi, \pi_{h_1}) &= E^{\pi}(\theta) - E^{\pi_{h_1}}(\theta), \\ KW(\pi, \pi_{h_2}) &= E^{\pi_{h_2}}(\theta) - E^{\pi}(\theta), \\ KW(\pi_x, \pi_{h_1,x}) &= E^{\pi_x}(\theta) - E^{\pi_{h_1,x}}(\theta), \\ KW(\pi_x, \pi_{h_2,x}) &= E^{\pi_{h_2,x}}(\theta) - E^{\pi_x}(\theta). \end{aligned} \tag{17}$$

We emphasize that uncertainty in both prior and posterior distorted bands can be evaluated by computing the difference between prior expectations or posterior expectations, respectively. It is worth mentioning that $KW(\pi_{h_1}, \pi_{h_2}) = KW(\pi, \pi_{h_2}) + KW(\pi, \pi_{h_1})$ so that it is possible to establish which bound, h_1 or h_2, is contributing the most to the uncertainty measure. A similar statement can be made for the posterior distributions. It is apparent that the difference between prior or posterior expectations depends strongly on the underlying prior belief and the choice of the distortion functions. Analogously to the Kolmogorov metric, we will see later on that simulation methods can be used to compute the posterior expectations.

Example 12. *Let us consider the distortion functions defined in* (7) *and a baseline prior belief given by* $\pi \sim U(0,1)$. *Using* (17)*, a straightforward computation shows that the Kantorovich metric is given by*

$$KW(\pi, \pi_{h_1}) = KW(\pi, \pi_{h_2}) = \frac{1}{2} - \frac{1}{\alpha + 1}.$$

It is clear that the distance increases when α *increases. It is also apparent that the distance between expectations cannot be greater than* 0.5.

At this point, it is natural to wonder how to choose distortion functions and how to elicit their parameters. The choice of the distortion functions and their parameters depends on the problem at hand and the level of uncertainty about the priors. In a financial context, the class might allow for both risk aversion and proneness through the use of convex and concave distortion functions, possibly one more than the other through a proper choice of parameters. For example, a convex distortion function gives more weight to higher risk events. Other aspects of the problem at hand could lead to specific choices, like the ones we have introduced earlier. In particular, the power functions given in (7) can satisfactorily represent uncertainty in the tails of the prior belief, the distortion functions given in (8) discard potentially information because they map some percentiles to a single point and the distortion functions given in (10) can be useful to elicit prior knowledge with "normal-like" shape but with lack of symmetry. All those families of distortion functions depend on parameters that represent the degree of distortion and they can be elicited by fixing a reasonable distance in terms of the Kolmogorov and Kantorovich metrics.

Following with the Bayesian perspective, we investigate now the relationship between lower and upper bounds of the posterior distorted band, i.e., $\pi_{h_1,x}$ and $\pi_{h_2,x}$ and Bayes actions. In the next section, we will show how in practice we can take advantage of Proposition 6 when we check for dominance among Bayes actions.

5 Ordering Bayes actions

First of all, we would like to emphasize that from (2) and (4) we easily obtain that

$$E^{\pi_{h_1}}(g(\theta)) \leq E^{\pi'}(g(\theta)) \leq E^{\pi_{h_2}}(g(\theta)), \forall \pi' \in \Gamma_{h_1,h_2,\pi},$$

for all increasing functions g for which the expectations exist. Analogously for the posterior distorted band,

$$E^{\pi_{h_1,x}}(g(\theta)) \leq E^{\pi'_x}(g(\theta)) \leq E^{\pi_{h_2,x}}(g(\theta)), \forall \pi'_x \in \Gamma_{x,h_1,h_2,\pi}.$$

Just taking $g(x) = x$, we easily obtain the range of posterior means. We will show that this result can be extended when we compute the range of other classical Bayes actions.

Before introducing the main result, we first recall the definition of a submodular function in the bivariate case. A function $L(a,\theta) : \mathbf{R}^2 \to \mathbf{R}$ is submodular if for all $(a_1,\theta_1),(a_2,\theta_2) \in \mathbf{R}^2$,

$$L((a_1,\theta_1) \vee (a_2,\theta_2)) + L((a_1,\theta_1) \wedge (a_2,\theta_2)) \le L(a_1,\theta_1) + L(a_2,\theta_2), \tag{18}$$

where

$$(a_1,\theta_1) \vee (a_2,\theta_2) = (\max\{a_1,a_2\}, \max\{\theta_1,\theta_2\})$$

and

$$(a_1,\theta_1) \wedge (a_2,\theta_2) = (\min\{a_1,a_2\}, \min\{\theta_1,\theta_2\}).$$

We also recall that if L is submodular then $-L$ is called supermodular. The concept of supermodularity is used in the social sciences to analyze how one agent's decision affects the incentives of others. It is also known that, if L is twice differentiable, then it is submodular if

$$\frac{\partial^2 L(a,\theta)}{\partial\theta\partial a} \le 0,\ \forall a,\theta. \tag{19}$$

On the other hand, given two real numbers a_1 and a_2 such that $a_1 \le a_2$, it is apparent from expression (18) that L is submodular if and only if the function

$$L(a_2,\theta) - L(a_1,\theta) \text{ is decreasing in } \theta. \tag{20}$$

For more details about submodular and supermodular functions see, e.g., Topkis (1978) and Topkis (1998).

From now on, we will consider the class of all convex loss functions L in $\mathcal{A}$ which satisfy the submodularity property and we will denote it by $\mathcal{L}_{sm}$. Note that widely used loss functions in the literature are included in the class $\mathcal{L}_{sm}$. As a direct consequence of (19) and (20) this class includes the class of L_p loss functions given by

$$\mathcal{L} = \{L_p(a,\theta) = |a-\theta|^p, p \ge 1\},$$

where $p = 1$ and $p = 2$ correspond to the absolute error loss and the squared error loss, respectively. The class of L_q quantile loss functions given by

$$\mathcal{L} = \{L_q(a,\theta) = |a-\theta| + a(2q-1), q \in [0,1]\},$$

where $q = 1/2$ corresponds to the absolute error loss, is included as well. The class of L_k LINEX loss functions given by

$$\mathcal{L} = \{L_k(a,\theta) = \exp(k(a-\theta)) - k(a-\theta) - 1,\ (k \ne 0)\}$$

is contained, as well as the class of $L_{\alpha,\beta}$ linear loss functions given by

$$\mathcal{L} = \{L_{\alpha,\beta}(a,\theta) = \begin{cases} \alpha(a-\theta) & a \ge \theta, \\ \beta(\theta-a) & a < \theta, \end{cases} \alpha,\beta > 0\}.$$

Martín et al. (1998) provided another interesting example of loss functions included in the class $\mathcal{L}_{sm}$ given by

$$\mathcal{L} = \{L_{\lambda(\cdot)}(a,\theta) = \int_0^{\theta - a} \lambda(t)dt\},$$

where $\lambda(t)$ is positive (negative, null) if and only if $t > 0$ ($t < 0, t = 0$) and $\lambda'(t) > 0$.

Finally, we will show that the class of loss functions expressed as

$$\mathcal{L} = \{L_{\phi(\cdot)}(a,\theta) = \phi(\theta - a)\},$$

where ϕ is a differentiable convex function is also included in the class $\mathcal{L}_{sm}$. Let a_1 and a_2 be two real numbers such that $a_1 \leq a_2$. Just taking derivatives, we obtain that

$$\frac{\partial\left(L_{\phi(\cdot)}(a_2,\theta) - L_{\phi(\cdot)}(a_1,\theta)\right)}{\partial\theta} = \phi'(\theta - a_2) - \phi'(\theta - a_1) \leq 0, \ \forall\theta,$$

where the last inequality holds from the fact that ϕ' is increasing. The result follows as direct consequence of (20).

Next we provide a useful result in order to check dominance among Bayes actions when the distorted band is considered. As commented earlier, relatively recent researches in Bayesian robustness have focused attention on computing the changes in posterior expected loss and optimal actions under classes of prior and/or loss functions. In particular, the search for optimal actions has lead, as a first approximation, to consider the set of nondominated actions; see Ríos Insua and Criado (2000). Using the range of this set as a measure of the robustness is a common practice in the literature; see Moreno (2000), among others. Moreover, using strictly convex loss functions, this range coincides with the range of the Bayes alternatives. This result is not true using non-strictly convex loss functions, see Arias-Nicolás et al. (2006). Therefore, our goal here will be to provide results which enable us to order the Bayes actions.

Remark 13. *First, we recall that, given a convex loss function in $\mathcal{A}$, the posterior expected loss $\rho(\pi, L, a)$ is also convex in $\mathcal{A}$. In addition, if the set of Bayes actions $B_{(L,\pi)}$ is not empty, the function $\rho(\pi, L, a)$ is strictly decreasing in $(-\infty, \underline{a}^*_{(L,\pi)})$ and strictly increasing in $(\overline{a}^*_{(L,\pi)}, +\infty)$, where*

$$\underline{a}^*_{(L,\pi)} = \inf_{a \in B_{(L,\pi)}} a,$$

$$\overline{a}^*_{(L,\pi)} = \sup_{a \in B_{(L,\pi)}} a,$$

with $B_{(L,\pi)} = [\underline{a}^*_{(L,\pi)}, \overline{a}^*_{(L,\pi)}]$.

Theorem 14. *Let π be a specific prior belief and let $\Gamma_{h_1,h_2,\pi}$ be the corresponding distorted band. Then*

$$\underline{a}^*_{(L,\pi_{h_1})} \leq \underline{a}^*_{(L,\pi')} \leq \underline{a}^*_{(L,\pi_{h_2})}$$

and

$$\overline{a}^*_{(L,\pi_{h_1})} \leq \overline{a}^*_{(L,\pi')} \leq \overline{a}^*_{(L,\pi_{h_2})},$$

for all $L \in \mathcal{L}_{sm}$ and for all $\pi' \in \Gamma_{h_1,h_2,\pi}$ such that the set of Bayes actions is not empty.

Proof. We will provide the inequalities for π_{h_1} and π'. Let us consider $a^*_{(L,\pi')} \in B_{(L,\pi')} = [\underline{a}^*_{(L,\pi')}, \overline{a}^*_{(L,\pi')}]$. Given an alternative $a \geq a^*_{(L,\pi')}$,

$$\begin{aligned} 0 &\leq \rho(\pi'_x, L, a) - \rho(\pi'_x, L, a^*_{(L,\pi')}) \\ &= E^{\pi'_x}(L(a,\theta) - L(a^*_{(L,\pi')}, \theta)) \\ &\leq E^{\pi_{h_1,x}}(L(a,\theta) - L(a^*_{(L,\pi')}, \theta)) \\ &= \rho(\pi_{h_1,x}, L, a) - \rho(\pi_{h_1,x}, L, a^*_{(L,\pi')}). \end{aligned}$$

The first inequality holds from the fact that $a^*_{(L,\pi')}$ is a Bayes action. From (20), the function $L(a,\theta) - L(a^*_{(L,\pi')}, \theta)$ is decreasing in θ. Due to fact that $\pi_{h_1,x} \leq_{st} \pi'_x$, the second inequality holds just using (2). Therefore, we conclude that $\rho(\pi_{h_1,x}, L, a) \geq \rho(\pi_{h_1,x}, L, a^*_{(L,\pi')})$ for all $a \geq a^*_{(L,\pi')}$ and for all $a^*_{(L,\pi')} \in B_{(L,\pi')}$. Hence, using Remark 13, $\underline{a}^*_{(L,\pi_{h_1})} \leq \underline{a}^*_{(L,\pi')}$ and $\overline{a}^*_{(L,\pi_{h_1})} \leq \overline{a}^*_{(L,\pi')}$ hold. An analogous result holds for π_{h_2} and π'. □

6 Obtaining a sample of the posterior distorted distributions

In practice, it is not easy in general to compute the exact distributions of the posterior distorted distributions. However, we can make use of the relationship between the posterior distribution and the distorted posterior distribution to apply simulation methods. We will assume that the prior π has a probability density function and the distortion functions are differentiable. Then the posterior distortion density $\pi_{h,x}$ we wish to simulate from has also a probability density which can be expressed as

$$\begin{aligned} \pi_{h,x}(\theta) = \frac{l(\theta)\pi_h(\theta)}{m_{\pi_h}(x)} &= \frac{l(\theta)\pi(\theta)h'(F_\pi(\theta))}{m_{\pi_h}(x)} \\ &= \frac{l(\theta)\pi(\theta)h'(F_\pi(\theta))}{m_{\pi_h}(x)} \frac{m_\pi(x)}{m_\pi(x)} \\ &= \pi_x(\theta)\frac{m_\pi(x)h'(F_\pi(\theta))}{m_{\pi_h}(x)}. \end{aligned} \tag{21}$$

Since convex and concave functions have increasing and decreasing derivative functions, respectively, just assuming h' to be a bounded function, it follows that

$$\pi_{h,x}(\theta) \leq \begin{cases} \pi_x(\theta)\frac{m_\pi(x)h'(1)}{m_{\pi_h}(x)} & \text{if } h \text{ is convex,} \\ \pi_x(\theta)\frac{m_\pi(x)h'(0)}{m_{\pi_h}(x)} & \text{if } h \text{ is concave.} \end{cases}$$

Therefore, $\pi_{h,x}$ seems to be "close" to π_x due to the fact that the ratio between the densities is bounded by a constant. Just considering that constant and applying the well known acceptance–rejection method in Simulation Theory by John von Neumann, if we are able to generate a sample from π_x, we will be also able to generate a sample from $\pi_{h,x}$. Next we provide two algorithms:

1. Algorithm 1, for h_2 convex.
   - Sample θ from π_x and u from $U(0,1)$, independently.
   - Check whether or not $u < \frac{h_2'(F_\pi(\theta))}{h_2'(1)}$
     - If this holds, accept θ as a realization of $\pi_{h_2,x}$;
     - If not, reject the value of θ and repeat the sampling step.
2. Algorithm 2, for h_1 concave.
   - Sample θ from π_x and u from $U(0,1)$, independently.
   - Check whether or not $u < \frac{h_1'(F_\pi(\theta))}{h_1'(0)}$
     - If this holds, accept θ as a realization of $\pi_{h_1,x}$;
     - If not, reject the value of θ and repeat the sampling step.

7 Numerical examples

7.1 Normal–normal model

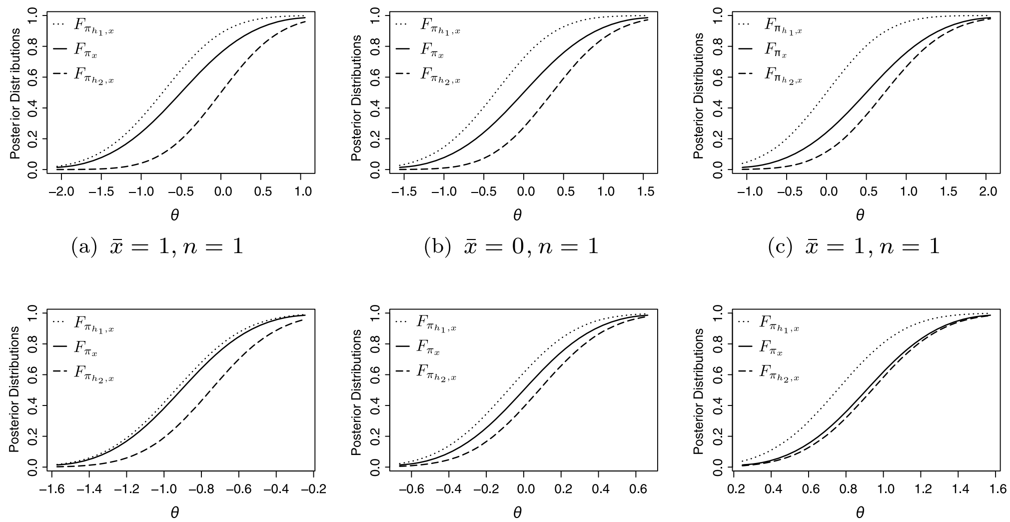
Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with normal distribution, $N(\theta, \sigma^2)$, where the mean θ is unknown and the variance σ^2 is known, and let π be a normal $N(\mu, \tau^2)$ prior distribution of θ. We are interested in computing the range of the Bayes actions when we consider the squared error loss function, L_2, and a distorted class $\Gamma_{h_1,h_2,\pi}$ defined by skewed distributions, where h_1 and h_2 are given by (10). Note that it is assumed that prior beliefs lead to a normal distribution whereas uncertainty is associated with some lack of symmetry. It is well-known that π_x is normally distributed with posterior mean $E^{\pi_x}(\theta) = (\sigma^2\mu + n\bar{x}\tau^2)/(\sigma^2 + n\tau^2)$ and variance $(\sigma^2\tau^2)/(\sigma^2 + n\tau^2)$. As a specific example, we will consider $\mu = 0$, $\tau^2 = 1$, $\sigma^2 = 1$ and $\beta = 1.2$ as the distortion parameter. In practice, the parameter β provides the degree of distortion and can be elicited by fixing a reasonable distance in terms of Kolmogorov and Kantorovich metrics. In our case, from expressions (15) and (17) and using the numerical integration command "*integrate*" available in software R, we obtain that

$$K(\pi_{h_2}, \pi_{h_1}) = 0.5577 \text{ and } KW(\pi_{h_2}, \pi_{h_1}) = 1.2259. \tag{22}$$

Since the exact distributions of the posterior distorted distributions are not known, all characteristics of both $\pi_{h_2,x}$ and $\pi_{h_1,x}$ have been estimated from their empirical distributions after simulating a sample of size $1{,}000{,}000$ using Algorithms 1 and 2 in Section 6.

We first show in Figure 3 the effect of the distortion functions on the posterior distributions combining several values of sample mean with sample sizes $n = 1$ and $n = 10$. It seems intuitive that uncertainty decreases when sample size increases.

Second, in a similar way as Sivaganesan and Berger (1989) and using a sample size equal to 1, we have computed in Table 1 all Bayes actions for different values of x, where we recall that Bayes actions are given by the posterior means.



(a) $\bar{x} = 1, n = 1$ (b) $\bar{x} = 0, n = 1$ (c) $\bar{x} = 1, n = 1$

(d) $\bar{x} = 1, n = 10$ (e) $\bar{x} = 0, n = 10$ (f) $\bar{x} = 1, n = 10$

Figure 3: Posterior Distorted Distributions for $\beta = 1.2$.

x	–3	–2	–1	0	1	2	3
$E^{\pi_x}(\theta)$	–1.5	–1	–0.5	0	0.5	1	1.5
$E^{\pi_{h_2,x}}(\theta)$	–0.660	–0.320	0.007	0.360	0.746	1.148	1.577
$E^{\pi_{h_1,x}}(\theta)$	–1.570	–1.140	–0.740	–0.360	–0.007	0.329	0.661
$K(\pi_{h_2,x}, \pi_{h_1,x})$	0.541	0.492	0.460	0.450	0.460	0.492	0.541
$KW(\pi_{h_2,x}, \pi_{h_1,x})$	0.913	0.812	0.752	0.720	0.752	0.813	0.913

Table 1: Range of posterior means, $n = 1$, $\beta = 1.2$.

It is shown graphically in Figure 4 that the range of Bayes actions decreases when $|x|$ decreases. Here it is worth mentioning that the range of Bayes actions coincides with the Kantorovich distance, see (17).

Sivaganesan and Berger (1989) use an ϵ-contamination prior class, $\epsilon = 0.1$, with symmetric unimodal contaminations sharing a common mode as that of the underlying baseline prior π, and conclude that the range can be considered fairly small for values of x such that $|x| < 3$. Here we reach a similar conclusion, since we see in Table 1 that both posterior distances $KW(\pi_{h_1,x}, \pi_{h_2,x})$ and $K(\pi_{h_1,x}, \pi_{h_2,x})$ show a decrease with respect to the prior distances given in (22).

Finally, Figure 5 shows the effect on the range of the posterior Bayes actions fixing different sample sizes, $n = 1$ and 10, and using different distorted prior bands given by $\beta = 0.5$, 1.2 and 1.5. It is apparent that the range of Bayes actions is larger when the uncertainty about the prior distribution π increases, i.e., when the value of β increases. Moreover, the range decreases when the sample size increases and/or $|x|$ decreases. It is also worth mentioning that the bound h_1 contributes the most to the uncertainty when

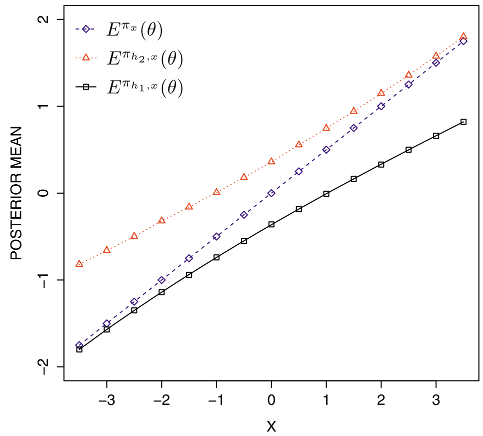


Figure 4: Range of the posterior means against x, $n = 1$ and $\beta = 1.2$.

negative values are sampled, like h_2 does for positive ones.

7.2 Pareto–exponential model

Let $X_1, X_2, \ldots X_n$ be an i.i.d. random sample from the Pareto income distribution $P(\theta, \beta)$ where θ is the unknown shape parameter and the scale parameter, β, is known.

We are interested in computing the range of the Bayes actions, when we consider a distorted class Γ defined by the classical power functions (see (7) in Section 2) and a LINEX loss function L_k defined in Section 6, where $k \neq 0$ is a known parameter.

The value of a that minimizes the posterior risk $E^{\pi_x}[L_k(a, \theta)]$, i.e., the Bayes alternative, is reached for

$$a^*_{(L_k,\pi)} = -\frac{1}{k} \log E^{\pi_x}[\exp(-k\theta)], \tag{23}$$

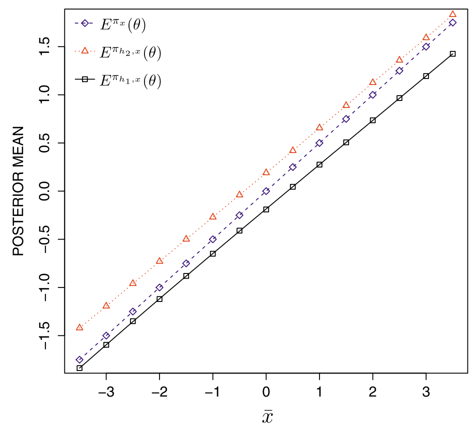
provided that $E^{\pi_x}[\exp(-k\theta)]$ exists and is finite.

Here we assume a fixed truncated two-parameter exponential prior distribution for θ, $\pi = \pi_{(\mu,\lambda)} \sim Exp(\lambda, \mu)$, where the hyperparameters λ and μ are assumed to be known. Then the posterior distribution of θ is a truncated gamma with parameters μ, t_λ and $n+1$, where $t_\lambda = Z + \lambda$, $Z = n \log(\frac{\bar{x}_G}{\beta})$ and $\bar{x}_G = (\prod_{i=1}^n x_i)^{\frac{1}{n}}$ is the geometric mean of the sample.

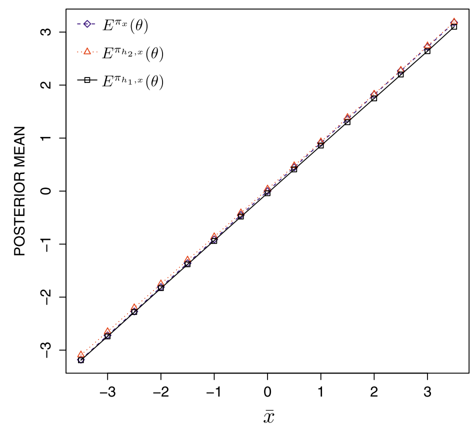
The Bayes actions of θ under LINEX loss functions are given by

$$a^*_{(L_k,\pi)} = \frac{1}{k} \log \left(\left(\frac{k + t_\lambda}{t_\lambda} \right)^{n+1} \frac{\Gamma(n+1, t_\lambda \mu)}{\Gamma(n+1, k\mu + t_\lambda \mu)} \right). \tag{24}$$

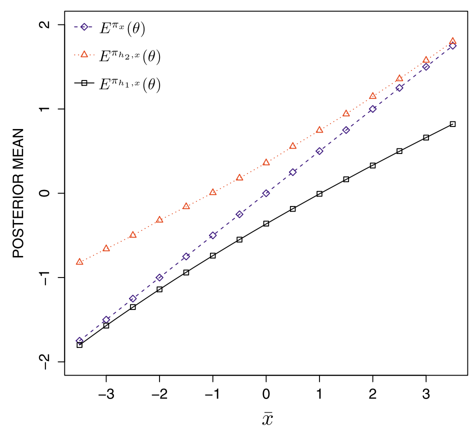
As a specific example, we consider a Pareto distribution with known parameter $\beta = 2$, a truncated exponential prior for θ with known parameter $\lambda = 1$ and $\mu = 1$, and



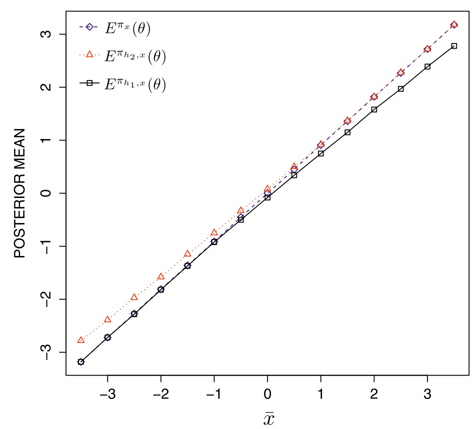
(a) $n = 1$, $\beta = 0.5$



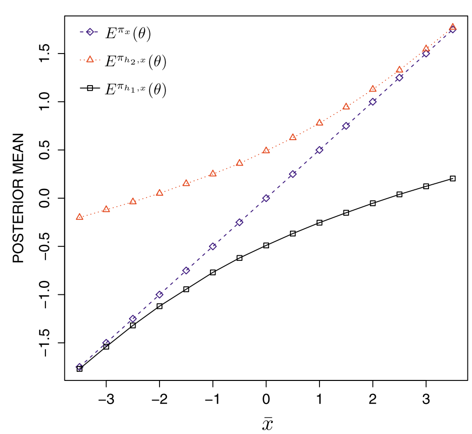
(b) $n = 10$, $\beta = 0.5$



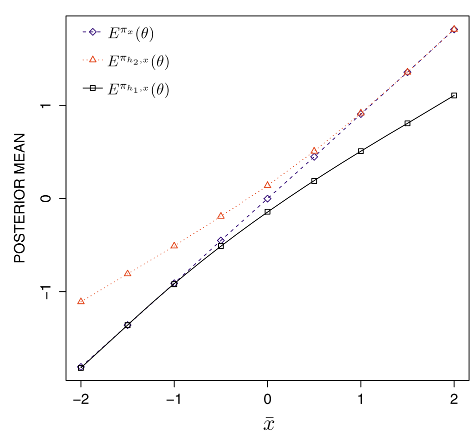
(c) $n = 1$, $\beta = 1.2$



(d) $n = 10$, $\beta = 1.2$

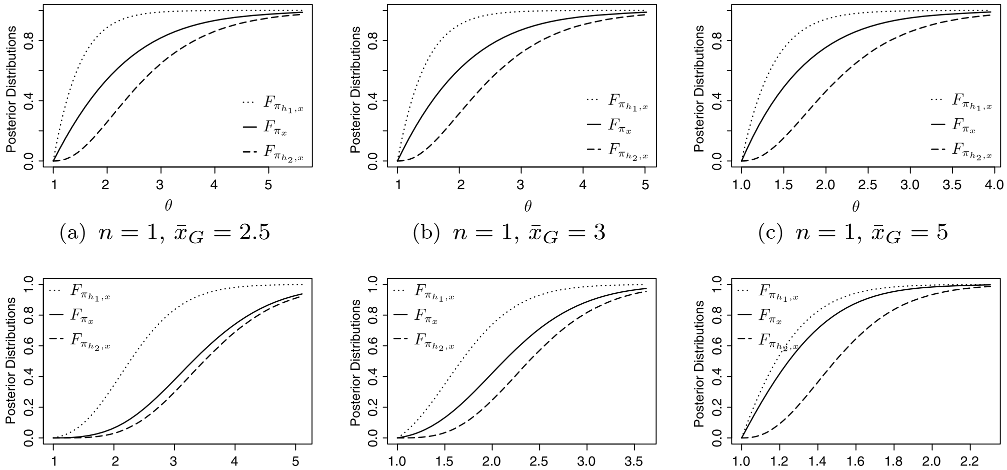


(e) $n = 1$, $\beta = 2.5$



(f) $n = 10$, $\beta = 2.5$

Figure 5: Range of the posterior means against $\bar{x}$, for $\beta = 0.5, 1.2$ and 2.5 and sample sizes $n = 1$ and 10.



(a) $n = 1$, $\bar{x}_G = 2.5$ (b) $n = 1$, $\bar{x}_G = 3$ (c) $n = 1$, $\bar{x}_G = 5$

(d) $n = 10$, $\bar{x}_G = 2.5$ (e) $n = 10$, $\bar{x}_G = 3$ (f) $n = 10$, $\bar{x}_G = 5$

Figure 6: Posterior Distorted Distributions for $\alpha = 2.5$.

2.05	2.23	2.26	2.26	2.27	2.27	2.27	2.28	2.33	2.35
2.35	2.42	2.46	2.57	2.58	2.59	2.63	2.81	3.01	3.35
3.43	3.77	4.95	5.07	5.10	5.36	6.09	6.47	8.33	14.33

Table 2: Range of Bayes actions.

a distorted class $\Gamma_{h_1,h_2,\pi}$ defined by the classical power functions, where h_1 and h_2 are given by (7) with $\alpha = 2.5$. Using the same argument as in the previous example, we obtain that

$$K(\pi_{h_2}, \pi_{h_1}) = 0.6464 \text{ and } KW(\pi_{h_2}, \pi_{h_1}) = 1.2804. \tag{25}$$

Remark 15. *Note that, if π is a truncated two-parameter exponential distribution, $\pi \sim Exp(\mu, \lambda)$, it is easy to verify that the distorted distribution with h_1 is a truncated two-parameter exponential distribution too, $\pi_{h_1} \sim Exp(\mu, \lambda\alpha)$. This is not verified for h_2.*

Since the exact distribution of the posterior distorted distribution associated to function h_2 is unknown, all characteristics of $\pi_{h_2,x}$ have been estimated from their empirical distribution after simulating a sample of size $1,000,000$ using Algorithm 1 in Section 6. Figure 6 shows the effect of the distortion function on the posterior distributions combining several values of sample geometric mean ($\bar{x}_G = 2.5, 3$ and 5) with sample sizes $n = 1$ and $n = 10$.

To discuss the sensitivity of the Bayes estimator considering the distorted band class of priors, we present in Table 2 the range of Bayes actions for the following random sample of 30 observations generated from a Pareto distribution with parameters $\theta = 2$ and $\beta = 2$, see Upadhyay and Shastri (1997): The sampling scheme considered in such paper and later in Saxena and Singh (2007) involves only those individuals whose

k	–3	–2	–1	–0.5	–0.2
$a^*_{(L_k,\pi)}$	2.0029	1.9180	1.8428	1.8083	1.7885
$a^*_{(L_k,\pi_{h_1})}$	1.7838	1.7182	1.6599	1.6331	1.6178
$a^*_{(L_k,\pi_{h_2})}$	2.1429	2.0656	1.9935	1.9632	1.9439
Min./Max.	0.8324	0.8318	0.8327	0.8319	0.8322
k	0.2	0.5	1	2	3
$a^*_{(L_k,\pi)}$	1.7632	1.7449	1.7158	1.6622	1.614170
$a^*_{(L_k,\pi_{h_1})}$	1.5981	1.5840	1.5614	1.5200	1.4831
$a^*_{(L_k,\pi_{h_2})}$	1.9199	1.9035	1.8746	1.8238	1.7797
Min./Max.	0.8324	0.8321	0.8329	0.8334	0.8334
$K(\pi_{h_2,x},\pi_{h_1,x})$			0.3786		
$KW(\pi_{h_2,x},\pi_{h_1,x})$			0.3221		

Table 3: Bayes actions of θ, for $k = \pm 3, \pm 2, \pm 1, \pm 0.5, \pm 0.2$ and for $\alpha = 2.5$.

annual incomes do not exceed the fixed value $w = 5$ (in the current paper, several values for $\lambda = 0.25, 0.5, 0.75, 1 - 5$ and $k = \pm 3, \pm 2, \pm 1, \pm 0.5, \pm 0.2$ are considered). They assume, among other distributions, a fixed truncated two-parameter exponential prior distribution for θ, $\pi_0 = \pi_{(\mu,\lambda)} \sim Exp(\lambda, \mu)$, where the hyperparameters λ and μ are assumed to be known.

Remark 16. *When censored sample data are considered (i.e., when exact incomes are available only when they do not exceed a prescribed value, say w ($w > \beta$)), the likelihood function can be expressed as $l(\theta) \propto \theta^r e^{-Z_w \theta}$, where r is the number of available incomes and $Z_w = \log(\beta^{-n} P_w)$, being $P_w = w^{n-r}(\prod_{i=1}^r x_i)$ the product income statistic introduced by Ganguly et al. (1992). The posterior distribution of θ is a truncated gamma with parameters μ, $t_{w,\lambda} = Z_w + \lambda$ and $r+1$. To compute the Bayes actions under LINEX loss functions, we may take $n = r$ and $t_\lambda = t_{w,\lambda}$ in (24).*

After thresholding, Saxena and Singh (2007) considered 23 ordered observations ($n = 30, r = 23$ and $Z_w = 12.56$) and found the Bayes estimates for LINEX and quadratic loss functions for different values of k and λ. Similarly, Table 3 shows all Bayes actions for different LINEX loss functions, where we recall that Bayes actions are computed using (23). The table also shows the ratio of minimum to maximum of Bayes estimates and the posterior distances between distorted distributions. Note that the Bayes actions associated to the baseline truncated exponential distribution π and the distorted distribution with h_1, are computed using (24) with $t_{w,\lambda}$ and $t_{w,\lambda\alpha}$, respectively. Obviously, for small values of $|k|$, optimal Bayes estimates under LINEX loss functions are not much different from those obtained with squared loss functions (the posterior mean associated to baseline prior π is 1.7757). We can see too that both posterior distances $KW(\pi_{h_1,x}, \pi_{h_2,x})$ and $K(\pi_{h_1,x}, \pi_{h_2,x})$ show a great decrease with respect to the prior distances given in (25).

Figure 7 shows how the range of Bayes actions has barely changed subject to variation in k. In Saxena and Singh (2007) (Table 4.3), it is displayed that the estimators only overestimate θ when both λ and k are quite small. Here we reach a similar conclusion.

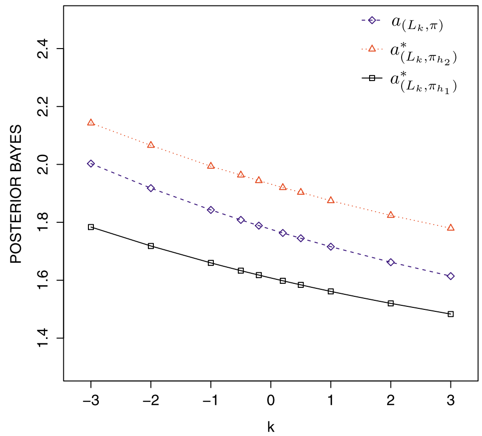


Figure 7: Range of Bayes action against k, for $\alpha = 2.5$ and $\mu = \lambda = 1$.

Moreover, all Bayes estimators obtained under LINEX loss functions are very robust, for variations in π within the class $\Gamma_{h_1,h_2,\pi}$, as the ratio of minimum to maximum is considerably close to unity (greater than 0.83). Note that Saxena and Singh (2007) consider different fixed truncated two-parameter exponential priors ($\lambda = 0.25, 0.5, \dots, 1-5$), whereas we consider a distorted band associated with prior $\pi \sim Exp(\mu, \lambda)$, with $\lambda = 1$.

8 Concluding remarks

In this paper we have illustrated a methodology which addresses the major criticisms about the Bayesian approach, i.e., arbitrariness in the choice of the prior distribution. For such a purpose, we have introduced a new class of prior (especially) and loss functions, based on notions from other fields in Statistics (e.g., stochastic ordering). In particular, we consider the distorted band class of priors (stemming from stochastic ordering), and a subclass of convex loss functions, the submodular ones, which contains the most widely used loss functions: quadratic, absolute value, quantile and LINEX loss functions, among others. The proposed class fulfills all the requirements that Berger (1994) discussed about the choice of a class. First of all, elicitation of the class should be easy, as well as its interpretation, and its size should reflect the prior uncertainty, with no exclusion of reasonable priors and inclusion of unreasonable ones (e.g., discrete ones in many problems). The major novelties of the proposed class are two: preservation of the same defining property (stochastic ordering) a posteriori and *distortion* of a baseline prior. The former aspect makes possible a quite straightforward extension of the properties found for the prior class to the posterior one, and vice versa. The latter aspect allows for natural and reasonable modifications of the baseline prior: as an example, compare the distortion of a prior allowing for risk aversion/proneness with the hardly interpretable mixture with a contaminating measure in the ϵ-contaminated class. Additionally, computations of sensitivity measures should be as easy as possible, possibly looking for the extremal distributions generating the class. Future works will be addressed to compute classes of priors based on particular distortion functions depending on their use. Finally, we plan to consider the case of a n-dimensional model parameter, stemming from the works by Shaked and Shanthikumar (2007) on the definitions of multivariate likelihood ratio order and multivariate stochastic order and by Di Bernardino and Rullière (2013) on the extension of the notion of distortion to the multivariate case.

References

Arias-Nicolás, J. P., Martín, J., Ruggeri, F., and Suárez-Llorens, A. (2009). "Optimal actions in problems with convex loss functions." *International Journal of Approximate Reasoning*, 50: 303–314. MR2514504. doi: `http://dx.doi.org/10.1016/j.ijar.2008.03.014`. 1108

Arias-Nicolás, J. P., Martín, J., and Suárez-Llorens, A. (2006). "The Nondominated set in Bayesian decision problems with Convex Loss Functions." *Communications in Statistics – Theory and Methods*, 35: 593–607. MR2282878. doi: `http://dx.doi.org/10.1080/03610920500257329`. 1108, 1124

Azzalini, A. (1985). "A class of distributions which includes the normal ones." *Scandinavian Journal of Statistics*, 12 (2): 171–178. MR0808153. 1114

Basu, S. (1994). "Variations of posterior expectations for symmetric unimodal priors in a distribution band." *Sankhyā A*, 56 (2): 320–334. MR1664917. 1107

Basu, S. and DasGupta, A. (1995). "Robust Bayesian analysis with distribution bands." *Statistics & Decisions*, 13: 333–349. MR1380253. 1108, 1118, 1119

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer, second edition. MR0804611. doi: `http://dx.doi.org/10.1007/978-1-4757-4286-2`. 1108

Berger, J. O. (1994). "An overview of robust Bayesian analysis (with discussion)." *Test*, 3: 5–124. MR1293110. doi: `http://dx.doi.org/10.1007/BF02562676`. 1108, 1132

Berger, J. O. and Berliner, L. M. (1986). "Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors." *The Annals of Statistics*, 14: 461–486. MR0840509. doi: `http://dx.doi.org/10.1214/aos/1176349933`. 1108, 1109

Betrò, B., Męczarski, M., and Ruggeri, F. (1994). "Robust Bayesian analysis under generalized moments conditions." *Journal of Statistical Planning and Inference*, 41: 257–266. MR1309612. doi: `http://dx.doi.org/10.1016/0378-3758(94)90022-1`. 1108

Chew, S. H., Karni, E., and Safra, S. (1987). "Risk aversion in the theory of expected utility with rank dependent probabilities." *Journal of Economic Theory*, 42: 370–381. MR0898526. doi: `http://dx.doi.org/10.1016/0022-0531(87)90093-7`. 1109

Cifarelli, D. M. and Regazzini, E. (1987). "On a general definition of concentration function." *Sankhyā B*, 49: 307–319. MR1056041. 1116

Denneberg, D. (1990). "Premium calculation: Why standard deviation should be replaced by absolute deviation." *ASTIN Bulletin*, 20: 181–190. 1108

Denuit, M., Dhaene, J., Goovaerts, M. J., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Chichester, UK: Wiley. 1111, 1119

Dey, D. K., Lou, K., and Bose, S. (1998). "A Bayesian approach to loss robustness." *Statistics & Decisions*, 16: 65–87. MR1622563. 1108

Di Bernardino, E. and Rullière, D. (2013). "Distortions of multivariate distribution functions and associated level curves: Applications in multivariate risk theory." *Insurance: Mathematics and Economics*, 53 (1): 190–205. MR3081472. doi: http://dx.doi.org/10.1016/j.insmatheco.2013.05.001. 1133

Ferreira, J. A. T. S. and Steel, M. F. J. (2006). "A constructive representation of univariate skewed distributions." *Journal of the American Statistical Association*, 101: 823–829. MR2256188. doi: http://dx.doi.org/10.1198/016214505000001212. 1114

Fortini, S. and Ruggeri, F. (1995). "On defining neighbourhoods of measures through the concentration function." *Sankhyā A*, 56: 444–457. 1115, 1117, 1118

Fortini, S. and Ruggeri, F. (2000). "On the Use of the Concentration Function in Bayesian Robustness." In Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 109–126. New York, NY, USA: Springer. MR1795212. doi: http://dx.doi.org/10.1007/978-1-4612-1306-2_6. 1108

French, S. and Ríos Insua, D. (2000). *Statistical Decision Theory*. London, UK: Arnold. 1110

Ganguly, A., Singh, N. K., Choudhuri, H., and Bhattacharya, S. K. (1992). "Bayesian estimation of the Gini index for the PID." *Test*, 1: 93–104. MR1266132. doi: http://dx.doi.org/10.1007/BF02562665. 1131

Gibbs, A. L. and Su, F. E. (2002). "On Choosing and Bounding Probability Metrics." *International Statistical Review*, 70 (3): 419–435. 1119

Goovaerts, M. J. and Laeven, R. J. A. (2008). "Actuarial risk measures for financial derivative pricing." *Insurance: Mathematics and Economics*, 42: 540–547. MR2404313. doi: http://dx.doi.org/10.1016/j.insmatheco.2007.04.001. 1108

Lehmann, E. L. and Rojo, J. (1992). "Invariant directional orderings." *Annals of Statistics*, 20: 2100–2110. MR1193328. doi: http://dx.doi.org/10.1214/aos/1176348905. 1112

López-Díaz, M., Sordo, M. A., and Suárez-Llorens, A. (2012). "On the L_p-metric between a probability distribution and its distortion." *Insurance: Mathematics and Economics*, 51: 257–264. 1119

Makov, U. (1994). "Some aspects of Bayesian loss robustness." *Journal of Statistical Planning and Inference*, 38: 359–370. MR1261809. doi: http://dx.doi.org/10.1016/0378-3758(94)90016-7. 1108

Martín, J., Ríos Insua, D., and Ruggeri, F. (1998). "Issues in Bayesian loss robustness." *Sankhyā A*, 60: 405–417. MR1718848. 1108, 1123

Moreno, E. (2000). "Global Bayesian robustness for some classes of prior distributions." In: Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 45–70. New York, NY, USA: Springer. MR1795209. doi: http://dx.doi.org/10.1007/978-1-4612-1306-2_3. 1108, 1124

Moreno, E. and Cano, J. A. (1991). "Bayesian analysis for ϵ-contaminations partially known." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53: 143–155. MR1094277. 1108, 1109

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. New York, NY, USA: Wiley. MR1889865. 1110

Quiggin, J. (1982). "A theory of anticipated utility." *Journal of Economic Behavior and Organization*, 3: 323–343. 1108

Ríos Insua, D. and Criado, R. (2000). "Topics on the Foundations of Robust Bayesian Analysis." In: Ríos Insua, D. and Ruggeri, F. (eds.), *Robust Bayesian Analysis*, 33–44. New York, NY, USA: Springer. MR1795208. doi: http://dx.doi.org/10.1007/978-1-4612-1306-2_2. 1124

Ríos Insua, D. and Martín, J. (1994). "Robustness issues under imprecise beliefs and preferences." *Journal of Statistical Planning and Inference*, 48: 383–389. MR1294991. doi: http://dx.doi.org/10.1016/0378-3758(94)90133-3. 1109

Ríos Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. New York, NY, USA: Springer. MR1795206. doi: http://dx.doi.org/10.1007/978-1-4612-1306-2. 1107

Saxena, S. and Singh, H. P. (2007). "Bayesian Estimation of Shape Parameter of Pareto Income Distribution Using LINEX loss Function." *The Korean Communications in Statistics*, 14 (1): 33–55. 1130, 1131, 1132

Schmeidler, D. (1989). "Subjective Probability and Expected Utility without Additivity." *Econometrica*, 57: 571–587. MR0999273. doi: http://dx.doi.org/10.2307/1911053. 1108

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. New York, NY, USA: Springer. MR2265633. doi: http://dx.doi.org/10.1007/978-0-387-34675-5. 1109, 1110, 1113, 1121, 1133

Shaked, M., Sordo, M. A., and Suárez-Llorens, A. (2010). "A class of location-independent variability orders, with applications." *Journal of Applied Probability*, 47: 407–425. MR2668497. doi: http://dx.doi.org/10.1239/jap/1276784900. 1109

Sivaganesan, S. (1989). "Sensitivity of posterior mean to unimodality preserving contaminations." *Statistics & Decisions*, 7: 77–93. MR1000928. 1109

Sivaganesan, S. and Berger, J. O. (1989). "Ranges of posterior measures for priors with unimodal contaminations." *The Annals of Statistics*, 17: 868–889. MR0994273. doi: http://dx.doi.org/10.1214/aos/1176347148. 1109, 1126, 1127

Sordo, M. A. and Suárez-Llorens, A. (2011). "Stochastic comparisons of distorted variability measures." *Insurance: Mathematics and Economics*, 49: 11–17. MR2811889. doi: http://dx.doi.org/10.1016/j.insmatheco.2011.01.014. 1109

Spizzichino, F. (2001). *Subjective Probability Models for Lifetimes*. Boca Raton, FL, USA: Chapman and Hall/CRC. MR1980207. doi: http://dx.doi.org/10.1201/9781420036138. 1114

Topkis, D. M. (1978). "Minimizing a submodular function on a lattice." *Operations Research*, 26: 305–321. MR0468177. 1123

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton, NJ, USA: Princeton University Press. MR1614637. 1123

Upadhyay, S. K. and Shastri, V. (1997). "Bayes results for classical Pareto distribution via Gibbs sampler with doubly censored observations." *IEEE Transactions on Reliability*, 46: 56–59. 1130

Wang, S. (1995). "Insurance pricing and increased limits ratemaking by proportional hazards transforms." *Insurance: Mathematics and Economics*, 17: 43–54. MR1363642. doi: http://dx.doi.org/10.1016/0167-6687(95)91054-P. 1108

Wang, S. (1996). "Premium calculation by transforming the layer premium density." *ASTIN Bulletin*, 26: 71–92. 1108

Wang, S. and Young, V. R. (1998). "Ordering risks: Expected utility theory versus Yaari's dual theory of risk." *Insurance: Mathematics and Economics*, 22: 145–161. MR1650583. doi: http://dx.doi.org/10.1016/S0167-6687(97)00036-X. 1109

Yaari, M. E. (1987). "The dual theory of choice under risk." *Econometrica*, 55: 95–115. MR0875518. doi: http://dx.doi.org/10.2307/1911158. 1108

Acknowledgments

The authors want to acknowledge the comments by two anonymous referees, and the Editor of the journal which have improved significantly the presentation of this paper. This research has been supported by *Ministerio de Economía y Competitividad*, Spain (Projects MTM2011-28983-C03-02, MTM2014-56949-C3-3-R and MTM2014-57559-P) and *European Union* (European Regional Development Funds).