On a New Class of Multivariate Prior Distributions: Theory and Application in Reliability

Fabrizio Ruggeri[*], Marta Sánchez-Sánchez[†], Miguel Ángel Sordo[‡], and Alfonso Suárez-Llorens[§]

Abstract. In the context of robust Bayesian analysis for multiparameter distributions, we introduce a new class of priors based on stochastic orders, multivariate total positivity of order 2 (MTP_2) and weighted distributions. 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 the Hellinger metric and the Kullback-Leibler divergence to measure the uncertainty induced by such a class, as well as its effect on the posterior distribution. Finally, we conclude the paper with a real example about train door reliability.

Keywords: robust Bayesian analysis, Bayesian sensitivity, class of priors, stochastic orders, multivariate total positivity, weighted distributions.

MSC2020 subject classifications: Primary 62F15; secondary 60E15.

1 Introduction

It is well known that Bayesian statistics provides an excellent theoretical framework for analyzing experimental data. The key of its success lies on its ability to incorporate prior knowledge about the quantity of interest as a distribution function. This prior distribution, together with experimental data, leads in general to a better estimation of the quantity under study. A thorough review of the Bayesian approach can be found in Berger (1985) and Bernardo (2003). As it is nicely described in Basu (1994), any elicitation process leading to a prior information is to some extent arbitrary. Therefore, the advantage given by the prior information becomes in some way a disadvantage, being in fact the critical and even polemical point of Bayesian methodology. This is exactly the problem addressed by robust Bayesian analysis, also called Bayesian sensitivity analysis, quantifying and interpreting the uncertainty induced by the partial knowledge of the prior information. Of course, this also holds for other classical elements in Bayesian analysis as likelihood and loss. A thorough review of the robust Bayesian approach can be

[*]CNR Istituto di Matematica Applicata e Tecnologie Informatiche. Via Alfonso Corti, 12, 20133 Milano, Italy, fabrizio@mi.imati.cnr.it

[†]Facultad de Ciencias, Universidad de Cádiz. Avda. Rep. Saharaui s/n, 11510 Puerto Real, Cádiz, Spain, marta.sanchez@uca.es

[‡]Facultad de Ciencias, Universidad de Cádiz. Avda. Rep. Saharaui s/n, 11510 Puerto Real, Cádiz, Spain, mangel.sordo@uca.es

[§]Facultad de Ciencias, Universidad de Cádiz. Avda. Rep. Saharaui s/n, 11510 Puerto Real, Cádiz, Spain, alfonso.suarez@uca.es

found in Berger (1994) and Ríos Insua and Ruggeri (2000). With respect to uncertainty in the loss functions we refer the readers to Makov (1994), Dey et al. (1998), Martín et al. (1998) and, more recently, to Arias-Nicolás et al. (2006, 2009), among others.

Focusing only on the prior uncertainty, it is a common practice in the literature to replace the specific prior distribution by a class of priors Γ. The use of this class somehow overcomes the classical criticism about the choice of a unique prior on the grounds of arbitrariness and bias. For such a reason, we find different valuable definitions in the literature depending on their use as parametric families, contamination classes, density bands, densities with a few specified percentiles, distributions bands, etc. Denoting in bold lowercase and bold uppercase vectors and random vectors, respectively, given a specific prior $\boldsymbol{\pi}$ and ϵ in $(0, 1)$, the classical ϵ-contamination class is defined as

$$\Gamma_\epsilon = \{\boldsymbol{\pi}' : \boldsymbol{\pi}' = (1 - \epsilon)\boldsymbol{\pi} + \epsilon \mathbf{Q}, \, \mathbf{Q} \in \mathcal{Q}\}, \tag{1.1}$$

where $\boldsymbol{\pi}'$ is given by a mixture and $\mathcal{Q}$ is a family of priors called the class of contaminations, see e.g. Berger and Berliner (1986), Moreno and Cano (1991), Betrò et al. (1994), among others. For a detailed illustration of classes of priors and topological neighborhoods we refer to Berger (1985, 1994), Fortini and Ruggeri (2000) and Moreno (2000).

Recently Arias-Nicolás et al. (2016) defined a new class of priors, called the distorted band, based on stochastic orders and distortion functions. That new class addresses the problem of uncertainty in models depending just on one-parameter. To better understand the concept of that class, let us first recall the classical Bayesian inference framework. We denote by X the underlying observation with probability density function (PDF) $p_{\boldsymbol{\theta}}(x)$, where $\boldsymbol{\theta}$ represents the unknown parameter defined in the set of states $\boldsymbol{\Theta} \subseteq \mathbb{R}^n$, $n \in \mathbb{N}, n \geq 1$. Then, a specific prior $\boldsymbol{\pi}$ over $\boldsymbol{\Theta}$ with PDF $\pi(\boldsymbol{\theta})$ represents the state of knowledge before any data X is observed. After taking into account the observed data, $\mathbf{x}$, and using Bayes's theorem, the posterior distribution is denoted by $\boldsymbol{\pi_x}$ and its density is given by

$$\pi_{\mathbf{x}}(\boldsymbol{\theta}) = \frac{l(\boldsymbol{\theta} \mid \mathbf{x})\pi(\boldsymbol{\theta})}{m_{\boldsymbol{\pi}}(\mathbf{x})}, \tag{1.2}$$

where $l(\boldsymbol{\theta} \mid \mathbf{x})$ and $m_{\boldsymbol{\pi}}(\mathbf{x})$ denote the likelihood function and the marginal density, respectively.

At this point we also recall the definition of the likelihood ratio order for univariate random variables. Let π_1 and π_2 be two prior distributions over $\Theta \subseteq \mathbb{R}$. The former is said to be smaller than the latter in the likelihood ratio order, if the ratio $\pi_2(\theta)/\pi_1(\theta)$ increases over the union of the supports of the two PDFs. We denote this occurrence by $\pi_1 \leq_{\mathrm{lr}} \pi_2$. Roughly speaking, this means that π_2 takes on larger values and more variability than π_1. For more details on the concept of likelihood ratio order see Müller and Stoyan (2002) and Shaked and Shanthikumar (2007).

We finally recall the concept of distortion functions. These are non-decreasing continuous functions $h : [0, 1] \to [0, 1]$ such that $h(0) = 0$ and $h(1) = 1$. If π is a univariate prior with cumulative distribution function (CDF) F_π, the distorted CDF is given by

$$F_{\pi_h}(\theta) = h \circ F_\pi(\theta) = h\left[F_\pi(\theta)\right], \forall \theta \in \Theta, \tag{1.3}$$

and represents the CDF of a particular random variable called the distorted random variable, with distribution π_h. If the distorted CDF is differentiable, then, taking the derivative in (1.3), we get the distorted density function given by

$$f_{\pi_h}(\theta) = h' \left[F_\pi \left(\theta \right) \right] \pi(\theta), \tag{1.4}$$

see Section 2.6 in Denuit et al. (2005) for a review of the distortion concept. Although priors can be distorted according to different criteria, Arias-Nicolás et al. (2016) argue for the use of convex and concave distortion functions based on two arguments. First, they represent satisfactorily a change in the weighting of the underlying prior. Thus, a convex (concave) distortion function gives more weight to higher (smaller) values of the variable. Second, they have desirable properties when we compare the original prior with the distorted one. If h_1 is concave and h_2 is convex, it follows from Lemma 1 in Arias-Nicolás et al. (2016) that:

$$\pi_{h_1} \leq_{lr} \pi \leq_{lr} \pi_{h_2}. \tag{1.5}$$

Inspired by property (1.5), the distortion band, denoted by $\Gamma_{h_1,h_2,\pi}$, associated with a specific prior π and based on h_1 and h_2, a concave distortion function and a convex distortion function, respectively, is defined as

$$\Gamma_{h_1,h_2,\pi} = \left\{ \pi' : \pi_{h_1} \leq_{lr} \pi' \leq_{lr} \pi_{h_2} \right\}. \tag{1.6}$$

As an immediate consequence, it is shown in Arias-Nicolás et al. (2016) that the posteriors corresponding to π_{h_1} and π_{h_2} are also lower and upper bounds for the class of all posteriors π'_x, in the likelihood ratio order sense, i.e., for all $\pi' \in \Gamma_{h_1,h_2,\pi}$ then

$$\pi_{h_1,x} \leq_{lr} \pi'_x \leq_{lr} \pi_{h_2,x}. \tag{1.7}$$

Other interesting properties of the distorted band are studied in detail in Arias-Nicolás et al. (2016), where the authors show that $\Gamma_{h_1,h_2,\pi}$ fulfills all the requirements that Berger (1994) discussed about the choice of a class of priors. First, its elicitation and interpretation is easy. Second, the prior uncertainty can be reflected by using different metrics. Finally, the range of quantities of interest can be computed by just looking for the extremal distributions generating the class. In conclusion, the distorted band is regarded as a "neighborhood" of the prior π. For further information regarding recent applications of the distorted band see Joshi et al. (2018), Sánchez-Sánchez et al. (2019) and Barrera et al. (2019). Also see Huang and Mi (2018) for applications of the likelihood ratio order in the Bayesian framework.

Note that the distorted band model defined in (1.6) is valid for prior distributions defined over $\Theta \subseteq \mathbb{R}$. As a natural extension, we propose in this paper a model for multiparameter distributions, i.e., for distributions such that the parameter set is $\boldsymbol{\Theta} \subseteq \mathbb{R}^n, n > 1$. Our first attempt is to extend the two key concepts used in the model (1.6), namely the likelihood ratio order and the distortion function, to the multivariate setting. Regarding the likelihood ratio order, this can be easily reached by using the

classical multivariate extension introduced in Karlin and Rinott (1980a) (see also Whitt (1980) and Shaked and Shanthikumar (2007)) that will be recalled later on. Regarding the extension of distortion function, although the univariate concept is widely used in different fields (see e.g., Quiggin (1982), Yaari (1987), Schmeidler (1989), Denneberg (1990), Wang (1995, 1996) and Goovaerts and Laeven (2008)), there is no unique way to extend it to the multivariate setting (see, e.g., Valdez and Xiao (2011), Di Bernardino and Rullière (2013) and Navarro et al. (2016)), for different multivariate extensions. Instead of facing the dilemma of choosing among these equally plausible candidates, we propose a new approach in this paper, based on weighted distributions. Weighted distributions, first introduced by Fisher (1934) and explicitly defined in Rao (1963) (see also Rao (1985)) provide a method to adjust a probability distribution in a way that is different from the one based on distortion functions and that can be unambiguously extended to the multivariate case (see, e.g., Jain and Nanda (1995) and Navarro et al. (2006, 2011)). Given a prior $\boldsymbol{\pi}$ with density function $\pi(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^n$, let w be a weight function, i.e., a non-negative function $w : \mathbb{R}^n \mapsto \mathbb{R}^+$ such that the expectation $E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]$ is strictly positive and finite. A weighted random vector $\boldsymbol{\pi}_w$ is a random vector with density function given by

$$\pi_w(\boldsymbol{\theta}) = \frac{w(\boldsymbol{\theta})}{E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]} \pi(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^n. \tag{1.8}$$

Note that for $n = 1$, in the case of an absolutely continuous prior π with CDF F_π, the distorted CDF associated to a differentiable distortion function h has a density function f_{π_h} that can be written as a weighted density function,

$$f_{\pi_h}(\theta) = h'[F_\pi(\theta)] \pi(\theta) = \omega(\theta)\pi(\theta), \tag{1.9}$$

where the weight function $\omega(\theta) = h'[F_\pi(\theta)]$ depends on F_π (this is noted, for example, in Furman and Zitikis (2008) and Blazej (2008)). In other words, when we consider absolutely continuous random variables, weighted distributions are more general objects than distorted distributions. Moreover, if this is the case, the distortion h is convex (resp. concave) if and only if the weight function w is increasing (resp. decreasing). In this new framework, we construct a weighted band (rather than a distorted band) by making a perturbation over the prior π by considering two weight functions: w_1 (decreasing) and w_2 (increasing). Under these assumptions, it is well-known (for $n = 1$) that

$$\pi_{w_1} \leq_{lr} \pi \leq_{lr} \pi_{w_2}. \tag{1.10}$$

Observe that although distorted distributions and weighted distributions are different in nature, they are used in a similar way to change the weighting of the underlying prior. Thus, an increasing (decreasing) weight function gives more weight to higher (smaller) values of the variable.

The multivariate weighted band is introduced in Section 2, along with its main properties. In particular, we show how this multivariate bands inherits some of the good properties of the univariate distorted band, including the relative easiness in specifying it and interpreting the posterior distributions. This fact is very uncommon in the multivariate case. In Section 3 the Hellinger metric and the Kullback-Leibler divergence

are introduced to incorporate the degree of uncertainty in the elicitation process. We present an application about failures of train doors in Section 4, discussing also how to obtain a sample from the class of posterior distributions. Concluding remarks and some ideas for future researches are finally presented in Section 5.

In this paper, for any random variable Z and an event A, we denote by $[Z|A]$ the random variable whose distribution is the conditional distribution of Z given A. 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. A set $\mathbf{U} \subseteq \mathbb{R}^n$ is called upper (lower) if $\mathbf{y} \in \mathbf{U}$ whenever $\mathbf{y} \geq (\leq)\mathbf{x}$ and $\mathbf{x} \in \mathbf{U}$. Given two real numbers x and y, we denote $x \vee y = \max\{x,y\}$ and $x \wedge y = \min\{x,y\}$. For two real vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \vee \mathbf{y}$ and $\mathbf{x} \wedge \mathbf{y}$ mean the componentwise maximum and the componentwise minimum, respectively. Finally, we use the terms increasing and decreasing in a wide sense, that is, a function $g : \mathbb{R}^n \mapsto \mathbb{R}$ is increasing (decreasing) if $g(\mathbf{x}) \leq (\geq)g(\mathbf{y})$ for all $\mathbf{x} \leq \mathbf{y}$. Let $\mathbf{X}$ be a n-dimensional random vector and let $I \subset \{1, \ldots, n\}$, then $\mathbf{X}_I$ denotes a random vector constructed by using only the coordinates of $\mathbf{X}$ in I.

2   A multivariate class of priors: the weighted band

In order to ease the exposition, we will only consider random vectors with absolutely continuous distributions having densities with respect to a product measure on $\mathbb{R}^n$. However, it is possible to define all key concepts for discrete distributions and the results are essentially the same. Let's start by recalling the definition of some multivariate stochastic orders that we will use in the paper.

Definition 2.1. *Let $\mathbf{X}$ and $\mathbf{Y}$ be two n-dimensional random vectors with CDFs F and G, respectively. Then, it is said that $\mathbf{X}$ is smaller than $\mathbf{Y}$ in the usual multivariate stochastic order, denoted by $\mathbf{X} \leq_{st} \mathbf{Y}$, if*

$$P\{\mathbf{X} \in \mathbf{U}\} \leq P\{\mathbf{Y} \in \mathbf{U}\}, \text{ for all upper sets } \mathbf{U} \subseteq \mathbb{R}^n.$$

Roughly speaking, $\mathbf{X}$ is less likely than $\mathbf{Y}$ to take on large values, where "large" means any value in an upper set $\mathbf{U}$, for any upper set $\mathbf{U}$. As an immediate consequence, just taking in account that the complement of any upper set is a lower set, and vice versa, we can also compare the CDFs, i.e.,

$$F(\mathbf{x}) \geq G(\mathbf{x}), \forall \mathbf{x} \in \mathbb{R}^n. \tag{2.1}$$

Note that condition (2.1) is also sufficient only in the univariate case. The multivariate stochastic order has a characterization in terms of the expectations of increasing transformations, i.e., $\mathbf{X} \leq_{st} \mathbf{Y}$ if, and only if,

$$E[\phi(\mathbf{X})] \leq E[\phi(\mathbf{Y})], \tag{2.2}$$

for all increasing real functions ϕ on $\mathbb{R}^n$ and provided the expectations exist.

Definition 2.2. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors with PDFs* f *and* g, *respectively. Then, it is said that* $\mathbf{X}$ *is smaller than* $\mathbf{Y}$ *in the multivariate likelihood ratio stochastic order, denoted by* $\mathbf{X} \leq_{lr} \mathbf{Y}$, *if*

$$f(\mathbf{x})g(\mathbf{y}) \leq f(\mathbf{x} \wedge \mathbf{y})g(\mathbf{x} \vee \mathbf{y}), \text{ for every } \boldsymbol{x} \text{ and } \boldsymbol{y} \text{ in } \mathbb{R}^n. \tag{2.3}$$

The following implication is also well known:

$$\mathbf{X} \leq_{lr} \mathbf{Y} \Rightarrow \quad \mathbf{X} \leq_{st} \mathbf{Y}. \tag{2.4}$$

The definition of multivariate total positivity of order 2 will be of key importance for our purpose.

Definition 2.3. *A function* $l : \mathbb{R}^n \mapsto \mathbb{R}^+$, *(*$n \in \mathbb{N}, n \geq 2$*) is said to be multivariate totally positive of order* 2 *(*MTP_2*),* *(*$TP2$ *when* $n = 2$*) if* l *satisfies*

$$l(\mathbf{x})l(\mathbf{y}) \leq l(\mathbf{x} \wedge \mathbf{y})l(\mathbf{x} \vee \mathbf{y}), \ \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n. \tag{2.5}$$

Additionally, a n*-dimensional random vector* $\mathbf{X}$ *with PDF* f *is said to be* MTP_2 *if its density* f *is* MTP_2 *or, equivalently, if* $\mathbf{X} \leq_{lr} \mathbf{X}$.

We will mention briefly some interesting results concerning MTP_2 and $\leq_{lr}$ concepts that we will use later on.

Lemma 2.4. *The product of* MTP_2 *functions is still* MTP_2. *Additionally, a function* $l : \mathbb{R}^n \mapsto \mathbb{R}^+$ *defined by* $l(\mathbf{x}) = \prod_{i=1}^n g(x_i)$, *where* g_i, $i = 1, \ldots, n$, *are univariate non-negative functions, is* MTP_2.

Lemma 2.5. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors with PDFs* f *and* g, *respectively. If either* $\mathbf{X}$ *or* $\mathbf{Y}$ *is* MTP_2 *then* $\mathbf{X} \leq_{lr} \mathbf{Y}$ *if and only if* $g(\mathbf{x})f(\mathbf{y}) \leq g(\mathbf{y})f(\mathbf{x})$, $\forall \mathbf{x} \leq \mathbf{y}$, *or, equivalently, if* $g(\mathbf{x})/f(\mathbf{x})$ *is increasing in the union of their supports.*

Lemma 2.6. *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be two random vectors such that* $\mathbf{X} \leq_{lr} \mathbf{Y}$, *then for any* $I \subset \{1, \ldots, n\}$ *we have* $\mathbf{X}_I \leq_{lr} \mathbf{Y}_I$.

It is worth mentioning that the relation $\mathbf{X} \leq_{lr} \mathbf{Y}$ always implies that the ratio $g(\mathbf{x})/f(\mathbf{x})$ is increasing in $\mathbf{x}$. The converse is only true in the univariate case.

Both concepts MTP_2 and $\leq_{lr}$ are explicitly defined in Karlin and Rinott (1980a) but they are also implicit in many classical papers in the literature, see for instance Lorentz (1953), Karlin (1968), Sarkar (1969), Rinott (1973), Holley (1974), Kemperman (1977) and Preston (1974). Additionally, an excellent review of all previous concepts and multivariate stochastic orders can be found in Müller and Stoyan (2002) and Shaked and Shanthikumar (2007). It is also remarkable the equivalence between log-supermodular and MTP_2 functions, where a function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is supermodular if $f(\mathbf{x} \wedge \mathbf{y}) + f(\mathbf{x} \vee \mathbf{y}) \geq f(\mathbf{x}) + f(\mathbf{y})$. This equivalence leads to an interesting analytical characterization of strictly positive MTP_2 functions, see for example Topkis (1978, 1998).

Lemma 2.7. *A function $l : \mathbb{R}^n \mapsto \mathbb{R}^+$ with 2 continuous derivatives is MTP_2 if and only if*

$$\frac{\partial^2}{\partial x_i \partial x_j} \ln(l(\mathbf{x})) \geq 0, \, \forall i \neq j.$$

Although at first glance MTP_2 property seems to be a technical condition, we find in Karlin and Rinott (1980a) a large list of multivariate distributions that satisfy the MTP_2 condition. For example, as a direct consequence of Lemma 2.4, it is apparent that independent random variables have a joint MTP_2 density. Also, the MTP_2 concept is fundamental for many probability inequalities that have important applications to multivariate analysis, multiple hypothesis testing and approximating probabilities, see for example Karlin and Rinott (1980a,b, 1981), Pearlman and Olkin (1980), Glaz and Johnson (1984), Sarkar and Chang (1997) and Sarkar (1998, 2008).

To conclude, it is also well-known that MTP_2 is a positive dependence property that implies the classical positive associated concept. If $\mathbf{X}$ is MTP_2 then the covariance

$$Cov(\phi(\mathbf{X}), \psi(\mathbf{X})) \geq 0,$$

provided ϕ and ψ are simultaneously monotone increasing or decreasing, see Esary et al. (1967), Sarkar (1969) and Karlin and Rinott (1980a).

As we have mentioned in the introduction, we find in the theory of weighted distributions a useful tool to incorporate the uncertainty about specification of a prior belief. Let $\boldsymbol{\pi}$ be a multivariate prior with density function $\pi(\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^n$ and let $w : \mathbb{R}^n \mapsto \mathbb{R}^+$ be a weight function. The weighted prior $\boldsymbol{\pi}_w$ having density function $\pi_w(\boldsymbol{\theta})$ defined in (1.8) represents a weighted density to measure the uncertainty about the specific prior density.

A weighted density can be defined according to different criteria, as we will show later on. We restrict our attention to increasing and decreasing weight functions for two reasons. First, as explained before, for $n = 1$ they represent a generalization of the convex and concave distortion functions used in the distorted band. And second, they enable an ordering when we compare the specific prior with the weighted one as we can see in the following result, which extends to $n \geq 2$ the property (1.10).

Lemma 2.8. *Let $\boldsymbol{\pi}$ be a specific MTP_2 prior belief. Let w be an increasing (decreasing) weight function. Then $\boldsymbol{\pi} \leq_{lr} (\geq_{lr})\boldsymbol{\pi}_w$.*

Proof. Let us consider an increasing weight function w and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ such that $\boldsymbol{\theta}_1 \leq \boldsymbol{\theta}_2$. The product $\pi_w(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2)$ satisfies that

$$
\begin{aligned}
\pi_w(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2) &= \frac{w(\boldsymbol{\theta}_1)}{E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]}\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2) \\
&\leq \frac{w(\boldsymbol{\theta}_2)}{E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]}\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2) \\
&= \pi_w(\boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_1),
\end{aligned}
$$

where the equalities follow directly from expression (1.8) and the inequality comes from the fact that w is increasing. The proof concludes just using Lemma 2.5. For a decreasing weight function the proof holds using a similar argument. □

It follows from (2.4) and Lemma 2.8 that, when the prior belief is MTP_2 and w is increasing (decreasing), the weighted prior belief $\boldsymbol{\pi}_w$ is more (less) likely than the prior belief $\boldsymbol{\pi}$ to take on larger values. This result extends the property (1.10) from the case $n = 1$ to $n \geq 2$. Moreover, for $n = 1$ this result plays an equivalent role to Lemma 1 in Arias-Nicolás et al. (2016) and Theorem 1 of Blazej (2008).

Now, let us suppose that the decision-maker is able to represent the changes in a prior belief, $\boldsymbol{\pi}$, by a decreasing weight function, w_1, and an increasing weight function, w_2. A direct consequence of this fact, jointly to Lemma 2.8, allows us to define two weighted priors, $\boldsymbol{\pi}_{w_1}$ and $\boldsymbol{\pi}_{w_2}$, such that $\boldsymbol{\pi}_{w_1} \leq_{lr} \boldsymbol{\pi} \leq_{lr} \boldsymbol{\pi}_{w_2}$. Then, we present the following neighborhood band for $\boldsymbol{\pi}$ inspired by these inequalities.

Definition 2.9. *Let $\boldsymbol{\pi}$ be a specific MTP_2 prior belief. We will define the weighted band $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$ associated with $\boldsymbol{\pi}$ based on w_1 and w_2, a decreasing weight function and an increasing weight function, respectively, (weighted band, for short), as*

$$\Gamma_{w_1,w_2,\boldsymbol{\pi}} = \left\{ \boldsymbol{\pi}' : \boldsymbol{\pi}_{w_1} \leq_{lr} \boldsymbol{\pi}' \leq_{lr} \boldsymbol{\pi}_{w_2} \right\}.$$

As a consequence of Lemma 2.8, it is evident that $\boldsymbol{\pi} \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}$. Therefore the weighted band can be considered as a particular "neighborhood" band of $\boldsymbol{\pi}$, where the weighted priors are the lower and upper bound distributions of the class. From Definition 2.9 uncertainty could be introduced just through an upper (lower) bound by considering w_2 (w_1) the identity function.

In order to give a meaningful understanding of the weighted band, we provide the following interpretations. First of all, from (2.4) and (2.1) it is apparent that the weighted band is a subclass of the following multivariate class of priors:

$$\begin{aligned}
\Gamma_{w_1,w_2,\boldsymbol{\pi}} &\subseteq \left\{ \boldsymbol{\pi}' : \boldsymbol{\pi}_{w_1} \leq_{st} \boldsymbol{\pi}' \leq_{st} \boldsymbol{\pi}_{w_2} \right\} \\
&\subseteq \left\{ \boldsymbol{\pi}' : F_{\boldsymbol{\pi}_{w_1}}(\boldsymbol{\theta}) \geq F_{\boldsymbol{\pi}'}(\boldsymbol{\theta}) \geq F_{\boldsymbol{\pi}_{w_2}}(\boldsymbol{\theta}), \forall \boldsymbol{\theta} \in \boldsymbol{\Theta} \right\}.
\end{aligned} \tag{2.6}$$

It is worth mentioning that the class given in (2.6) is a multivariate generalization of the classical band class defined in Basu and DasGupta (1995) for univariate priors. Additionally, the weighted band has also some interesting interpretations in terms of prior probability sets. For example, given a prior $\boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}$ we obtain that

$$\boldsymbol{\pi}_{w_1}(\cdot|\mathbf{A}) \leq_{lr} \boldsymbol{\pi}'(\cdot|\mathbf{A}) \leq_{lr} \boldsymbol{\pi}_{w_2}(\cdot|\mathbf{A}),$$

for any measurable set $\mathbf{A} \subseteq \boldsymbol{\Theta} \subseteq \mathbb{R}^n$ of the form $A = A_1 \times \cdots \times A_n$ where $A_i \subseteq \mathbb{R}$, $i = 1, \ldots, n$, see Rinott and Scarsini (2006) and Shaked and Shanthikumar (2007) for further information in this sense.

It is worth mentioning that the likelihood ratio order does not apply, in general, when comparing two priors $\boldsymbol{\pi}'_1$ and $\boldsymbol{\pi}'_2$ in $\Gamma_{w_1,w_2,\pi}$. Each of them is just ordered w.r.t. $\boldsymbol{\pi}_{w_1}$ and $\boldsymbol{\pi}_{w_2}$. However, there are infinitely many priors in the weighted band as it is

shown in the following result. The next proposition connects somehow the weighted band with the classical contamination class.

Proposition 2.10. *The class* $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$ *contains all the priors* $\boldsymbol{\pi}_\epsilon = (1 - \epsilon)\boldsymbol{\pi}_1 + \epsilon\boldsymbol{\pi}_2$, *for any pair of priors* $\boldsymbol{\pi}_1$ *and* $\boldsymbol{\pi}_2$ *in* $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$, *obtained as a mixture of* $\boldsymbol{\pi}_1$ *and* $\boldsymbol{\pi}_2$ *and for any* $0 \le \epsilon \le 1$. *In particular, it contains the mixture between the underlying prior* $\boldsymbol{\pi}$ *and any other prior in the band.*

Proof. In order to prove that $\boldsymbol{\pi}_\epsilon \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}$ we will first prove that $\boldsymbol{\pi}_{w_1} \le_{lr} \boldsymbol{\pi}_\epsilon$. From Definition 2.2 and expression (1.2) a straightforward computation shows that

$$
\begin{aligned}
\pi_{w_1}(\boldsymbol{\theta}_1)\pi_\epsilon(\boldsymbol{\theta}_2) &= (1 - \epsilon)\pi_{w_1}(\boldsymbol{\theta}_1)\pi_1(\boldsymbol{\theta}_2) + \epsilon\pi_{w_1}(\boldsymbol{\theta}_1)\pi_2(\boldsymbol{\theta}_2) \\
&\le (1 - \epsilon)\pi_{w_1}(\boldsymbol{\theta}_1 \wedge \boldsymbol{\theta}_2)\pi_1(\boldsymbol{\theta}_1 \vee \boldsymbol{\theta}_2) + \epsilon\pi_{w_1}(\boldsymbol{\theta}_1 \wedge \boldsymbol{\theta}_2)\pi_2(\boldsymbol{\theta}_1 \vee \boldsymbol{\theta}_2) \\
&= \pi_{w_1}(\boldsymbol{\theta}_1 \wedge \boldsymbol{\theta}_2)\pi_\epsilon(\boldsymbol{\theta}_1 \vee \boldsymbol{\theta}_2),
\end{aligned}
$$

where the equalities follow from the expression of the mixture density and the inequality from the fact that $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ belong to $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$. Similarly, it is possible to prove that $\boldsymbol{\pi}_\epsilon \le_{lr} \boldsymbol{\pi}_{w_2}$. □

Since Definition 2.9 is based on w_1 and w_2, there are many possible bands just by considering different increasing and decreasing weight functions. It is apparent the choice of those weight functions cannot be arbitrary and it should adjust perfectly to the problem we are studying. Some useful weight functions can be found in Jain and Nanda (1995), Nanda and Jain (1997) and Navarro et al. (2006, 2011). In those papers, the authors study properties related to reliability measures, orderings, characterizations and dependency analysis. Of special interest is the multivariate size biased distribution derived from the weight function

$$
w(\boldsymbol{\theta}) = \prod_{i=1}^{n} \theta_i,
$$

which represents sampling procedures where a vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$ has a sampling probability proportional to θ_i, for $i = 1, \dots, n$. Another classical case is given by

$$
w(\boldsymbol{\theta}) = \frac{1 - F_{\boldsymbol{\pi}}(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta})},
$$

which leads to equilibrium distributions in reliability. Also, other interesting weight functions are given by

$$
w(\boldsymbol{\theta}) = \prod_{i=1}^{n} g_i(\theta_i) \quad \text{and} \quad w(\boldsymbol{\theta}) = \sum_{i=1}^{n} g_i(\theta_i),
$$

where, if g_i's are non-negative increasing (decreasing) functions, then w will be also increasing (decreasing). For example, the weight functions

$$
w(\boldsymbol{\theta}) = \prod_{i=1}^{n} \theta_i^{a_i-1} \quad \text{and} \quad w(\boldsymbol{\theta}) = \sum_{i=1}^{n} \theta_i^{a_i-1}, \tag{2.7}
$$

are increasing (decreasing) when $a_i > 1$ ($a_i < 1$), for $i = 1, \dots, n$.

Particular attention about the effect of the weight functions in the dependence structure is discussed in Navarro et al. (2006). For example, it is clear from Lemma 2.4 that if $\boldsymbol{\pi}$, $l(\boldsymbol{\theta}|\mathbf{x})$ and w are MTP_2, then the weighted prior and the weighted posterior are also MTP_2. Additionally, if $\boldsymbol{\pi}$ has independent components and w is of the form $w(\boldsymbol{\theta}) = \prod_{i=1}^{n} g_i(\theta_i)$, then the weighted prior has also independent components.

Finally, we present two bivariate weight functions that we will use later on. We will use them for mathematical convenience. We will be able to compute the conditional distributions to estimate the characteristics of the weighted priors and posteriors using Markov chain Monte Carlo (MCMC) methods in the real example

$$w_1(\boldsymbol{\theta}) = \theta_1^{a-1}\theta_2^{b-1}\exp[-c\theta_1\theta_2] \ \ \text{and} \ \ w_2(\boldsymbol{\theta}) = \theta_1^{a'-1}\theta_2^{b'-1}(\theta_1^{c'} + \theta_2^{c'}), \qquad (2.8)$$

where it is apparent that, if $a, b < 1$ and $c > 0$, then w_1 is decreasing and, if $a', b' > 1$ and $c' > 0$, then w_2 is increasing.

In order to clarify the previous ideas we present the following example.

Example 2.11. *Suppose that the specific prior belief $\boldsymbol{\pi}$ over $\boldsymbol{\Theta} = \mathbb{R}^+ \times \mathbb{R}^+$ is a bivariate random vector having independent exponential marginal distributions $Exp(\lambda_i)$, $i = 1, 2$, with joint PDF given by*

$$\pi(\theta_1, \theta_2) = \lambda_1\lambda_2\exp(-\lambda_1\theta_1)\exp(-\lambda_2\theta_2), \ (\theta_1, \theta_2) \in \boldsymbol{\Theta},$$

where λ_i, $i = 1, 2$, are the hyperparameters. It is apparent from Lemma 2.4 that $\pi(\theta_1, \theta_2)$ is MTP_2. Let consider the weight function w defined in the left-hand side of (2.7) *as*

$$w(\theta_1, \theta_2) = \theta_1^{a_1-1}\theta_2^{a_2-1}, \qquad (2.9)$$

where $a_i > 0$, $i = 1, 2$. A straightforward computation shows that the weighted prior $\boldsymbol{\pi}_w$ is a bivariate random variable having independent gamma marginal distributions $G(a_i, \lambda_i)$, $i = 1, 2$, with joint PDF given by

$$\pi_w(\theta_1, \theta_2) \ \ = \ \ \frac{\lambda_1^{a_1}}{\Gamma(a_1)}\frac{\lambda_2^{a_2}}{\Gamma(a_2)}\theta_1^{a_1-1}\theta_2^{a_2-1}exp(-\lambda_1\theta_1)\exp(-\lambda_2\theta_2), \ (\theta_1, \theta_2) \in \boldsymbol{\Theta}.$$

Let us consider w_1 and w_2 defined as in (2.9)*, taking $(a_1, a_2) = (1/4, 1/4)$ and $(a_1, a_2) = (7/4, 7/4)$, respectively, and also consider the weighted band $\Gamma_{w_1, w_2, \boldsymbol{\pi}}$. By taking the values of the hyperparameters $\lambda_1 = 2$ and $\lambda_2 = 3$, the weighted PDFs $\pi_{w_1}(\boldsymbol{\theta})$ (in blue) and $\pi_{w_2}(\boldsymbol{\theta})$ (in green) are represented in Figure 1, where we can see how they differ from the baseline prior PDF (a) and from the baseline prior CDF (b) (in yellow). From the inclusion given in* (2.6)*, the CDF of the specific prior lies between the CDFs of the weighted versions.*

The next result shows that posterior distributions inherit the likelihood ratio order under MTP_2 likelihood functions in the parameters. That proposition will be the key to study the robustness in practice as we will see later on.

Proposition 2.12. *Let $\boldsymbol{\pi}$ be a specific MTP_2 prior and let $\Gamma_{w_1, w_2, \boldsymbol{\pi}}$ be a weighted band associated with $\boldsymbol{\pi}$ based on w_1 and w_2. Given the observed data $\mathbf{x}$, if $l(\boldsymbol{\theta} \mid \mathbf{x})$ is MTP_2 in $\boldsymbol{\theta}$, then for all $\boldsymbol{\pi}' \in \Gamma_{w_1, w_2, \boldsymbol{\pi}}$ we obtain that $\boldsymbol{\pi}_{w_1, \mathbf{x}} \leq_{lr} \boldsymbol{\pi}'_{\mathbf{x}} \leq_{lr} \boldsymbol{\pi}_{w_2, \mathbf{x}}$.*
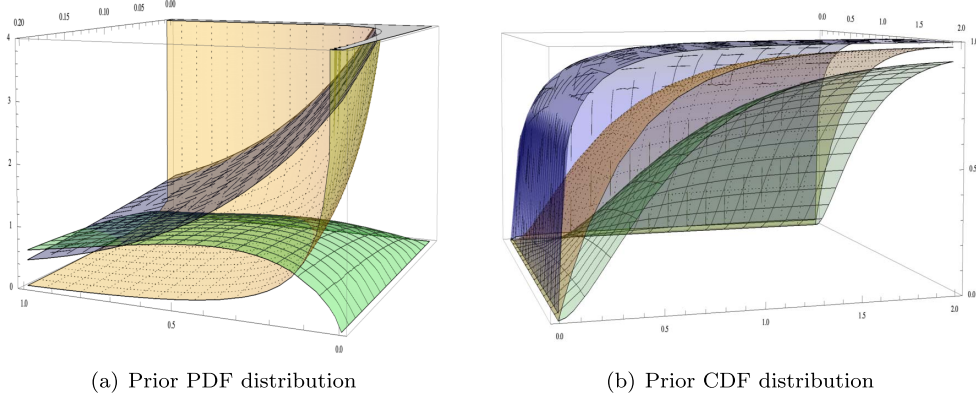
(a) Prior PDF distribution   (b) Prior CDF distribution

Figure 1: Prior knowledge: $\boldsymbol{\pi}_{w_1}$ (in blue), $\boldsymbol{\pi}$ (in yellow) and $\boldsymbol{\pi}_{w_2}$ (in green).

Proof. We will first prove that $\boldsymbol{\pi}_{w_1,\mathbf{x}} \leq_{lr} \boldsymbol{\pi}'_{\mathbf{x}}$. From Definition 2.2 and expression (1.2) a straightforward computation shows that

$$
\begin{aligned}
\pi_{w_1,\mathbf{x}}(\boldsymbol{\theta}_1)\pi'_{\mathbf{x}}(\boldsymbol{\theta}_2) &= \frac{l(\boldsymbol{\theta}_1\mid\mathbf{x})\pi_{w_1}(\boldsymbol{\theta}_1)}{m_{\boldsymbol{\pi}_{w_1}}(\mathbf{x})}\frac{l(\boldsymbol{\theta}_2\mid\mathbf{x})\pi'(\boldsymbol{\theta}_2)}{m_{\boldsymbol{\pi}'}(\mathbf{x})} \\
&\leq \frac{l(\boldsymbol{\theta}_1\wedge\boldsymbol{\theta}_2\mid\mathbf{x})\pi_{w_1}(\boldsymbol{\theta}_1\wedge\boldsymbol{\theta}_2)}{m_{\boldsymbol{\pi}_{w_1}}(\mathbf{x})}\frac{l(\boldsymbol{\theta}_1\vee\boldsymbol{\theta}_2\mid\mathbf{x})\pi'(\boldsymbol{\theta}_1\vee\boldsymbol{\theta}_2)}{m_{\boldsymbol{\pi}'}(\mathbf{x})} \\
&= \pi_{w_1,\mathbf{x}}(\boldsymbol{\theta}_1\wedge\boldsymbol{\theta}_2)\pi'_{\mathbf{x}}(\boldsymbol{\theta}_1\vee\boldsymbol{\theta}_2),
\end{aligned}
$$

where the inequality follows from the fact that $l(\boldsymbol{\theta}\mid\mathbf{x})$ is MTP_2 in $\boldsymbol{\theta}$ and $\boldsymbol{\pi}_{w_1}\leq_{lr}\boldsymbol{\pi}'$. Using a similar argument it is shown that $\boldsymbol{\pi}'_{\mathbf{x}}\leq_{lr}\boldsymbol{\pi}_{w_2,\mathbf{x}}$. □

Proposition 2.12 says that the posterior distributions of the lower and upper bound of the weighted band are also lower and upper bounds for the family of all posterior distributions in the $\leq_{lr}$ order sense. It is apparent that the posteriors $\boldsymbol{\pi}_{w_1,\mathbf{x}}$ and $\boldsymbol{\pi}_{w_2,\mathbf{x}}$ can be also interpreted as two weighted distributions of the specific posterior $\boldsymbol{\pi}_{\mathbf{x}}$ based on w_1 and w_2, respectively.

Remark 2.13. *From Lemma 2.6, it is clear that the orderings in Proposition 2.12 are also valid for the marginal distributions of the posteriors.*

Remark 2.14. *If we consider another model with a MTP_2 likelihood function, denoted by $l^*(\boldsymbol{\theta}\mid\mathbf{x})$, such that the ratio $l(\boldsymbol{\theta}\mid\mathbf{x})/l^*(\boldsymbol{\theta}\mid\mathbf{x})$ is increasing in $\boldsymbol{\theta}$ in the union of their supports, it is apparent just using jointly Lemma 2.4 and 2.5 that the posterior distribution associated with the specific prior $\boldsymbol{\pi}$ and the new model $l^*(\boldsymbol{\theta},\mathbf{x})$, denoted by $\boldsymbol{\pi}^*_{\mathbf{x}}$, satisfies that $\boldsymbol{\pi}_{w_1,\mathbf{x}}\leq_{lr}\boldsymbol{\pi}^*_{\mathbf{x}}\leq_{lr}\boldsymbol{\pi}_{w_2,\mathbf{x}}$. Therefore, in some way the weighted band also provides sensitivity to the model specification. This fact could be the start point for a future research line.*

Remark 2.15. *For the one-parameter model, i.e., when $\theta\in\Theta\subseteq\mathbb{R}$, we only need a decreasing weight function w_1 and an increasing weight function w_2 to apply Proposition*

2.12. As explained before, this procedure provides a band that is different in nature from the distorted band defined in Arias-Nicolás et al. (2016).

At this point one may wonder if assumptions in Proposition 2.12 are very strict. To answer this question we first emphasize that density priors having independent components are MTP_2 and most of the applications in literature are concerned with those priors. Second, as it is argued in Gâsemyr and Natvig (1995, 1997), we find situations where some particular MTP_2 priors can be more realistic than one assuming independent components. Finally, with respect to the likelihood function we find some well-known models that satisfy the MTP_2 property as we can see in the following two examples and in the final application.

Example 2.16. *Let $X_1, \ldots, X_n$ be an i.i.d. random sample from the Pareto income distribution $P(x_0, \alpha)$ with density function $f(x) = \alpha x_0^\alpha / x^{\alpha+1}$, $x \geq x_0$, where $\boldsymbol{\theta} = (x_0, \alpha) \in \boldsymbol{\Theta} = \mathbb{R}^+ \times \mathbb{R}^+$ represents the unknown parameter and x_0 and α are the mode and shape parameters, respectively. It is well-known that the likelihood function is given by*

$$l(\boldsymbol{\theta} \mid \mathbf{x}) = l(\alpha, x_0 \mid \mathbf{x}) = \alpha^n \frac{x_0^{n\alpha}}{(\prod_{i=1}^n x_i^{\alpha+1})} I_{[x_0, +\infty)}(x_{(1)}),$$

where $I_{[x_0, +\infty)}$ and $x_{(1)}$ are the indicator function of $[x_0, +\infty)$ and the sample minimum, respectively. From Lemma 2.4, in order to check if $l(\boldsymbol{\theta} \mid \mathbf{x})$ is MTP_2 in $\boldsymbol{\theta}$, it is sufficient to show that the function $g(\alpha, x_0) = x_0^{n\alpha}$ defined in $\mathbb{R}^+ \times \mathbb{R}^+$ is MTP_2. This last follows easily by computing $\partial^2 \ln(g(\alpha, x_0))/\partial\alpha\partial x_0 = n/x_0$ and using Lemma 2.7.

Example 2.17. *Let $X_1, \ldots, X_n$ be an i.i.d. random sample from a Gamma distribution $\mathbb{G}(\alpha, \lambda)$ with density function $f(x) = \lambda^\alpha x^{\alpha-1} e^{-\lambda x}/\Gamma(\alpha)$, $x \geq 0$, where $\boldsymbol{\theta} = (\alpha, \lambda) \in \boldsymbol{\Theta} = \mathbb{R}^+ \times \mathbb{R}^+$ represents the unknown parameter and α and λ are the shape and scale parameters, respectively. The likelihood function is given by*

$$l(\boldsymbol{\theta} \mid \mathbf{x}) = l(\alpha, \lambda \mid \mathbf{x}) = \frac{\lambda^{n\alpha}}{\Gamma(\alpha)^n}(\Pi_{i=1}^n x_i^{\alpha-1})e^{-\lambda \sum_{i=1}^n x_i}.$$

Using the same argument as in the previous example, it is sufficient to show that the function $g(\alpha, \lambda) = \lambda^{n\alpha}$ defined in $\mathbb{R}^+ \times \mathbb{R}^+$ is MTP_2. This last follows easily by computing $\partial^2 \ln(g(\alpha, \lambda))/\partial\alpha\partial\lambda = n/\lambda$ and using Lemma 2.7.

To conclude, Corollary 2.18 will allow us to quantify and interpret the uncertainty induced by the partial knowledge of the prior. Let us suppose we have a functional H that maps the underlying distribution X to a real number of interest. This functional obviously inherits the dependency of the parameter. We will denote by $H_X(\boldsymbol{\theta})$ that number. Under the previous assumption we obtain the following result.

Corollary 2.18. *Let X be the underlying random variable and let H be a functional of interest such that $H_X(\boldsymbol{\theta})$ is non-decreasing in $\boldsymbol{\theta}$. Given $\boldsymbol{\pi}$, a specific MTP_2 prior, and the corresponding weighted band $\Gamma_{w_1, w_2, \boldsymbol{\pi}}$ based on w_1 and w_2, if $l(\boldsymbol{\theta} \mid \mathbf{x})$ is MTP_2 in $\boldsymbol{\theta}$, then the univariate random variables obtained by mapping the posterior distributions by the functional $H_X(\boldsymbol{\theta})$ satisfy*

$$H_X(\boldsymbol{\pi}_{w_1, \mathbf{x}}) \leq_{st} H_X(\boldsymbol{\pi}'_{\mathbf{x}}) \leq_{st} H_X(\boldsymbol{\pi}_{w_2, \mathbf{x}}), \ \forall \boldsymbol{\pi}' \in \Gamma_{w_1, w_2, \boldsymbol{\pi}}. \tag{2.10}$$

Proof. From Proposition 2.12 and using (2.4) we obtain that

$$\boldsymbol{\pi}_{w_1,\mathbf{x}} \leq_{st} \boldsymbol{\pi}'_{\mathbf{x}} \leq_{st} \boldsymbol{\pi}_{w_2,\mathbf{x}}.$$

The rest of the proof follows easily from Theorem 6.B.16 in Shaked and Shanthikumar (2007). □

Under the assumptions of Corollary 2.18 we can obtain bounds for some characteristic of interest in the Bayesian framework. For example, using the characterization of the stochastic order given in (2.2) and taking ϕ as the identity function, we obtain that the predictive expectations of H_X satisfy

$$E^{\boldsymbol{\pi}_{w_1,\mathbf{x}}}[H_X(\boldsymbol{\theta})] \leq E^{\boldsymbol{\pi}_{\mathbf{x}}}[H_X(\boldsymbol{\theta})] \leq E^{\boldsymbol{\pi}_{w_2,\mathbf{x}}}[H_X(\boldsymbol{\theta})], \ \forall \boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}. \tag{2.11}$$

Also, using expression (1.A.12) in Shaked and Shanthikumar (2007), we obtain that quantiles of H_X are also ordered

$$F^{-1}_{H_X(\boldsymbol{\pi}_{w_1,\mathbf{x}})}(p) \leq F^{-1}_{H_X(\boldsymbol{\pi}'_{\mathbf{x}})}(p) \leq F^{-1}_{H_X(\boldsymbol{\pi}_{w_2,\mathbf{x}})}(p), \ \forall p \in (0,1), \forall \boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}. \tag{2.12}$$

It is remarkable to observe the range of quantities of interest which can be computed just looking for the extremal distributions generating the weighted class.

Remark 2.19. *From Remark 2.13, all inequalities are also valid for the marginal distributions of the posteriors.*

3   Using metrics to measure uncertainty

In robust Bayesian analysis, probability metrics are used to incorporate uncertainty in the elicitation process by considering an error in the specification, see e.g. Basu and DasGupta (1995). An excellent review of the most common probability metrics and divergences and the relationships among them is given in Gibbs and Su (2002). For one-parameter distributions, Arias-Nicolás et al. (2016) use the Kolmogorov and Kantorovich metrics due to their mathematical tractability when comparing a specific prior and its distorted version. However, in our multiparameter context, we find in the Hellinger distance and the Kullback-Leibler divergence excellent tools to evaluate the degree of uncertainty between multivariate distributions and their weighted versions.

We first recall the definition of the Hellinger metric. Given two multivariate random vectors $\mathbf{X}$ and $\mathbf{Y}$ over $\boldsymbol{\Omega} \subseteq \mathbb{R}^n$ with PDFs $f_{\mathbf{X}}$ and $f_{\mathbf{Y}}$, respectively, the Hellinger metric is defined by

$$\begin{aligned} H(\mathbf{X}, \mathbf{Y}) &= \frac{1}{2} \int_{\Omega} (\sqrt{f_{\mathbf{X}}(\mathbf{x})} - \sqrt{f_{\mathbf{Y}}(\mathbf{x})})^2 d\mathbf{x} \\ &= 1 - \int_{\Omega} \sqrt{f_{\mathbf{X}}(\mathbf{x}) f_{\mathbf{Y}}(\mathbf{x})} d\mathbf{x}, \end{aligned} \tag{3.1}$$

which is used to quantify the similarity between two probability distributions. On the other hand, the Kullback-Leibler divergence (KL) (also called relative entropy) is defined by

$$KL(\mathbf{X}, \mathbf{Y}) \quad = \quad \int_\Omega f_{\mathbf{X}}(\mathbf{x}) \ln \frac{f_{\mathbf{X}}(\mathbf{x})}{f_{\mathbf{Y}}(\mathbf{x})} d\mathbf{x}, \tag{3.2}$$

which is a measure of how the PDF of $\mathbf{X}$ is different from the corresponding one of $\mathbf{Y}$.

Let $\boldsymbol{\pi}$ be a specific prior and let w, w_1 and w_2 be different weight functions. From the fact that posteriors of the weighted versions can be also interpreted as weighted distributions of the specific posterior having the same weight function, we easily obtain the distances given in Table 1.

Distributions	Hellinger metric	KL divergence
$\boldsymbol{\pi}, \boldsymbol{\pi}_w$	$1 - \dfrac{E^{\boldsymbol{\pi}}[\sqrt{w(\boldsymbol{\theta})}]}{\sqrt{E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]}}$	$\ln\left(\dfrac{E^{\boldsymbol{\pi}}[w(\boldsymbol{\theta})]}{\exp(E^{\boldsymbol{\pi}}[\ln(w(\boldsymbol{\theta}))])}\right).$
$\boldsymbol{\pi}_{\mathbf{x}}, \boldsymbol{\pi}_{w,\mathbf{x}}$	$1 - \dfrac{E^{\boldsymbol{\pi}_{\mathbf{x}}}[\sqrt{w(\boldsymbol{\theta})}]}{\sqrt{E^{\boldsymbol{\pi}_{\mathbf{x}}}[w(\boldsymbol{\theta})]}}$	$\ln\left(\dfrac{E^{\boldsymbol{\pi}_{\mathbf{x}}}[w(\boldsymbol{\theta})]}{\exp(E^{\boldsymbol{\pi}_{\mathbf{x}}}[\ln(w(\boldsymbol{\theta}))])}\right).$
$\boldsymbol{\pi}_{w_1}, \boldsymbol{\pi}_{w_2}$	$1 - \dfrac{E^{\boldsymbol{\pi}}[\sqrt{w_1(\boldsymbol{\theta})w_2(\boldsymbol{\theta})}]}{\sqrt{E^{\boldsymbol{\pi}}[w_1(\boldsymbol{\theta})]E^{\boldsymbol{\pi}}[w_2(\boldsymbol{\theta})]}}$	$\ln\left(\dfrac{E^{\boldsymbol{\pi}}[w_2(\boldsymbol{\theta})]/E^{\boldsymbol{\pi}}[w_1(\boldsymbol{\theta})]}{\exp(E^{\boldsymbol{\pi}}[\ln(w_2(\boldsymbol{\theta})/w_1(\boldsymbol{\theta}))])}\right)$
$\boldsymbol{\pi}_{w_1,\mathbf{x}}, \boldsymbol{\pi}_{w_2,\mathbf{x}}$	$1 - \dfrac{E^{\boldsymbol{\pi}_{\mathbf{x}}}[\sqrt{w_1(\boldsymbol{\theta})w_2(\boldsymbol{\theta})}]}{\sqrt{E^{\boldsymbol{\pi}_{\mathbf{x}}}[w_1(\boldsymbol{\theta})]E^{\boldsymbol{\pi}_{\mathbf{x}}}[w_2(\boldsymbol{\theta})]}}$	$\ln\left(\dfrac{E^{\boldsymbol{\pi}_{\mathbf{x}}}[w_2(\boldsymbol{\theta})]/E^{\boldsymbol{\pi}_{\mathbf{x}}}[w_1(\boldsymbol{\theta})]}{\exp(E^{\boldsymbol{\pi}_{\mathbf{x}}}[\ln(w_2(\boldsymbol{\theta})/w_1(\boldsymbol{\theta}))])}\right).$

Table 1: The Hellinger metric and KL divergence between priors and posteriors.

Expressions (3.1) and (3.2) have been widely used in the literature in many contexts. For example, in Bayesian experimental inference it is a common goal to maximize the expected KL divergence between the prior and the posterior. In other words, the KL divergence is a measure of the information gained when one revises one's beliefs from the prior probability distribution to the posterior one. The log-score suffers from the shortcoming of focusing on the tails of the distribution by heavily penalizing observations that would have been predicted with low probability. However, the Hellinger metric is a surrogate divergence for the Total-Variation and is more robust than the KL divergence in practice. Further information about the use of these measures in Bayesian inference can be found in Hooker and Vidyshankar (2014) and more recently in Jewson et al. (2018).

Example 3.1. *Let us consider the prior and the weighted function defined in Example 2.11. We will compute the Hellinger metric and the KL divergence to measure the uncertainty induced by the weight function. From Table 1 we obtain that*

$$\begin{aligned}
H(\boldsymbol{\pi}, \boldsymbol{\pi}_w) &= 1 - \frac{E^{\boldsymbol{\pi}}[\sqrt{\theta_1^{a_1-1}\theta_2^{a_2-1}}]}{\sqrt{E^{\boldsymbol{\pi}}[\theta_1^{a_1-1}\theta_2^{a_2-1}]}} \\
&= 1 - \frac{E^{Exp(\lambda_1)}[\theta_1^{(a_1-1)/2}]E^{Exp(\lambda_2)}[\theta_2^{(a_2-1)/2}]}{\sqrt{E^{Exp(\lambda_1)}[\theta_1^{a_1-1}]E^{Exp(\lambda_2)}[\theta_2^{a_2-1}]}} \\
&= 1 - \frac{\Gamma((a_1+1)/2)\Gamma((a_2+1)/2)}{\sqrt{\Gamma(a_1)\Gamma(a_2)}},
\end{aligned}$$

$$
\begin{aligned}
KL(\boldsymbol{\pi}, \boldsymbol{\pi}_w) &= \ln\left(\frac{E^{\boldsymbol{\pi}}[\theta_1^{a_1-1}\theta_2^{a_2-1}]}{\exp(E^{\boldsymbol{\pi}}[\ln(\theta_1^{a_1-1}\theta_2^{a_2-1})])}\right) \\
&= \ln\left(\frac{E^{Exp(\lambda_1)}[\theta_1^{a_1-1}]E^{Exp(\lambda_2)}[\theta_2^{a_2-1}]}{\exp((a_1-1)E^{Exp(\lambda_1)}[\ln(\theta_1)] + (a_2-1)E^{Exp(\lambda_2)}[\ln(\theta_2)])}\right) \\
&= \ln(\Gamma(a_1)\Gamma(a_2)) + \gamma(a_1 + a_2 - 2),
\end{aligned}
$$

where equalities follow jointly from the expectation of the product of two independent random variables, from the expression of the r-th moment of an exponential distribution $X \sim Exp(\lambda)$ given by $\Gamma(r+1)/\lambda^r$ and, finally, from the well-known fact that the transformation $-\ln(X)$ has a Gumbel distribution with expectation $\gamma + \ln(\lambda)$, where γ is the Euler-Mascheroni constant. It is worthwhile noting that both the Hellinger metric and the KL divergence do not depend on the hyperparameters. Finally, by considering w_1 and w_2 with $(a_1, a_2) = (1/4, 1/4)$ and $(a_1, a_2) = (7/4, 7/4)$, respectively, we obtain that $H(\boldsymbol{\pi}, \boldsymbol{\pi}_{w_1}) = 0.432$, $H(\boldsymbol{\pi}, \boldsymbol{\pi}_{w_2}) = 0.14$, $KL(\boldsymbol{\pi}, \boldsymbol{\pi}_{w_1}) = 1.71$ and $KL(\boldsymbol{\pi}, \boldsymbol{\pi}_{w_2}) = 0.697$.

Of course, it is not easy in general to find a closed expression for the Hellinger metric and KL divergence. However, just observing Table 1 they can be estimated by using simulation methods as we will see later on. It is remarkable they only depend on the specific prior and its posterior and the corresponding weight functions.

Finally, it is natural to wonder how to choose weight functions and how to elicit their parameters. Obviously, the choice of the weight functions and their parameters depends on the problem at hand and the level of uncertainty about the priors. First, we should make sure to understand how the weight function modifies the underlying prior density. In general, increasing (decreasing) weight functions gives more weight to higher (smaller) values of the variable but that can be done in many ways. Note that any increasing or decreasing function is a potential weight function for our purpose. We would like to emphasize that there are particular weight functions associated with classical multivariate distributions in the literature, see for example Kim (2008) and references therein. To summarize, weight functions can induce asymmetry, heavy tailed distributions, truncated distributions, etc. Finally, in order to choose the parameters we should fix a reasonable distance in terms of the Hellinger metric and/or the KL divergence.

4 A real example about train door reliability

In this section we will illustrate our method analyzing real data in reliability. Stemming from a consulting project, we consider failure data from 40 underground trains associated with the door opening system. All trains were delivered to an European transportation company between November 1989 and March 1991 and all of them were put in service from 20th March 1990 to 20th July 1992. Failure monitoring ended on 31st December 1998. When a failure took place, both odometer reading and failure date were recorded, along with the code of the failed component. Those data were first analyzed in Pievatolo et al. (2003) without making any distinction among seven different failure modes. They were analyzed a second time in Pievatolo and Ruggeri (2010) where the authors determined that two failure modes (mechanical and electrical) were the most

relevant ones. Those components were failing for many different causes, implying that the two point processes describing the failures were actually obtained superposing the processes due to each cause. The regularity of the pattern and a theorem due to Grigelionis (see, e.g., Thompson (1988), p. 69) about the superposition of many processes into one justified the use of a nonhomogeneous Poisson process. For our purpose, we will consider one of those two failure modes, namely the mode associated with electrical opening commands where 530 failures were recorded from 19th September 1991 to 31st December 1998. For further information about data see Pievatolo and Ruggeri (2010). Some related works about Bayesian inference and reliability can be found in Ríos Insua et al. (2019).

Our interest is focused on finding a model to describe the failure history of electrical opening commands and predicting the number of failures in future time intervals in an Bayesian framework.

The model

Based on the interpretation of the model as a complex system, it is argued in Pievatolo and Ruggeri (2010) that the total number of failures that occur in the electrical opening system in the interval $(0, t]$, denoted by $N(t)$, follows a nonhomogeneous Poisson process (NHPP) with a common intensity function, $\lambda(t)$, and an increasing and invertible mean value function,

$$m(t) = E[N(t)] = \int_0^t \lambda(x)dx,$$

such that $m(\infty) = \infty$. From the plot of the cumulative number of failures versus failure time, see Figure 11.2 in Pievatolo and Ruggeri (2010), the authors postulate three different forms of the intensity function $\lambda(t)$, where all of them assume an improvement in the rate of occurrence of failures. For our purpose we will choose one of those models, namely the popular power law process (PLP) with intensity function

$$\lambda(t|\boldsymbol{\theta}) = M\beta t^{\beta-1}, \ \boldsymbol{\theta} = (M, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+, \tag{4.1}$$

such that

$$m(t|\boldsymbol{\theta}) = Mt^\beta, \ \boldsymbol{\theta} = (M, \beta) \in \mathbb{R}^+ \times \mathbb{R}^+. \tag{4.2}$$

Our choice is based on the attractive physical explanation of both of its parameters M and β and the fact that the PLP model is the mostly used NHPP in reliability due to its simplicity. In our model, we assume $0 < \beta < 1$, indicating a reliability growth information, excluding the cases of constant ($\beta = 1$) and decaying ($\beta > 1$) reliability. More details on the statistical analysis of repairable systems and NHPPs can be found in Thompson (1988), Kingman (1993), Rigdon and Basu (2000), Aven and Jensen (2000) and Ríos Insua et al. (2012) and a comprehensive catalogue of intensity functions is given in McCollin (2014).

The likelihood function

Let $N(t)$ be a PLP process with intensity function $\lambda(t|\boldsymbol{\theta})$ given in (4.1). Suppose that the vector of observed failure times $\mathbf{t}^* = (T_1, \ldots, T_n)$ recorded in the interval $(0, T]$,

where T is a known value, satisfies $T_1 < \ldots < T_n$, then, from Theorem 5.4 in Ríos Insua et al. (2012), the likelihood function is given by

$$
\begin{aligned}
l(\boldsymbol{\theta}|\mathbf{t}^*) &= \exp(-m(T|\boldsymbol{\theta})) \prod_{i=1}^{n} \lambda(T_i) \\
&= \exp(-MT^\beta) \prod_{i=1}^{n} M\beta T_i^{\beta-1} \\
&= M^n \beta^n \exp((\beta-1) \exp(-MT^\beta) \sum_{i=1}^{n} \ln(T_i)).
\end{aligned}
$$

From Lemma 2.7, just computing the partial derivatives of the logarithm of $l(\boldsymbol{\theta}|\mathbf{t}^*)$ given by

$$
\frac{\partial^2 \ln(l(\boldsymbol{\theta} \mid \mathbf{t}^*))}{\partial M \partial \beta} = -\ln(T)T^\beta,
$$

it is apparent that $l(\boldsymbol{\theta} \mid \mathbf{t}^*)$ is MTP_2 in $\boldsymbol{\theta}$ if, and only if, $T \le 1$. Then, by changing the scale of the time from the original record T_i to T_i/T, i.e., considering T as the unit, and denoting the new vector of observed failure times by $\mathbf{t} = (T_1/T, \ldots, T_n/T)$, we obtain a new likelihood function expressed as

$$
l(\boldsymbol{\theta}|\mathbf{t}) = M^n \beta^n \exp((\beta-1) \sum_{i=1}^{n} \ln(\frac{T_i}{T})) \exp(-M). \tag{4.3}
$$

It is clear from Lemma 2.4 that the likelihood function $l(\boldsymbol{\theta} \mid \mathbf{t})$ given in (4.3) is MTP_2 in $\boldsymbol{\theta}$.

Remark 4.1. *In order to validate our model we will first consider T the 31st December 1997. We will predict the number of failures in 1998 and we compare it with the actual value.*

The specific prior and posterior

It is assumed that the specific prior belief $\boldsymbol{\pi}$ over $\boldsymbol{\Theta} = \mathbb{R}^+ \times \mathbb{R}^+$ is a bivariate random vector having independent exponential marginal distributions $Exp(\lambda)$ and $Exp(\mu)$, associated with the parameters M and β, respectively, with joint PDF given by

$$
\pi(\boldsymbol{\theta}) = \lambda\mu \exp(-\lambda M) \exp(-\mu\beta), \ (M, \beta) \in \boldsymbol{\Theta},
$$

where the hyperparameters λ and μ are assumed to be known; namely they are obtained from the initial estimates $M_0 = 495.5$ and $\beta_0 = 0.79$. Those estimates are based on an expert judgment combining the regression analysis between the empirical cumulative number of failures versus the failure times and the maximum likelihood estimates, MLEs, when using only some old trains which were not included in the final study. The values M_0 and β_0 are very reasonable, denoting the latter a clear, although not excessive, reliability growth, whereas the former expresses the expected number of failures in the

unit interval since $m(1|\boldsymbol{\theta}) = M$. From the initial values M_0 and β_0, we take the values $\lambda = 1/M_0$ and $\mu = 1/\beta_0$ due to the fact that $E^{\boldsymbol{\pi}}[\boldsymbol{\theta}] = (1/\lambda, 1/\mu)$. It is apparent from Lemma 2.4 that $\pi(M, \beta)$ is MTP_2 in $\boldsymbol{\theta} = (M, \beta)$.

On the other hand, just considering the likelihood function given by (4.3) and using expression (1.2) to update the state of knowledge, we easily obtain that the posterior distribution is a bivariate random vector having independent gamma marginal distributions $G(n+1, 1+\lambda)$ and $G(n+1, \mu - \sum_{i=1}^{n} \ln(\frac{T_i}{T}))$, associated with the parameters M and β, respectively, with joint PDF given by

$$\pi_{\mathbf{t}}(\boldsymbol{\theta}) = \frac{(1+\lambda)^{n+1}}{\Gamma(n+1)} M^n \exp(-(1+\lambda)M) \frac{\left(\mu - \sum_{i=1}^{n} \ln(\frac{T_i}{T})\right)^{n+1}}{\Gamma(n+1)} \beta^n \exp(-(\mu - \sum_{i=1}^{n} \ln(\frac{T_i}{T}))\beta).$$

The weighted band

Now, we introduce a perturbation scheme on the prior distribution by considering the weighted band $\Gamma_{w_1, w_2, \boldsymbol{\pi}}$ where w_1 and w_2 are defined by expression (2.8). Then,

$$\begin{aligned} \pi_{w_1}(\boldsymbol{\theta}) &\propto \lambda\mu M^{a-1} \beta^{b-1} \exp(-\lambda M) \exp(-\mu\beta) \exp[-cM\beta], \\ \pi_{w_2}(\boldsymbol{\theta}) &\propto \lambda\mu M^{a'-1} \beta^{b'-1} \exp(-\lambda M) \exp(-\mu\beta)(M^{c'} + \beta^{c'}). \end{aligned} \quad (4.4)$$

The exact distributions associated with both weighted priors, and having PDFs $\pi_{w_1}(\boldsymbol{\theta})$ and $\pi_{w_2}(\boldsymbol{\theta})$, are unknown. However, after a straightforward computation in (4.4), we can easily compute the conditional distributions, as shown in Table 2.

Priors	Conditional	Distribution
$\boldsymbol{\pi}_{w_1}$	$\boldsymbol{\pi}_{w_1}(M\|\beta)$	$G(a, \lambda + c\beta)$.
	$\boldsymbol{\pi}_{w_1}(\beta\|M)$	$G(b, \mu + cM)$.
$\boldsymbol{\pi}_{w_2}$	$\boldsymbol{\pi}_{w_2}(M\|\beta)$	$(1-\epsilon)G(a', \lambda) + \epsilon G(a'+c', \lambda)$
		$\epsilon = \frac{\Gamma(a'+c')}{\Gamma(a'+c')+(\lambda\beta)^{c'}\Gamma(a')}$.
	$\boldsymbol{\pi}_{w_2}(\beta\|M)$	$(1-\epsilon)G(b', \mu) + \epsilon G(b'+c', \mu)$
		$\epsilon = \frac{\Gamma(b'+c')}{\Gamma(b'+c')+(\mu M)^{c'}\Gamma(b')}$.

Table 2: The conditional distributions of the weighted priors.

We notice that the conditional distributions of the weighted prior $\boldsymbol{\pi}_{w_2}$ are given by a mixture of gamma distributions.

The posterior weighted distributions

Also the joint posterior distributions do not have closed-form expressions. However, just taking in account the conditional distributions given in Table 2, we can also compute their conditional distributions, as shown in Table 3.

Posteriors	Conditional	Distribution
$\boldsymbol{\pi}_{w_1,\mathbf{t}}$	$\boldsymbol{\pi}_{w_1,\mathbf{t}}(M\mid\beta)$	$G(n+a, 1+\lambda+c\beta).$
	$\boldsymbol{\pi}_{w_1,\mathbf{t}}(\beta\mid M)$	$G(n+b, \mu+cM-\sum_{i=1}^{n}\ln(\frac{T_i}{T})).$
$\boldsymbol{\pi}_{w_2,\mathbf{t}}$	$\boldsymbol{\pi}_{w_2,\mathbf{t}}(M\mid\beta)$	$(1-\epsilon)G(n+a', 1+\lambda) + \epsilon G(n+a'+c', 1+\lambda),$
		$\epsilon = \frac{\Gamma(n+a'+c')}{\Gamma(n+a'+c')+\Gamma(n+a')(\beta(1+\lambda))^{c'}}.$
	$\boldsymbol{\pi}_{w_2,\mathbf{t}}(\beta\mid M)$	$(1-\epsilon)G(n+b', \mu-t') + \epsilon G(n+b'+c', \mu-t'),$
		$\epsilon = \frac{\Gamma(n+b'+c')}{\Gamma(n+b'+c')+\Gamma(n+b')(M(\mu-t'))^{c'}}, t' = \sum_{i=1}^{n}\ln\left(\frac{T_i}{T}\right).$

Table 3: The conditional distributions of the weighted posteriors.

We notice again that the conditional distributions of the weighted posterior $\boldsymbol{\pi}_{w_2,\mathbf{t}}$ are given by a mixture of gamma distributions. The conditional distributions shown in Table 2 and 3 can be used to estimate the characteristics of the weighted priors and posteriors using Markov chain Monte Carlo (MCMC) methods, using the Gibbs sampling method. In general, w_1 has more effect on $\boldsymbol{\pi}$ than the corresponding w_2 as a consequence of the positive skewness of the specific prior. Finally, it is apparent from Table 2 and 3 that w_1 and w_2 induce a dependence structure in both weighted priors and weighted posteriors.

Choosing the degree of uncertainty

At this moment, a natural question is how to choose the parameters a, b and c and a', b', c' for w_1 and w_2 of the two weight functions, respectively. One possibility is to require that the resulting weighted priors are symmetrically close to the specified prior in terms of Hellinger metric or KL divergence. Due to the complexity of our real problem, we decided to consider a great amount of uncertainty in order to evaluate the effect on the final decision. Having set a goal of symmetric Hellinger distance approximately equal to 0.7, we found, after some trials, that it can be achieved taking $a = 0.8$, $b = 0.4$ and $c = 0.17$ and $a' = 3.8$, $b' = 3.4$ and $c' = 1.17$ as we can see in Table 4. We first notice that the Hellinger metric is a symmetric measure. The maximum uncertainty 1 in the Hellinger metric is achieved when one density assigns probability zero to every set to which the other density assigns a positive probability, and vice versa. With respect to the KL divergence, we recall it is an asymmetric measure. Then, we can only interpret that the amount of information lost when $\boldsymbol{\pi}_{w_1}$ is used to approximate $\boldsymbol{\pi}$ is larger than the corresponding information lost when we use $\boldsymbol{\pi}_{w_2}$. We notice that all measures in Table 4 have been estimated from a sample of the exact distribution of the specific prior using jointly the Markov chain Monte Carlo (MCMC) method and Table 1.

| Priors	$\boldsymbol{\pi}, \boldsymbol{\pi}_{w_1}$	$\boldsymbol{\pi}, \boldsymbol{\pi}_{w_2}$	$\boldsymbol{\pi}_{w_1}, \boldsymbol{\pi}_{w_2}$
Hellinger metric	0.69916	0.69590	0.99982
KL divergence	66.724	7.552	27.101

Table 4: The Hellinger metric and the KL divergence for priors.

From the exact distribution of the specific posterior and using again the MCMC method and Table 1, we obtain in Table 5 the Hellinger metric and the KL divergence for the posterior weighted distributions.

Posteriors	$\boldsymbol{\pi_t}, \boldsymbol{\pi_{w_1,t}}$	$\boldsymbol{\pi_t}, \boldsymbol{\pi_{w_2,t}}$	$\boldsymbol{\pi_{w_1,t}}, \boldsymbol{\pi_{w,t}}$
Hellinger metric	0.72191	0.00713	0.77096
KL divergence	5.1661	0.02869	4.43343

Table 5: The Hellinger metric and the KL divergence for posteriors.

Finally, in order to graphically see the effect of the weighted functions, we represent in Figure 2 (a) and (b) the histograms and the CDFs for the prior distributions $\boldsymbol{\pi_{w_1}}$ (in blue), $\boldsymbol{\pi}$ (in yellow) and $\boldsymbol{\pi_{w_2}}$ (in green), respectively. We also represent in Figure 2 (c) and (d) the corresponding ones for the posterior distributions. As expected, the uncertainty decreases when we incorporate experimental data. At first glance, the contribution to uncertainty of the weighted function w_1 is greater than the corresponding induced by w_2. This result is coherent with a greater improvement in the failure rate with respect to the expected one by the PLP model, as we will see later on.
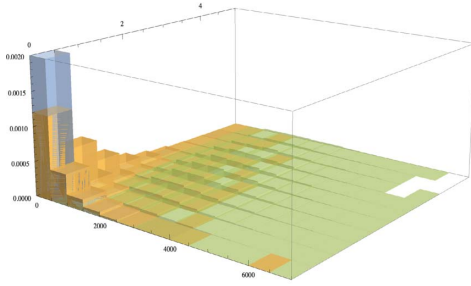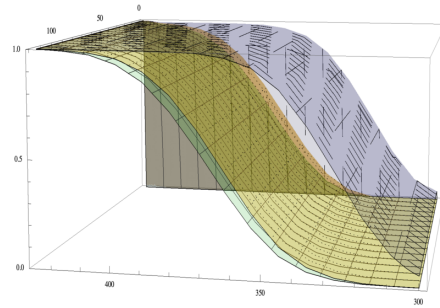

(a) Histograms for prior distributions
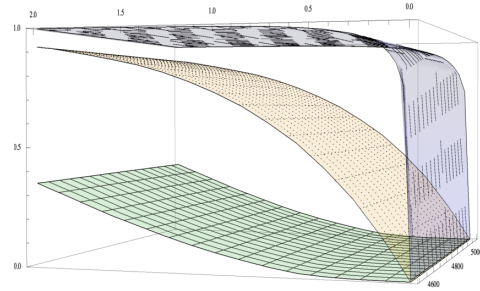
(b) The CDFs for prior distributions
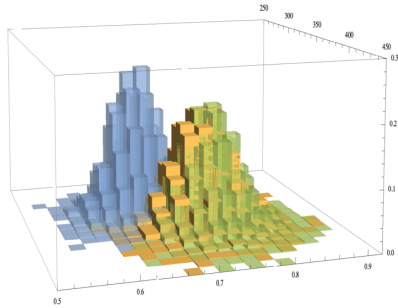
(c) Histograms for posterior distributions

(d) The CDFs for posterior distributions

Figure 2: The weighted prior distributions and the posterior weighted distributions.

Forecasts for the expected number of failures

As we previously mentioned, we have first considered 31st December 1997 as the (ending) calendar day T. We are interested in predicting the expected number of failures in future time intervals of the form $[T, T+u]$. Such functional, conditional on $\boldsymbol{\theta}$, is easily computed by the mean value function of the PLP model given in expression (4.2) and takes the form

$$H_X(\boldsymbol{\theta}) = E[N(1+h) - N(1)] = \int_1^{1+h} \lambda(t|\boldsymbol{\theta})dt = M((1+h)^\beta - 1),$$

where T is represented by the unit in the above expression and h is computed proportionally in the time scale, i.e., $h = u/T$. A straightforward computation shows that $H_X(\boldsymbol{\theta})$ is increasing in $\boldsymbol{\theta}$. Then, recalling that both $\boldsymbol{\pi}$ and $l(\boldsymbol{\theta}|\mathbf{t})$ are MTP_2 in $\boldsymbol{\theta}$ and using Corollary 2.18, we obtain that

$$H_X(\boldsymbol{\pi}_{w_1,\mathbf{t}}) \leq_{st} H_X(\boldsymbol{\pi}'_\mathbf{t}) \leq_{st} H_X(\boldsymbol{\pi}_{w_2,\mathbf{t}}), \ \forall \boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}.$$

From the above orderings we obtain bounds for some characteristics of interest in the Bayesian framework. First, from expression (2.11), we know the predictive expectations of the number of failures of the weighted posterior distributions are bounds for the corresponding expectation of the specific posterior, i.e.,

$$E^{\boldsymbol{\pi}_{w_1,\mathbf{t}}}[H_X(\boldsymbol{\theta})] \leq E^{\boldsymbol{\pi}'_\mathbf{t}}[H_X(\boldsymbol{\theta})] \leq E^{\boldsymbol{\pi}_{w_2,\mathbf{t}}}[H_X(\boldsymbol{\theta})], \ \forall \boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}.$$

Second, from (2.12) the endpoints of the 95% Bayesian credible quantile-based intervals are ordered, i.e.,

$$F^{-1}_{H_X(\boldsymbol{\pi}_{w_1,\mathbf{t}})}(p) \leq F^{-1}_{H_X(\boldsymbol{\pi}'_\mathbf{t})}(p) \leq F^{-1}_{H_X(\boldsymbol{\pi}_{w_2,\mathbf{t}})}(p), \ p \in \{0.025, 0.975\}, \forall \boldsymbol{\pi}' \in \Gamma_{w_1,w_2,\boldsymbol{\pi}}.$$

From the mentioned fact that the posterior weighted distributions are unknown, the bounds in the above inequalities have been estimated from the conditional distributions given in Table 3 using the Gibbs sampling. Table 6 shows the effect of the weighted functions on the prediction of the expected number of failures in 1998, i.e., when T represents the 31st December 1997 and $u = 1$. We notice that we ran the corresponding MCMC algorithm for 10000 iterations using a burn-in of 1000. The iteration numbers were chosen after experimentation to deliver stable results over multiple runs.

True value	95% credibility Int. $\boldsymbol{\pi}_{w_1,\mathbf{t}}$	Post. mean $\boldsymbol{\pi}_{w_1,\mathbf{t}}$	95% credibility Int. $\boldsymbol{\pi}_\mathbf{t}$	Post. mean $\boldsymbol{\pi}_\mathbf{t}$	95% credibility Int. $\boldsymbol{\pi}_{w_2,\mathbf{t}}$	Post. mean $\boldsymbol{\pi}_{w_2,\mathbf{t}}$
23	[22.75, 45.40]	34	[29.73, 55.04]	42	[30.46, 56.03]	43

Table 6: Forecasts and credibility intervals of the number of failures in 1998. Forecasts and intervals are shown for the specific posterior and the weighted posteriors.

We first note that weighting the prior information causes a change in the final Bayesian prediction. Therefore, the model can not be considered robust from a Bayesian point of view. From another point of view, a similar conclusion is pointed out in Pievatolo and Ruggeri (2010). It is also interesting to observe that the weighted prior $\boldsymbol{\pi}_{w_1}$

provides a more accurate result than the PLP model and the latter is close to the result predicted by the weighted prior $\boldsymbol{\pi}_{w_2}$. A reason for this behavior is that the PLP model overestimates the number of failures in 1998 and does not reflect the "true" failure process closely enough. It is also remarkable that the credibility interval for the decreasing weighted posterior contains the true value which solves somehow the problem detected in Pievatolo and Ruggeri (2010) about the sharp change in slope of the curve of cumulative number of failures in the last years. Finally, we know from Proposition 2.10 that all priors of the form $\boldsymbol{\pi}_{\epsilon} = (1 - \epsilon)\boldsymbol{\pi}_{w_1} + \epsilon\pi_{w_2}$ (obtained as a mixture of $\boldsymbol{\pi}_{w_1}$ and $\boldsymbol{\pi}_{w_2}$) belong to the class $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$, for all $0 \le \epsilon \le 1$. Since $\Gamma_{w_1,w_2,\boldsymbol{\pi}}$ is a convex class of distributions and $\boldsymbol{\pi}_{\epsilon}$ is continuous (see Lemma 3.1 in Ruggeri and Sivaganesan (2005)) it follows that any value in any interval of interest obtained for the weighted priors can be expressed as a value from a particular prior for some ϵ.

To end up, and using a similar procedure than above, we present in Table 7 forecasts after forecasting horizons of one, two and three years beyond different ending periods, T, which is set on 31st December of the indicated years, denoted by year-i, $i = 1, 2$ and 3.

T	True value	95% credibility Int. $\boldsymbol{\pi}_{w_1,\mathbf{t}}$	Post. mean $\boldsymbol{\pi}_{w_1,\mathbf{t}}$	95% credibility Int. $\boldsymbol{\pi}_{\mathbf{t}}$	Post. mean $\boldsymbol{\pi}_{\mathbf{t}}$	95% credibility Int. $\boldsymbol{\pi}_{w_2,\mathbf{t}}$	Post. mean $\boldsymbol{\pi}_{w_2,\mathbf{t}}$
1992-1	83	[47.63, 78.53]	63	[66.25, 101.95]	84	[70.95, 107.79]	89
1992-2	72	[42.81, 72.30]	57	[63.20, 98.12]	80	[68.45, 104.64]	86
1992-3	62	[39.83, 68.37]	54	[61.25, 95.66]	78	[66.84, 102.62]	84
1993-1	72	[35.90, 63.29]	49	[48.22, 79.32]	63	[50.67, 82.44]	66
1993-2	62	[35.52, 58.77]	45	[45.28, 75.51]	60	[47.82, 78.78]	63
1993-3	42	[30.15, 55.58]	42	[43.16, 72.78]	57	[45.76, 76.13]	60
1994-1	62	[31.59, 57.54]	44	[42.01, 71.31]	56	[43.53, 73.29]	58
1994-2	42	[29.12, 54.22]	41	[39.77, 68.38]	54	[41.35, 70.45]	55
1994-3	35	[27.28, 51.68]	39	[38.05, 66.12]	52	[39.66, 68.24]	53
1995-1	42	[30.84, 56.54]	43	[40.92, 69.91]	55	[42.10, 71.45]	56
1995-2	35	[28.92, 53.94]	41	[39.21, 67.66]	53	[40.42, 69.26]	54
1995-3	23	[27.40, 51.86]	39	[37.82, 62.84]	51	[39.08, 67.48]	53
1996-1	35	[26.20, 50.21]	38	[34.49, 61.42]	47	[35.35, 62.58]	48
1996-2	23	[24.65, 48.05]	36	[32.99, 59.43]	46	[33.88, 60.61]	47
1997-1	23	[22.75, 45.40]	34	[29.73, 55.04]	42	[30.46, 56.03]	43

Table 7: Forecasts and credibility intervals of the number of failures from 1993 to 1998. Forecasts and intervals are shown for the specific posterior and the weighted posteriors.

Just observing all predictions, it is clear that uncertainty decreases when the sample size increases. Also, in only two cases the true value does not belong to any credibility interval, namely when the ending periods are on 31st December 1995 and 31st December 1996 and we estimate the expected number of failures in 1998, i.e., in forecasting periods of three and two years, respectively. As we have previously analyzed in Table 6, the forecasts in 1998 slightly improve when the recording period is on 31st December 1997. Finally, Figure 3 summarizes graphically the information about posterior means contained in Table 7. In general, it is interesting to note that the weighted function

w_1 provides more accurate results than the PLP model when this latter overestimates enough the expected number of failures. On the contrary, when the PLP underestimates enough the expected number of failures the weighted function w_2 will lead to better results.
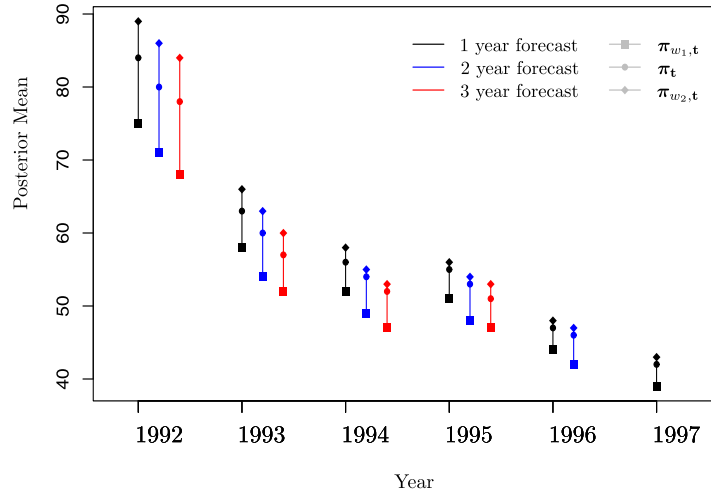

Figure 3: Range of the posterior means of the number of failures from 1993 to 1998.

5   Concluding remarks

In this paper we have provided a methodology to induce uncertainty in the choice of a specific prior by using a particular class of priors called the weighted band. We have followed the typical approach of Bayesian robustness, as outlined in Ríos Insua and Ruggeri (2000) which provides a thorough review of the motivations and the methods which lead to consider classes of priors. In such work classes of likelihoods or sampling models are considered as well, but most of interest of the researchers (and ours, as well) has concentrated on priors. Stemming from the practical impossibility of specifying an exact, unique prior distribution, Bayesian robustness proposes to weaken the request considering a distribution which is deemed as a good approximation to the prior beliefs and then building a neighborhood around it to express the uncertainty about such prior. Our work goes in such direction since it embeds a prior in a class where weight functions, with adequate properties, are used to give more or less weight to part of the parameter space. This theoretical construction corresponds also to practical situations, like the one we have considered in our work about train door failures. Of course many classes of priors can be considered (see Ríos Insua and Ruggeri (2000)) and the choice should be based on the problem at hand and the available expert knowledge (as an example, consider the class based on generalized moments presented in Betrò et al. (1994)). Whereas the choice of a class of priors could be motivated by practical problems where the experts are unable to specify a unique prior distributions, more theoretical

studies are interested in specifying probability distributions over prior distributions, like in Bayesian nonparametrics. The idea of combining Bayesian nonparametrics and Bayesian robustness lead to some works, like Ruggeri (2010, 2014). Our methodology combines practical motivations and theoretical contributions involving stochastic orders. Although it is inspired by the distorted band introduced by Arias-Nicolás et al. (2016), it provides a different way of inducing uncertainty in the univariate model and can be easily generalized for multiparameter distributions. The proposed prior class is also relatively easy to specify, an inheritance from the selected weight functions used to specify the distorted band. These weight functions also help when interpreting the resulting class of posterior distributions.

We have also shown how the proposed methodology has not only a theoretical interest but it can be also applied to real problems. On purpose, we have chosen a complex problem in reliability which had been already investigated elsewhere with not completely satisfactory results. Here we have shown how a weighted priors on the parameters of the very popular power law process can lead to better forecasting of the expected number of failures.

The models considered in Pievatolo and Ruggeri (2010) pose new challenges, also from a methodological viewpoint, which could be addressed in future works. In the current paper we have considered only failure dates whereas the available data consist also of odometer readings upon failures. The nonhomogeneous Poisson process considered in Pievatolo and Ruggeri (2010) is one of the few examples of bivariate intensity function in reliability. Those authors performed an informal sensitivity analysis with respect to the model considering three different baseline univariate intensity functions for the failure times. Sensitivity with respect to changes in the model has been rarely considered in the Bayesian framework, apart from informal approaches. The reasons are manifold: computational complexity and major focus on the most critical aspect of the Bayesian approach, i.e. the prior choice, are the most relevant ones.

The extension of our approach to a class of models, instead of priors, is very challenging, especially when considering a class of Poisson processes, possibly with bivariate intensity functions. Also the notion of class of Poisson processes should be considered carefully.

The train door failure data suggest another methodological approach: the extension of our approach to hierarchical models. Data are from 40 trains and each train could have its own model with exchangeable parameters, leading to a hierarchical model. Such model has a clear physical interpretation: differences in construction, operation and maintenance make trains more similar rather than equal. In this case the distorted class would be on the common prior for the distinct parameters of each model.

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: https://doi.org/10.1016/j.ijar.2008.03.014. 32

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: https://doi.org/10.1080/03610920500257329.   32

Arias-Nicolás, J. P., Ruggeri, F., and Suárez-Llorens, A. (2016). "New Classes of Priors Based on Stochastic Orders and Distortion Functions." *Bayesian Analysis*, 11 (4): 1107–1136. MR3545475. doi: https://doi.org/10.1214/15-BA984.   32, 33, 38, 42, 43, 54

Aven, T. and Jensen, U. (2000). "A general minimal repair model." *Journal of Applied Probability*, 37: 187–197. MR1761669. doi: https://doi.org/10.1017/s0021900200015333.   46

Barrera, M., Lira, I., Sánchez-Sánchez, M., and Suárez-Llorens, A. (2019). "Bayesian treatment of results from radioanalytical measurements. Effect of prior information modification in the final value of the activity." *Radiation Physics and Chemistry*, 156: 266–271.   33

Basu, S. (1994). "Variations of posterior expectations for symmetric unimodal priors in a distribution band." *Sankhyā A*, 56 (2): 320–334. MR1664917.   31

Basu, S. and DasGupta, A. (1995). "Robust Bayesian analysis with distribution bands." *Statistics & Decisions*, 13: 333–349. MR1380253.   38, 43

Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY, USA: Springer, second edition. MR0804611. doi: https://doi.org/10.1007/978-1-4757-4286-2.   31, 32

Berger, J. O. (1994). "An overview of robust Bayesian analysis (with discussion)." *Test*, 3: 5–124. MR1293110. doi: https://doi.org/10.1007/BF02562676.   32, 33

Berger, J. O. and Berliner, L. M. (1986). "Robust Bayes and empirical Bayes analysis with ϵ-contaminated priors." *Annals of Statistics*, 14: 461–486. MR0840509. doi: https://doi.org/10.1214/aos/1176349933.   32

Bernardo, J. M. (2003). "Bayesian Statitistics". In Viertl, R. (ed.), *Probability and Statistics. Encyclopedia of Life Support Systems (EOLSS)*. Oxford, UK: UNESCO.   31

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: https://doi.org/10.1016/0378-3758(94)90022-1.   32, 53

Blazej, P. (2008). "Preservation of classes of life distributions under weighting with a general weight function." *Statistics & Probability Letters*, 78: 3056–3061. MR2474397. doi: https://doi.org/10.1016/j.spl.2008.05.028.   34, 38

Denneberg, D. (1990). "Premium calculation: Why standard deviation should be replaced by absolute deviation." *ASTIN Bulletin*, 20: 181–190.   34

Denuit, M., Dhaene, J., Goovaerts, M. J., and Kaas, R. (2005). *Actuarial Theory for Dependent Risks*. Chichester, UK: Wiley.   33

Dey, D. K., Lou, K., and Bose, S. (1998). "A Bayesian approach to loss robustness." *Statistics & Decisions*, 16: 65–87. MR1622563.   32

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: https://doi.org/10.1016/j.insmatheco.2013.05.001.   34

Esary, J. D., Proschan, F., and Walkaup, D. W. (1967). "Association of random variables, with applications." *Annals of Mathematical Statistics*, 38: 1466–1474. MR0217826. doi: https://doi.org/10.1214/aoms/1177698701.   37

Fisher, R. A. (1934). "The effects of methods of ascertainment upon the estimation of frequencies." *Annals of Eugenics*, 6: 13–25.   34

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: https://doi.org/10.1007/978-1-4612-1306-2_6.   32

Furman, E. and Zitikis, R. (2008). "Weighted premium calculation principle." *Insurance: Mathematics and Economics*, 42: 459–465. MR2392102. doi: https://doi.org/10.1016/j.insmatheco.2007.10.006.   34

Gâsemyr, J. and Natvig, B. (1995). "Using expert opinions in Bayesian prediction of component lifetimes in a shock modely." *Mathematics of Operations Research*, 20: 227–242. MR1320457. doi: https://doi.org/10.1287/moor.20.1.227.   42

Gâsemyr, J. and Natvig, B. (1997). "An application of the MTP2 property on bounds on system reliability." *Naval Research Logistics*, 44: 741–755. MR1482469. doi: https://doi.org/10.1002/(SICI)1520-6750(199712)44:8<741::AID-NAV3>3.0.CO;2-A.   42

Gibbs, A. L. and Su, F. E. (2002). "On Choosing and Bounding Probability Metrics." *International Statistical Review*, 70 (3): 419–435.   43

Glaz, J. and Johnson, B. M. (1984). "Probability inequalities for multivariate distributions with dependence structures." *Journal of the American Statistical Association*, 79: 436–440. MR0755099.   37

Goovaerts, M. J. and Laeven, R. J. A. (2008). "Actuarial risk measures for financial derivative pricing." *Insurance: Matematics and Economics*, 42: 540–547. MR2404313. doi: https://doi.org/10.1016/j.insmatheco.2007.04.001.   34

Holley, R. (1974). "Remarks on the FKG inequalitiess." *Communications in Mathematical Physic*, 36: 227–231. MR0341552.   36

Hooker, G. and Vidyshankar, A. N. (2014). "Bayesian model robustness via disparitiess." *Test*, 23: 556–584. MR3252095. doi: https://doi.org/10.1007/s11749-014-0360-z.   44

Huang, K. and Mi, J. (2018). "Applications of likelihood ratio order in Bayesian

inferences." *Probability in the Engineering and Informational Sciences*, In press. doi: https://doi.org/10.1017/S026996481800027X.   33

Jain, K. and Nanda, A. (1995). "On multivariate weighted distributions." *Communications in Statistics – Theory and Methods*, 24 (10): 2517–2539. MR1354470. doi: https://doi.org/10.1080/03610929508831631.   34, 39

Jewson, J., Smith, J. Q., and Holmes, C. (2018). "Principles of Bayesian Inference Using General Divergence Criteria." *Entropy*, 20 (6): 442. MR3879894. doi: https://doi.org/10.3390/e20060442.   44

Joshi, C., Ruggeri, F., and Wilson, S. P. (2018). "Prior Robustness for Bayesian Implementation of the Fault Tree Analysis." *IEEE Transactions on Reliability*, 67 (1): 170–183.   33

Karlin, S. (1968). *Total Positivity*. Stanford, Calif.: Stanford University Press.   36

Karlin, S. and Rinott, Y. (1980a). "Classes of orderings of measures and related correlation inequalities I. Multivariate totally positive distributions." *Journal of Multivariate Analysis*, 10: 467–498. MR0599685. doi: https://doi.org/10.1016/0047-259X(80)90065-2.   34, 36, 37

Karlin, S. and Rinott, Y. (1980b). "Classes of orderings of measures and related correlation inequalities II. Multivariate reverse rule distributions." *Journal of Multivariate Analysis*, 10: 499–516.   37

Karlin, S. and Rinott, Y. (1981). "Total positivity properties of absolute value multinormal variables with applications to confidence interval estimates and related probabilistic inequalities." *The Annals of Statistics*, 9: 1035–1049. MR0628759.   37

Kemperman, J. H. B. (1977). "On the FKG inequality for measures on a partially ordered space." *Indagationes Mathematicae*, 39: 313–331. MR0467867.   36

Kim, H.-J. (2008). "A class of weighted multivariate normal distributions and its properties." *Journal of Multivariate Analysis*, 99: 1758–1771. MR2444818. doi: https://doi.org/10.1016/j.jmva.2008.01.008.   45

Kingman, J. (1993). *Poisson Processes*. Oxford, UK: Clarendon Press. MR1207584.   46

Lorentz, G. G. (1953). "An inequality for rearrangement." *The American Mathematical Monthly*, 60: 176–179. MR0052476. doi: https://doi.org/10.2307/2307574.   36

Makov, U. (1994). "Some aspects of Bayesian loss robustness." *Journal of Statistical Planning and Inference*, 38: 359–370. MR1261809. doi: https://doi.org/10.1016/0378-3758(94)90016-7.   32

Martín, J., Ríos Insua, D., and Ruggeri, F. (1998). "Issues in Bayesian loss robustness." *Sankhyā A*, 60: 405–417. MR1718848.   32

McCollin, C. (2014). *Intensity functions for nonhomogeneous Poisson processes*. Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd.   46

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: https://doi.org/10.1007/978-1-4612-1306-2_3. 32

Moreno, E. and Cano, J. A. (1991). "Bayesian analysis for ϵ-contaminations partially known." *Journal of the Royal Statistical Society B*, 53: 143–155. MR1094277. 32

Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks*. New York, NY, USA: Wiley. MR1889865. 32, 36

Nanda, A. and Jain, K. (1997). "Some weighted distribution results on univariate and bivariate cases." *Journal of Statistical Planning and Inference*, 77: 169–180. MR1687954. doi: https://doi.org/10.1016/S0378-3758(98)00190-6. 39

Navarro, J., Ruiz, J., and del Aguila, Y. (2006). "Multivariate weighted distributions: a review and some extensions." *Statistics*, 40 (1): 51–64. MR2207404. doi: https://doi.org/10.1080/02331180500439691. 34, 39, 40

Navarro, J., Sunoj, S., and Linu, M. (2011). "Characterizations of bivariate models using dynamic Kullback-Leibler discrimination measures." *Statistics and Probability Letters*, 81: 1594–1598. MR2832917. doi: https://doi.org/10.1016/j.spl.2011.05.016. 34, 39

Navarro, J., Águila, Y. D., Sordo, M., and Suárez-Llorens, A. (2016). "Preservation of Stochastic Orders under the Formation of Generalized Distorted Distributions. Applications to Coherent Systems." *Methodology and Computing in Applied Probability*, 18: 529–545. MR3488591. doi: https://doi.org/10.1007/s11009-015-9441-z. 34

Pearlman, M. D. and Olkin, I. (1980). "Unbiasedness of invariant tests for MANOVA and other multivariate problems." *Annals of Statistics*, 8: 1326–1341. MR0594648. 37

Pievatolo, A. and Ruggeri, F. (2010). "Bayesian modelling of train doors reliability." In O'Hagan, T. and West, M. (eds.), *The Oxford Handbook of Applied Bayesian Analysis*. Oxford, UK: Oxford University Press. MR2790347. 45, 46, 51, 52, 54

Pievatolo, A., Ruggeri, F., and Argiento, R. (2003). "Bayesian analysis and prediction of failures underground trains." *Quality Reliability Engineering International*, 19: 327–336. 45

Preston, C. J. (1974). "A generalization of the KFG inequalities." *Communications in Mathematical Physic*, 36: 223–241. MR0341553. 36

Quiggin, J. (1982). "A theory of anticipated utility." *Journal of Economic Behavior and Organization*, 3: 323–343. 34

Rao, C. R. (1963). "On discrete distributions arising out of methods of ascertainment." *Sankya A*, 25: 311–324. MR0208736. 34

Rao, C. R. (1985). "Weighted distributions arising out of methods of ascertainment: what population does a sample represent." In Atkinson, A. C., Fienberg, S. E. (eds.), *A Celebration of Statistics: The ISI Centenary Volume*, 543–569. New York: Springer. MR0816158. 34

Rigdon, S. E. and Basu, A. (2000). *Statistical methods for the reliability of repairable systems*. New York: John Wiley. MR1781523. 46

Rinott, Y. (1973). "Multivariate majorization and rearrangement inequalities with some applications to probability and statistics." *Israel Journal of Mathematics*, 15: 60–77. MR0368308. doi: https://doi.org/10.1007/BF02771774. 36

Rinott, Y. and Scarsini, M. (2006). "Total positivity order and the normal distribution." *Journal of Multivariate Analysis*, 97: 1251–1261. MR2276157. doi: https://doi.org/10.1016/j.jmva.2005.07.008. 38

Ríos Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Lecture notes in statistics 152, New York, NY, USA: Springer-Verlag. MR1795207. doi: https://doi.org/10.1007/978-1-4612-1306-2_1. 32, 53

Ríos Insua, D., Ruggeri, F., Soyer, R., and Wilson, S. (2019). "Advances in Bayesian decision making in reliability." *European Journal of Operational Research*, In press. MR4041567. doi: https://doi.org/10.1016/j.ejor.2019.03.018. 46

Ríos Insua, D., Ruggeri, F., and Wiper, M. (2012). *Bayesian Analysis of Stochastic Process Models*. New York, NY, USA: Wiley. MR2963563. doi: https://doi.org/10.1002/9780470975916. 46, 47

Ruggeri, F. (2010). "Nonparametric Bayesian robustness." *Chilean Journal of Statistics*, 1 (2): 51–68. MR2756119. 54

Ruggeri, F. (2014). "On some optimal Bayesian nonparametric rules for estimating distribution functions." *Econometric Reviews*, 33: 289–304. MR3170850. doi: https://doi.org/10.1080/07474938.2013.807183. 54

Ruggeri, F. and Sivaganesan, S. (2005). "On modeling change points in non-homogeneous Poisson processes." *Statistical Inference for Stochastic Processes*, 8: 311–329. MR2177316. doi: https://doi.org/10.1007/s11203-005-6100-y. 52

Sánchez-Sánchez, M., Sordo, M. A., Suárez-Llorens, A., and Gómez-Déniz, E. (2019). "Deriving Robust Bayesian Premiums under Bands of Prior Distributions with Applications." *ASTIN Bulletin*, 49: 147–168. MR3919848. doi: https://doi.org/10.1017/asb.2018.36. 33

Sarkar, S. K. (1998). "Some probability inequalities for ordered MTP2 random variables: a proof of the Simes conjecture." *The Annals of Statistics*, 26: 494–504. MR1626047. doi: https://doi.org/10.1214/aos/1028144846. 37

Sarkar, S. K. (2008). "Generalizing Simes' test and Hochberg's stepup procedure." *The Annals of Statistics*, 36: 337–363. MR2387974. doi: https://doi.org/10.1214/009053607000000550. 37

Sarkar, S. K. and Chang, C.-K. (1997). "The Simes method for multiple hypothesis testing with positively dependent test statistics." *Journal of the American Statistical Association*, 92: 1601–1608. MR1615269. doi: https://doi.org/10.2307/2965431. 37

Sarkar, T. K. (1969). "Some lower bounds of realiability." *Technical Report No. 124*, Dept. of Operations Research and Statistics, Stanford University, Stanford, CA. MR2618992. 36, 37

Schmeidler, D. (1989). "Subjective Probability and Expected Utility without Additivity." *Econometrica*, 57: 571–587. MR0999273. doi: https://doi.org/10.2307/1911053. 34

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic orders*. New York, NY, USA: Springer. MR2265633. doi: https://doi.org/10.1007/978-0-387-34675-5. 32, 34, 36, 38, 43

Thompson, W. A. J. (1988). *Point Process Models with Applications to Safety and Reliability*. New York: Chapman & Hall. MR0958126. doi: https://doi.org/10.1007/978-1-4613-1067-9. 46

Topkis, D. M. (1978). "Minimizing a submodular function on a lattice." *Operations Research*, 26: 305–321. MR0468177. doi: https://doi.org/10.1287/opre.26.2.305. 36

Topkis, D. M. (1998). *Supermodularity and Complementarity*. Princeton, NJ, USA: Princeton University Press. 36

Valdez, E. A. and Xiao, Y. (2011). "On the Distortion of a Copula and its Margins." *Scandinavian Actuarial Journal*, 4: 292–317. MR2864198. doi: https://doi.org/10.1080/03461238.2010.490021. 34

Wang, S. (1995). "Insurance pricing and increased limits ratemaking by proportional hazards transforms." *Insurance: Mathematics and Economics*, 17: 43–54. MR1363642. doi: https://doi.org/10.1016/0167-6687(95)91054-P. 34

Wang, S. (1996). "Premium calculation by transforming the layer premium density." *ASTIN Bulletin*, 26: 71–92. 34

Whitt, W. (1980). "The Effect of Variability in the GI/G/s Queue." *Journal of Applied Probability*, 17: 1062–1071. MR0587206. doi: https://doi.org/10.2307/3213215. 34

Yaari, M. E. (1987). "The dual theory of choice under risk." *Econometrica*, 55: 95–115. MR0875518. doi: https://doi.org/10.2307/1911158. 34