GENERALIZED MOMENT THEORY AND BAYESIAN ROBUSTNESS ANALYSIS FOR HIERARCHICAL MIXTURE MODELS

Bruno Betrò, Antonella Bodini and Alessandra Guglielmi

CNR-IMATI, via Bassini 15, 20133 Milano, Italy

BAYESIAN ROBUSTNESS IN HIERARCHICAL MODELS

Abstract. In applications of Bayesian analysis one problem that arises is the evaluation of the sensitivity, or robustness, of the adopted inferential procedure with respect to the components of the formulated statistical model. In particular, it is of interest to study robustness with respect to the prior, when this latter cannot be uniquely elicitated, but a whole class Γ of probability measures, agreeing with the available information, can be identified. In this situation, the analysis of robustness consists of finding the extrema of posterior functionals under Γ. In this paper we provide a theoretical framework for the treatment of a global robustness problem in the context of hierarchical mixture modeling, where the mixing distribution is a random probability whose law belongs to a generalized moment class Γ. Under suitable conditions on the functions describing the problem, the solution of this latter coincides with the solution of a linear semi-infinite programming problem.

Key words and phrases: Bayesian robustness analysis, hierarchical mixture models, nonparametric prior, moment theory, linear semi-infinite programming.

1. Introduction

Robustness analysis is concerned with the sensitivity of the results of the inference to the assumptions of the adopted model. In particular, in Bayesian inference a robustness problem arises for instance when, due to lack of information, the prior is difficult to be elicitated.

The state of the art, up to 2000, of robustness issues in Bayesian analysis is exhibited in the papers collected in Rios Insua and Ruggeri (2000); the opening paper by Berger *et al.* (2000) presents an overview of the robust Bayesian approach, which usually includes the global robustness approach, where the class of all priors coherent with the elicited prior information is considered, and the local robustness approach, where the interest is in the rate of change in inferences with respect to small changes in the prior.

When considering global robustness (in short, robustness henceforth), the analysis usually can be expressed as follows

$$\sup_{\pi \in \Gamma} \frac{\int_\Theta g(\theta) l(\theta) \pi(d\theta)}{\int_\Theta l(\theta) \pi(d\theta)} - \inf_{\pi \in \Gamma} \frac{\int_\Theta g(\theta) l(\theta) \pi(d\theta)}{\int_\Theta l(\theta) \pi(d\theta)}$$

where $g : \Theta \to \mathbb{R}$ is some function of interest, $l(\theta)$ is the likelihood function, Θ is the parameter space and π is the prior distribution which is assumed to belong to a class Γ of probability distributions. Of course, without loss of generality, the analysis can be focused on the supremum. In this framework, the possibility of providing effective algorithms for the analysis of robustness is available in the case of classes of priors defined by generalized moment conditions, since here the problem of analyzing Bayesian robustness is reduced to a problem of Linear Semi-Infinite Programming.

Generalized moment classes have been considered in connection with robustness first by Betrò *et al.* (1994) and then by Betrò and Guglielmi (1994), Goutis (1994), Dall'Aglio (1995), Smith (1995), Betrò *et al.* (1996), Betrò and Guglielmi (1997, 2000). Such classes incorporate a number of interesting situations – the most common being the one in which bounds on quantiles of the prior distribution are available – and have been widely studied in other contexts. Consequently, a rather comprehensive theory exists for optimization of linear functionals defined over them, mainly due to Kemperman (1971, 1983, 1987). In robustness analysis, the functional to be optimized is not linear but, as first noticed in Betrò and Guglielmi (1994), it is possible to obtain linearity by a suitable transformation, so that the above theory can be applied, as extensively studied in Betrò and Guglielmi (2000). We remark that, as shown in Hoff (2003), probability measures belonging to a generalized moment class can be represented as convex combinations of extremal probability measures. This property can be exploited as an alternative way for solving the robustness problem, as in Betrò *et al.* (1994); however, the corresponding optimization problem turns out to be a global nonlinear one, and its numerical solution seems more difficult to be obtained.

The aim of this paper is to provide a theoretical framework for the treatment of a global robustness problem within nonparametric hierarchical mixture modeling. By their flexibility, combined with the development of suitable sample techniques, Bayesian hierarchical models based on Dirichlet processes or other random probability measures have greatly increased their popularity. Here, extending the results of Betrò and Guglielmi (2000), the random probability measure defining the mixing distribution is assumed to vary in a generalized moment class, as described in Section 2. The resulting robustness problem will be referred to as nonparametric robustness problem. Since the extension requires to work with the set $\mathcal{M}$ of all finite Borel measures on a separable metric space, in Section 3 we establish the fundamental results of the generalized moment problem in

$\mathcal{M}$. The key result is Theorem 3.1, which is similar to Theorem 5 in Kemperman (1983); proof of this latter was never published, while the former is proved here using a classical result in convex analysis known as Farkas' lemma. The application of the general theory to robustness analysis is described in Section 4. Finally, two examples illustrate the approach.

2. The problem

In the last 10 years, a large amount of papers in the nonparametric Bayesian literature have been devoted to study inferences in the context of hierarchical mixture modeling, described as follows:

$$\begin{gathered} X_1, \ldots, X_r | Y_1, \ldots, Y_r \text{ are independent} \\ X_i | Y_i \text{ is distributed according to } \mathcal{L}(X_i | Y_i),\ i = 1, \ldots, r \\ Y_1, \ldots, Y_r | \tilde{\pi} \text{ are i.i.d. according to } \tilde{\pi}, \end{gathered} \tag{2.1}$$

and

$$\tilde{\pi} \sim q, \tag{2.2}$$

where q is a nonparametric prior, i.e. the distribution of a *random probability measure* $\tilde{\pi}$. The most popular choice for $\tilde{\pi}$ is the Dirichlet process, and the resulting model, introduced by Lo (1984), is known as "mixture of Dirichlet process models" (MDP) or "Dirichlet mixture of kernels". Relevant contributions in the context of nonparametric Bayesian hierarchical mixture modeling include those by, among others, Escobar and West (1995, 1998) for the MDP model, Petrone (1995) for Bernstein polynomials, Lijoi *et al.* (2005) for mixtures of normalized inverse-Gaussian processes, and Nieto-Barajas *et al.* (2004) for mixtures of normalized random distribution functions with independent increments. For an overview on the Bayesian nonparametric approach, see Ghosh and

Ramamoorthi (2003).

This article considers a more general framework, assuming that the prior q cannot be uniquely specified but belongs to a generalized moment class. According to (2.1), the data are represented by r random vectors $X_1, \ldots, X_r$ with values in $\mathbb{R}^k$, and the conditional distribution of X_i given Y_i is defined by a transition probability density $k(x; y)$, i.e. $k : \mathbb{R}^k \times \mathcal{Y} \to [0, +\infty)$, where $\mathcal{Y}$ is a measurable subset of $\mathbb{R}^n$, such that

- $y \mapsto k(x; y)$ is π-measurable for all x in $\mathbb{R}^k$;
- $x \mapsto k(x; y)$ is a probability density on $\mathbb{R}^k$, for all y in $\mathcal{Y}$, with respect to a σ-finite measure λ on $\mathbb{R}^k$.

Observe that, by (2.1), whatever distribution q is chosen, the prior is concentrated on the space of densities. We assume that $\tilde{\pi}$ takes values in a subset $\mathcal{S}$ of $\mathcal{P}(\mathcal{Y})$, the set of all probability measures on $\mathcal{Y}$, so that q belongs to $\mathcal{P}(\mathcal{S})$. If $x = (x_1, \ldots, x_r)$ is a sample from $(X_1, \ldots, X_r)$, the posterior distribution of $\tilde{\pi}$ is given by the Bayes theorem

$$q(d\pi|x) = \frac{l(\pi)q(d\pi)}{\int_{\mathcal{S}} l(\pi)q(d\pi)},$$

where $l(\pi) := l(\pi; x) = \prod_{i=1}^{r} p(x_i; \pi)$ is the likelihood function, and $p(x; \pi) := \int_{\mathcal{Y}} k(x; y)\pi(dy)$. We assume that q in (2.2) belongs to a (nonempty) generalized moment class Γ,

$$\Gamma = \{q \in \mathcal{P}(\mathcal{S}) : \int_{\mathcal{S}} f_i(\pi)q(d\pi) \leq \alpha_i,\ i = 1, \ldots, m\}, \tag{2.3}$$

where f_i are given q-integrable functions, and α_i are fixed real constants, $i = 1, \ldots, m$.

The global robustness problem consists of determining

$$\sup_{q \in \Gamma} \frac{\int_{\mathcal{S}} g(\pi)l(\pi)q(d\pi)}{\int_{\mathcal{S}} l(\pi)q(d\pi)} \tag{2.4}$$

where $g : \mathcal{S} \to \mathbb{R}$ is a given function such that $\int_{\mathcal{S}} g(\pi)q(d\pi|x)$ exists for all q.

Observe that, if $\mathcal{S}$ is the space of degenerate probability measures on $\mathcal{Y}$, then $p(x;\pi) = p(x;\delta_y) = k(x;y)$, and the problem considered here coincides with the parametric problem considered in Betrò and Guglielmi (1997, 2000). For this reason, (2.3)-(2.4) can be viewed as an extension to the nonparametric setting of the robustness parametric analysis under generalized moment conditions.

As concern the functions defining Γ, possible choices for the function g are the following:

- $g(\pi) = I_{\mathcal{S}_1}(\pi)$, the indicator function of some measurable subset $\mathcal{S}_1$ of interest; in this case we are concerned with the posterior probability that the random mixing probability measure $\tilde{\pi}$ belongs to $\mathcal{S}_1$;
- $g(\pi) = \pi(A)$, $A \in \mathcal{B}(\mathcal{Y})$; in this case, we want to compute the supremum of the *a posteriori* expected value of the random variable $\tilde{\pi}(A)$, when q varies in Γ;
- $g(\pi) = p(x;\pi)$, $x \in \mathbb{R}^k$, i.e. we deal with the predictive density of a future observation.

As far as the constraints are concerned, possible functions f_i's are:

- $f_i(\pi) = \int_{K_i} p(x;\pi)\lambda(dx)$, $K_i \in \mathcal{B}(\mathbb{R}^k)$, so that $\int_{\mathcal{S}} f_i(\pi)q(d\pi) = \int_{K_i} m_{X_1}(x)\lambda(dx)$, where m_{X_1} denotes the marginal density of a single observation; such a constraint specifies a bound on the marginal distribution;
- for $k = 1$, $f_i(\pi) = \int_{\mathbb{R}} x^i\, p(x;\pi)\lambda(dx)$ so that $\int_{\mathcal{S}} f_i(\pi)q(d\pi) = \int_{\mathbb{R}} x^i\, m_{X_1}(x)\lambda(dx)$, here the bound is on the moments of the marginal distribution;
- $f_i(\pi) = \int_{\mathcal{Y}} y^i \pi(dy)$, if $\mathcal{Y} \in \mathcal{B}(\mathbb{R})$; such a constraint represents a bound on the expected value of the i-th moment functional of the random probability measure $\tilde{\pi}$.

Analogously to the parametric problem, the functional to be optimized in (2.4) is not linear in the argument q; however by means of a *variable transformation approach*, it is easy to obtain an equivalent linear optimization problem (see Betrò and Guglielmi, 2000). Indeed, consider the map ψ from $\mathcal{P}(\mathcal{S})$ to $\mathcal{M}(\mathcal{S})$ which associates to the probability measure q the finite measure μ defined by

$$\psi(q)(A) =: \mu(A) = \frac{q(A)}{\int_{\mathcal{S}} l(\pi) q(d\pi)}, \quad A \in \mathcal{B}(\mathcal{S}), \tag{2.5}$$

assuming that $0 < \int_{\mathcal{S}} l(\pi) q(d\pi) < +\infty$. It is easily seen that $\int_{\mathcal{S}} l(\pi)\mu(d\pi) = 1$ if, and only if, $\mu \in \psi(\mathcal{P}(\mathcal{S}))$. Moreover, the map ψ is injective. Indeed, if $\psi(q_1) = \psi(q_2)$ then, by (2.5) with $A = \mathcal{S}$ it holds $\int_{\mathcal{S}} l(\pi) q_1(d\pi) = \int_{\mathcal{S}} l(\pi) q_2(d\pi)$, so that $q_1(A) = q_2(A)$ for all $A \in \mathcal{B}(\mathcal{S})$, i.e. $q_1 = q_2$. By this transformation, problem (2.3)-(2.4) turns into

$$\sup_{\mu \in \mathcal{M}_1} \int_{\mathcal{S}} g(\pi) l(\pi) \mu(d\pi), \tag{2.6}$$

where

$$\mathcal{M}_1 = \mathcal{M}_1(\mathcal{S}) := \{\mu \in \mathcal{M}(\mathcal{S}) : \int_{\mathcal{S}} \tilde{f}_i(\pi)\mu(d\pi) \leq 0,\ i = 1, \ldots, m, \int_{\mathcal{S}} l(\pi)\mu(d\pi) = 1\} \tag{2.7}$$

and $\tilde{f}_i(\pi) := f_i(\pi) - \alpha_i$, $i = 1, \ldots, m$.

Problem (2.6)-(2.7) is an instance of the class of generalized moment problems widely studied by Kemperman (see e.g. Kemperman, 1971; 1983; 1987). The main results which are useful for our purposes are reported in the following section.

3. Some theory on the generalized moment problem

Consider a general separable metric space $\mathcal{S}$, and let $\mathcal{M} = \mathcal{M}(\mathcal{S})$ be the set of all finite Borel measures on $\mathcal{S}$; see Appendix A for preliminaries on the space $\mathcal{M}$. Let $h_i : \mathcal{S} \to \mathbb{R}$, $u : \mathcal{S} \to \mathbb{R}$ be measurable functions and η_i be real constants, $i = 1, \ldots, n$.

The generalized moment problem consists in determining the upper bound

$$U := \sup_{\mu \in \mathcal{M}_1} \int_{\mathcal{S}} u d\mu, \tag{3.1}$$

where $\mathcal{M}_1 = \mathcal{M}_1(\mathcal{S}) := \{\mu \in \mathcal{M} : \int_{\mathcal{S}} h_i^+ d\mu < +\infty, \int_{\mathcal{S}} h_i d\mu \leq \eta_i, i = 1, \ldots, n\}$. We assume the existence of μ^* in $\mathcal{M}_1$ such that $\int u^- d\mu^* > -\infty$, so that $U > -\infty$. From now on, $\mathbb{R}_+^n$ will denote the set $\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_i \geq 0 \ \forall \ i = 1, \ldots, n\}$.

It is easy to show that, for any convex subset $\mathcal{M}_0$ of $\mathcal{M}$ containing $\mathcal{M}_1$, it holds

$$U \leq U^* := \inf \Big\{ \sum_{i=1}^n \beta_i \eta_i + \sup_{\mu \in \mathcal{M}_0} \int_{\mathcal{S}} (u - \sum_{i=1}^n \beta_i h_i) d\mu, \ \beta \in \mathbb{R}_+^n \Big\}.$$

At this point we aim to determine: (a) conditions under which U^* is finite and $U = U^*$; (b) a tractable form for U^*. Note that if $\mathcal{M}_0 = \mathcal{M}$, then

$$U^* = \inf \Big\{ \sum_{i=1}^n \beta_i \eta_i : \ \beta \in \mathbb{R}_+^n, \ \sum_{i=1}^n \beta_i h_i(s) \geq u(s) \ \forall \ s \in \mathcal{S} \Big\} \tag{3.2}$$

and the problem of calculating U^* as in (3.2) is usually called *Linear Semi-Infinite Programming* (LSIP) problem; see Goberna and López (1998). However, assuming $\mathcal{M}_0 = \mathcal{M}$ will not guarantee that U and U^* coincide. Theorem 3.2, representing the main result of this section, answers to issue (a) and permits to obtain (3.2) when $\mathcal{M}_0$ is a proper compact subset of $\mathcal{M}$.

We introduce the following conditions:

H1: h_i is lower semicontinuous (l.s.c.), i.e. $\{s \in \mathcal{S} : h_i(s) > b\}$ is an open set for any $b \in \mathbb{R}$, for all $i = 1, \ldots, n$;

H2: $h_i(s) \geq C_i$ for all $s \in \mathcal{S}$, for some constants $C_i < 0, \quad i = 1, \ldots, n$;

H3: there exists $\tilde{\beta} \in \mathbb{R}_+^n$ such that $\sum_{i=1}^n \tilde{\beta}_i h_i(s) > 1$ for all $s \in \mathcal{S}$;

H4: for any $\varepsilon > 0$ there exist a compact set K_ε in $\mathcal{S}$ and $\beta^\varepsilon \in \mathbb{R}_+^n$, with $\sum_{i=1}^n \beta_i^\varepsilon \leq B$ for some positive constant B, such that $\sum_{i=1}^n \beta_i^\varepsilon h_i(s) > \frac{1}{\varepsilon}$ for all $s \in K_\varepsilon^\complement$.

H5: u is upper semicontinuous (u.s.c.), i.e. $-u$ is l.s.c.;

H6: $u(s) \leq G$ for all $s \in \mathcal{S}$, for some constant $G > 0$;

H7: there exists $\tilde{\beta} \in \mathbb{R}_+^{n+1}$, such that $\sum_{i=1}^n \tilde{\beta}_i h_i(s) - \tilde{\beta}_{n+1} u(s) > 1$ for all $s \in \mathcal{S}$;

H8: for any $\varepsilon > 0$ there exist a compact set K_ε in $\mathcal{S}$ and $\beta^\varepsilon \in \mathbb{R}_+^{n+1}$ with $\sum_{i=1}^{n+1} \beta_i^\varepsilon \leq B$ for some positive constant B, such that $\sum_{i=1}^n \beta_i^\varepsilon h_i(s) - \beta_{n+1}^\varepsilon u(s) > \frac{1}{\varepsilon}$ for all $s \in K_\varepsilon^\complement$.

Observe that H3 (H4) is a special case of H7 (H8).

Next theorem is the key result of the section. It is analogous to Theorem 5 in Kemperman (1983), which, however, was stated without proof. Condition H3 used here appears to be easier to verify than formula (4.4) in Kemperman's theorem.

Theorem 3.1. *Under H1-H4, $\mathcal{M}_1 \neq \emptyset$ if, and only if, the following condition holds:*

$$\text{if } \beta \in \mathbb{R}_+^n \text{ is such that } \sum_{i=1}^n \beta_i h_i(s) \geq 0 \text{ for any } s \in \mathcal{S}, \text{ then} \sum_{i=1}^n \beta_i \eta_i \geq 0. \tag{3.3}$$

Proof. See Appendix B.

Theorem 3.2 stems now from Theorem 3.1; the proof is rather technical and uses standard arguments of generalized moment theory. The interested reader is referred to Betrò *et al.* (2002).

Theorem 3.2. *Under H1-H2 and H5-H8, if $\mathcal{M}_1 \neq \emptyset$, then*

$$U = \inf\{\sum_{i=1}^n \beta_i \eta_i : \ \beta \in \mathbb{R}_+^n, \ \sum_{i=1}^n \beta_i h_i(s) \geq u(s), \ \forall \ s \in \mathcal{S}\}. \tag{3.4}$$

Moreover, the supremum U is finite and attained.

If the metric space $\mathcal{S}$ is compact, some of the hypotheses in Theorem 3.2 are automatically satisfied so we can state the following

Corollary 3.1. *Let $\mathcal{S}$ be compact. Under H1, H5 and H7, if $\mathcal{M}_1 \neq \emptyset$, then*

$$U := \sup_{\mu \in \mathcal{M}_1} \int_{\mathcal{S}} u d\mu = \inf\{\sum_1^n \beta_i \eta_i : \beta \in \mathrm{IR}_+^n, \sum_1^n \beta_i h_i(s) \geq u(s) \ \forall \ s \in \mathcal{S}\}. \qquad (3.5)$$

Moreover, the supremum U is finite and attained.

4. Solving nonparametric robustness problems

In this section we apply the results of the previous one to problem (2.6)-(2.7), using the same notation as in Section 2. Indeed, any subset $\mathcal{S}$ of $\mathcal{P}(\mathcal{Y})$ is separable by the separability of $\mathcal{Y} \subset \mathrm{IR}^k$. Conditions H1-H2 and H5-H8 are immediately transposed into the following I1-I7, exploiting the nonnegativity of $l(\pi)$ in the formulation of I6 and I7.

I1: $\tilde{f}_j$ is l.s.c., $j = 1, \ldots, m$, and l is continuous;

I2: g is u.s.c.;

I3: $\tilde{f}_j \geq C_j$, $j = 1, \ldots, m$, where C_j's are real (negative) constants;

I4: l is bounded;

I5: gl is bounded from above;

I6: there exist $\tilde{\beta}_0 \in \mathrm{IR}_+$, $\tilde{\beta} \in \mathrm{IR}_+^{m+1}$ such that $\tilde{\beta}_0 l(\pi) + \sum_{i=1}^m \tilde{\beta}_i \tilde{f}_i(\pi) - \tilde{\beta}_{m+1} g(\pi) l(\pi) > 1$ for all π in $\mathcal{S}$;

I7: for any $\varepsilon > 0$ there exist a compact set K_ε in $\mathcal{S}$ and $\beta_0^\varepsilon \in \mathrm{IR}_+$, $\beta^\varepsilon \in \mathrm{IR}_+^{m+1}$ with $\sum_{i=0}^{m+1} \beta_i^\varepsilon \leq B$ for some positive constant B, such that $\beta_0^\varepsilon l(\pi) + \sum_{i=1}^m \beta_i^\varepsilon \tilde{f}_i(\pi) - \beta_{m+1}^\varepsilon g(\pi) l(\pi) > \frac{1}{\varepsilon}$ for all π in $K_\varepsilon^{\mathsf{C}}$.

Remark 1. If I6 holds with $\tilde{\beta}_{m+1} = 0$, then choosing $\tilde{\varepsilon}$ such that $1 - \tilde{\beta}_0 \tilde{\varepsilon} > 0$ and $\tilde{\beta}'_i = \tilde{\beta}_i / (1 - \tilde{\beta}_0 \tilde{\varepsilon})$, it is also $\sum_{i=1}^m \tilde{\beta}'_i \tilde{f}_i(\pi) > 1$ for all π in $\mathcal{S}$ such that $l(\pi) \leq \tilde{\varepsilon}$. Roughly speaking, when l is small, at least one of the $\tilde{f}_i$'s is positive and far from 0. The converse is also true: if there exist $\tilde{\varepsilon} > 0$ and $\tilde{\beta} \in \mathbb{R}_+^m$ such that $\sum_{i=1}^m \tilde{\beta}_i \tilde{f}_i(\pi) > 1$ for all $\pi \in \mathcal{S}$ at which $l(\pi) \leq \tilde{\varepsilon}$, then I6 holds with $\tilde{\beta}_{m+1} = 0$ and $\tilde{\beta}_0 > (1 - \sum_{i=1}^m \tilde{\beta}_i C_i) / \tilde{\varepsilon}$.

The result follows immediately:

THEOREM 4.1. *Under I1-I7, if $\mathcal{M}_1(\mathcal{S}) \neq \emptyset$, then*

$$\begin{aligned} \sup_{\mu \in \mathcal{M}_1} \int_{\mathcal{S}} g(\pi) l(\pi) \mu(d\pi) = \inf\{\ & \beta_0 : \beta_0 \in \mathbb{R}, \beta \in \mathbb{R}_+^m, \\ & \beta_0 l(\pi) + \textstyle\sum_1^m \beta_i \tilde{f}_i(\pi) \geq g(\pi) l(\pi) \ \ \forall\, \pi \in \mathcal{S}\}. \end{aligned} \tag{4.1}$$

and the supremum in the left hand-side of (4.1) is finite and attained.

PROOF. Apply Theorem 3.2 with $n = m+2$, $u = gl$, $\mathcal{M}_1 = \mathcal{M}_1(\mathcal{S})$, $h_1 = l$, $h_2 = -l$, $\eta_1 = 1$, $\eta_2 = -1$, $h_{i+2} = \tilde{f}_i$, $\eta_{i+2} = 0$, $i = 1, \ldots, m$.

Remark 2. According to Corollary 3.1, when $\mathcal{S}$ is compact, only assumptions I1-I2 and I6 are to be verified.

Remark 3. If any of the constraint functions $f_j(\pi)$ has the form $\int_{\mathcal{Y}} z(y) \pi(dy)$, and $z : \mathcal{Y} \to \mathbb{R}$ is bounded from below and l.s.c, then it is easy to show that f_j is l.s.c. as well (see Lemma 3 in Kemperman, 1983). Indeed, the result holds not only for probabilities, but also for finite measures.

Here follow some considerations about the solution of (4.1). Theorem 4.1 shows that the required maximum can be obtained solving a LSIP problem. Although algorithms for this latter exist, regardless of the structure of $\mathcal{S}$ (see e.g. Betrò, 2004), in practice the infinite dimensionality of $\mathcal{S}$ must be dealt with, so that a finite approximation is necessary. Treatment of such computational aspects is beyond the scope of this paper.

Once it has been ensured that the supremum in (2.6) is reached by some measure μ^*, then it can be assumed that μ^* has finite support of at most $m+1$ points. Indeed, setting

$$z_0 = \int_{\mathcal{S}} g l d\mu^*, \quad z_i = \int_{\mathcal{S}} \tilde{f}_i d\mu^*, \; i = 1, \dots m,$$

and recalling that $1 = \int_{\mathcal{S}} l d\mu^*$, Theorem 1 in Rogosinsky (1958) (see also Lemma 1 in Kemperman, 1983) states that there exists a measure μ, having finite support of at most $m+2$ points, and such that $\int_{\mathcal{S}} l d\mu = 1$, $\int_{\mathcal{S}} g l d\mu = z_0$, $\int_{\mathcal{S}} \tilde{f}_i d\mu = z_i$, $i = 1, \dots, m$, so that such a measure is still optimal. Consequently, the problem of determining $\sup_{\mu \in \mathcal{M}_1} \int_{\mathcal{S}} g l d\mu$ turns out be an ordinary (finite) linear programming problem in $m+2$ variables and $m+1$ constraints for which it is well known that, if a solution exists, then it can be assumed to have at most $m+1$ non-null coordinates.

Denoting by μ^* an optimal measure with atoms $\pi_1^*, \dots, \pi_k^*$ and corresponding masses $\mu_1^*, \dots, \mu_k^*$, $\mu_j^* > 0$, $j = 1, \dots, k$, $k \le m+1$, if the infimum in (4.1) is a minimum achieved by, say, $\beta_0^* \in \mathbb{R}$ and corresponding coefficients $\beta_i^* \ge 0$ $(i = 1 \dots, m)$, then it is

$$\begin{aligned} 0 \le \; & \int \left(\beta_0^* l + \textstyle\sum_{i=1}^m \beta_i^* \tilde{f}_i - g l \right) d\mu^* \\ & = \beta_0^* + \textstyle\sum_{i=1}^m \beta_i^* \int \tilde{f}_i d\mu^* - \beta_0^* \le 0 \end{aligned}$$

so that

$$\beta_0^* l(\pi_j^*) + \sum_{i=1}^m \beta_i^* \tilde{f}_i(\pi_j^*) = g(\pi_j^*) l(\pi_j^*), \; j = 1, \dots, k \tag{4.2}$$

and

$$\sum_{j=1}^k \tilde{f}_i(\pi_j^*) \mu_j^* = 0 \tag{4.3}$$

for all i's such that $i \in \{1, \ldots, m\}$ and $\beta_i^* > 0$.

Conversely, if μ^* in $\mathcal{M}_1$ is an atomic measure with support $\pi_1^*, \ldots, \pi_k^*$, $k \leq m+1$ which satisfies (4.2)–(4.3), then μ^* is an optimal measure. Indeed

$$\beta_0^* \geq \int gl d\mu^* = \beta_0^* \int l d\mu^* + \sum_{i=1}^{m} \beta_i^* \int \tilde{f}_i d\mu^* = \beta_0^*$$

so that $\int gl d\mu^* = \beta_0^*$, i.e. μ^* is optimal.

The above characterization of an optimal measure is useful for practical computations, as illustrated in the next Section.

Finally, observe that in order to apply Theorem 4.1 we need to either provide conditions under which $\mathcal{S}$ is compact, or single out compact subsets of $\mathcal{S}$ (see I7). If we assume that $\mathcal{Y}$ is compact and $\mathcal{S}$ is closed, then $\mathcal{S}$ is compact; indeed, if $\mathcal{Y}$ is compact, then $\mathcal{S} \subset \mathcal{P}(\mathcal{Y})$ is trivially tight and, being closed, is compact too by Prohorov's theorem (see, e.g., Parthasarathy, 1967; Theorem 6.7, p. 47). If $\mathcal{S}$ is not compact, but $\mathcal{Y}$ and $\mathcal{S}$ are closed subsets in their corresponding spaces, then $\mathcal{S}$ is a Polish space too, so by Prohorov's theorem we can characterize compact sets in $\mathcal{S}$ as the closures of tight sets is $\mathcal{S}$. For instance, condition I7 holds when one of the functions $\tilde{f}_i$'s, say $\tilde{f}_1$, has the form

$$\tilde{f}_1(\pi) = \int_{\mathcal{Y}} z(y)\pi(dy) - \alpha_1, \quad z(y) \geq C, \; C < 0,$$

δ_0 is a positive real number such that $1/\delta_0 - C + \alpha_1 > 0$, and for some family of nonempty compact sets $\{A_\delta \subset \mathcal{Y}; 0 < \delta \leq \delta_0\}$ nondecreasing when δ decreases to zero and such that $\pi(A_{\delta_0}^{\complement}) = 0$ for some π in $\mathcal{S}$, there exists a function $v(\delta)$ such that

$$\lim_{\delta \to 0+} v(\delta) = 0, \quad v(\delta) \inf_{y \in A_\delta^{\complement}} z(y) \geq \frac{1}{\delta} - C + \alpha_1 \;\text{ for } \delta \leq \delta_0.$$

This is the case, e.g., if $\mathcal{Y} = [0, +\infty)$, $z(y) \geq y^\alpha$ and nondecreasing in $[K, +\infty)$ for some $\alpha > 0$ and $K > 0$, provided that $\pi((K, +\infty)) = 0$ for some π in $\mathcal{S}$; indeed, it is easily verified that we can take $A_\delta = [0, (z(1/\delta)(1/\delta - C + \alpha_1))^{\frac{1}{\alpha}}]$ and $v(\delta) = z^{-1}(1/\delta)$ for $\delta \le \delta_0 = 1/K$, assuming $1/K - C + \alpha_1 \ge 1$.

For any $\varepsilon \le \delta_0$, define $K_\varepsilon = \{\pi \in \mathcal{S} : \pi(A_\delta^{\mathsf{C}}) \le v(\delta)\ \forall\ \delta \le \varepsilon\ \} \ne \emptyset$. By definition, K_ε is tight and is easily seen to be closed, so that it is compact. For any π in $K_\varepsilon^{\mathsf{C}}$, there exists $\delta_\pi \le \varepsilon$ such that $\pi(A_{\delta_\pi}^{\mathsf{C}}) > v(\delta_\pi)$, and hence

$$\begin{aligned}\tilde{f}_1(\pi) &= \textstyle\int_{A_{\delta_\pi}^{\mathsf{C}}} z(y)\pi(dy) + \int_{A_{\delta_\pi}} z(y)\pi(dy) - \alpha_1 \\ &\ge \inf_{y \in A_{\delta_\pi}^{\mathsf{C}}} z(y) v(\delta_\pi) + C - \alpha_1 \ge \tfrac{1}{\delta_\pi} \ge \tfrac{1}{\varepsilon} \quad \forall\ \pi \in K_\varepsilon^{\mathsf{C}},\end{aligned}$$

so that I7 is fulfilled by choosing $\tilde{\beta}_1^\varepsilon = 1$ and all other coefficients equal to 0. An analogous argument can be developed when $g(\pi)$ has the form $\int_{\mathcal{S}} z d\pi$. Essentially, this example shows that condition I7 is a request for "divergence to infinity" for some of the constraint functions $\tilde{f}_i$'s or for $-gl$.

5. Examples

We apply Theorem 4.1 to the study of robustness in the following two simple situations, which aim at illustrating a technique that can be applied to more general cases.

Example 1. ($\mathcal{S}$ compact). According to notation in Sections 2 and 4, assume that, for any i, $X_i|Y_i$ in (2.1) is Bernoulli distributed with parameter Y_i so that

$$p(x;\pi) = \int_{\mathcal{Y}} y^x (1-y)^{1-x} \pi(dy),\ \ x \in \{0,1\},$$

where $\mathcal{Y} = [0,1]$ and $\pi \in \mathcal{S} = \mathcal{P}(\mathcal{Y})$. The likelihood can be expressed simply as

$$l(\pi) = \Big(\int_{\mathcal{Y}} y\pi(dy)\Big)^{\sum_1^r x_i} \cdot \Big(1 - \int_{\mathcal{Y}} y\pi(dy)\Big)^{r - \sum_1^r x_i}, \quad \pi \in \mathcal{S}.$$

Let $\mathcal{S}_1 = \{(1-\alpha)\pi_1 + \alpha\pi_2 : \alpha \in [0,1]\}$ and $\mathcal{S}_2 = \{(1-\alpha)\pi_1 + \alpha\pi_2 : \alpha \in [\frac{1}{4}, \frac{3}{4}]\}$, where π_1 and π_2 are probability measures in $\mathcal{S}$ such that $\int_0^1 y\pi_1(dy) = 1/4$, $\int_0^1 y\pi_2(dy) = 3/8$.

Assume $r = 2$, $x_1 = x_2 = 1$, $m = 3$, and let

$$\tilde{f}_1(\pi) = \int_{\mathcal{Y}} y\pi(dy) - \frac{1}{2}, \quad \tilde{f}_2(\pi) = -\int_{\mathcal{Y}} y\pi(dy) + \frac{1}{2}, \quad \tilde{f}_3(\pi) = \frac{1}{8} - I_{\mathcal{S}_2}(\pi).$$

This means that q belongs to Γ if, and only if,

$$P(X_1 = 1) = \int_{\mathcal{S}} f_1(\pi) q(d\pi) = \int_{\mathcal{S}} \int_0^1 y\pi(dy) q(d\pi) = \frac{1}{2} \text{ and } q(\mathcal{S}_2) \geq \frac{1}{8}.$$

Finally, consider

$$g(\pi) = I_{\mathcal{S}_1}(\pi),$$

assuming we are interested in the supremum of the posterior probability that $\tilde{\pi}$ belongs to $\mathcal{S}_1$.

Since $\mathcal{S}$ is compact by the compactness of $\mathcal{Y}$, in order to apply Theorem 4.1 it is sufficient to verify I1, I2 and I6. Note that $\mathcal{M}_1(\mathcal{S})$ is a nonempty set as, for example, it contains the measure with atoms $(\pi_1 + \pi_2)/2$ and $\delta_{3/4}$ and weights, respectively, 0.1696 and 0.1272. Moreover, the function $\tilde{f}_3$ is l.s.c. since $\mathcal{S}_2$ is closed in $\mathcal{S}$, g is u.s.c. since $\mathcal{S}_1$ is closed in $\mathcal{S}$, while $\tilde{f}_1$ is continuous and bounded and l is continuous and positive on $\mathcal{S}$. By Remark 1, condition I6 is verified since, if $l(\pi) \leq \varepsilon < \frac{1}{4}$, then $\tilde{f}_2(\pi) \geq \frac{1}{2} - \varepsilon^{1/2} > 0$, so that $\sum \tilde{\beta}_i \tilde{f}_i(\pi) > 1$ for $\tilde{\beta}_1 = \tilde{\beta}_3 = 0$, $\tilde{\beta}_2 \geq (1/2 - \varepsilon^{1/2})^{-1}$ and $\tilde{\beta}_0 > (1 + \tilde{\beta}_2/2)/\varepsilon$. Observe that, if $\pi \in \mathcal{S}_1$, then $\int_0^1 y\pi(dy) = \frac{1}{4} + \frac{\lambda}{8}$ for some λ in [0,1]; if $\pi \notin \mathcal{S}_1$, denote $\int_0^1 y\pi(dy)$ by z. Therefore, by Theorem 4.1, the solution of (2.3)-(2.4) is the solution of the LSIP problem

$$\beta_0^* = \inf \beta_0$$

$$(\beta_0 - 1)\left(\tfrac{1}{4} + \tfrac{\lambda}{8}\right)^2 + b\left(-\tfrac{1}{4} + \tfrac{\lambda}{8}\right) - \tfrac{7}{8}\beta_3 \geq 0, \qquad \text{for all } \tfrac{1}{4} \leq \lambda \leq \tfrac{3}{4}\ ,$$

$$(\beta_0 - 1)\left(\tfrac{1}{4} + \tfrac{\lambda}{8}\right)^2 + b\left(-\tfrac{1}{4} + \tfrac{\lambda}{8}\right) + \tfrac{1}{8}\beta_3 \geq 0, \qquad \text{for all } 0 \leq \lambda < \tfrac{1}{4} \text{ or } \tfrac{3}{4} < \lambda \leq 1, \tag{5.1}$$

$$\beta_0 z^2 + b\left(z - \tfrac{1}{2}\right) + \tfrac{1}{8}\beta_3 \geq 0, \text{ for all } z \in [0,1],$$

$$0 \leq \beta_0 \leq 1, b = \beta_1 - \beta_2 \in \mathbb{R}, \beta_3 \in \mathbb{R}_+.$$

As $0 \leq \beta_0 \leq 1$, the first set of inequalities, when $\lambda = \frac{3}{4}$, yields $b \leq 0$. Therefore, $(\beta_0 - 1)\left(\frac{1}{4} + \frac{\lambda}{8}\right)^2 + b\left(-\frac{1}{4} + \frac{\lambda}{8}\right) - \frac{7}{8}\beta_3$, as a function of λ, is nonincreasing on [0,1], so that the first and the second sets of inequalities hold if, and only if, they hold for $\lambda = \frac{3}{4}$ and $\lambda = 1$, respectively. As far as the third condition is concerned, it can be clearly seen that it holds when $2\beta_0 + b \geq 0$; if $2\beta_0 + b < 0$, it holds if, and only if, $\beta_0 + \frac{b}{2} + \frac{1}{8}\beta_3 \geq 0$. Summing up, solving (5.1) is equivalent to solve the LP problems

$$\begin{array}{ll}
\min \beta_0 & \min \beta_0 \\
(\beta_0 - 1)\left(\frac{11}{32}\right)^2 + b\left(-\frac{5}{32}\right) - \frac{7}{8}\beta_3 \geq 0 & (\beta_0 - 1)\left(\frac{11}{32}\right)^2 + b\left(-\frac{5}{32}\right) - \frac{7}{8}\beta_3 \geq 0 \\
(\beta_0 - 1)\left(\frac{9}{64}\right)^2 + b\left(-\frac{1}{8}\right) + \frac{1}{8}\beta_3 \geq 0 & (\beta_0 - 1)\left(\frac{9}{64}\right)^2 + b\left(-\frac{1}{8}\right) + \frac{1}{8}\beta_3 \geq 0 \\
2\beta_0 + b \geq 0 & 2\beta_0 + b < 0 \\
0 \leq \beta_0 \leq 1 & \beta_0 + \frac{1}{2}b + \frac{1}{8}\beta_3 \geq 0 \\
b \leq 0 & 0 \leq \beta_0 \leq 1 \\
\beta_3 \geq 0, & b \leq 0 \\
 & \beta_3 \geq 0.
\end{array}$$

The corresponding solutions are, respectively, $\beta_0 = 0.3483$, $b = -0.6967$, $\beta_3 = 0.0364$, and $\beta_0 = 0.3454$, $b = -0.6999$, $\beta_3 = 0.0366$, so that $\beta_0^* = 0.3454$. As observed in Section

4, the optimal measure μ^* is discrete with at most $m+1$ points $\{\pi_1^*, \ldots, \pi_{m+1}^*\}$ in its support. This latter set can be found via (4.2) with $\beta_0^* = 0.3454$, $b^* = \beta_2^* - \beta_1^* = 0.6999$, $\beta_3^* = 0.0366$; the set of the corresponding weights is determined solving the finite system given by (4.3) together with $\int l d\mu^* = 1$ as seen in Section 4. It turns out that

$$\mu^* = 0.6546\,\delta_{\pi_1^*} + 0.3967\,\delta_{\pi_2^*} + 2.1226\,\delta_{\pi_3^*},$$

where $\pi_1^* = \delta_1$, $\pi_2^* = \frac{1}{4}\pi_1 + \frac{3}{4}\pi_2$ and $\pi_3^* = \pi_2$.

To complete the analysis of robustness, we need to determine

$$\inf_{q\in\Gamma} \frac{\int_{\mathcal{S}} I_{\mathcal{S}_1}(\pi) l(\pi) q(d\pi)}{\int_{\mathcal{S}} l(\pi) q(d\pi)} = -\sup_{q\in\Gamma} \frac{\int_{\mathcal{S}} -I_{\mathcal{S}_1}(\pi) l(\pi) q(d\pi)}{\int_{\mathcal{S}} l(\pi) q(d\pi)}. \tag{5.2}$$

Unfortunately $-I_{\mathcal{S}_1}$ is not u.s.c., as $\mathcal{S}_1$ is a closed set, so that we cannot apply directly Theorem 4.1. However, denoting the interior of $\mathcal{S}_1$ by $\mathcal{S}_1'$, $-I_{\mathcal{S}_1'}$ is u.s.c., and

$$\inf_{q\in\Gamma} \frac{\int_{\mathcal{S}} I_{\mathcal{S}_1}(\pi) l(\pi) q(d\pi)}{\int_{\mathcal{S}} l(\pi) q(d\pi)} \geq \inf_{q\in\Gamma} \frac{\int_{\mathcal{S}} I_{\mathcal{S}_1'}(\pi) l(\pi) q(d\pi)}{\int_{\mathcal{S}} l(\pi) q(d\pi)}.$$

We will determine the right hand-side in the above inequality and show that it is equal to the left hand-side.

We note that condition I6 is satisfied here too. Using the same argument as before, we obtain an LP problem having solution $\beta_0^* = -0.0208$, $b^* = -\beta_0$ and $\beta_3^* = 0.0833$, so that

$$\mu^* = 0.6546\,\delta_{\pi_1^*} + 0.0793\,\delta_{\pi_2^*} + 2.1226\,\delta_{\pi_3^*},$$

where $\pi_1^* = \delta_1$, $\pi_2^* = \frac{1}{4}\pi_1 + \frac{3}{4}\pi_2$ and $\pi_3^* = \pi_2$. Since the atoms of μ^* do not belong to $\mathcal{S}_1 \setminus \mathcal{S}_1'$, then μ^* is an optimal measure for (5.2) too. Summing up, for all q in Γ, it is

$$0.0208 \leq \frac{\int_{\mathcal{S}} I_{\mathcal{S}_1}(\pi) l(\pi) q(d\pi)}{\int_{\mathcal{S}} l(\pi) q(d\pi)} \leq 0.3454.$$

Example 2. ($\mathcal{S}$ not compact). We assume that $X_1, \ldots, X_r$ are i.i.d., given π, with density

$$p(x;\pi) = \int_{\mathcal{Y}} \sum_{k=1}^{y} \lambda_k^{(y)} k\mathrm{e}^{-kx} \pi(dy), \quad x > 0, \tag{5.3}$$

where $\mathcal{Y} = [1, +\infty)$, $\mathcal{S} = \mathcal{P}(\mathbb{N})$, and $\lambda_1^{(y)}, \ldots, \lambda_y^{(y)}$ are fixed probability weights for all $y = 1, 2, \ldots$, i.e. $X_i|Y_i$ is distributed according to a mixture of y exponentials.

We consider $g(\pi) = I_{\mathcal{S}_1}(\pi)$, with $\mathcal{S}_1 = \{\delta_1\}$, and

$$\tilde{f}_1(\pi) = \int_{\mathcal{Y}} y\pi(dy) - \alpha_1, \quad \tilde{f}_2(\pi) = -\pi(\{2\}) - \pi(\{3\}) - \alpha_2,$$

where α_1, α_2 are fixed real constants, $\alpha_1 \geq 1$, $-1 \leq \alpha_2 \leq 0$, i.e., we are interested here in the supremum of the posterior probability that $X_1, \ldots, X_r$ is a sample from the exponential distribution with parameter equal to 1, under the constraints

$$E_q\Big(\int_{\mathcal{Y}} y\tilde{\pi}(dy)\Big) \leq \alpha_1, \quad E_q\Big(\tilde{\pi}(\{2\}) + \tilde{\pi}(\{3\})\Big) \geq \alpha_2.$$

Observe that a more general example would result assuming that the weights $\lambda_1^{(y)}, \ldots, \lambda_y^{(y)}$ in (5.3) are random as well, so that π should represent the distribution of $(y, \lambda_1^{(y)}, \ldots, \lambda_y^{(y)})$. However, in this framework, fixing $\lambda_1^{(y)}, \ldots, \lambda_y^{(y)}$ for all $y = 1, 2, \ldots$ in (5.3) corresponds to averaging with respect to a fixed distribution for $(\lambda_1^{(y)}, \ldots, \lambda_y^{(y)})$ given y.

In this example, $\mathcal{S} = \mathcal{P}(\mathbb{N})$ is not compact, but it is easily seen to be closed. Therefore, due to the arguments introduced at the end of Section 4, I7 holds since f_1 has the form $\int_{\mathcal{Y}} z(y)\pi(dy)$. Moreover, $\tilde{f}_1$ is l.s.c. (see Remark 3), $\tilde{f}_2$ and l are continuous on $\mathcal{S}$, g is u.s.c. since $\mathcal{S}_1$ is trivially closed, and it is easy to verify that I3, I4 and I5 hold. As far as I6 is concerned, simple calculations show that the condition is satisfied, applying Remark 1 with $\tilde{\beta}_3 = \tilde{\beta}_1 = 0$ and $\tilde{\beta}_0$, $\tilde{\beta}_2$ large enough. Finally, it can be checked that, if $-\alpha_2 \leq \alpha_1 - 1$, then $\mathcal{M}_1(\mathcal{S}) \neq \emptyset$. Therefore, the required maximum is equal to

$$\inf\{\beta_0 : 0 \leq \beta_0 \leq 1, \beta_1, \beta_2 \geq 0, \beta_0 l(\pi) + \beta_1 \tilde{f}_1(\pi) + \beta_2 \tilde{f}_2(\pi) \geq g(\pi) l(\pi) \; \forall \pi \in \mathcal{S}\}. \tag{5.4}$$

Now, assuming $r = 1$, and denoting $\pi(\{i\})$ by π_i, the feasible region is given by

$$(\beta_0 - 1)e^{-x_1} + \beta_1(1 - \alpha_1) + \beta_2(-\alpha_2) \geq 0 \tag{5.5}$$

$$\beta_0 \sum_{=1}^{+\infty} \pi_i \Big(\sum_{k=1}^{i} \lambda_k^{(i)} k\mathrm{e}^{-kx_1} \Big) + \beta_1 \Big(\sum_{1}^{+\infty} i\pi_i - \alpha_1 \Big) + \beta_2(-\pi_2 - \pi_3 - \alpha_2) \geq 0, \tag{5.6}$$

where $0 \leq \pi_1 < 1$, $0 \leq \sum_1^{+\infty} \pi_i \leq 1$ in (6.6), which is linear in $\pi := \{\pi_1, \pi_2, \ldots\}$. Evaluating (5.6) at the simplex vertices, together with (5.5), gives the following equivalent formulation of (5.4)

$\min_{0 \leq \beta_0 \leq 1} \beta_0$

$\beta_1 \geq 0$, $\beta_2 \geq 0$

$(\beta_0 - 1)e^{-x_1} + \beta_1(1 - \alpha_1) + \beta_2(-\alpha_2) \geq 0$

$\beta_0(\lambda_1^{(2)}\mathrm{e}^{-x_1} + \lambda_2^{(2)}2\mathrm{e}^{-2x_1}) + \beta_1(2 - \alpha_1) + \beta_2(-1 - \alpha_2) \geq 0$

$\beta_0(\lambda_1^{(3)}\mathrm{e}^{-x_1} + \lambda_2^{(3)}2\mathrm{e}^{-2x_1} + \lambda_3^{(3)}3\mathrm{e}^{-3x_1}) + \beta_1(3 - \alpha_1) + \beta_2(-1 - \alpha_2) \geq 0.$

For example, if $\alpha_1 = 1.5$, $-\alpha_2 = 0.5$, $x_1 = 4.6$ and $\lambda_1^{(2)} = 0.2$, $\lambda_1^{(3)} = 0.01$, $\lambda_2^{(3)} = 0.02$, we found that the infimum is equal to $\beta_0^* = 0.8223$, while, if $\alpha_1 = 3.5$, $-\alpha_2 = 0.75$, $x_1 = 0.6931$ and $\lambda_1^{(2)} = 0.5$, $\lambda_1^{(3)} = \frac{1}{3}$, $\lambda_2^{(3)} = \frac{1}{3}$, then $\beta_0^* = 0.2667$.

To complete the example, we note that the infimum in (5.2) is equal to 0 in this case; indeed, there exists a degenerate measure $\hat{q} = \delta_{\hat{\pi}}$ belonging to Γ, where the support of $\hat{\pi}$ is contained in $\{1, 2, 3\}$ and $\hat{\pi}(\{1\}) < 1$.

Acknowledgements

We wish to thank the Associated Editor for his valuable suggestions, which lead to an improvement in the organization of the paper.

Appendix A

Let $\mathcal{S}$ be a metric space, endowed with the class $\mathcal{B}(\mathcal{S})$ of Borel sets in $\mathcal{S}$. Let $\mathcal{P}(\mathcal{S})$ be the space of all probability measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, with the topology of weak convergence: $\pi_n \xrightarrow{w} \pi$ in $\mathcal{P}(\mathcal{S})$ if $\int_{\mathcal{S}} f d\pi_n \to \int_{\mathcal{S}} f d\pi$ for every bounded, continuous real function f on $\mathcal{S}$. Let $\mathcal{M}(\mathcal{S})$ be the space of all finite positive measures on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, equipped with the topology of weak convergence as well (see Parthasaraty, 1967, p. 40). Over $\mathcal{P}(\mathcal{S})$ and $\mathcal{M}(\mathcal{S})$ we consider the corresponding Borel σ-fields. It is well known that $\mathcal{P}(\mathcal{S})$ and $\mathcal{M}(\mathcal{S})$, both endowed with the topology of weak convergence, are separable metric spaces if, and only if, $\mathcal{S}$ is separable; distances metrizing the topology of weak convergence in $\mathcal{P}(\mathcal{S})$ and $\mathcal{M}(\mathcal{S})$ are the Prohorov distance and the one described in Doob (1994), p. 139-140, respectively. Moreover, observe that if $\mathcal{S}$ is a Polish space then both $\mathcal{P}(\mathcal{S})$ and $\mathcal{M}(\mathcal{S})$ are Polish in their corresponding topologies (see Prohorov, 1956).

When dealing with finite measures on a topological space, an important notion is the *tightness* of a family of finite measures. A subset A of $\mathcal{M}(\mathcal{S})$ is *tight* if for all $\varepsilon > 0$ there exists a compact set K_ε in $\mathcal{S}$ such that $\mu(K_\varepsilon^{\complement}) < \varepsilon$ for each μ in A (Ash, 1972, p. 330). Recalling that a subset A of a metric space is *relatively compact* if every sequence of elements in A contains a weakly convergent subsequence, the following theorem states that tightness is a sufficient condition for relative compactness of a set of uniformly bounded measures.

THEOREM A.1 *Let $\mathcal{S}$ be separable. If $A \subset \mathcal{M}(\mathcal{S})$ is tight and* $\sup_{\mu \in A} \mu(\mathcal{S}) \leq C$ *for some positive constant C, then A is relatively compact.*

This is a part of the well-known Prohorov's theorem; see, for instance, Ash (1972), where the result is stated when $\mathcal{S} = \mathbb{R}$.

Appendix B

PROOF OF THEOREM 3.1. The "only if" part is obvious. Indeed, if $\mathcal{M}_1 \neq \emptyset$, there exists a finite measure $\tilde{\mu}$ in $\mathcal{M}_1$; therefore, if we choose $\beta \in \mathrm{I\!R}^n_+$ such that $\sum_{i=1}^n \beta_i h_i(s) \geq 0$ for any $s \in \mathcal{S}$, we have

$$0 \leq \int_{\mathcal{S}} \sum_{i=1}^n \beta_i h_i(s) \tilde{\mu}(ds) \leq \sum_{i=1}^n \beta_i \eta_i.$$

We prove the "if" part in several steps. We show that the measures in $\mathcal{M}_1$ are uniformly bounded (Step 1), so that $\mathcal{M}_1$ is tight (Step 2). As $\mathcal{M}_1$ is contained in a convex and compact subset too, we will provide a necessary and sufficient condition for $\mathcal{M}_1$ to be not empty (Step 3) that is implied by (4.2) (Step 4).

STEP 1. $\mathcal{M}_1 \subset \mathcal{M}_D := \{\mu \in \mathcal{M}(\mathcal{S}) : \mu(\mathcal{S}) \leq D\}$ for some positive D.

Indeed, if $\mu \in \mathcal{M}_1$, by H3,

$$\mu(\mathcal{S}) = \int_{\mathcal{S}} \mu(ds) \leq \int_{\mathcal{S}} \sum_{i=1}^n \tilde{\beta}_i h_i(s) \mu(ds) \leq \sum_{i=1}^n \tilde{\beta}_i \eta_i;$$

therefore, if $D := 1 + \sum_{i=1}^n \tilde{\beta}_i \eta_i$, then $D > 0$ and $\mu(\mathcal{S}) \leq D$. Of course, if $\mathcal{M}_1 = \emptyset$, trivially $\mathcal{M}_1 \subset \mathcal{M}_D$ for all $D > 0$.

STEP 2. $\mathcal{M}_1$ is tight.

This is obvious when $\mathcal{M}_1 = \emptyset$. Otherwise, let $\mu \in \mathcal{M}_1$ and $\sigma > 0$; by H2 and H4 we have

$$\begin{aligned}\textstyle\sum_{i=1}^n \beta_i^\sigma \eta_i &\geq \textstyle\int_{\mathcal{S}} \sum_{i=1}^n \beta_i^\sigma h_i(s)\mu(ds) = \sum_{i=1}^n \beta_i^\sigma \int_{K_\sigma} h_i(s)\mu(ds) + \int_{K_\sigma^{\complement}} \sum_{i=1}^n \beta_i^\sigma h_i(s)\mu(ds) \\ &\geq \textstyle\sum_{i=1}^n \beta_i^\sigma C_i \mu(K_\sigma) + \frac{1}{\sigma}\mu(K_\sigma^{\complement}) = \sum_{i=1}^n \beta_i^\sigma C_i \mu(\mathcal{S}) + \Big(\frac{1}{\sigma} - \sum_{i=1}^n \beta_i^\sigma C_i\Big)\mu(K_\sigma^{\complement}),\end{aligned}$$

so that

$$\mu(K_\sigma^{\complement}) \leq \frac{\sum_{i=1}^n \beta_i^\sigma \eta_i - \sum_{i=1}^n \beta_i^\sigma C_i \mu(\mathcal{S})}{\frac{1}{\sigma} - \sum_{i=1}^n \beta_i^\sigma C_i} \leq \sigma \sum_{i=1}^n \beta_i^\sigma (\eta_i - C_i \mu(\mathcal{S})).$$

By H4

$$\sum_{i=1}^{n} \beta_i^\sigma (\eta_i - C_i \mu(\mathcal{S})) \leq \sum_{i=1}^{n} \beta_i^\sigma (\eta_{max} - C_{min} D) \leq \max\{0, \eta_{max} - C_{min} D\} B$$

where $\eta_{max} := \max_i \eta_i$ and $C_{min} := \min_i C_i$, and this yields

$$\mu(K_\sigma^{\complement}) \leq \sigma B \max\{0, \eta_{max} - C_{min} D\} =: \delta(\sigma) \to 0, \text{ for } \sigma \to 0,$$

i.e. $\mathcal{M}_1$ is tight.

STEP 3. $\mathcal{M}_1 \neq \emptyset$ if, and only if, the following condition holds:

$$\inf_{\mu \in \mathcal{N}_D} \int_{\mathcal{S}} \sum_{i=1}^{n} \beta_i h_i(s) \mu(ds) \leq \sum_{i=1}^{n} \beta_i \eta_i, \text{ for all } \beta \in \mathbb{R}_+^n, \tag{B.1}$$

where $\mathcal{N}_D := \{\mu \in \mathcal{M}_D : \mu(K_\sigma^{\complement}) \leq \delta(\sigma) \ \forall \sigma > 0\}$, and $\mathcal{M}_D$ and $\delta(\sigma)$ are as in previous Steps.

Indeed, first of all, observe that $\mathcal{M}_1 \subset \mathcal{N}_D$; moreover, if $\mathcal{N}_\sigma := \{\mu \in \mathcal{M}_D : \mu(K_\sigma^{\complement}) \leq \delta(\sigma)\}$, it is $\mathcal{N}_D = \cap_{\sigma > 0} \mathcal{N}_\sigma$. We prove that, for all $\sigma > 0$, if $\{\mu_n\}_n \subset \mathcal{N}_\sigma$, $\mu_n \xrightarrow{w} \mu$, then $\mu \in \mathcal{N}_\sigma$ and, hence, $\mathcal{N}_\sigma$ is closed. By Lemma 3 in Kemperman (1983), for each n, $\sigma > 0$, $\mu_n(K_\sigma^{\complement}) = \int_{\mathcal{S}} I_{K_\sigma^{\complement}} \, d\mu = \sup\{\int_{\mathcal{S}} f d\mu_n : f$ bounded and continuous, $f \leq I_{K_\sigma^{\complement}}\}$, since $I_{K_\sigma^{\complement}}$ is lower semicontinuous and bounded from below. Hence, if f is bounded and continuous and $f \leq I_{K_\sigma^{\complement}}$, it is

$$\int_{\mathcal{S}} f d\mu = \lim_{n \to +\infty} \int_{\mathcal{S}} f d\mu_n \leq \limsup_{n \to +\infty} \int_{\mathcal{S}} I_{K_\sigma^{\complement}} \, d\mu_n \leq \delta(\sigma)$$

since $\mu_n \in \mathcal{N}_\sigma$, so that

$$\mu(K_\sigma^{\complement}) \leq \delta(\sigma),$$

i.e. $\mu \in \mathcal{N}_\sigma$. Therefore, $\mathcal{N}_D$ is closed and, since it is tight, it is compact too by Theorem A.1. As $\mathcal{M}_1 \subset \mathcal{N}_D$ and $\mathcal{N}_D$ is convex, applying Theorem 1 in Kemperman (1983) with $\mathcal{M}_0 = \mathcal{N}_D$, we obtain that (B.1) is a necessary and sufficient condition for $\mathcal{M}_1$ to be nonempty.

Step 4. (3.3) yields (B.1), concluding our proof.

Introducing the notation $\eta = (\eta_1 \ldots, \eta_n)'$, $h_s = (h_1(s), \ldots, h_n(s))'$, (3.3) can be rewritten as

$$h_s'\beta \geq 0, \beta \geq 0 \Rightarrow \eta'\beta \geq 0, \beta \geq 0$$

where y' denotes the transpose of the vector y. By Farkas' lemma (see Goberna and López, 1998, p. 71) this is true if, and only if, η belongs to the closure of the convex cone generated by $\{h_s, s \in \mathcal{S}\}$ and the canonical basis in $\mathbb{R}^n$. It is easy to show (see Goberna and López, 1998, Exercise 3.5(iii), p. 76) that η belongs to the closure of this cone if, and only if, there exists a vector $\nu \in \mathbb{R}^n$ such that for any $\varepsilon > 0$ there exist constants $\mu_s^\varepsilon \geq 0$ (all but finitely many, say k, equal to 0) and a vector $\lambda_\varepsilon \in \mathbb{R}_+^n$ such that

$$\lambda_\epsilon + \sum_s \mu_s^\epsilon h_s = \eta + \epsilon\nu. \tag{B.2}$$

Multiplying (5.) by $\beta \in \mathbb{R}_+^n$, we obtain

$$\sum_s \mu_s^\varepsilon h_s'\beta \leq \lambda_\varepsilon'\beta + \sum_s \mu_s^\varepsilon h_s'\beta = \eta'\beta + \varepsilon\nu'\beta, \tag{B.3}$$

and the left hand-side can be written as $\int_{\mathcal{S}} h_s'\beta\bar{\mu}_\varepsilon(ds)$, where $\bar{\mu}_\varepsilon$ is the measure on $\mathcal{S}$ with masses $\mu_{s_1}^\varepsilon, \ldots, \mu_{s_k}^\varepsilon$ at some points $s_1, \ldots, s_k$ in $\mathcal{S}$, respectively. If $\bar{\mu}_\varepsilon \in \mathcal{N}_D$, then by (B.3) we have

$$\inf_{\mu \in \mathcal{N}_D} \int_{\mathcal{S}} h_s'\beta\mu(ds) \leq \eta'\beta + \varepsilon\nu'\beta \to \eta'\beta \text{ as } \varepsilon \to 0,$$

i.e. condition (B.1) holds, so that $\mathcal{M}_1 \neq \emptyset$.

It remains to show that $\bar{\mu}_\varepsilon \in \mathcal{N}_D$. It is sufficient to consider $\varepsilon < 1$.

By H3 e (B.3) it is

$$\bar{\mu}_\varepsilon(\mathcal{S}) = \sum_s \mu_s^\varepsilon \leq \sum_s \mu_s^\varepsilon h_s'\tilde{\beta} \leq \eta'\tilde{\beta} + \varepsilon\nu'\tilde{\beta} \leq \eta'\tilde{\beta} + \nu'\tilde{\beta}.$$

Redefining $D = \max\{D, \eta'\tilde{\beta} + \nu'\tilde{\beta}\}$, it is $\bar{\mu}_\varepsilon \in \mathcal{M}_D$. Multiplying (B.2) by β^σ, by H2 and H4 we have

$$\eta'\beta^\sigma + \varepsilon\nu'\beta^\sigma \geq \sum_s \mu_s^\varepsilon h_s'\beta^\sigma = \sum_{s\in K_\sigma} \mu_s^\varepsilon h_s'\beta^\sigma + \sum_{s\notin K_\sigma} \mu_s^\varepsilon h_s'\beta^\sigma \geq \sum_{s\in K_\sigma} \mu_s^\varepsilon h_s'\beta^\sigma + \frac{1}{\sigma}\sum_{s\notin K_\sigma} \mu_s^\varepsilon$$

$$\geq \sum_{s\in K_\sigma} \mu_s^\varepsilon C_{min}(\textstyle\sum_j \beta_j^\sigma) + \frac{1}{\sigma}\sum_{s\notin K_\sigma} \mu_s^\varepsilon \geq BC_{min}\sum_{s\in K_\sigma} \mu_s^\varepsilon + \frac{1}{\sigma}\bar{\mu}_\epsilon(K_\sigma^\complement)$$

$$\geq BC_{min}\bar{\mu}_\varepsilon(\mathcal{S}) + \frac{1}{\sigma}\bar{\mu}_\epsilon(K_\sigma^\complement) \geq BC_{min}D + \frac{1}{\sigma}\bar{\mu}_\varepsilon(K_\sigma^\complement).$$

Hence,

$$\bar{\mu}_\varepsilon(K_\sigma^\complement) \leq \sigma(\eta'\beta^\sigma + \varepsilon\nu'\beta^\sigma - BC_{min}D) \leq \sigma\textstyle\sum_j(\eta_j + \varepsilon\nu_j)\beta_j^\sigma - \sigma BC_{min}D$$

$$\leq \sigma(\nu_{max} + \eta_{max})B - \sigma BC_{min}D \leq \sigma B(\max\{0, \eta_{max} + \nu_{max} - C_{min}D\} = \delta^*(\sigma).$$

Redefining $\delta(\sigma)$ as $\max\{\delta(\sigma), \delta^*(\sigma)\}$ we have that $\bar{\mu}_\varepsilon \in \mathcal{N}_\sigma$ for all $\sigma > 0$ and, hence, $\bar{\mu}_\varepsilon \in \mathcal{N}_D$.

References

Ash, R. B. (1972). *Real Analysis and Probability*, Academic Press, San Diego.

Berger, J.O., Rios Insua D. and Ruggeri, F. (2000). Bayesian robustness, *Robust Bayesian analysis, Lecture Notes in Statistics*, (eds. D. Rios Insua, F. Ruggeri),Vol. 152, 1–32, Springer, New York.

Betrò, B. (2004). An Accelerated Central Cutting Plane Algorithm for Semi-Infinite Linear Programming. *Mathematical Programming*, series A, **101**, 479-495.

Betrò, B. and Guglielmi, A. (1994). An algorithm for robust Bayesian analysis under generalized moment conditions. Quaderno IAMI, 94.6, CNR-IAMI, Milan.

Betrò, B. and Guglielmi, A. (1997). Numerical robust Bayesian analysis under generalized moment conditions. *Bayesian Robustness, IMS Lecture Notes - Monograph Series*, (eds. J.O. Berger, B. Betrò, E. Moreno, L.R. Pericchi, F. Ruggeri, G. Salinetti, L. Wasserman), 3-20, IMS, Hayward.

Betrò, B. and Guglielmi, A. (2000). Methods for global prior robustness under generalized moment conditions. *Robust Bayesian analysis, Lecture Notes in Statistics*, Vol. 152, (eds. D. Rios Insua, F. Ruggeri), 273–294, Springer, New York.

Betrò B., Bodini, A. and Guglielmi, A. (2002). Generalized Moment Theory and Robust Bayesian Analysis: the Nonparametric Case. Technical Report 02-06, CNR-IMATI, Milan.

Betrò B., Guglielmi, A. and Rossi, F. (1996). Robust Bayesian Analysis for the Power Law Process. *ASA 1996 Proceedings of the Section on Bayesian Statistical Science*, 288-291, ASA, Alexandria.

Betrò, B., Męczarski, M. and Ruggeri, F. (1994). Robust Bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference*, **41**, 257-266.

Dall'Aglio, M. (1995). Problema dei Momenti e Programmazione Lineare Semi-infinita nella Robustezza Bayesiana. Ph.D. Thesis, Dip. Statistica, Probabilità e Statistiche Applicate, Università "La Sapienza", Roma.

Doob, J.L. (1994). *Measure Theory.* Springer-Verlag. New York.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, **90**, 577-588.

Escobar, M. D. and West, M. (1998). Computing nonparametric hierarchical models, *Practical nonparametric and semiparametric Bayesian statistics* (eds. Dey D., Muller P., Sinha D.), 1-22, Springer, New York.

Ghosh J.K. and Ramamoorthi R.V. (2003). *Bayesian Nonparametrics*, Springer, New York.

Goberna, M.A. and López, M.A. (1998). *Linear Semi-Infinite Optimization*, Wiley, New York.

Goutis, C. (1994). Ranges of posterior measures for some classes of priors with specified moments, *International Statistical Review*, **62**, 245-357.

Hoff, P.D. (2003). Nonparametric estimation of convex models via mixtures. *The Annals of Statistics*, **31**, 174-200.

Kemperman, J.H.B. (1971). On a class of moment problems. *Proceedings Sixth Berkeley Symposium on Mathematical Statistics and Probability*, (eds. L.M. LeCam, J. Neyman, and E.L. Scott), Vol. 2, 101-106, University of California Press, Berkeley.

Kemperman, J.H.B. (1983). On the role of duality in the theory of moments, *Lecture Notes in Economics and Mathematical Systems*, Vol. 215, 63-92, Springer-Verlag, Berlin.

Kemperman, J.H.B. (1987). Geometry of the moment problem, *Moments in mathematics, Proceedings of Symposia in Applied Mathematics*, (eds. Landau H.J.), 16-53, AMS, Providence.

Lijoi, A., Mena, R. and Prünster, I. (2005). Hierarchical mixture modeling with normalized inverse-Gaussian priors. *Journal of the Americal Statistical Association*. To appear.

Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* **12**, 351-357.

Nieto-Barajas, L.E., Prünster, I., and Walker, S.G. (2004). Normalized random measures driven by increasing additive processes. *The Annals of Statistics* **32**, 2343-2360.

Parthasarathy, K.R. (1967). *Probability measures on metric spaces*, Academic Press, San Diego.

Petrone, S. (1999). Bayesian density estimation using Bernstein polynomials. *Canadian Journal of Statistics*, **26**, 373-393.

Prohorov, Yu.V. (1956). Convergence of random processes and limit theorems in probability theory, *Theory of Probability and its Applications*, **1**, 157-214.

Rios Insua D. and Ruggeri, F. (Eds.) (2000). *Robust Bayesian analysis*, Lecture Notes in Statistics, Vol. 152, Springer, New York.

Rogosinsky, W.W. (1958). Moments of non-negative mass, *Proceedings of the Royal Society of London*, Ser. A, **245**, 1-27.

Smith, J.E. (1995). Generalized Chebychev Inequalities: Theory and Applications in Decision Analysis. *Operations Research*, **43**, 807-825.