RACSAM
Rev. R. Acad. Cien. Serie A. Mat.
VOL. **97** (3), 2003, pp. 407–412
Estadística e Investigación Operativa / Statistics and Operations Research
Comunicación Preliminar / Preliminary Communication

Optimal alternative robustness in Bayesian Decision Theory

Fabrizio Ruggeri, Jacinto Martín and David Ríos Insua

Abstract. In Martin et al (2003), we suggested an approach to general robustness studies in Bayesian Decision Theory and Inference, based on ϵ-contamination neighborhoods. In this note, we generalise the results considering neighborhoods based on norms, specifically, the supremum norm for utilities and the total variation norm for probability distributions. We provide tools to detect changes in preferences between alternatives under perturbations of the prior and/or the utility and the most sensitive direction.

Robustez de preferencias en Teoría de la Decisión Bayesiana

Resumen. En Martin et al (2003) propusimos una aproximación a estudios de robustez en análisis bayesiano, basada en entornos ϵ-contaminados. En esta nota, generalizamos los resultados considerando entornos basados en normas, empleando, específicamente, la norma del supremo para las utilidades y la norma de variación total para las distribuciones de probabilidad. Proporcionamos herramientas para determinar cuánto podemos perturbar la utilidad o probabilidad originales sin que cambie la alternativa óptima y la dirección de perturbación más sensible.

1. Introduction

We consider the standard Bayesian decision theoretical framework, see e.g. French and Ríos Insua (2000) for a review. A decision maker (DM) makes decisions $a \in \mathcal{A}$, the space of alternatives. We associate a consequence $c \in \mathcal{C}$ with each pair (a, θ), where θ denotes the state of nature. We model the DM's beliefs about the states $\theta \in \Theta$ with a probability distribution π, which is updated to the posterior $\pi(\cdot|x)$ in presence of additional information x provided by an experiment with likelihood $l(x|\theta)$. We also model his preferences over consequences with a utility function u, and we associate to each alternative a its posterior expected utility

$$T(u, \pi, a) = \frac{\int u(a, \theta) l(x|\theta) \pi(\theta) d\theta}{\int l(x|\theta) \pi(\theta) d\theta} = \frac{N(u, \pi, a)}{D(\pi)}. \tag{1}$$

The optimal alternative a^* maximises $T(u, \pi, a)$ in a. Since the output a^* of the analysis depends on the inputs π and u, the DM may demand ways of checking the impact of these inputs on the output. This is the motivation for much of the recent work in Bayesian robustness, see Ros Insua and Ruggeri (2000) for a review, and our interest here.

Presentado por Sixto Ríos.
Recibido: 14 de Enero de 2004. *Aceptado*: 5 de Mayo de 2004.
Palabras clave / Keywords: Statistical Decision Theory, Decision Analysis, Sensitivity Analysis, Class of utilities, Class of priors, Dominance.
Mathematics Subject Classifications: 62F15, 62C10

Specifically, we are interested here in decision robustness, within Bayesian decision theory: our objective is to find which changes in prior distributions or utility functions produce a change in the optimal alternative, generalising work in Martin et al (2003). Our starting point is the comparison of the optimal alternative with possible competitors. We assume that, for two alternatives a and b, we have

$$T(u, \pi, b) \leq T(u, \pi, a), \tag{2}$$

suggesting $b \preceq a$ (b at most as preferred as a). We are interested in studying whether this preference of a over b holds when there are changes in u and π. a could be the optimal alternative and b a competitor. Specifically, we address these issues:

- How much can we perturb (u, π) in a certain direction until $\neg(b \preceq a)$?
- Is there a specially sensitive direction, so that the preference dilutes more rapidly, if u and π are perturbed in such direction?

We view (u, π) as the initial assessment of preferences and beliefs to be criticised. Perturbations of (u, π) are constrained to a class $\mathcal{U} \times \Gamma$ of pairs utility function-prior distribution, which model imprecision in beliefs and preferences. For our purposes, the classes have to be convex, without loss of generality. We assume also that utility functions are normalised between 0 and 1. The use of classes for priors and utility functions is standard in robust Bayesian analysis.

In Section 2 we introduce basic definitions. Sensitivity of preferences with respect to the utility is studied in Section 3. Section 4 refers to prior sensitivity, whereas joint sensitivity is addressed in Section 5. We end up with some conclusions.

2. Basic definitions

We shall analyse properties of the operator

$$T^{ab}(w, P) = T(w, P, a) - T(w, P, b),\ w \in \mathcal{U},\ P \in \Gamma$$

for a and b fixed, since it explains the preference relation between a and b. For example, if $T^{ab}(u, \pi) \geq 0$, a is preferred to b, for the current assessment (u, π), and we aim at criticising such information, when such assessment is perturbed. We assume that $D(P) > 0$ for any $P \in \Gamma$, so we can study

$$N^{ab}(w, P) = N(w, P, a) - N(w, P, b),\ w \in \mathcal{U},\ P \in \Gamma.$$

instead of $T^{ab}(w, P)$

The results shown in Martin et al (2003) apply when considering ε–contaminations of the current assessment (u, π). We can extend the notion of ε–robustness to general classes of priors and utilities in which neighbourhoods of (u, π) are given by spheres in a metric, topological space. To do so, we consider a distance $d(\cdot, \cdot)$ in the space of priors and/or utilities and modify, accordingly, the definitions in Martin et al (2003).

Definition 1 *(u, π) is ε-robust for $b \preceq a$ within $\mathcal{U} \times \Gamma$, if $b \preceq a$ for all $v \in \mathcal{U}$ and $Q \in \Gamma$ such that $d((u, \pi), (v, Q)) \leq \varepsilon$.*

ε–robustness under changes in either the prior or the utility are defined similarly, when we consider only imprecision in preferences and beliefs. We have also the following definition:

Definition 2 *Given (u, π), we say that $(v, Q) \in \mathcal{U} \times \Gamma$ is ε-sensitive for $b \preceq a$, when $b \preceq a$ does not hold for (v, Q) and $d((u, \pi), (v, Q)) = \varepsilon$.*

We aim at finding (v, Q) for which ε is minimum. (v, Q) is the perturbation of (u, π) leading to faster reductions in T^{ab}. We call it *most sensitive.*

Definition 3 *Given $(u,\pi), (v,Q) \in \mathcal{U}\times\Gamma$ is the most sensitive pair for $b \preceq a$ within $\mathcal{U}\times\Gamma$, if it is ε-sensitive and for other ε'-sensitive $(v',Q') \in \mathcal{U}\times\Gamma$, $\varepsilon \leq \varepsilon'$.*

Similarly, we define ε-sensitive and most sensitive utility functions and priors.

A reasonable choice for d is given by the norms: in $\mathcal{U}$, the supremum norm, i.e. $\|u\| = \sup_{c\in\mathcal{C}}|u(c)|$; in $\mathcal{M}$, the total variation norm, i.e., $\|\delta\| = \sup_{A\in\beta}|\delta(A)|$; in $\mathcal{U}\times\mathcal{M}$, $\|(m,\delta)\|_\infty = \max\{\|m\|,\|\delta\|\}$, where $\mathcal{U}$ is the space of utilities and $\mathcal{M}$ is the space of signed measures, which includes the prior probability measures. We should notice that neighbourhoods of the utility u under the supremum norm over $\mathcal{U}$ contain functions which do not fulfill the assumptions of convexity and normalisation between 0 and 1 of the utility. At the same time, neighbourhoods of a probability measure π under the norm in $\mathcal{M}$ contain measures which are not probabilities. We will restrict the neighbourhoods, to get normalised utilities and probability measures and illustrate some results about the most sensitive utility/prior under this more general definition.

3. Utility Sensitivity

We consider the important class $\mathcal{U}$ of all utility functions v such that $v(a,\theta) = v(\theta - a)$, for all $a,\theta \in \Re$. This includes the standard utility functions in statistical decision theory, related with the quadratic loss, the absolute loss and others. As stated, we look for the largest size ε of a neighbourhood around $u(a,\theta)$ which does not reverse the ranking $b \preceq a$, together with the most sensitive utility. We start with unrestricted, topological neighbourhoods.

Proposition 1 *The most sensitive utility in $\mathcal{U}$, under the above conditions, is given by*

$$w(c,\theta) = \begin{cases} u(c,\theta) - \hat{\varepsilon}, & if \quad \Delta(\theta) \geq 0 \\ u(c,\theta) + \hat{\varepsilon}, & if \quad \Delta(\theta) < 0 \end{cases},$$

where $\Delta(\theta) = l(x|\theta+a)\pi(\theta+a) - l(x|\theta+b)\pi(\theta+b)$ and $\hat{\varepsilon} = \dfrac{N^{ab}(u)}{\int|\Delta(t)|dt}$, which is the size of the largest neighbourhood of u in which the ranking is preserved.

PROOF. Let $l(\theta) = l(x|\theta)$. Consider $\int[u(a,\theta) - u(b,\theta)]l(\theta)\pi(\theta)d\theta \geq 0$, and look for the smallest ε such that

$$\inf_v \int [v(a,\theta) - v(b,\theta)]l(\theta)\pi(\theta)d\theta \leq 0, \; v \in \mathcal{U}, d(u,v) \leq \varepsilon.$$

A change of variable leads to

$$\int[v(a,\theta) - v(b,\theta)]l(\theta)\pi(\theta)d\theta = \int v(t)[l(t+a)\pi(t+a) - l(t+b)\pi(t+b)]dt = \int v(t)\Delta(t)dt.$$

As a consequence, we have

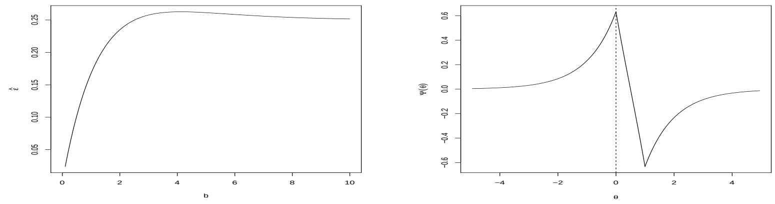
$$\inf_v \int v(t)\Delta(t)dt = \int_{\Delta\geq 0}[u(t)-\varepsilon]\Delta(t)dt + \int_{\Delta<0}[u(t)+\varepsilon]\Delta(t)dt = N^{ab}(u) - \varepsilon\int|\Delta(t)|dt, \qquad (3)$$

and (3) is equal to 0 when considering $\hat{\varepsilon}$. Therefore, $\hat{\varepsilon}$ is the threshold neighbourhood size: we have ε–robustness only for values of ε not greater than $\hat{\varepsilon}$ and the corresponding w is the most sensitive utility. ■

We provide an example in which the threshold ε is computed.

Example 1 *Suppose that $u(t) = e^{-|t|}$, $\pi(\theta) = \frac{e^{-|\theta|}}{2}$ and $l(\theta) = 1$ (there is no additional information). We have that*

$$\Delta(\theta) = \pi(\theta+a) - \pi(\theta+b) = \frac{e^{-|\theta+a|} - e^{-|\theta+b|}}{2}.$$



a) b)

Figure 1. a) $\hat{\varepsilon}$ for different values of b, with $a = 0$ in Example 1. b) Representation of function $\Psi(\theta)$ in Example 3.

Without loss of generality, assume that $0 \le a < b$. *Then,* $\Delta(\theta) \ge 0$ *if and only if* $\theta \ge -(a+b)/2$. *Moreover,*

$$2N^{ab}(u) = \int_{-\infty}^{\infty} \left(e^{-|\theta-a|} - e^{|\theta-b|}\right) e^{-|\theta|} d\theta = e^{-a}(a+1) - e^{-b}(b+1)$$

and $2\int |\Delta(\theta)| d\theta = 4(1 - e^{(a-b)/2})$. *Then* $\hat{\varepsilon} = \frac{e^{-a}(a+1) - e^{-b}(b+1)}{4(1-e^{(a-b)/2})}$ *and*

$$w(c,\theta) = \begin{cases} e^{-|\theta-c|} - \hat{\varepsilon}, & if \quad \theta \ge -(a+b)/2 \\ e^{-|\theta-c|} + \hat{\varepsilon}, & if \quad \theta < -(a+b)/2 \end{cases}.$$

Suppose, for example, that $a = 0$ *and* $b = 1$. *Then, it follows that* $\hat{\varepsilon} = .1679$. *Figure 1 a) represents the values of* $\hat{\varepsilon}$ *when we compare* $a = 0$ *with* $b > 0$. □

Frequently, we want utilities normalised between 0 and 1. If we designate by $\mathcal{U}'$ the subset of utilities in $\mathcal{U}$ bounded between 0 and 1, we have that equation (3) becomes

$$\begin{aligned} \inf_v \int v(t)\Delta(t)dt &= \int_{\Delta \ge 0} \max\{0, u(t) - \varepsilon\}\Delta(t)dt + \int_{\Delta < 0} \min\{1, u(t) + \varepsilon\}\Delta(t)dt \\ &= N^{ab}(u) - \int_{\Delta \ge 0} \min\{u(t), \varepsilon\}\Delta(t)dt + \int_{\Delta < 0} \min\{1 - u(t), \varepsilon\}\Delta(t)dt \end{aligned}$$

Example 2 (Continuation of Example 1) *Consider* $a = 0$ *and* $b = 1$. *Using numerical methods to solve the optimisation problem, we find* $\hat{\varepsilon} = 0.1739$. *Note that we have found a value* $\hat{\varepsilon}$ *larger than before. It was expected since we are considering a subset of the class presented earlier.* □

4. Prior Sensitivity

We turn now our attention to changes in the prior and search for the largest neighbourhood of a prior π preserving the ranking $b \preceq a$, under the topology induced by the total variation norm.

Proposition 2 **a)** *Let* $\Psi(\theta) = [u(a,\theta) - u(b,\theta)]l(x|\theta)$. *If there is* $\theta^* : \Psi(\theta^*) = \inf_\theta \Psi(\theta)$, *then the most sensitive prior in* Γ, *under the above conditions, is given by* $Q^* = \tilde{\varepsilon}\delta_{\theta^*} + (1-\tilde{\varepsilon})\pi I_{A_{\tilde{\varepsilon}}}$, *where* δ_x *is a Dirac measure at* x, $A^C_{\tilde{\varepsilon}}$ *is a measurable subset of* π*–measure* $\tilde{\varepsilon}$ *with the largest values of* $\Psi(\theta)$ *and* $\tilde{\varepsilon}$ *is the smallest* ε *such that*

$$N^{ab}(\pi) - \int_{A^C_{\tilde{\varepsilon}}} [\Psi(\theta) - \inf_\theta \Psi(\theta)]\pi(\theta)d(\theta) \le 0. \qquad (4)$$

b) *If there is no such θ^*, for the $\tilde{\varepsilon}$ defined in* **a)**, *there is a sequence $\{Q_i\}_{i=1}^{\infty}$ such that $\forall\delta > 0$ there is n_δ such that Q_{n_δ} is t-sensitive, with $t \in (\varepsilon, \varepsilon + \delta)$ and Q_n is at most $(\varepsilon_\pi + \delta)$-sensitive.*

c) *π is $\tilde{\varepsilon}$-sensitive.*

PROOF. **a)** We search for the smallest ε such that

$$\inf_Q \int [u(a,\theta) - u(b,\theta)] l(x|\theta) Q(d\theta) = \inf_Q \int \Psi(\theta) Q(d\theta) \leq 0, \; d(\pi, Q) \leq \varepsilon \tag{5}$$

It is known, see, e.g., Fortini and Ruggeri (1994), that infima of expectations within a total variation neighbourhood are attained at measures with a point mass ε at a point and coinciding with the prior π on a subset A of measure $1-\varepsilon$. We have

$$\begin{aligned} \inf_Q \int \Psi(\theta) Q(d\theta) &= \inf_A \int_A \Psi(\theta)\pi(\theta)d\theta + \varepsilon \inf_\theta \Psi(\theta) \\ &= N^{ab}(\pi) - \int_{A^C} \Psi(\theta)\pi(\theta)d\theta + \varepsilon \inf_\theta \Psi(\theta). \end{aligned} \tag{6}$$

Since (6) is monotonic, nonincreasing in ε, there is a smallest ε such that (6) is nonnegative.

b) Let $\{t_i\} \in A$ such that $\Psi(t_i) \rightarrow \inf_t \Psi(t)$ and define Q_i by $Q_i = \tilde{\varepsilon}\delta_{t_i} + (1-\tilde{\varepsilon})\pi I_{A_{\tilde{\varepsilon}}}$.

c) A consequence of **a)** or **b)**. ■

When the inequality in (5) becomes an equality, its solution gives $\tilde{\varepsilon}$, as in the next example.

Example 3 (Continuation of Example 1) *Consider $a = 0$ and $b = 1$. We have that*

$$\Psi(\theta) = \begin{cases} e^{|\theta|} - e^{\theta-1} & \theta < 1 \\ e^{-\theta} - e^{-\theta+1} & 1 \leq \theta \end{cases} .$$

Observe that $\inf_\theta \Psi(\theta) = \Psi(1) = e^{-1} - 1 = -\Psi(0) = -\sup_\theta \Psi(\theta)$. Because of the shape of $\Psi(\theta)$, see Figure 1 b), we look for an interval $A^C = (t_1, t_2)$, with $t_1 < 0 < t_2$ and $\Psi(t_1) = \Psi(t_2)$. We find, numerically, that the solution to (5) is given by $A^C = (-.1618, .0701)$ with $\tilde{\varepsilon} = .1086$.

□

5. Prior-Utility Sensitivity

Based on the previous results, we present a strategy to compute $\hat{\varepsilon}$ and the most sensitive prior and utility in the case of a neighbourhood of (u, π) as defined in Sections 2, 3 and 4. We know that

$$N^{ab}(v, q) = \int v(\theta) \left[l(x|\theta + a) q(\theta + a) - l(x|\theta + b) q(\theta + b) \right] d\theta = \int v(\theta) \Delta_q(\theta) d\theta.$$

where (v, q) is in an ε-neighbourhood of (u, π). Then, we could proceed as follows:

1. `Set` $\varepsilon_l = 0,\ \varepsilon_u = \min\{\varepsilon$`-robustness for` u`,` ε`-robustness for` $\pi\}$.
2. `Set` $\varepsilon = (\varepsilon_u + \varepsilon_l)/2$.
3. `Search for` $\hat{\theta}_\varepsilon$ `minimising` $\min \int v(\theta)\Delta_\delta(\theta)d\theta$`, where` δ `is Dirac measure at` $\hat{\theta}_\varepsilon$.
4. `Take` $v_\varepsilon(\theta) = \max\{0, u(\theta) - \varepsilon\}$ `where` $\Delta_\pi(\theta) \geq 0$ `and` $v_\varepsilon(\theta) = \min\{1, u(t) + \varepsilon\}$ `elsewhere.`
5. `Find subset` A_ε `of` π`-measure` $1-\varepsilon$ `with the smallest value of` $v_\varepsilon(\theta)\Delta_\pi(\theta)$ `and consider the prior` q_ε `which coincids with` π `on` A_ε`, has point mass` ε `at` $\hat{\theta}_\varepsilon$ `and vanishes elsewhere.`
6. `Compute` $N^{ab}(v_\varepsilon, q_\varepsilon)$`. If` $|N^{ab}(v_\varepsilon, q_\varepsilon)| <$ `eps (for a given eps), then stop; otherwise, change` ε_l `or` ε_u `according to sign of` $N^{ab}(v_\varepsilon, q_\varepsilon)$`, go back to 2.`

Note that, in step 3, $\min \int v(\theta)\Delta_\delta(\theta)d\theta$ is equivalent to

$$\min_{\theta\in\Theta_-} \{\max\{0, u(\theta)-\varepsilon\}l(x|\theta+a) - \min\{1, u(\theta)+\varepsilon\}l(x|\theta+b)\} = \min_{\theta\in\Theta_-} \phi(\theta) \tag{7}$$

where $\Theta_- = \{\theta : \Delta(\theta) \leq 0\}$

Example 4 (Continuation of Example 1) *We consider now prior-utility robustness, within our problem. The following table contains the results obtained from the applications of the algorithm above. We take* `eps`*=0.0001 and solve steps 3,4 and 5 numerically:*

ε_l	ε_u	ε	$\hat{\theta}_e$	A_ε^C	$N^{ab}(v_\varepsilon, q_\varepsilon)$
0	*0.1086*	*0.0543*	*-1.05583*	*(-0.0429,0.0689)*	*0.0480361*
0.0543	*0.1086*	*0.0815*	*-1.08496*	*(-0.0651,0.10527)*	*0.0148017*
0.0815	*0.1086*	*0.0950*	*-1.09985*	*(-0.0763,0.1240)*	*-0.0035509*
0.0815	*0.0950*	*0.0882*	*-1.09237*	*(-0.0707,0.1146)*	*0.0053917*
0.0882	*0.0950*	*0.0916*	*-1.09610*	*(-0.0735,0.1192)*	*-0.0011112*
0.0882	*0.0916*	*0.0899*	*-1.09424*	*(-0.0721,0.1169)*	*0.0003098*
0.0899	*0.0916*	*0.9078*	*-1.09517*	*(-0.0728,0.1181)*	*-0.0000980*

Then, the joint ε-robustness is 0.9078.

6. Conclusions

Here, following Martin et al (2003), we have addressed issues concerning joint sensitivity with respect to the prior and the utility, considering distances based on norms. By analysing changes in differences in expected utility among alternatives, we are able to detect directions in which perturbations of the assessed utility and/or probability lead to fastest changes in differences in expected utility and, as a consequence, to directions in which assessments should be considered more carefully.

The results concerning ε-contaminations, discussed in Martin et al (2003), can be incorporated in this general approach, by changing the distance d by a set function over the space of priors and/or utilities.

Acknowledgement. Research supported by grants from MCYT, URJC, the URJC-DMR Foundation Decision Engineering Lab and the ESF TED programme. We are grateful to the referees' comments.

References

[1] Fortini, S. and Ruggeri, F. (1994). Concentration functions and Bayesian robustness. *Journal of Statistical Planning and Inference*, **40**, 205–220.

[2] French, S. and Ríos Insua, D. (2000) *Statistical Decision Theory*, Arnold, London.

[3] Martin, J., Ríos Insua, D. and Ruggeri, F. (2003) Joint Sensitivity in Bayesian Decision Theory, *Test*, **12**, 173-194.

[4] Ríos Insua, D. and Ruggeri, F. (2000). *Robust Bayesian Analysis*. Springer Verlag, New York.

Fabrizio Ruggeri
Istituto di Matematica Applicata
e Tecnologie Informatiche
CNR, Milano

Jacinto Martin
Departamento de
Matemáticas
Universidad de Extremadura

David Ríos Insua
Grupo de Estadística y
Ciencias de la Decisión
Universidad Rey Juan Carlos