BAYESIAN ROBUSTNESS FOR CLASSES OF BIDIMENSIONAL PRIORS WITH GIVEN MARGINALS

BY BRUNERO LISEO[1], ELIAS MORENO[2] AND GABRIELLA SALINETTI
Università di Roma "La Sapienza" and Universidad de Granada

We address the problem of finding the range of the posterior expectation of an arbitrary function of the parameters when the prior distribution varies in an ε-contamination class and the resulting priors have specified marginals. This problem, which is an example of the Monge-Kantorovich problem has not yet received a complete solution. We provide an accurate approximation, by considering, as the contamination class, the set of priors with one specified marginal and an arbitrary number n of specified quantiles on the other coordinate. We show that, by using Moment Problem Theory, this problem can be restated in a more tractable form, and provide some interesting illustrations in which posterior robustness is achieved.

1. Introduction and problem setting. In problems involving vector valued parameters, the elicitation of multivariate prior distributions is extremely challenging and any choice should be carefully investigated from the sensitivity viewpoint. Elicitation often proceeds by eliciting the one dimensional marginal distributions, and it is then natural to consider the class of all joint prior distributions with that given set of univariate marginals, namely the Fréchet class $\mathcal{Q}$.

As already discussed in Lavine, Wasserman and Wolpert (1991) and Moreno and Cano (1995), the use of $\mathcal{Q}$ as the class of plausible priors will typically give uselessly large ranges for the posterior expectation of a given quantity of interest, due to the extremely huge size of $\mathcal{Q}$ [see Walley (1991) p.298, for an example].

However the class $\mathcal{Q}$ is particularly useful as the contaminating class when using an ε-contaminated neighbourhood of priors

$$(1)\quad \Gamma(\mathcal{Q},\varepsilon)=\{\Pi:\Pi(d\theta_1,d\theta_2)=(1-\varepsilon)\Pi_0(d\theta_1,d\theta_2)+\varepsilon Q(d\theta_1,d\theta_2),Q\in\mathcal{Q}\}$$

where Π_0 is the base elicited prior (usually $\Pi_0\in\mathcal{Q}$), $(1-\varepsilon)$ quantifies the confidence in Π_0 and $\mathcal{Q}$ is the class of allowed contaminations.

The class $\Gamma(\mathcal{Q},\varepsilon)$ would model those situations where one is rather confident in the marginals elicitation, but a sensitivity analysis is still necessary, with respect to departures from Π_0 which preserve the marginals.

[1] Research partially supported by the CNR.
[2] Research partially supported by DGICIT (PB93-1154)
[3] *AMS Subject classification*: Primary 62F15; Secondary 62A15.
[4] *Key words and phrases*: Multivariate robustness; Priors with given marginals, ε-contamination classes, Moment Problem.

Finding the range of the posterior expectation of a given measurable function when the prior probability measure varies over $\Gamma(\mathcal{Q}, \varepsilon)$ reduces (once appropriately linearized) to the so called *Monge-Kantorovich mass transference* problem [see for example Rachev (1985)]. Its role in robust Bayesian inference was first addressed in Lavine, Wasserman and Wolpert (1991). They proposed an approximation of the contaminating class $\mathcal{Q}$ based on a discretization of the marginals. Moreno and Cano (1995) provided a more accurate approximation in the bivariate case: one of the marginals is completely specified and a rough approximation of the other is assumed. Namely they consider the contaminating class

$$\mathcal{Q}_a = \left\{ Q(d\theta_1, d\theta_2) : Q_2(d\theta_2) \text{ fixed}, \int_A Q_1(d\theta_1) = \alpha \right\}, \tag{2}$$

where α and A are given and Q_1 and Q_2 denote the marginal measures of Q for θ_1 and θ_2 respectively. This only specifies the prior probability of two subsets of Θ_1, i.e. A and $\Theta_1 - A$. The mass transference problem that naturally arises with the computation of the extrema of the posterior expectation of a function $\psi(\theta_1, \theta_2)$ when the prior varies over $\Gamma(\mathcal{Q}_a, \varepsilon)$ is solved in Moreno and Cano (1995) using arguments which represent an extension of the usual methodology for quantile classes [Moreno and Cano (1991)].

Unfortunately this approach does not seem to work when a more refined partition of Θ_1, say $(A_1, A_2, \cdots, A_n)$, is considered, that is when we consider the class of contaminations

$$\mathcal{Q}_1 = \left\{ Q(d\theta_1, d\theta_2) : Q_2(d\theta_2) \text{ fixed}, \int_{A_i} Q_1(d\theta_1) = \alpha_i, i = 1, \cdots, n \right\}. \tag{3}$$

In this paper we address the problem of finding the range of a posterior functional over the class $\Gamma(\mathcal{Q}_1, \varepsilon)$ using the general Moment Problem Theory.

Let $\theta = (\theta_1, \theta_2)$ be a vector of real parameters, with $\theta \in \Theta = \Theta_1 \times \Theta_2$. The likelihood function of the data x will be denoted $f(\theta_1, \theta_2)$ while $\psi(\theta_1, \theta_2)$ is a measurable function which represents our quantity of interest. The symbol $E^P(Y)$ will denote the expected value of Y with respect to P.

The problem consists in finding

$$\inf \left\{ \frac{\int \psi(\theta) f(\theta) \Pi(d\theta)}{\int f(\theta) \Pi(d\theta)}, \Pi \in \Gamma(\mathcal{Q}_1, \varepsilon) \right\}$$

$$= \inf \left\{ \frac{(1-\varepsilon) \int \psi(\theta) f(\theta) \Pi_0(d\theta) + \varepsilon \int \psi(\theta) f(\theta) Q(d\theta)}{(1-\varepsilon) \int f(\theta) \Pi_0(d\theta) + \varepsilon \int f(\theta) Q(d\theta)} : Q \in \mathcal{Q}_1 \right\} \tag{4}$$

(of course the supremum of the same quantity will be treated in a similar way); here $\mathcal{Q}_1$ is as in (3), the sets A_i are given and the probabilities α_i are fixed coherently with the Π_0-marginal on Θ_1.

We show that problem (4) can be restated in a more tractable form, namely a linear $n-1$ dimensional optimization problem and a standard root-finding operation, through a direct application of Moment Problem Theory (MPT, henceforth). The approximation approach here proposed moves along the lines of Lavine, Wasserman and Wolpert (1991) and Moreno and Cano (1995), extending Moreno and Cano (1995). The results, possibly complemented with convergence and error analyses, could also be used to guide the refinement of the partition towards better approximations.

Kemperman (1987) will be used as a reference for the MPT. The relevance of this theory to solve variational problems in Robust Bayesian inference was first noticed in Sivaganesan and Berger (1989) for several constrained contamination classes. Salinetti (1994) discussed the use of MPT for most of the usual classes of priors and combinations thereof. It is relevant to observe that, in this proposed approximation of the fixed marginals problem, the MPT approach is not only powerful from the operational point of view: also, it helps to clarify "how" the solution to the mass transference problem actually works (see Section 2 for details). Section 4 deals with some particular choices of the function of interest, where substantial computational simplifications are obtained. Illustrations and discussion of the results are given in the last section.

2. Preliminary results: linear robustness and MPT. Let $\mathcal{Q}_0$ be the class of all probability measures with marginal Q_2 fixed. Minimizing a linear functional over $\mathcal{Q}_0$ is easy, as it can be directly derived or obtained as a particular case of Lemma 1 in Moreno and Cano (1995):

LEMMA 1. *Let $\psi : \Theta \to \mathbb{R}$ be a measurable function. To avoid trivial degenerate cases assume that, for every $Q \in \mathcal{Q}_0$, $\int \psi(\theta_1, \theta_2)\, Q(d\theta_1, d\theta_2)$ and $\int_{\Theta_2} \inf_{\theta_1 \in \Theta_1} \psi(\theta_1, \theta_2)\, Q_2(d\theta_2)$ exist, possibly infinite. Then*

$$(5) \qquad \inf_{Q \in \mathcal{Q}_0} \left\{ \int_\Theta \psi(\theta_1, \theta_2) Q(d\theta_1, d\theta_2) \right\} = \int_{\Theta_2} \inf_{\theta_1 \in \Theta_1} \{\psi(\theta_1, \theta_2)\}\, Q_2(d\theta_2).$$

The class $\mathcal{Q}_1$ given in (3) with fixed marginal $Q_2 = \Pi_{02}$ can be expressed as a subclass of $\mathcal{Q}_0$, in the form

$$\mathcal{Q}_1 = \left\{ Q \in \mathcal{Q}_0 : \int_{A_i \times \Theta_2} Q(d\theta_1, d\theta_2) = \alpha_i;\ i = 1, \dots, n \right\},$$

where $\alpha_i = \Pi_{01}(A_i)$, $0 < \alpha_i < 1$, $i = 1, \dots, n$, $(A_1, A_2, \dots, A_n)$ is a partition of Θ_1, and Π_{0j} denotes the Π_0 marginal distribution over θ_j, $j = 1, 2$.

Consider the following linear minimization problem over $\mathcal{Q}_1$:

$$\underline{\psi} = \inf_{Q \in \mathcal{Q}_1} \left\{ \int \psi dQ \right\} = \inf_{Q \in \mathcal{Q}_0} \left\{ \int \psi\, dQ; \int_\Theta I_{C_i} dQ = \alpha_i, i = 1, \dots, n-1 \right\}$$

where $C_i = A_i \times \Theta_2$, $i = 1, \ldots, n-1$, and I_A is the indicator function of a set A, and $\underline{\psi}$ is assumed to be finite.

In this form $\underline{\psi}$ is a generalized moment problem. According to MPT [Kemperman (1987), specifically Theorem 2.20 restricted to the single problem $\underline{\psi}$], since $\mathcal{Q}_0$ is a convex class, the subclass $\mathcal{Q}_1$ is non empty, the constraints are linearly independent and, as a consequence of the assumptions, the vector $\alpha = (\alpha_1, \ldots, \alpha_{n-1})$ is an interior point of the set

$$\mathcal{Y} = \left\{ y = (y_1, \ldots, y_{n-1}) \in \mathbb{R}^{n-1}; \int I_{C_i} dQ = y_i = Q(C_i); Q \in \mathcal{Q}_0 \right\},$$

where $i = 1, \ldots, n-1$, we obtain

$$\underline{\psi} = \sup_{d \in \mathbb{R}^{n-1}} \left\{ \sum_{i=1}^{n-1} d_i \alpha_i + \inf_{Q \in \mathcal{Q}_0} \int \left(\psi(\theta_1, \theta_2) - \sum_{i=1}^{n-1} d_i I_{C_i}(\theta_1, \theta_2) \right) Q(d\theta_1, d\theta_2) \right\}.$$

Then, by (5),

$$\underline{\psi} = \sup_{d \in \mathbb{R}^{n-1}} \left\{ \sum_{i=1}^{n-1} d_i \alpha_i + \int \inf_{\theta_1 \in \Theta_1} \left(\psi(\theta_1, \theta_2) - \sum_{i=1}^{n-1} d_i I_{C_i}(\theta_1, \theta_2) \right) Q_2(d\theta_2) \right\},$$

or, equivalently,

$$\begin{aligned} (6) \qquad \underline{\psi} \quad = \quad & \sup_{d \in \mathbb{R}^n, d_n = 0} \left\{ \sum_{i=1}^{n} d_i \alpha_i \right. \\ & \left. + \int \inf_{\theta_1 \in \Theta_1} \left(\psi(\theta_1, \theta_2) - \sum_{i=1}^{n} d_i I_{C_i}(\theta_1, \theta_2) \right) Q_2(d\theta_2) \right\}, \end{aligned}$$

The presence of the indicator functions in the integrand of (6) allows a considerable simplification. For each θ_2 and for each A_i set

$$\psi_i(\theta_2) = \inf_{\theta_1 \in A_i} \psi(\theta_1, \theta_2),$$

and, for fixed $d = (d_1, \ldots, d_{n-1}, 0)$ let

$$S_{ij} = S(d_i, d_j) = \{\theta_2 : \psi_i(\theta_2) - \psi_j(\theta_2) < d_i - d_j\}.$$

Set

$$S_i = \left(\bigcap_{j=1}^{i-1} S_{ji}^c \right) \bigcap \left(\bigcap_{j=i+1}^{n} S_{ij} \right), i = 1, \ldots, n,$$

where S_{ji}^c denotes the complement of S_{ji} in Θ_2. Also we will adopt the convention that $\bigcap_{j=1}^{0} S_{ij} = \bigcap_{j=n+1}^{n} S_{ij} = \Theta_2$. It can be shown that the sets $\{S_1, \ldots, S_n\}$ form a partition of Θ_2 (see the Appendix).

Then the integrand in (6), as a function of θ_2, can be written as

$$\text{(7)} \qquad \inf_{\theta_1 \in \Theta_1} \left(\psi(\theta_1, \theta_2) - \sum_{i=1}^{n} d_i I_{C_i}(\theta_1, \theta_2) \right) = \sum_{i=1}^{n} (\psi_i(\theta_2) - d_i) I_{S_i}(\theta_2)$$

Using this in (6) allows one to conclude

THEOREM 1. *The minimum value,* $\underline{\psi}$, *is*

$$\begin{aligned} \underline{\psi} &= \inf_{Q \in \mathcal{Q}_1} \int \psi dQ \\ &= \sup_{d \in \mathrm{R}^n, d_n = 0} \left\{ \int \sum_{i=1}^{n} \psi_i(\theta_2) I_{S_i}(\theta_2) Q_2(d\theta_2) + \sum_{i=1}^{n} d_i(\alpha_i - Q_2(S_i)) \right\} \\ &= \sup_{d \in \mathrm{R}^n, d_n = 0} \left\{ \int \sum_{i=1}^{n} \psi_i(\theta_2) I_{S_i}(\theta_2) Q_2(d\theta_2) + \sum_{i=1}^{n-1} (d_i - d_n)(\alpha_i - Q_2(S_i)) \right\}. \end{aligned}$$

PROOF. The first expression directly follows from using (7) in (6). The second expression is derived observing that $\sum_i \alpha_i = \sum_i Q_2(S_i) = 1.\square$

REMARK 1. Theorem 1 shows how the solution to the mass transference problem operates. First, curves are determined over $C_1, C_2, \ldots, C_n$ through the functions $\psi_1, \psi_2, \ldots, \psi_n$. Second, pieces of these curves are chosen, taking into account the differences $\psi_i(\theta_2) - \psi_j(\theta_2)$. The selection is optimal as a consequence of the MPT. Thus, the extremal prior $Q^*(d\theta_1, d\theta_2)$ will put all its mass on these pieces of curves.

3. Ratio-linear robustness on $\Gamma(\mathcal{Q}_1, \varepsilon)$. Consider the problem stated in Section 1 of determining

$$\text{(8)} \qquad \underline{\lambda} = \inf \left\{ \frac{\int \psi(\theta_1, \theta_2) f(\theta_1, \theta_2) \Pi(d\theta_1, d\theta_2)}{\int f(\theta_1, \theta_2) \Pi(d\theta_1, d\theta_2)}; \Pi \in \Gamma(\mathcal{Q}_1, \varepsilon) \right\}$$

where $\psi(\theta_1, \theta_2)$ is the function of interest, and $f(\theta_1, \theta_2)$ is the likelihood function. We assume that $\int f d\Pi_0 > 0$ and the likelihood function is bounded, so that

$$\inf \left\{ \int f d\Pi, \Pi \in \Gamma(\mathcal{Q}_1, \varepsilon) \right\} > 0 \text{ and } \sup \left\{ \int f d\Pi, \Pi \in \Gamma(\mathcal{Q}_1, \varepsilon) \right\} < +\infty.$$

This makes the linearization technique [Wasserman, Lavine and Wolpert (1993)] applicable in solving (8) and the value $\underline{\lambda}$ is the unique solution of the equation

$$G(\lambda) = \inf \left\{ \int (\psi(\theta_1, \theta_2) - \lambda) f(\theta_1, \theta_2) \Pi(d\theta_1, d\theta_2);\ \Pi \in \Gamma(\mathcal{Q}_1, \varepsilon) \right\} = 0,$$

or, equivalently,

$$
\begin{aligned}
G(\lambda) &= (1-\varepsilon)E^{\Pi_0}\left(\psi(\theta_1,\theta_2)-\lambda|x\right)m(x|\Pi_0) \\
&+ \varepsilon \inf\left\{E^{Q}\left[(\psi(\theta_1,\theta_2)-\lambda)f(\theta_1,\theta_2)\right]; Q\in\mathcal{Q}_1\right\} = 0,
\end{aligned} \tag{9}
$$

where $m(x|\Pi_0) = E^{\Pi_0} f(\theta_1,\theta_2)$. In accordance with Theorem 1 the equation (9) can be expressed as

$$
\begin{aligned}
G(\lambda) &= (1-\varepsilon)E^{\Pi_0}\left(\psi(\theta_1,\theta_2)-\lambda|x\right)m(x|\Pi_0) \\
&+ \varepsilon \sup_{d\in\mathbb{R}^n, d_n=0}\left\{\int\sum_{i=1}^{n}\underline{\psi}_i(\theta_2,\lambda)I_{S_i(d,\lambda)}(\theta_2)Q_2(d\theta_2)\right. \\
&+ \left.\sum_{i=1}^{n}(d_i-d_n)\left(\alpha_i - Q_2(S_i(d,\lambda))\right)\right\}
\end{aligned}
$$

where, for $i=1,\ldots,n$,

$$
\underline{\psi}_i(\theta_2,\lambda) = \inf_{\theta_1\in A_i}\left\{\left[\psi(\theta_1,\theta_2)-\lambda\right]f(\theta_1,\theta_2)\right\},
$$

while

$$
S_i(d,\lambda) = \left(\bigcap_{j=1}^{i-1} S_{ji}^{c}\right)\bigcap\left(\bigcap_{j=i+1}^{n} S_{ij}\right)
$$

and

$$
S_{ij} = S_{ij}(d,\lambda) = \left\{\theta_2 : \underline{\psi}_i(\theta_2,\lambda) - \underline{\psi}_j(\theta_2,\lambda) < d_i - d_j\right\}.
$$

4. Some particular cases. The proposed approach is quite general and it can be applied to any function $\psi : \Theta \to \mathbb{R}$ such that its posterior mean is a linear functional or a ratio of linear functionals of Π.

Neverthless there are particular cases which deserve special attention for several different reasons, specifically:

1. $\psi(\theta_1,\theta_2) = I_B(\theta_1,\theta_2)$. In this case we are interested in the posterior probability of a given set B, with $B\in\mathcal{B}(\Theta_1\times\Theta_2)$, the Borel σ-field of subsets of the parameter space. We will show that, in this case, the computional effort is considerably reduced. Important choices of B are

   1a) $B = \{\theta : \theta_1 < \theta_2\}$ which arises in, say, the hypotheses testing settings.

   1b) $B = B(t_1,t_2) = \{\theta : \theta_1 < t_1, \theta_2 < t_2\}$, which is the case of the posterior range of the joint posterior cumulative distribution function of (θ_1,θ_2).

2. $\psi(\theta_1, \theta_2) = \psi(\theta_1)$ or $\psi(\theta_2)$. In this case we are actually interested in a function of one parameter only, the other being a nuisance parameter. This case is very common in statistical applications and turns out to be very important in our context because the elicitation of the marginals is typically much more informative here than in the general case. We will discuss this case in Examples 2 and 3.

In this section we discuss detailed solutions of the particular cases 1*a* and 1*b*. It should be noted that the major computational simplifications are obtained when we are able to write, in a closed form, the functions $\underline{\psi}_i(\theta_2, \lambda)$, $\overline{\psi}_i(\theta_2, \lambda)$, and the sets $S_i(d, \lambda)$.

Consider now the general case where the function of interest $\psi(\theta_1, \theta_2)$ is the indicator function of some set B, and suppose one chooses, as a quantile partition on Θ_1, the sets

$$A_1 = (-\infty, a_1)\,,\ A_2 = [a_1, a_2)\,, \ldots, A_n = [a_{n-1}, +\infty)\,. \tag{10}$$

We define the following sets: for each $i = 1, \ldots, n$, let

$$A_i(\theta_2, B) = \{\theta_1 \in A_i : (\theta_1, \theta_2) \in B\}$$

and its A_i-complementary set

$$\overline{A}_i(\theta_2, B) = \{\theta_1 \in A_i : (\theta_1, \theta_2) \notin B\}\,.$$

Then, for each $\lambda \in \mathbb{R}$, it can be shown that

$$\underline{\psi}_i(\theta_2, \lambda) = \begin{cases} -\lambda \overline{f}_i^{sup}(\theta_2) & \text{if } \overline{A}_i(\theta_2, B) \neq \emptyset \\ (1-\lambda) f_i^{inf}(\theta_2) & \text{if } \overline{A}_i(\theta_2, B) = \emptyset \end{cases}$$

where

$$\overline{f}_i^{sup}(\theta_2) = \sup_{\theta_1 \in \overline{A}_i(\theta_2, B)} f(\theta_1, \theta_2)$$

and

$$f_i^{inf}(\theta_2) = \inf_{\theta_1 \in A_i(\theta_2, B)} f(\theta_1, \theta_2).$$

Similar results hold in the supremum case, and

$$\overline{\psi}_i(\theta_2, \lambda) = \begin{cases} (1-\lambda) f_i^{sup}(\theta_2) & \text{if } A_i(\theta_2, B) \neq \emptyset \\ -\lambda \overline{f}_i^{inf}(\theta_2) & \text{if } A_i(\theta_2, B) = \emptyset \end{cases},$$

where

$$\overline{f}_i^{inf}(\theta_2) = \inf_{\theta_1 \in \overline{A}_i(\theta_2, B)} f(\theta_1, \theta_2)$$

and

$$f_i^{sup}(\theta_2) = \sup_{\theta_1 \in A_i(\theta_2, B)} f(\theta_1, \theta_2).$$

We now give the explicit form of the $\underline{\psi}_i$'s in the case $B = \{(\theta_1, \theta_2) : \theta_1 < \theta_2\}$. Similar results hold for the $\overline{\psi}_i$'s and they will therefore be omitted.

THEOREM 2. *For each real value* λ *one has*

$$\underline{\psi}_1(\theta_2, \lambda) = \begin{cases} -\lambda \sup_{\theta_1 \in (\theta_2, a_1)} f(\theta_1, \theta_2) & \text{if } \theta_2 \leq a_1 \\ (1-\lambda) \inf_{\theta_1 \in A_1} f(\theta_1, \theta_2) & \text{if } \theta_2 > a_1 \end{cases};$$

for $1 < i < n$,

$$\underline{\psi}_i(\theta_2, \lambda) = \begin{cases} -\lambda \sup_{\theta_1 \in A_i} f(\theta_1, \theta_2) & \text{if } \theta_2 \leq a_{i-1} \\ -\lambda \sup_{\theta_1 \in (\theta_2, a_i)} f(\theta_1, \theta_2) & \text{if } a_{i-1} < \theta_2 \leq a_i \\ (1-\lambda) \inf_{\theta_1 \in A_i} f(\theta_1, \theta_2) & \text{if } \theta_2 > a_i \end{cases};$$

$$\underline{\psi}_n(\theta_2, \lambda) = \begin{cases} -\lambda \sup_{\theta_1 \in A_n} f(\theta_1, \theta_2) & \text{if } \theta_2 \leq a_{n-1} \\ -\lambda \sup_{\theta_1 \in (\theta_2, +\infty)} f(\theta_1, \theta_2) & \text{if } \theta_2 > a_{n-1} \end{cases}.$$

As a second application, consider the case $B = B(t_1, t_2) = \{(\theta_1, \theta_2) : \theta_1 < t_1, \theta_2 < t_2\}$.

THEOREM 3. *For any* $(t_1, t_2) \in \mathbb{R}^2$ *and for each real value* λ, *given the partition (10), one has*

i) for each index i such that $a_i \leq t_1$

$$\underline{\psi}_i(\theta_2, \lambda) = \begin{cases} (1-\lambda) \inf_{\theta_1 \in A_i} f(\theta_1, \theta_2) & \theta_2 \leq t_2 \\ -\lambda \sup_{\theta_1 \in A_i} f(\theta_1, \theta_2) & \theta_2 > t_2 \end{cases};$$

ii) if $a_{i-1} < t_1 \leq a_i$

$$\underline{\psi}_i(\theta_2, \lambda) = \begin{cases} -\lambda \sup_{\theta_1 \in (t_1, a_i)} f(\theta_1, \theta_2) & \theta_2 \leq t_2 \\ -\lambda \sup_{\theta_1 \in A_i} f(\theta_1, \theta_2) & \theta_2 > t_2 \end{cases};$$

iii) for each index i such that $a_{i-1} \geq t_1$

$$\underline{\psi}_i(\theta_2, \lambda) = -\lambda \sup_{\theta_1 \in A_i} f(\theta_1, \theta_2).$$

5. Examples. EXAMPLE 1. Let (X_1, X_2) be a bivariate normal random variable $N((\theta_1, \theta_2), I_2))$, where I_k is a $k \times k$ identity matrix, and suppose the base prior $\Pi_0(\theta_1, \theta_2)$ is the standard normal bivariate distribution $N((0,0), I_2)$. We are interested in the robustness of the posterior probability of $H_0 : \theta_1 < \theta_2$ to departures from Π_0 which preserve the marginals. We will consider several nested contamination classes, namely $\Gamma(\mathcal{P}, \varepsilon)$, $\Gamma(\mathcal{Q}_0, \varepsilon)$, $\Gamma(\mathcal{Q}_1, \varepsilon)$ and $\Gamma(\mathcal{Q}_3, \varepsilon)$, where

- $\mathcal{P} = \{$*All the distributions over* $\mathcal{B}(\Theta_1, \Theta_2)\}$
- $\mathcal{Q}_0 = \{Q \in \mathcal{P} :$ *the* Θ_2*-marginal* $Q_2(d\theta_2) = \Pi_{02}(d\theta_2)\}$
- $\mathcal{Q}_1 = \{Q \in \mathcal{Q}_0 : \; Q_1(-\infty, 0) = \frac{1}{2}\}$
- $\mathcal{Q}_3 = \{Q \in \mathcal{Q}_1 : \; Q_1(-\infty, -\nu) = Q_1(\nu, \infty) = \frac{3}{8};\; Q_1(-\nu, 0) = Q_1(0, \nu) = \frac{1}{8}\}$;

here $\nu = .3186$ is the .625-percentile of a standard normal distribution. Tables 1,2 and 3 shows the results for different representative data sets, $(x_1, x_2) = (0,0), (0,3)$ and $(3,0)$, respectively, and several values of ε. Note that $\Pi_0(B|(0,0)) = .5$, $\Pi_0(B|(3,0)) = .933$ and $\Pi_0(B|(0,3)) = .067$.

TABLE 1. *Posterior bounds for the four classes when* $(x_1, x_2) = (0,0)$.

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\, \Gamma(\mathcal{P}, \varepsilon)$	.409	.333	.270	.214	.166	.125	.089	.050	.026	0
$inf\, \Gamma(\mathcal{Q}_0, \varepsilon)$	.437	.378	.322	.268	.218	.170	.125	.081	.039	0
$inf\, \Gamma(\mathcal{Q}_1, \varepsilon)$	.437	.378	.322	.268	.218	.170	.125	.081	.039	0
$inf\, \Gamma(\mathcal{Q}_3, \varepsilon)$	.439	.378	.322	.269	.219	.171	.127	.082	.040	0
$sup\, \Gamma(\mathcal{Q}_3, \varepsilon)$	.561	.621	.677	.730	.780	.828	.872	.918	.959	1
$sup\, \Gamma(\mathcal{Q}_1, \varepsilon)$	.563	.621	.677	.731	.781	.829	.874	.919	.96	1
$sup\, \Gamma(\mathcal{Q}_0, \varepsilon)$	.563	.621	.678	.731	.781	.829	.874	.919	.96	1
$sup\, \Gamma(\mathcal{P}, \varepsilon)$	.590	.666	.730	.785	.833	.875	.911	.944	.973	1

TABLE 2. *Posterior bounds for the four classes when* $(x_1, x_2) = (0,3)$.

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\, \Gamma(\mathcal{P}, \varepsilon)$	.764	.623	.503	.401	.312	.234	.165	.104	.049	0
$inf\, \Gamma(\mathcal{Q}_0, \varepsilon)$	.880	.821	.756	.684	.604	.514	.411	.294	.158	0
$inf\, \Gamma(\mathcal{Q}_1, \varepsilon)$	.880	.821	.756	.684	.604	.516	.414	.296	.160	0
$inf\, \Gamma(\mathcal{Q}_3, \varepsilon)$	.880	.821	.756	.684	.604	.516	.414	.296	.160	0
$sup\, \Gamma(\mathcal{Q}_3, \varepsilon)$	.941	.948	.956	.963	.970	.976	.982	.987	.992	1
$sup\, \Gamma(\mathcal{Q}_1, \varepsilon)$	.941	.948	.956	.963	.970	.976	.982	.987	.992	1
$sup\, \Gamma(\mathcal{Q}_0, \varepsilon)$	.942	.950	.958	.965	.972	.978	.984	.989	.995	1
$sup\, \Gamma(\mathcal{P}, \varepsilon)$	.978	.988	.992	.995	.996	.997	.998	.999	.999	1

The tables show a sensible reduction in the posterior range when we reduce the contaminating class from $\mathcal{P}$ to $\mathcal{Q}_0$, that is, when we fix a marginal.

Also, improvements are significant only for small to moderate values of ε. When $\varepsilon = 1$ the range is always, hopelessly, (0,1).

This is an example where knowledge of the marginals need not dramatically reduce the posterior range: it depends on the shape of B. Consider, for example, the case $(x_1 = 0, x_2 = 0)$: roughly speaking the situation is that, for each value of θ_2, there are points $(\theta_1, \theta_2) \in B$ and points $(\theta_1, \theta_2) \notin B$ with similar likelihood values and it is thus possible to construct two priors, Q_L and Q_U, with the prescribed marginals, such that $Q_L(B|x) = 0$ and $Q_U(B|x) = 1$. With different shapes of the set of interest, the behaviour can be quite different. We will show examples where the posterior range of the probability of a set is substantially less than one, even in the case $\varepsilon = 1$ (i.e. no base prior).

TABLE 3. *Posterior bounds for the four classes when $(x_1, x_2) = (3, 0)$.*

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\,\Gamma(\mathcal{P},\varepsilon)$	.021	.011	.007	.005	.003	.002	.001	.0008	.0003	0
$inf\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.026	.015	.010	.006	.004	.003	.002	.001	.0005	0
$inf\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.033	.020	.013	.010	.006	.005	.003	.002	.0008	0
$inf\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.038	.025	.017	.013	.009	.007	.005	.0023	.0011	0
$sup\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.114	.174	.234	.309	.390	.482	.584	.700	.835	1
$sup\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.119	.178	.240	.315	.395	.486	.589	.706	.841	1
$sup\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.120	.180	.242	.309	.390	.482	.584	.700	.841	1
$sup\,\Gamma(\mathcal{P},\varepsilon)$	.230	.370	.500	.600	.688	.766	.835	.896	.950	1

EXAMPLE 2. Let (X_1, X_2) be a bivariate Cauchy distribution with density

$$f(x_1, x_2|\theta_1, \theta_2) = \frac{1}{\pi(1 + (x_1 - \theta_1)^2)} \frac{1}{\pi(1 + (x_2 - \theta_2)^2)}$$

and suppose the base prior is a product of Cauchy(0,1) distributions, with density

$$\pi_0(\theta_1, \theta_2) = \frac{1}{\pi(1 + \theta_1^2)} \frac{1}{\pi(1 + \theta_2^2)}.$$

We are now interested in the posterior robustness of the probability of $B = \{(\theta_1, \theta_2) : \theta_1 > 0\}$. Also in this case we will consider four possible classes of contaminations, namely

- $\mathcal{P} = \{$*All the distributions on* $\mathcal{B}(\Theta_1, \Theta_2)\}$
- $\mathcal{Q}_0 = \{Q \in \mathcal{P} :$ *the* Θ_2 *-marginal* $Q_2(d\theta_2) = \Pi_{02}(d\theta_2)\}$
- $\mathcal{Q}_1 = \{Q \in \mathcal{Q}_0 : \ Q_1(-\infty, 0) = .5\}$
- $\mathcal{Q}_3 = \{Q \in \mathcal{Q}_1 : \ Q_1(-\infty, -\nu) = Q_1(-\nu, 0) = Q_1(0, \nu) = Q_1(\nu, \infty) = .25\}$;

here $\nu = 1$ is the .75-percentile of a Cauchy(0,1) distribution. We show the results (Table 4 and 5) for the data sets ($x_1 = 0, x_2 = 0$) and ($x_1 = 2, x_2 = 0$). Note that $\Pi_0(B|(0,0)) = .5$ and $\Pi_0(B|(2,0)) = .804$

TABLE 4. *Posterior bounds for the four classes, in the Cauchy example, when* $(x_1, x_2) = (0,0)$.

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\,\Gamma(\mathcal{P},\varepsilon)$	.346	.250	.184	.136	.100	.072	.048	.030	.014	0
$inf\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.409	.333	.269	.214	.166	.125	.088	.055	.026	0
$inf\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.423	.355	.294	.240	.190	.145	.104	.066	.032	0
$inf\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.450	.400	.350	.300	.250	.200	.150	.100	.050	0
$sup\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.550	.600	.650	.700	.750	.800	.850	.900	.950	1
$sup\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.577	.645	.706	.760	.810	.855	.896	.934	.968	1
$sup\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.591	.667	.731	.786	.837	.875	.912	.945	.974	1
$sup\,\Gamma(\mathcal{P},\varepsilon)$	.653	.750	.816	.864	.900	.928	.952	.970	.986	1

Improvements in terms of posterior ranges through the classes are dramatically more significant in this example, especially in the case $(x_1, x_2) = (2,0)$ (see Table 5). Note that, in this case, the posterior range has been reduced even for ε=1.

TABLE 5. *Posterior bounds for the four classes, in the Cauchy example, when* $(x_1, x_2) = (2,0)$.

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\,\Gamma(\mathcal{P},\varepsilon)$	.682	.574	.477	.389	.309	.236	.169	.108	.052	0
$inf\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.738	.670	.598	.524	.446	.365	.280	.191	.098	0
$inf\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.749	.691	.628	.560	.486	.406	.318	.222	.117	0
$inf\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.760	.710	.664	.607	.544	.472	.390	.295	.185	.117
$sup\,\Gamma(\mathcal{Q}_3,\varepsilon)$	.824	.846	.868	.888	.908	.926	.944	.961	.978	.994
$sup\,\Gamma(\mathcal{Q}_1,\varepsilon)$	.856	.892	.918	.938	.954	.967	.977	.986	.993	1
$sup\,\Gamma(\mathcal{Q}_0,\varepsilon)$	.864	.902	.927	.946	.960	.972	.981	.988	.994	1
$sup\,\Gamma(\mathcal{P},\varepsilon)$	.895	.934	.955	.969	.978	.985	.990	.994	.997	1

The reason of this phenomenon is relatively clear. If the indicator function of interest $I_B(\theta_1, \theta_2)$ does not depend (or weakly depends) on θ_2 and the bulk of the likelihood function is contained in B, then the degree of elicitation over θ_1 becomes critical.

In other words, the role of the marginal priors can be more or less effective according to:

- the shape of the set of interest B; typically, if B depends on one coordinate only, then the marginal elicitation is important. This is not the only case, however.

- the relative weights of the likelihood function on B and $\overline{B}$
- the shape of the marginals priors.

According to different combinations of the above scenarios the effect of adding more prior constraints can be less or more dramatic. For example, the use of a thin-tailed distribution like the Normal instead of a thick-tailed one like a Cauchy (either as likelihood function or as marginal priors) can sensibly modify the posterior ranges. For a thorough discussion of the tail effects on reported inference see O'Hagan (1988). We analyze this aspect by considering the following simple modifications of Example 2.

EXAMPLE 3. Suppose we observe a random variable (X_1, X_2) whose density function is either a bivariate normal $N((\theta_1, \theta_2), 2.19 I_2)$ or the product of two Cauchy $(\theta_i, 1)$, $i = 1, 2$, as in Example 2. Suppose further that the base prior is either Π_{0N}, a bivariate Normal distribution with vector mean zero and covariance matrix $2.19\, I_2$ or Π_{0C}, the product of two standard Cauchy distributions. In this example the value 2.19 for the normal variances is chosen in order to match the quartiles of a standard Cauchy distribution [Berger (1985),p.195]. Assume that we observe $(x_1, x_2) = (2, 0)$ and that we are interested, as in Example 2, in the range of the posterior probability of the set $B = \{(\theta_1, \theta_2) : \theta_1 > 0\}$.

We already discussed the case where both the marginal priors and the likelihood function were Cauchy distributions (CC case). Now we compare those results with the ones obtained by using the other 3 possible combinations (CN, NC and NN, respectively) of marginal priors and likelihood, when the class of allowed contaminations is either $\mathcal{Q}_{3N}$, when using Π_{0N}, or $\mathcal{Q}_{3C}$, when using Π_{0C} and

$$\begin{aligned} \mathcal{Q}_{3N} &= \{Q \in \mathcal{P} : Q \text{ has the same } Q_1\text{-marginal as } \Pi_{0N} \\ &\quad \text{and } Q_2(-\infty, -1) = Q_2(-1, 0) = Q_2(0, 1) = Q_2(1, +\infty) = .25\} \end{aligned}$$

and

$$\begin{aligned} \mathcal{Q}_{3C} &= \{Q \in \mathcal{P} : Q \text{ has the same } Q_1\text{-marginal as } \Pi_{0C} \\ &\quad \text{and } Q_2(-\infty, -1) = Q_2(-1, 0) = Q_2(0, 1) = Q_2(1, +\infty) = .25\} \end{aligned}$$

TABLE 6. *Posterior bounds when using $\Gamma(\mathcal{Q}_{3N}, \varepsilon)$ with a normal likelihood.*

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$\inf \Gamma(\mathcal{Q}_{3N}, \varepsilon)$	.805	.777	.747	.706	.659	.608	.532	.450	.328	.322
$\sup \Gamma(\mathcal{Q}_{3N}, \varepsilon)$	.847	.859	.873	.888	.902	.917	.932	.944	.945	.946

TABLE 7. *Posterior bounds when using $\Gamma(Q_{3C},\varepsilon)$ with a normal likelihood.*

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\, \Gamma(Q_{3C},\varepsilon)$	.774	.742	.705	.663	.612	.550	.472	.393	.249	.078
$sup\, \Gamma(Q_{3C},\varepsilon)$	.821	.842	.861	.881	.900	.919	.938	.956	.979	.982

TABLE 8. *Posterior bounds when using $\Gamma(Q_{3N},\varepsilon)$ with a Cauchy likelihood.*

	Values of ε									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$inf\, \Gamma(Q_{3N},\varepsilon)$	.816	.801	.772	.742	.705	.651	.591	.498	.330	.112
$sup\, \Gamma(Q_{3N},\varepsilon)$	.848	.863	.879	.894	.908	.925	.941	.957	.973	.979

Both in the CN and in the CC cases the posterior probability of B under the base prior is approximately equal to .801. From Table 7 and the 4th and 5th rows of Table 5 we see that the ranges in the CC case are larger than in the CN case for small to moderate values of ε. That means that, as long as the confidence on the base prior is strong, the results shows a behaviour similar to that described in O'Hagan (1988): the normal tails of the likelihood dominate the Cauchy prior and the ranges are smaller. However, for $\varepsilon = 1$ the results are reversed: here the role of the prior assumptions is weaker and the ranges are typically very large. Both in the NC and in the NN cases the posterior probability of B under the base prior is approximately .83; from Table 6 and Table 8 we see that the ranges in the NC case are shifted to the right with respect to those obtained in the NN case and they are smaller for $\varepsilon < .9$. Even in this case the relation is reversed when $\varepsilon = 1$.

We can summarize the results by saying that, given the same class of priors, comparisons of ranges depend on ε. For small to moderate values of ε the conflicts between prior and likelihood are resolved according to the thickness of the tails. When ε gets large, the sizes of the classes of priors become huge and the ranges are typically very large. On the other hand, it is worthwhile to note that, given the same class of priors, the "conjugate" analyses (CC and NN) provide larger ranges compared with the "non conjugate" analyses (CN and NC, respectively), for small to moderate values of ε. It is also interesting to compare the ranges, given the same likelihood, taking into account the different values of the base posterior probabilities of B, ($\approx .801$ with a Cauchy prior and $\approx .830$ with a Normal prior). From this point of view the use of Normal marginal priors sensibly reduces the ranges for any values of ε. We did calculations also for larger values of x_1: the behaviour was consistent with the other results. When x_1 gets larger and ε is not too large the results reflect O'Hagan conclusions even more closely.

6. Final remarks. The approximation of classes with given marginals, illustrated here, could be improved, actually it should be complemented, by monitoring the error. This aspect is particularly challenging in the general case, mainly when the parameter space is not compact. A guideline in this direction could be to adopt strong quantiles assessments on the tails of the marginal distribution of θ_1, in such a way that tightness could recover the non compactness of the parameter space.

Also, it is important to stress that the approach described here depends on the chosen partition of Θ_1, where the marginal distribution is actually known. This information allows, and the authors are working in this direction, to choose an appropriate initial partition, and to drive the refinement procedure taking into account the "degree of sensitivity" of the sets of the partition. Details of this approach are in Liseo, Moreno and Salinetti (1996).

Acknowledgements. The authors are grateful to two anonymous referees for their helpful comments and suggestions. We are particularly indebted to one of the referees for pointing out the importance of error and convergence analyses for the approximation procedure.

APPENDIX

We now show that, for any fixed d, the sets $\{S_1, S_2, \ldots, S_n\}$ form a partition of Θ_2. In fact, by construction, they are pairwise disjoint. Moreover, for any $\theta_2 \in \Theta_2$, there exists an index k such that $\theta_2 \in S_k$. For, let $\theta_2 \in \Theta_2$ and

$$k = \max\left\{i : \psi_i(\theta_2) - d_i = \min_{1\le j\le n} \{\psi_j(\theta_2) - d_j\}\right\}.$$

Then for $j < k$ we have that $\psi_j(\theta_2) - d_j \ge \psi_k(\theta_2) - d_k$, i.e. $\theta_2 \in S^c_{jk}$. For $j > k$ we have $\psi_k(\theta_2) - d_k < \psi_j(\theta_2) - d_j$, i.e. $\theta_2 \in S_{kj}$; it follows that $\theta_2 \in \left(\bigcap_{j=1}^{k-1} S^c_{jk}\right) \bigcap \left(\bigcap_{j=k+1}^{n} S_{kj}\right) = S_k$.

REFERENCES

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

KEMPERMAN J.H.B.(1987). Geometry of the moment problem. In *Proceedings of Symposia in Applied Mathematics* **37** 16-53.

LAVINE, M., WASSERMAN, L. AND WOLPERT, R.L.(1991). Bayesian inference with specified prior marginals. *J. Amer. Statist. Assoc.* **86** 964-971.

LISEO, B., MORENO, E. AND SALINETTI,G.(1996). Further on Bayesian robustness for classes of bidimensional priors with given marginals. *Manuscript under preparation.*

MORENO, E. AND CANO, J.A. (1991). Robust Bayesian analysis for ε-contaminations partially known. *J. Roy. Statist. Soc. Ser. B* **53** 143-155.

MORENO, E. AND CANO, J.A. (1995). Classes of bidimensional prior specified on a collection of sets: Bayesian robustness. *J. Statist. Plann. Inference* **46** 325-334.

O'HAGAN, A. (1988). Modelling with heavy tails. In *Bayesian Statistics 3* (J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith, eds.) 345-359. Oxford University Press.

RACHEV, S.T. (1985). The Monge-Kantorovich mass transference problem and its stochastic applications. *Theory of Probability and its Applications* **29** 647-671.

SALINETTI, G. (1994) Discussion on "An overview of Robust Bayesian Analysis" by J.O. Berger. *Test* **3** 5-125.

SIVAGANESAN, S. AND BERGER, J.O. (1989). Ranges of posterior measures for priors with unimodal contaminations. *Ann. Statist.* **17** 868-889.

WALLEY, P. (1991) *Statistical Reasoning with Imprecise Probabilities.* Chapman and Hall, London.

WASSERMAN, L., LAVINE, M. AND WOLPERT, R.L. (1993). Linearization of Bayesian robustness problems. *J. Statist. Plann. Inference* **37** 307-316.

B. LISEO and G. SALINETTI
DIP. DI STATIST., PROBAB. E STAT. APPL.
UNIVERSITA' DI ROMA
ROMA 00185
ITALY

E. MORENO
DEPT. DE ESTADISTICA E I.O.
UNIVERSIDAD DE GRANADA
GRANADA 18071
SPAIN

Bayesian Robustness for Classes of Bidimensional Priors with Given Marginals

discussion by
SANDRA FORTINI
Università Bocconi, Milano.

Relevance of MPT, in solving variational problems, connected with robust Bayesian inference, was not fully explored yet, although its first suggestion dates back to 1989; the present paper constitutes a step forward in this direction, in that it gives approximate solutions to minimization problems, over classes of bidimensional priors, based on a direct application of MPT.

My comments will focus on two points, namely the lack of accuracy in the approximation which can occur in particular cases and the choice of ε-contamination classes of priors in multivariate robustness analysis.

Lack of accuracy in particular cases. Suppose that the problem of finding the range of some posterior quantity of interest reduces to that of minimizing

$$\int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2),$$

over the class $\mathcal{Q}$ of the probability measures on Θ with specified marginals Q_1 and Q_2. The authors suggest approximating

$$\inf_{Q \in \mathcal{Q}} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2) \tag{1}$$

by

$$\inf_{Q \in \mathcal{Q}_1} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2),$$

where $\mathcal{Q}_1$ is the class of the probability measures on Θ with Q_1 as first marginal and some quantiles of the second marginal fixed according to Q_2. It may happen (e.g. when ψ is not bounded, as a function of θ_2 and Θ is not compact) that

$$\inf_{Q \in \mathcal{Q}_1} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2) = -\infty$$

while

$$\inf_{Q \in \mathcal{Q}} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2) > -\infty.$$

Although the use of numerical methods results in a finite bound for (1), the accuracy of the approximation might be strongly compromised, in this case.

The choice of ε-contamination classes of priors in multivariate rubustness analysis. In problems involving vector valued parameters it is usually difficult to assign the prior law, even when the opinion on the marginal distributions is quite precise. There can be practical situations where a specific prior distribution π_0 is suggested by the knowledge of the connections between θ_1 and θ_2, deriving from the empirical meaning of the parameters in the concrete context. This typically occurs when, for some reasons, θ_1 and θ_2 are thought to be independent. More often the idea on the way that the different hypotheses on θ_1 influence one's prior opinion on the law of θ_2 is too vague to be translated into a precise mathematical form, so that it would be arbitrary to fix one prior probability law and it is not clear how this could be actually done. In this second case ε-contamination classes are not appropriate and a different choice should be made. In particular, if the class can be specified through generalized moment conditions, then the approximation method based on MPT can still be applied. As an example, suppose that some information on the concordance of θ_1 and θ_2 is available, in addition to the marginal laws Q_1 and Q_2 of θ_1 and θ_2. Typically, such information can be expressed through the inequality

$$c_1 \leq C(Q) \leq c_2,$$

where c_1 and c_2 are real numbers and $C(Q) = \int k(\theta_1, \theta_2) Q(d\theta_1 d\theta_2)$ is a concordance index, k being a quasi-monotone function (cfr. Cambanis *et al.* (1976), Tchen (1980)). In this situation

$$\mathcal{Q}_C = \{Q \in \mathcal{Q} : c_1 \leq C(Q) \leq c_2\}$$

is a sensible choice for the class of prior distributions and, for any function ψ on Θ,

$$\inf_{Q \in \mathcal{Q}_C} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2)$$

can be approximated by

$$\inf_{\{Q \in \mathcal{Q}_1 : c_1 \leq C(Q) \leq c_2\}} \int \psi(\theta_1, \theta_2) Q(d\theta_1 d\theta_2)$$

which, because of the linearity of C, can be obtained by a direct application of MPT.

References

Cambanis, S., Simons, G., Stout, W. (1976). Inequalities for $E(K(X, Y))$ when the marginals are fixed. *Z. Wahrscheinlichkeitstheorie und Vern. Gebiete*, **36**, 285-294.

Tchen, A. H. (1980). Inequalities for distributions with given marginals. *Annals of Probab.* **8**, 814-827.

REJOINDER

Brunero Liseo, Elias Moreno and Gabriella Salinetti

Let us thank Sandra Fortini for her interesting discussion of the paper. She emphasized two points. The first one is related to the accuracy of the approximation we used and the second to the choice of the ϵ-contamination class of priors.

The approximation of the class $\mathcal{Q}$ with the class $\mathcal{Q}_1$ can certainly result in an enlargement of the range of posterior quantities of interest up to the extreme case where the inf over $\mathcal{Q}$ is finite and the inf over $\mathcal{Q}_1$ is infinite, compromising the accuracy of the approximation. This extreme fact could actually be taken into account, for example, with an appropriate choice of the quantiles on the tails of Θ_1. However it seems to us that the major fact to stress is that the class $\mathcal{Q}$ itself will typically give a vacuous posterior range even when, as emphasized in our paper and in Walley (1985), the parameter space is compact and the quantity of interest is bounded.

Different conclusions are obtained when $\mathcal{Q}$ is a contamination class: here, with reasonable choices of ϵ, robustness is achieved. It is true, however, that the possible lack of robustness of $\Gamma(\mathcal{Q}_1, \epsilon)$ will require more information on $\mathcal{Q}_1$ compatible with $\mathcal{Q}$: not only quantiles but possibly shape constraints, which could be suggested by the marginal Q_1. This latter analysis is being considered in Liseo, Moreno and Salinetti (1995).

In this direction, the information on the concordance structure, as suggested by Fortini in the second part of her discussion is indeed interesting; certainly it might be accessible from the experts; technically, it simply requires to add one more linear constraint over the class $\mathcal{Q}_1$ and, as the discussant noticed, MPT still applies.

However it seems that this is not enough to alleviate the lack of robustness of $\mathcal{Q}$. Bayesian robustness reduces the elicitation effort indeed, but if very weak information is processed (as $\mathcal{Q}$ does, even with the concordance constraint), very weak (or vacuous) posterior answers will be obtained.

Thus, when confidence on the one dimensional marginals is assumed, it is still necessary to elicit a base prior π_0 which matches the given marginals and to allow a certain degree of uncertainty on any other features of the priors, as the ϵ-contamination classes do. It is clear that if a given concordance or covariance structure is deemed elicitable, then π_0 should be chosen according this new input.

REFERENCES

Liseo, B., Moreno, E., Salinetti, G. (1996). On Frechét classes with shape constraints. (*Manuscript under preparation*).