

Part I

INVITED PAPERS

NUMERICAL ROBUST BAYESIAN ANALYSIS UNDER GENERALIZED MOMENT CONDITIONS

BY BRUNO BETRÒ AND ALESSANDRA GUGLIELMI

CNR-IAMI and Università di Milano

In Betrò *et al.* (1994) the optimization of posterior functionals $E^\pi(g | x)$ with respect to prior measures π has been considered for the class of priors defined by a number of generalized moment conditions $\int H_i d\pi \leq \alpha_i$. Constraints of this type are very general as they include for instance, besides ordinary moment conditions, bounds on prior quantiles or bounds on marginal probabilities of data. This paper presents an algorithm for the numerical solution of the above optimization problem based on ideas suggested by the interval approach to numerical optimization as well as from semi-infinite linear programming.

1. Introduction. The interest for robust Bayesian analysis is steadily increasing (see Berger, 1994 for a recent survey). In particular the attention is focussed on robustness with respect to the prior and various examples of classes of priors which are convenient for the analysis have been studied.

In Betrò *et al.* (1994) the class of priors defined through the so called generalized moment conditions have been analyzed. It has been shown in particular that the infinite dimensional problem of finding the extrema of posterior functionals under such a prior class reduces, under rather general conditions, to a finite dimensional one. This latter, however, is typically a global optimization problem, and therefore its solution is in any case a difficult task from an algorithmic point of view. It has to be noted that the solution of a global optimization problem is usually searched for repeatedly starting a local search routine from different points within the feasible region; this gives, in general, only an approximation from below to the global maximum and from above to the global minimum, and therefore it might erroneously indicate that robustness holds.

An interesting approach that is able to provide intervals which are guaranteed to contain the global extrema in global optimization problems has been proposed in the literature (see e.g. Hansen, 1992; Ratscheck and Rokne, 1988; for a survey, see Ratscheck and Rokne, 1995). The approach exploits the ideas of Interval Analysis for building up functions mapping intervals I in \mathbb{R}^n into intervals in \mathbb{R} containing the range over I of the function to be optimized. On the other hand, the connection recently pointed out (see Dall'Aglio and Salinetti, 1994) between robust Bayesian analysis with generalized moment conditions and semi-infinite linear programming has disclosed

AMS 1991 subject classification. Primary 62F15; secondary 62F35.

Key words. Bayesian robustness, generalized moment conditions, semi-infinite linear programming, interval analysis.

the possibility of prompting solution approaches developed in this latter context.

Merging Interval Analysis and semi-infinite linear programming ideas, we shall construct here, after a suitable transformation of the original problem into a new one in which only linear functionals on a measure space are concerned, a general algorithm based on finite dimensional linear programming problems whose solutions bracket the supremum of the posterior function of interest. We remark that the linearization here proposed requires a single step, unlike the one in Lavine (1991). Obviously in a quite analogous way the infimum of posterior functionals can be dealt with.

We start in Section 2 recalling some features of the generalized moment classes of priors, stressing the fact that such classes extend far beyond ordinary moment classes and are also able to accomplish restrictions to the prior as unimodality or ϵ -contamination. In Section 3 we exhibit the linearization of the problem, and how to build up the linear programming problems whose optimal solutions bound from above and from below the required extremum, while in Section 4 we show how to refine the bounds. After the definitions of basic notions in Interval Analysis, in Section 6 we sketch a proof of convergence of the bounds to the extremum considered, when the parameter is real, under some rather general conditions. In Section 7 the implementation of the bounding procedure in the Maple programming language, exploiting its symbolic manipulation facilities, is discussed. Finally, in Section 8, two numerical examples are given.

2. The problem. Let X be a random variable on a dominated statistical space $(\mathcal{X}, \mathcal{F}_X, \{P_\theta, \theta \in (\Theta, \mathcal{F})\})$, with density $f(x|\theta)$ with respect to a dominant measure λ , where the parameter space (Θ, \mathcal{F}) is such that Θ is a subset of a finite-dimensional Euclidean space; denote by $l_x(\theta) = f(x|\theta)$ the likelihood function, which we assume $\mathcal{F}_X \otimes \mathcal{F}$ -measurable.

Let Q and M be respectively the space of all probability measures and the space of all finite measures on (Θ, \mathcal{F}) and define

$$(1) \quad \Gamma = \left\{ \pi \in Q : \int_{\Theta} H_i(\theta) \pi(d\theta) \leq \alpha_i, i = 1, \dots, m \right\},$$

where H_i are given π -integrable functions, and α_i are fixed real constants, $i = 1, \dots, m$. Suppose that Γ is nonempty.

Let

$$(2) \quad F(\pi) = \frac{\int_{\Theta} g(\theta) l_x(\theta) \pi(d\theta)}{\int_{\Theta} l_x(\theta) \pi(d\theta)}$$

be a functional of the posterior density, where $g : \Theta \rightarrow \mathbb{R}$ is a given function such that F exists.

Conditions of the type (1) will be referred to as generalized moment conditions. The simplest case is obviously $H_i(\theta) = \theta^i$, but we might consider

$$H_i(\theta) = \int_{K_i} l_x(\theta) dx,$$

where K_i is a measurable subset of \mathcal{X} , so that

$$\int_{\Theta} H_i(\theta) \pi(d\theta) = \int_{K_i} m_\pi(x) dx,$$

where $m_\pi(x) = \int_{\Theta} l_x(\theta) \pi(d\theta)$, and consequently deal with bounds on the marginal probabilities; bounds on conditional marginal probabilities

$$P(x_1 \in A_i | x_2) \leq \alpha_i$$

can be considered as well as they can be written in the form:

$$\int_{A_i} m_\pi(x_1, x_2) dx_1 - \alpha_i \int_{\mathcal{X}} m_\pi(x_1, x_2) dx_1 \leq 0,$$

where $m_\pi(x_1, x_2) = \int_{\Theta} l_{(x_1, x_2)}(\theta) \pi(d\theta)$. On the other hand, restrictions of Q in (1) to a class of ϵ -contaminated priors or to a class of mixture priors can be accomplished within a suitable generalized moment class. Indeed, in the first case, if $\pi = (1 - \epsilon)\pi_0 + \epsilon q$, $q \in Q$, then for a generic function f it holds:

$$(3) \quad \int f d\pi = \int \{ \int [(1 - \epsilon)f d\pi_0] + \epsilon f \} dq$$

so that we still deal with constraints of the form (1) and with a functional of the form (2). In the second case, we have for any function f

$$\int f(\theta) \pi(d\theta) = \int \int f(\theta) \pi(d\theta | \alpha) d\nu(\alpha)$$

and therefore generalized moment conditions are still preserved. We remark that, according to Khintchine's Theorem (Feller, 1971) the class of unimodal priors π in \mathbb{R} is a particular mixture class.

In any case, our aim is to find out algorithms to calculate

$$(4) \quad \inf_{\pi \in \Gamma} F(\pi)$$

$$(5) \quad \sup_{\pi \in \Gamma} F(\pi).$$

3. Approximation by finite dimensional problems. We consider here the following transformation from the linear ratio form (2) into a linear one without altering the linearity of the constraints. We refer to problem (5),

but obviously problem (4) can be addressed as well, considering $-g$ instead of g .

Consider $\nu \in M$ given by

$$(6) \quad \nu(A) = \frac{\int_A \pi(d\theta)}{\int_{\Theta} l_x(\theta) \pi(d\theta)}, \quad A \in \mathcal{F},$$

assuming that $\int_{\Theta} l_x(\theta) \pi(d\theta) > 0$; (6) yields

$$(7) \quad \int_{\Theta} l_x(\theta) \nu(d\theta) = 1.$$

Thus, by (6) and (7), it is easy to show that π can be written as

$$(8) \quad \pi(A) = \frac{\int_A \nu(d\theta)}{\int_{\Theta} \nu(d\theta)}, \quad A \in \mathcal{F}$$

and the problem turns into

$$(9) \quad G^* = \sup_{\nu \in \Gamma'} \int_{\Theta} g(\theta) l_x(\theta) \nu(d\theta) = \sup_{\nu \in \Gamma'} F'(\nu),$$

where

$$(10) \quad \Gamma' = \left\{ \nu \in M : \int_{\Theta} f_i(\theta) \nu(d\theta) \leq 0, i = 1, \dots, m, \int_{\Theta} l_x(\theta) \nu(d\theta) = 1 \right\}$$

and $f_i(\theta) = H_i(\theta) - \alpha_i$, $i = 1, \dots, m$. We observe that linearization can be also achieved through the posterior measure $\mu(A) = \int_A l_x(\theta) \pi(d\theta) / \int_{\Theta} l_x(\theta) \pi(d\theta)$, $A \in \mathcal{F}$ (see Betrò and Guglielmi, 1994); however this introduces division by $l_x(\theta)$ which typically yields numerical instabilities. For a review of linearization techniques in robust Bayesian analysis see Perone Pacifico *et al.* (1994).

The basic idea in the algorithm we propose is to formulate two finite dimensional optimization problems whose solutions are an upper and a lower bound, respectively, of the optimum value (9) under the constraints (10). Let us start from the upper bound.

Consider a partition Π_k of Θ into a finite number of subsets, say $\Theta = \Theta_1^k \cup \dots \cup \Theta_k^k$. If $\phi : \Theta \rightarrow \mathbb{R}$, then $\overline{\phi}_j^k$ and $\underline{\phi}_j^k$ will denote an upper and a lower bound, respectively, of the function ϕ over the subset Θ_j^k . We assume that the bounds are such that if k, k' are two integers, $k' > k$, and, for some l and j , $\Theta_l^{k'} \subset \Theta_j^k$, then

$$(11) \quad \begin{aligned} \overline{\phi}_l^{k'} &\leq \overline{\phi}_j^k \\ \underline{\phi}_l^{k'} &\geq \underline{\phi}_j^k. \end{aligned}$$

If ϕ is indexed by some index, say i , then we write $\bar{\phi}_{i,j}^k$ and $\underline{\phi}_{i,j}^k$. We set $\nu_j^k = \nu(\Theta_j^k)$. When not required for clearness, the upper index relative to the partition Π_k will be omitted.

If $\nu \in \Gamma'$, then $0 \geq \int_{\Theta} f_i(\theta)\nu(d\theta) \geq \sum_{j=1}^k \underline{f}_{i,j}\nu_j$, $i = 1, \dots, m$ and $1 \geq \int_{\Theta} l_x(\theta)\nu(d\theta) \geq \sum_{j=1}^k \underline{l}_{x,j}\nu_j$, $1 \leq \int_{\Theta} l_x(\theta)\nu(d\theta) \leq \sum_{j=1}^k \bar{l}_{x,j}\nu_j$. Then

$$\Gamma' \subseteq \left\{ \nu \in M : \nu_j \geq 0, j = 1, \dots, k, \sum_{j=1}^k \underline{l}_{x,j}\nu_j \leq 1, \sum_{j=1}^k \bar{l}_{x,j}\nu_j \geq 1, \right. \\ \left. \sum_{j=1}^k \underline{f}_{i,j}\nu_j \leq 0, i = 1, \dots, m \right\} \stackrel{\text{def}}{=} \tilde{\Gamma}_k$$

and $\sup_{\nu \in \Gamma'} F'(\nu) \leq \sup_{\nu \in \tilde{\Gamma}_k} F'(\nu)$. Analogously

$$\sup_{\nu \in \tilde{\Gamma}_k} \int_{\Theta} g(\theta)l_x(\theta)\nu(d\theta) \leq \sup_{\nu \in \tilde{\Gamma}_k} \sum_{j=1}^k \overline{(g(\theta)l_x(\theta))}_j \nu_j$$

For simplicity, to avoid handling coefficients with infinite values, we will assume that all the functions considered in the problem are bounded. Now we observe that $\sup_{\nu \in \tilde{\Gamma}_k} \sum_{j=1}^k \overline{(g(\theta)l_x(\theta))}_j \nu_j$ is not necessarily smaller than $\max_{j=1, \dots, k} \bar{g}_j$. But, if g^* is an upper bound for g , then it is possible to find non positive upper bounds for $g(\theta) - g^*$ on Θ_j , so that

$$\int_{\Theta} g(\theta)l_x(\theta)\nu(d\theta) = \int_{\Theta} g^*l_x(\theta)\nu(d\theta) + \int_{\Theta} (g(\theta) - g^*)l_x(\theta)\nu(d\theta) \\ \leq g^* + \sum_{j=1}^k \overline{[g(\theta) - g^*]}_j \underline{l}_{x,j}\nu_j \leq g^*.$$

Thus, the supremum (9) is not larger than

$$(12) \quad \sup_{\nu \in \tilde{\Gamma}_k} \left(g^* + \sum_{j=1}^k \overline{[g(\theta) - g^*]}_j \underline{l}_{x,j}\nu_j \right),$$

and $\tilde{\Gamma}_k$ can be easily written as

$$(13) \quad \tilde{\Gamma}_k = \{(\nu_1, \dots, \nu_k) \in \mathbb{R}^k : \nu_j \geq 0, j = 1, \dots, k, \sum_{j=1}^k \bar{l}_{x,j}\nu_j = 1, \\ \sum_{j=1}^k \underline{f}_{i,j}\nu_j \leq 0, i = 1, \dots, m\}.$$

To conclude, the solution of (12)-(13) provides an upper bound to the optimal value of G^* in (9).

Coming to the construction of the lower bound of (9), it is natural to consider the following finite dimensional problem

$$(14) \quad \sup_{\hat{\Gamma}_l} \sum_{j=1}^l g(\theta_j) l_x(\theta_j) \eta_j$$

$$(15) \quad \hat{\Gamma}_l = \{(\eta_1, \dots, \eta_l) \in \mathbb{R}^l : \eta_j \geq 0, j = 1, \dots, l, \sum_{j=1}^l l_x(\theta_j) \eta_j = 1, \sum_{j=1}^l f_i(\theta_j) \eta_j \leq 0, i = 1, \dots, m\},$$

where $\theta_1, \dots, \theta_l$ are given points in Θ .

Indeed, (12)-(13) and (14)-(15) are ordinary finite dimensional problems for which efficient solution techniques exist.

We assume that $\hat{\Gamma}_l$ in (15) is not empty for $l \geq l_0$, which implies that the feasible regions Γ' in (10) and $\tilde{\Gamma}_k$, for any integer k , in (13) are not empty too.

4. Refining the bounds. In the previous section we showed how to build an “upper” and a “lower” problem, whose corresponding optimal values bracket the optimal value (9). We want now to construct two sequences of such finite dimensional problems so that the corresponding sequences of upper and lower bounds of G^* will converge, under suitable conditions, to the optimal value G^* .

Here we assume for simplicity that Θ is an interval in \mathbb{R} ; extensions to multidimensional intervals can be easily worked out. A refinement procedure is here proposed whose general stage is as follows. Let Π_{k_n} be the current partition of Θ into k_n intervals as in Section 3. As (13) defines $m+1$ constraints, a suitable value for k_1 is $m+1$. Then the linear programming problem (12)-(13) is solved obtaining an optimum value $G^{(k_n)} = (g^* + \sum_{j=1}^{k_n} [g(\theta) - g^*]_j l_{x_j} \tilde{\nu}_j)$, $G^{(k_n)} \geq G^*$. Let us consider now the at most $m+1$ intervals corresponding to positive $\tilde{\nu}_j$. They are candidates for containing the atoms of a solution $\tilde{\nu}$ of the original problem (9)-(10). We remind that if an optimal solution to such a problem exists, then the measure $\tilde{\nu}$ achieving the supremum is a discrete one with finite support (see Betrò *et al.*, 1994). Such intervals are then split into two parts, obtaining a new partition with a finite number of intervals greater than k_n . Let $\{\theta_1, \dots, \theta_{l_n}\}$ ($l_n > k_n$) be the points in Θ defining this new partition. We solve the corresponding problem (14)-(15), obtaining $G_{(l_n)} = \sum_{j=1}^{l_n} g(\theta_j) l_x(\theta_j) \hat{\eta}_j$, $G_{(l_n)} \leq G^* \leq G^{(k_n)}$. Then, according to the strategy described below, we choose new points in Θ to be added to $\{\theta_1, \dots, \theta_{l_n}\}$, getting a new partition of Θ , $\Pi_{k_{n+1}}$, $k_{n+1} > l_n$. This completes the current stage and we can proceed to the next one, computing a new upper bound $G^{(k_{n+1})}$ to the optimal solution.

The choice of the points in Θ to be added to $\{\theta_1, \dots, \theta_{l_n}\}$ after $G_{(l_n)}$ has been computed is accomplished exploiting the semi-infinite linear programming approach. Indeed, the dual formulation of (9)-(10) (see Kemperman, 1987) is given by

$$(16) \quad \inf_{(y_0, \dots, y_m)} y_0$$

$$(17) \quad v(y; \theta) = y_0 l_x(\theta) + \sum_{i=1}^m y_i f_i(\theta) - g(\theta) l_x(\theta) \geq 0, \quad \forall \theta \in \Theta,$$

$$y_1 \geq 0, \dots, y_m \geq 0.$$

(16)-(17) is referred to as a semi-infinite linear programming problem and for its solution a vast literature is available (for a recent survey, see Hettich and Kortanek, 1993). The connection between Bayesian robust analysis under generalized moments and semi-infinite linear programming has been first pointed out in Dall'Aglio and Salinetti (1994).

A widely studied class of methods in this contest is the class of *exchange methods* based on the idea of considering a finite number of constraints $v(y; \theta_j) \geq 0, j = 1, \dots, l$ and solving the resulting ordinary linear programming problem, obtaining a solution \hat{y} . If $v(\hat{y}; \theta) \geq 0$, for each $\theta \in \Theta$, then \hat{y} solves (16)-(17), otherwise the list of discretization points $\theta_1, \dots, \theta_l$ is updated by a suitable point θ^* which achieves a “small enough” (negative) value of the function $v(\hat{y}; \theta)$. A criterion for the choice of θ^* , making precise the above concept of smallness, is given in Hu (1990): for a fixed $0 < \alpha < 1$, θ^* is any point such that

$$(18) \quad \begin{array}{ll} \text{either} & v(\hat{y}; \theta^*) \leq -\alpha \\ \text{or} & v(\hat{y}; \theta^*) \leq \alpha \inf_{\Theta} v(\hat{y}; \theta). \end{array}$$

This criterion ensures convergence of the resulting exchange algorithm under wide conditions (see Theorem 1 in Hu's paper).

Coming back to our situation, we consider the dual solution \hat{y} of (14)-(15). If Θ_j denotes the general interval in the Θ -partition defined by $\{\theta_1, \dots, \theta_{l_n}\}$ and \underline{v}_j denotes the lower bound of $v(\hat{y}; \theta)$ over Θ_j , then \underline{v}_j is easily computed as we assume the availability of the bounds on g, l_x, f_i 's. Then, new points are added to the previous l_n ones, iteratively considering, for $r = 1, 2, \dots$, one or more points of the interval, say Θ_r^* , corresponding to $\underline{v}_r^* = \min_j \underline{v}_j$ and consequently updating the grid of points in Θ , the corresponding partition, the bounds \underline{v}_j 's and r . Let θ_r^* denote the point, among the ones considered in Θ_r^* , with the smallest values of $v(\hat{y}; \cdot)$. Iteration is stopped when

$$(19) \quad \begin{array}{ll} \text{either} & v(\hat{y}; \theta_r^*) \leq -\alpha \\ \text{or} & v(\hat{y}; \theta_r^*) \leq \alpha \underline{v}_r^* \end{array}$$

is met. In this way, the new partition $\Pi_{k_{n+1}}$ is determined by $\{\theta_1, \dots, \theta_{l_n}\}$ plus the points considered in the above iteration.

Summing up, the algorithm we propose is an iterative one, whose general stage provides the computation of an upper and a lower bound. We switch from the current stage to the next one updating adaptively the grid in Θ that enables to build up alternatively the two finite dimensional problems (12)-(13) and (14)-(15). In this way, two sequences $(G_{(l_n)})_n$ and $(G^{(k_n)})_n$ are obtained such that $G_{(l_n)} \leq G^* \leq G^{(k_n)}$ and, by (11) it is easily seen that $(G_{(l_n)})_n$ is non-decreasing while $(G^{(k_n)})_n$ is non-increasing. The convergence of the sequences will be discussed in Section 6. We need before to find a proper way of computing the sequences $(\bar{g}_j^{k_n})_{k_n}$, $(\underline{f}_{j,j}^{k_n})_{k_n}$, $(\bar{l}_{x_j}^{k_n})_{k_n}$, $(\underline{l}_{x_j}^{k_n})_{k_n}$. This can be achieved by techniques derived from Interval Analysis.

5. Interval Analysis. Interval analysis is a generalization of real analysis in which interval numbers replace real numbers and interval arithmetics replace real arithmetics. See Hansen (1992) and Ratschek and Rokne (1988, 1995) to have an idea of Interval Analysis applied to global optimization problems.

An interval number X is a real interval $[a, b]$. An interval function Φ is an interval-valued function, defined on a class of intervals; it is called inclusion function for a real-valued function ϕ on \mathbb{R}^n if

$$\{\phi(x) : x \in X\} \subset \Phi(X).$$

Then it is possible to calculate an upper bound of the supremum of ϕ , and a lower bound of the infimum of ϕ over an interval X in the following ways:

a) if ϕ can be written, using arithmetic operations, in terms of elementary functions whose range on X is known, the range of ϕ is obtained, via interval arithmetics, from the ranges of the elementary functions;

b) if ϕ is differentiable and $D\phi$ denotes its gradient, as well writable in terms of elementary functions of known range on X , then the following inclusion formula can be used to obtain the range of ϕ on X via interval arithmetics

$$(20) \quad \phi(x) \in \phi(\text{mid}(X)) + (X - \text{mid}(X))D\Phi(X)$$

where $\text{mid}(X)$ is the midpoint of the interval X , to be defined in an appropriate way when X is unbounded, and $D\Phi$ is an inclusion function for $D\phi$.

The outputs of both a) and b) are intervals, whose left endpoints are lower bounds for $\underline{\phi}$ and whose right endpoints are upper bounds for $\bar{\phi}$. Comparing the two procedures, b) is expected to be better than a) when the length of X is small. It is easily seen that a) and b) give inclusion functions for which (11) holds. Besides, we assume that the following condition holds

$$(21) \quad w(\Phi(X)) - w(\square\phi(X)) \rightarrow 0, \text{ as } w(X) \rightarrow 0,$$

where Φ is an inclusion function for ϕ obtained by a) or b), w denotes the length of an interval and $\square\phi(X)$ the range of ϕ over X (see Ratschek and Rokne, 1988, p.87). As stated in Ratschek and Rokne (1995), p. 783, assumption (21) is not very restrictive and it is satisfied in a wide range of situations. We remark that it does not imply the continuity of ϕ .

6. Convergence of the bounds. In this section we sketch a proof of convergence of the sequences $G^{(k_n)}$ and $G_{(l_n)}$ under the following conditions:

C1: Θ is a compact interval in \mathbb{R} .

C2: g is upper semicontinuous, f_i , $i = 1, \dots, n$, are lower semicontinuous, l_x is continuous (this implies that $v(y; \cdot)$ is lower semicontinuous for every $y \in \mathbb{R}^{m+1}$) and they have only a finite number of points of discontinuity of the first kind.

C3: for every partition Π_{k_n} , there exists a couple (ε, δ) of positive numbers such that an index $s = s(k_n) \in \{1, \dots, m\}$ can be found with $\underline{f}_{s,j} > \delta$ for all j for which $\bar{l}_{xj} < \varepsilon$.

C4: interval splitting in the “upper” substage occurs at the midpoint of the interval. In the “lower” substage the points considered in Θ_r^* are its midpoint and the possible points of discontinuity of $v(\hat{y}; \cdot)$ in Θ_r^* . We recall that θ_r^* in (19) is such a point with the smallest value of $v(\hat{y}; \cdot)$.

Let us consider, first, the convergence of $G_{(l_n)}$. If a point θ_r^* satisfies (19), then this point satisfies (18) too, so that convergence of $G_{(l_n)}$ is ensured by the same conditions giving the convergence of Hu’s algorithm, keeping in mind that such convergence implies the absence of duality gap. We remark that Hu’s results are easily seen to hold under the additional requirement $y_1 \geq 0, \dots, y_m \geq 0$. Let us suppose that, at the n -th stage, (19) is violated at each iteration (so that an infinite number of points is added). By Lemma 1 in Ratschek and Rokne (1988), p.85, it follows that $w(\Theta_r^*) \rightarrow 0$ as $r \rightarrow +\infty$. By the compactness of Θ and the splitting strategy we adopt, it is possible to choose a subsequence of Θ_r^* , that we still denote by Θ_r^* for simplicity, such that $\Theta_r^* \rightarrow c^*$ as $r \rightarrow +\infty$, and $c^* \in \Theta_r^*$ for an infinite number of indexes r . By $\Theta_r^* \rightarrow c^*$ we mean that for each sequence θ_r , $r = 1, 2, \dots$, $\theta_r \in \Theta_r^*$, it holds $\theta_r \rightarrow c^*$ as $r \rightarrow +\infty$. By Theorem 3 in Ratschek and Rokne (1988), p. 87, thanks to condition (21), we have that $\lim_{r \rightarrow +\infty} \underline{v}_r^* = \inf_{\Theta} v(\hat{y}; \theta)$. Besides, by the lower semicontinuity of v , we can see that $\lim_{r \rightarrow +\infty} \underline{v}_r^* = v(\hat{y}; c^*)$, so that $\lim_{r \rightarrow +\infty} \underline{v}_r^* = \inf_{\Theta} v(\hat{y}; \theta) = v(\hat{y}; c^*)$. We show that $\liminf_{r \rightarrow +\infty} v(\hat{y}; \theta_r^*) = v(\hat{y}; c^*)$. Indeed, this is trivial if c^* is a continuity point for $v(\hat{y}; \cdot)$. On the other hand, if c^* is a discontinuity point for $v(\hat{y}; \cdot)$, since the condition $c^* \in \Theta_r^*$ implies $v(\hat{y}; \theta_r^*) =$

$v(\hat{y}; c^*)$, it follows that $c^* \in \Theta_r^*$ for an infinite number of indexes r implies $\liminf_{r \rightarrow +\infty} v(\hat{y}; \theta_r^*) = v(\hat{y}; c^*) = \inf_{\Theta} v(\hat{y}; \theta)$. If $v(\hat{y}; \theta_r^*) > \alpha \max(-1, \underline{v}_r^*)$ for every r , then $\inf_{\Theta} v(\hat{y}; \theta) = \liminf_r v(\hat{y}; \theta_r^*) \geq \alpha \max(-1, \lim_r \underline{v}_r^*) = \alpha \max(-1, \inf_{\Theta} v(\hat{y}; \theta))$. Then, $\inf_{\Theta} v(\hat{y}; \theta) < 0$ (if $\inf_{\Theta} v(\hat{y}; \theta) \geq 0$, \hat{y} is the dual solution of (9)-(10) implies $\inf_{\Theta} v(\hat{y}; \theta) > \max(-1, \inf_{\Theta} v(\hat{y}; \theta))$, that is absurd).

Coming to the convergence of $G^{(k_n)}$, let us first show that, under condition C3, the discrete measure $\tilde{\nu}$, solution of the finite dimensional problem (12)-(13) is finite and its whole mass is bounded, for every k_n . Indeed, $\sum_{j=1}^{k_n} \bar{l}_{x_j} \nu_j = 1$ implies $\sum_{j: \bar{l}_{x_j} \geq \varepsilon} \nu_j \leq \frac{1}{\varepsilon}$. Besides, $0 \geq \sum \underline{f}_{s,j} \nu_j \geq \delta \sum_{j: \bar{l}_{x_j} < \varepsilon} \nu_j + \frac{1}{\varepsilon} \min_j \underline{f}_{s,j}$, so that $\sum_{j: \bar{l}_{x_j} < \varepsilon} \nu_j \leq -\frac{1}{\delta \varepsilon} \min_j \underline{f}_{s,j} \leq -\frac{1}{\delta \varepsilon} \min_{i,j} \underline{f}_{i,j} < C$, C independent on k_n (of course, $\min_{i,j} \underline{f}_{s,j}$ must be negative, if the vector (ν_j) belongs to the feasible region, that we suppose not empty). Thus, $\sum_j \tilde{\nu}_j$ is bounded by a positive constant independent on n .

Without loss of generality, we can suppose $g^* = 0$, so that $G^{(k_n)} = \sum_{j=1}^{k_n} \bar{g}_j \underline{l}_{x_j} \tilde{\nu}_j^{k_n}$. For every k_n , denote by $\tilde{\nu}^{k_n} = (\tilde{\nu}_{i_1}^{k_n}, \dots, \tilde{\nu}_{i_{m+1}}^{k_n}) \in \mathbb{R}^{m+1}$ a vector containing the at most $m+1$ positive $\tilde{\nu}_j^{k_n}$ and the remaining components set to 0. Denote by $c_j^{k_n}$ and $d_j^{k_n}$ the midpoints and the lengths of the interval $\Theta_j^{k_n}$ corresponding to $\tilde{\nu}_j^{k_n}$ respectively (c^{k_n} and d^{k_n} are the corresponding vectors) and, if ϕ is one of the functions considered, by $\phi_j^{k_n}, \bar{\phi}_j^{k_n}$ the bounds of the function ϕ on the same interval. By condition C3, the compactness of Θ and the boundedness of the functions considered, there exists a subsequence of $\{k_n\}$, that for simplicity we still denote by $\{k_n\}$, such that $\bar{g}_j^{k_n}, \underline{l}_{x_j}^{k_n}, \bar{l}_{x_j}^{k_n}, \underline{f}_{1j}^{k_n}, \dots, \underline{f}_{nj}^{k_n}, \tilde{\nu}_j^{k_n}, c_j^{k_n}, d_j^{k_n}$ converge as $n \rightarrow +\infty$, $j = i_1, \dots, i_{m+1}$. Let ν^*, c^*, d^* be the corresponding limit for $\tilde{\nu}^{k_n}, c^{k_n}, d^{k_n}$ and denote by $(\nu_1^*, \dots, \nu_{m'}^*)$, $m' \leq m+1$ the positive components of ν^* .

We want to show that the measure $\sum_{j=1}^{m'} \nu_j^* \delta_{c_j^*}$ is a feasible measure for problem (9) – (10). First, $d_j^* = 0$, $j = 1, \dots, m'$. Indeed, should an index $\hat{j} \in \{1, \dots, m'\}$ exist such that $d_{\hat{j}}^* > 0$, this would mean that the interval centered at $c_{\hat{j}}^*$ with length $d_{\hat{j}}^*$ was not definitively split. This happens only if $\nu_{\hat{j}}^* = 0$, against the definition of $\nu_{\hat{j}}^*$. Thus, $\Theta_j^{k_n} \rightarrow c_j^*$, $j = 1, \dots, m'$. From condition (21), we have $\lim_{n \rightarrow +\infty} \bar{g}_j^{k_n} - \sup_{\Theta_j^{k_n}} g(\theta) = \lim_{n \rightarrow +\infty} \bar{g}_j^{k_n} - g(\hat{\theta}_j^{k_n}) = 0$, $\hat{\theta}_j^{k_n} \in \Theta_j^{k_n}$, $j = 1, \dots, m'$. By the upper semicontinuity of g , $\limsup_{n \rightarrow +\infty} g(\hat{\theta}_j^{k_n}) \leq g(c_j^*)$, $j = 1, \dots, m'$. Thus, $\lim_{n \rightarrow +\infty} G^{(k_n)} = \lim_{n \rightarrow +\infty} \sum_{j=1}^{m'} \bar{g}_j^{k_n} \underline{l}_{x_j} \tilde{\nu}_j^{k_n} \leq \sum_{j=1}^{m'} g(c_j^*) \underline{l}_x(c_j^*) \nu_j^*$, and analogously $0 \geq \lim_{n \rightarrow +\infty} \sum_{j=1}^{m'} \underline{f}_{i,j}^{k_n} \nu_j^{k_n} \geq \sum_{j=1}^{m'} f_i(c_j^*) \nu_j^*$, $i = 1, \dots, m$. Moreover, by continuity of \underline{l}_x , $1 = \lim_{n \rightarrow +\infty} \sum_{j=1}^{m'} \bar{l}_{x_j}^{k_n} \tilde{\nu}_j^{k_n} = \sum_{j=1}^{m'} \underline{l}_x(c_j^*) \nu_j^*$. This

means that the measure $\sum_{j=1}^{m'} \nu_j^* \delta_{c_j^*}$ belongs to Γ' , so that $\lim_{n \rightarrow +\infty} G^{(k_n)} \leq G^*$. As $G^* \leq G^{(k_n)}$ for each n , it follows $G^* = \lim_{n \rightarrow +\infty} G^{(k_n)}$.

We remark that $\sum_{j=1}^{m'} \nu_j^* \delta_{c_j^*}$ turns out to be an optimal solution of problem (9)-(10).

7. Implementation of the bounding procedure. The procedure outlined in the previous sections was implemented in the Maple V programming language (see Char *et al.*, 1991 for reference).

The upper and lower bounds of the functions considered are evaluated by:

a) direct use of the Maple procedure **evalr**, which handles interval arithmetics;

b) application of **evalr** to the Taylor form (20); the capability of performing symbolic differentiation is exploited here.

Switching from a) to b) occurs when the interval length is less than 0.5.

At each stage, two linear programming problems must be solved (see Section 4). Since the Maple procedure **simplex** was found to be unacceptably slow for our purposes, the Maple **system** command was used to call the public domain program “**loqo**” retrievable from <ftp://elib.zib-berlin.de/pub/opt-net/software/loqo> which was found to be satisfactorily fast and reliable. A drawback is that the method **loqo** uses is an interior-point one, so that null variables in the solution are computed as small positive numbers. To remedy to this in the “upper” substage, where it is required to deal with intervals corresponding to non-null solution variables, we considered as positive variables greater than a fraction β of the largest one.

The procedure is stopped when the difference between the upper and lower bound of (9) gets smaller than some specified threshold δ .

It is possible to specify the list of intervals of the initial partition when necessary, for instance when the functions g , l_x or f_i are not differentiable at certain points. In this case, taking such points to form the initial partition ensures that (20) can be applied.

We remark that, if $\inf_{\Theta} v(\hat{y}; \theta)$ in (18) is close to 0, then the search of a point satisfying (19) may require a large number of trial points. Thus, it is sensible to stop the search after a prefixed number of trials M .

8. Numerical examples. In the following two examples, the procedure was run with $\alpha = 0.1$, $\beta = 10^{-3}$, $\delta = 10^{-3}$, $M = 50$; a maximum of 300 stages were performed. In both examples we report the values of the upper bound for $\sup_{\pi \in \Gamma} F(\pi)$ and the values of the lower bound for $\inf_{\pi \in \Gamma} F(\pi)$ obtained at the stopping of the procedure.

EXAMPLE 1. Assume that the probability distribution for the data is

$\mathcal{N}(\theta, \sigma^2)$, σ^2 known and that the prior for θ is an ϵ -contamination of π_0 given by $\mathcal{N}(\mu, \tau^2)$. We consider

$$\Gamma = \{\pi \in Q : \pi = (1 - \epsilon)\pi_0 + \epsilon q, q \in Q, \text{ and } \int_{K_i} m_\pi(u) du = \alpha_i, i = 1, 2, 3\}$$

for $K_1 = (-\infty, \gamma_1)$, $\alpha_1 = 0.25$, $K_2 = (-\infty, \gamma_2)$, $\alpha_2 = 0.5$, $K_3 = (-\infty, \gamma_3)$, $\alpha_3 = 0.75$, where γ_i 's are the quartiles of m_{π_0} , the marginal density of data corresponding to π_0 . This choice of Γ corresponds to the assumption that we are confident on the quartiles of m_{π_0} and we wish that such features are retained in Γ too. We want to investigate the robustness of the probability of $100(1 - \rho)\%$ Bayes credible region $C(x) = [c_1(x), c_2(x)]$, i.e. $g(\theta) = I_{C(x)}$ at an observation x . It is easy to see that (C2)-(C4) hold.

After transforming the problem into a new one in which, according to (3), the ϵ -contamination restriction is removed, we formulate the corresponding problem (12)-(13). The results reported in Table I were obtained for $\sigma^2 = 1$, $\mu = 0$, $\tau^2 = 2$, $\rho = 0.05$, $\epsilon = 0.2$, and $x = 0.5, 1.0, \dots, 4.0$ as in Betrò *et al.* (1994). See Berger and Berliner (1986), Moreno and Cano (1991) for further reference.

TABLE I. *Ranges of $F(\pi)$*

x	$\inf_{\pi \in \Gamma} F(\pi)$	$\sup_{\pi \in \Gamma} F(\pi)$
0.5	.904894	.961248
1.0	.906332	.961310
1.5	.891356	.961987
2.0	.855498	.961555
2.5	.794774	.961636
3.0	.734300	.962881
3.5	.623294	.966724
4.0	.468170	.975661

Results in Table I are in good agreement with the ones reported in Table 4 in Betrò *et al.* (1994). They show a good level in robustness for the smaller values of x . On the contrary, large values of x point out lack of robustness. In particular, when $x = 4$, it is not possible to conclude which event between $\{\theta \in C(x)\}$ and $\{\theta \notin C(x)\}$ is a posteriori more likely. In spite of the fact that Θ is not compact, convergence of the bounds was achieved.

EXAMPLE 2. Reliability analysis is a major field of application of Bayesian procedures. It is therefore attractive to apply robust Bayesian ideas to this field. The situation considered here is the one describing the occurrence of failures resulting from repeated application of correcting actions (like in the process of prototype development). A suitable mathematical model is the

so called Power Law Process, i.e. a non-homogeneous Poisson process with intensity

$$(22) \quad \lambda(t) = (\beta/\alpha)(t/\alpha)^{\beta-1}, t > 0, \alpha > 0, 0 < \beta < 1.$$

Let $(0, \tau]$ be the time interval in which (instantaneous) modifications are introduced into the prototype. After τ let no further improvement be incorporated into the system. Then the system reliability in $(\tau, \tau + \Delta\tau]$ is

$$R(\Delta\tau) = \exp(-\lambda(\tau)\Delta\tau).$$

Assume that we are interested into the robustness of the posterior mean of $R(\Delta\tau)$. Let t_1, \dots, t_n be the times at which intervention on the system occurs. We take $\tau = t_n$. As in Calabria *et al.* (1990), we reparametrize (22) in terms of

$$M_T = E(N(T) | \alpha, \beta) = (T/\alpha)^\beta,$$

where T is a fixed time and $N(T)$ is the number of failures in $(0, T]$, so that the likelihood can be written as

$$l(M_T, \beta; t_1, \dots, t_n) = T^{-n\beta} u^{\beta-1} \beta^n M_T^n \exp\left(-(\tau/T)^\beta M_T\right),$$

where $u = \prod_1^n t_i$. Let us assume that M_T and β are independent and that for M_T a Gamma distribution with parameters (a, b) is appropriate. In order to establish a class for the prior distribution of β , it is reasonable to think that the elicitation of the prior leads to consider the distribution of the first failure time, conditionally to M_T and β ,

$$P(T_1 > t | M_T, \beta) = \exp\left(-(\tau/T)^\beta M_T\right)$$

and to assess that upper and lower bounds of the marginal distribution $1 - P(T_1 > t)$ are given at some time instants, say s_1, s_2, s_3 . Then, after integration with respect to the distribution of M_T , we obtain

$$\begin{aligned} l_x(\beta) &= T^{-n\beta} u^\beta \beta^n \left(\left(\frac{\tau}{T} \right)^\beta + b \right)^{-(n+a)}, \\ g(\beta) &= \left(\left(\frac{\tau}{T} \right)^\beta + b \right)^{(n+a)} \cdot \left(\beta \frac{\Delta\tau}{\tau} \left(\frac{\tau}{T} \right)^\beta + \left(\frac{\tau}{T} \right)^\beta + b \right)^{-(n+a)}, \\ H_i(\beta) &= b^a \left(\left(\frac{s_i}{T} \right)^\beta + b \right)^{-a}, \quad i = 1, 2, 3. \end{aligned}$$

Let us assume $a = b = 2$; as in Finkelstein (1976) (see also Calabria *et al.*, 1992), let the data be given by $n = 10$, $t_1 = 13$, $t_2 = 48$, $t_3 = 89$, $t_4 =$

$121, t_5 = 189, t_6 = 262, t_7 = 323, t_8 = 395, t_9 = 499, t_{10} = 626$. We set $T = 10; \Delta\tau = 20; s_1 = 25, s_2 = 5, s_3 = 3, \Gamma = \{0.2 \leq \int_0^1 H_1 d\pi \leq 0.3, 0.55 \leq \int_0^1 H_2 d\pi \leq 0.6, 0.65 \leq \int_0^1 H_3 d\pi \leq 0.75\}$. Besides, it is easy to check that conditions (C1)-(C4) hold.

The range of the posterior mean of $R(\Delta\tau)$ has been computed as:

$$[.693284, .853333],$$

showing a certain degree of robustness in the Bayes estimate of $R(\Delta\tau)$, in spite of the restricted number of constraints.

9. Conclusions. We have shown how it is possible to build up an algorithm which represents a general tool for robust Bayesian analysis when the class of priors is specified in terms of generalized moment conditions. It has been pointed out that such a class covers a great variety of situations, including ϵ -contaminations and unimodal priors. There is obviously no claim that the limited numerical experience here reported is a proof of the effectiveness of the algorithm; however this latter appears useful in practical situations. We remark that tools of general availability like **Maple** and **loqo** have been exploited in the implementation of the algorithm.

Acknowledgements. The authors wish to thank referees for their insightful comments on the first draft of the paper.

REFERENCES

- BERGER, J.O. (1994). An Overview of Robust Bayesian Analysis. *Test* **3**, 5-59.
- BERGER, J.O. and BERLINER, L.M. (1986). Robust Bayes and empirical Bayes analysis with ϵ -contamination priors. *Ann. Statist.* **14**, 461-486.
- BETRÒ, B. and GUGLIELMI, A. (1994). An algorithm for robust Bayesian analysis under generalized moment conditions. *Quaderno IAMI 94.6*, IAMI-CNR Milano.
- BETRÒ, B., MĘCZARSKI, M. and RUGGERI, F. (1994). Robust Bayesian analysis under generalized moments conditions. *J. Statist. Plann. Infer.*, **41**, 257-266.
- CALABRIA, R., GUIDA, M. and PULCINI, G. (1990). Bayes estimation of prediction intervals for a power law process. *Commun. Stat. - Theory & Methods*, **19**, 3023-3035.
- CALABRIA, R., GUIDA, M. and PULCINI, G. (1992). A Bayes procedure for estimation of current system reliability. *IEEE Transactions on Reliability*, **41**, 616-620.
- CHAR, B., GEDDES, K.O., GONNET, G.H., LEONG, B.L., MONAGAN, M.B. and WATT, S.M. (1991). *Maple V Language Reference Manual*. Springer Verlag, New York.
- DALL'AGLIO, M. and SALINETTI, G. (1994). Bayesian robustness, moment problem and semi-infinite linear programming. *Poster presented at the Fifth Valencia Meeting*, Alicante (Spain), 5-9 June 1994.

FELLER, W. (1971). *An introduction to Probability Theory and its Applications*, vol.II, J. Wiley & Sons, New York.

FINKELSTEIN, J.M. (1976). Confidence bounds on the parameters of the Weibull process. *Technometrics*, **18**, 115-117.

HANSEN, E. (1992). *Global Optimization using Interval Analysis*. M. Dekker, New York.

HETTICH, R. and KORTANEK, K.O. (1993). Semi-infinite programming: theory, methods and applications. *SIAM Review*, **35**, 380-429.

HU, H. (1990). A one-phase algorithm for semi-infinite linear programming. *Math. Progr.*, **46**, 85-103.

KEMPERMAN, J.H.B. (1987). Geometry of the moment problem. In *Moments in probability, Proc. Symp. Appl. Math.* (LANDAU H.J., eds.), **37**. AMS, Providence, Rhode Island.

LAVINE, M. (1991). An approach to robust Bayesian analysis for multidimensional parameter spaces. *J. Amer. Statist. Assoc.*, **86**, 400-403.

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

PERONE PACIFICO, M., SALINETTI, G. and TARDELLA, L. (1994). Fractional optimization in Bayesian robustness. *Pubbl. Scient. Dip. Stat., Prob. e Stat. Appl., Università di Roma "La Sapienza"*, ser. A, **23**.

RATSCHEK, H. and ROKNE, J. (1988). *New Computer Methods for Global Optimization*. Ellis Horwood Lmt., Chichester (England).

RATSCHEK, H. and ROKNE, J. (1995). *Interval Methods*. In *Handbook of Global Optimization*, R. Horst and P.M. Pardalos eds., Kluwer, Dordrecht.

ISTITUTO PER LE APPLICAZIONI DELLA
MATEMATICA E DELL'INFORMATICA
CNR
VIA AMPERE, 56
I-20131 MILANO, ITALY

DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ DEGLI STUDI DI MILANO
VIA SALDINI, 50
I-20133 MILANO, ITALY

Numerical Robust Bayesian Analysis Under Generalized Moment Conditions

discussion by

ELÍAS MORENO.

University of Granada. Spain

Let me first congratulate the authors for the paper in which a very important problem in the field of Robust Bayesian Analysis is considered. In fact, most of the classes of prior distributions we have studied up to now can be formulated as “generalized moment problem”. The resulting numerical problems in the solution of the generalized moment problems are really very hard and therefore any effort in this direction is very welcome. In particular this proposed by Betrò and Guglielmi.

To begin, let me briefly review the problem. Let $f(x|\theta)$ be the likelihood of the statistical problem, $g(\theta)$ the quantity of interest and Γ the class of prior distributions

$$\begin{aligned}\Gamma &= \{\pi(\theta) : \int_{\Theta} \pi(d\theta) = 1, \int_{\Theta} H_i(\theta)\pi(d\theta) \leq \alpha_i, i = 1, 2, \dots, n\} \\ &= \{\pi(\theta) : \int_{\Theta} (H_i(\theta) - \alpha_i)\pi(d\theta) \leq 0, i = 1, 2, \dots, n\}\end{aligned}$$

where the functions $H_i(\theta)$, $i \geq 1$, and the real numbers α_i , $i \geq 1$, are specified. Given the observations $\mathbf{x} = (x_1, x_2, \dots, x_n)$, the goal is to find

$$\inf_{\pi \in \Gamma} E^{\pi}(g(\theta) | x) = \inf_{\pi \in \Gamma} \frac{\int_{\Theta} g(\theta)f(x | \theta)\pi(d\theta)}{\int_{\Theta} f(x | \theta)\pi(d\theta)}$$

The first nice idea in the paper is to set that

$$\inf_{\pi \in \Gamma} E^{\pi}(g(\theta) | x) = \inf_{\nu \in \Gamma_1} E^{\nu}\{g(\theta)f(\mathbf{x} | \theta)\},$$

where

$$\begin{aligned}\Gamma_1 &= \{\nu(\theta) : \int_{\Theta} \nu(d\theta) = k, k \geq 0, \int_{\Theta} (H_i(\theta) - \alpha_i)\nu(d\theta) \leq 0, i = 1, 2, \dots, n, \\ &\quad \int_{\Theta} f(\mathbf{x} | \theta)\nu(d\theta) = 1\}.\end{aligned}$$

Note that $E^{\pi}(g(\theta) | \mathbf{x})$ is a ratio of linear function in π but $E^{\nu}\{g(\theta)f(\mathbf{x} | \theta)\}$ is a linear functional in ν . The price to pay is that Γ involves n linear constraints on π and Γ_1 involves $n + 2$ constraints on ν .

On the other hand, since the extreme priors in Γ_1 are discrete measures a natural thing to do is to discretize the space Θ . This is the idea tried in the paper.

This discretization of the problem, however, may be done in many different ways. My first question to the authors is: why the discretization followed in the paper is particularly attractive?

Furthermore, as the authors pointed out in the section 4 of the paper, there exists a dual formulation of the problem which contains a finite number of parameter (see Kemperman (1987), Perone, Salinetti, Tardella (1994), Dall'Aglio and Salinetti (1994), Liseo, Moreno and Salinetti (1995)), and allows to formulate the problem as

$$\begin{aligned} & \inf_{\nu \in \Gamma_1} E^\nu \{g(\theta)f(\mathbf{x} | \theta)\} = \\ & \sup_{\substack{d_i \geq 0 \\ d_{n+1} \in \mathcal{R}}} \{d_{n+1} + \inf_{\nu \in \Gamma_1} \int_{\Theta} [g(\theta)f(\mathbf{x} | \theta) - \sum_1^n d_i(H_i(\theta) - \alpha_i) - d_{n+1}f(\mathbf{x} | \theta)]\nu(d\theta)\} \\ & = \sup_{\substack{d_i \geq 0 \\ d_{n+1} \in \mathcal{R}}} \{d_{n+1} + \inf_{k \geq 0} \inf_{\theta \in \Theta} [g(\theta)f(\mathbf{x} | \theta) - \sum_1^n d_i(H_i(\theta) - \alpha_i) - d_{n+1}f(\mathbf{x} | \theta)]k\}. \end{aligned}$$

In this setting my second question is: why not to concentrate numerical efforts in solving this dual problem that is a finite dimensional problem from the beginning? I feel that the authors could be more efficient if this problem is focused instead of the former one.

ADDITIONAL REFERENCES

- LISEO B., MORENO, E. AND SALINETTI G. (1995). Bayesian Robustness for the class of bidimensional priors with given marginals (with discussion). *In This Volume*.
- KEMPERMAN, J.H.K. (1987). Geometry of the moment problem. In *Semi-infinite Programming and Applications*, A.V. Fiacco and K.O. Kortanek, Eds., Lectures Notes in Econometrics and Mathematical Systems, vol. **215**, 63-92.

REJOINDER

BY BRUNO BETRÒ AND ALESSANDRA GUGLIELMI

We thank the discussant for his appreciation of our effort to contribute to the numerical solution of a “really very hard” problem. Indeed, we are convinced that, in view of the wideness of the Generalized Moment class, its effective numerical treatment represents a major step towards making robust Bayesian analysis more appealing to the statistical practitioner.

Concerning his observation that the price to pay for linearization via (6) is an increase of the number of constraints, this does not appear as a serious drawback, since (7) does not alter the linear structure of the constraint set. So the price is worthy to be paid.

About his first question, we do not obviously claim that our discretization is the “best” one. We propose an adaptive discretization scheme which eventually leads to isolate the atoms of the extreme measure. Computing upper and lower bounds over the intervals of the discretization enables to control the accuracy so far reached. Any improvement over the simple scheme adopted, which however is borrowed from standard techniques in the interval analysis approach to global optimization, is welcome.

Coming to the second question, we observe that the problem

$$\begin{aligned} & \sup_{d_i \geq 0} \{d_{n+1} + \inf_{k \geq 0} \inf_{\theta \in \Theta} \{g(\theta)f(x|\theta) + \sum_{i=1}^n d_i(H_i(\theta) - \alpha_i) - d_{n+1}f(x|\theta)\}k\} \\ & d_{n+1} \in \mathbb{R} \end{aligned}$$

(notice that a positive sign is needed before the sum - see Kemperman, 1987, p. 26) is easily seen to reduce to the problem $\sup d_{n+1}$ under the constraints $g(\theta)f(x|\theta) + \sum_{i=1}^n d_i(H_i(\theta) - \alpha_i) - d_{n+1}f(x|\theta) \geq 0$, for all $\theta \in \Theta$, $d_i \geq 0$, $i = 1, \dots, n$, $d_{n+1} \in \mathbb{R}$, that is the analogue to the dual formulation (16)-(17), when we look for the inf instead of the sup. Thus, it is natural to approach the problem by algorithms developed in the frame of semi-infinite linear programming, as we did using Hu’s procedure. This latter was modified in a more operative way, adopting interval analysis techniques, in order to solve the global optimization problem it involves.

REFERENCES

- KEMPERMAN, J.H.B. (1987). Geometry of the moment problem. In *Moments in probability, Proc. Symp. Appl. Math.* (LANDAU H.J., eds.), **37**. AMS, Providence, Rhode Island.