Bayesian Robustness
IMS Lecture Notes - Monograph Series (1996) Volume 29

COMPARATIVE PROBABILITY AND ROBUSTNESS[1]

BY GIULIANA REGOLI

Università di Perugia

In this paper we propose the comparative approach as a natural approach in a probability elicitation process and we show that Bayesian robustness analysis is the way to quantify comparative probabilities. Special cases of comparative probability assessments, derived from Savage's almost-uniform partitions, are considered.

1. Introduction. In a probability elicitation process, comparative probability judgments seem to be particularly natural and close to common language and mental categories. Therefore such judgments seem more reliable than numerical ones, typically either when the experts of the field under consideration are unaware of statistics or probability theory, or when the considered problem is new. Imagine, for example, to ask a physician about the transmission rate of Ebola virus. Depending on his/her knowledge of the epidemic, he/she can easily and firmly state some comparisons among the probabilities of possible values (or interval of values) of the rate, while numerical judgments could be nearly impossible.

The classical comparative probability theory essentially deals with two problems: 1) Are the assessments compatible with a probability measure? 2) If this is the case, are they so detailed that they can single out a probability measure without ambiguity? and which one?

When thinking about applications, the unrealistic traditional axiom of completeness in comparative judgments has to be abandoned. The consequent unicity of the numerical representation fails and a new problem arises: are the comparative assessments enough to grant a sufficiently accurate answer to a given problem? Following the "Doogian" point of view (see Good 1950 and following), Giron and Rios (1980) seem to be the first who explicitly propose to carry out a Bayesian analysis with the whole family of probabilities compatible with a set of comparisons and join the Robust Bayesian point of view with Comparative Probability.

In this paper, we will see that the family of probabilities representing a set of comparisons turns out to be a particular moment class, namely the convex closure of a set of quantile classes. Therefore it can be treated with all the facilities tuned up for this (see Moreno and Cano 1991, Berger 1994). Here a direct treatment of a robust analysis is suggested and then used to examine some special cases: the *n-almost uniform partitions.* These are realistic and quite natural to assess. Furthermore, it is shown that they are

---

[1] *AMS 1991 subject classification.* Primary 62A15; secondary 62F35, 60A05.

Key words and phrases. Comparative probability, robust bayesian inference, group opinion, classes of priors.

computationally cheaper than the general case and they can give a quite accurate analysis.

2. Preliminaries and notation. Let $\mathcal{F}$ be a σ-algebra of subsets of a set Θ and $Pr(\Theta)$ the set of all probability measures on $\mathcal{F}$. A *comparative probability* on $\mathcal{F}$ is a finite or infinite list, $\mathcal{C}$, of comparisons among events, equivalently it is a partial binary relation on $\mathcal{F}$. For every $(B, A) \in \mathcal{C}$ we denote by $B \preceq A$ the assertion that "*B is not more probable than A*". A comparative probability is said to be *consistent on $\mathcal{F}$* if there exists a probability measure P on $\mathcal{F}$ such that

$$B \preceq A \Rightarrow P(B) \leq P(A).$$

In such a case we say that P *represents* (or is *compatible with*) $\mathcal{C}$.

Starting from de Finetti (1931) and Savage (1954), the traditional literature on this field gives different conditions for a comparative probability to be representable by a finitely additive probability. Villegas' continuity condition makes the probability σ-additive (for reviews see Fishburn, 1986, and Regoli, 1994). We will use the condition given by Buehler (1976):

PROPOSITION: *A set of comparisons, $\mathcal{C} = \{B_j \preceq A_j, j \in J\}$, is* **not** *representable by any finitely additive probability measure if and only if there exist in $\mathcal{C}$ a finite set of comparisons, $\{B_j \preceq A_j, j \in F\}$, $F \subset J$, and $y \in \mathcal{R}^F$, $y > 0$, such that*

$$\sup \sum_{j \in F} y_j (I_{A_j} - I_{B_j}) < 0, \tag{1}$$

where I_A denotes the indicator of the event A.

When $\mathcal{C}$ is finite, the corresponding events B_j, A_j $(j \in J)$ generate a finite algebra $\mathcal{A}_\mathcal{C}$. If $\mathcal{C}$ is consistent on $\mathcal{A}_\mathcal{C}$, then it is clear that a σ-additive extension of P to $\mathcal{F} \neq \mathcal{A}_\mathcal{C}$ exists but need not be unique. Therefore the problem of robustness arises naturally.

In the next section, we will determine the range of a posterior expectation of a given function, using the following Lemma which is essentially known (see for instance Berger 1985, Sivaganesan and Berger 1989).

LEMMA **1** *Let ψ be defined by*

$$\psi(P) = \frac{\int_\Theta h(\theta) dP(\theta)}{\int_\Theta k(\theta) dP(\theta)}, \ \textit{with } h \textit{ and } k \textit{ bounded}, \ k > 0.$$

Let P be a mixture of a family of probability measures, $\{P_t,\ t \in T\}$, for some set T. Then $\psi(P)$ is a mixture of $\psi(P_t)$, and vice versa.

PROOF: Let P be given by:

$$P(A) = \int_T P_t(A)d\mu(t) \quad \forall A \in \mathcal{F}. \tag{2}$$

First note that if

$$\xi(P) = \int_\Theta k(\theta)dP(\theta),$$

then the function $\frac{\xi(P_t)}{\xi(P)}$ is positive and $\mathcal{T}$-measurable, with $\int_T \frac{\xi(P_t)}{\xi(P)}d\mu(t) = 1$.

If ν is defined by

$$\nu(S) = \int_S \frac{\xi(P_t)}{\xi(P)}d\mu(t),$$

then ν is a probability measure on $\mathcal{T}$ and verifies

$$\psi(P) = \int_T \psi(P_t)d\nu(t). \tag{3}$$

Vice versa, given ν, define a probability measure, μ, by setting

$$\mu(S) = \int_S 1/\xi(P_t)\left\{\int_T 1/\xi(P_t)d\nu(t)\right\}^{-1} d\nu(t).\square$$

3. The comparative class. In this section we consider the general case of a finite comparative probability: we describe the whole family of compatible probability measures, which is called *comparative class*, and we determine the range of given posterior quantities of interest.

For a finite set of comparisons $\mathcal{C} = \{B_j \preceq A_j, j = 1, 2, ...N\}$, where $B_j, A_j \in \mathcal{F}$, the family, Γ, of probability measures representing $\mathcal{C}$ is given by

$$\Gamma = \{P \in Pr(\Theta),\ P(B_j) \leq P(A_j),\ j = 1, 2, ...N\}.$$

Let $\{C_1, C_2, ...C_n\}$, be the partition generated by $\{A_j, B_j, j = 1, ...N\}$. Then $P \in \Gamma$ if and only if

$$\sum_{C_i \subset B_j} P(C_i) \leq \sum_{C_i \subset A_j} P(C_i),\ j = 1, 2, ...N.$$

Let $S \subset \mathcal{R}^n$ be the set of non negative normalized solutions to the system

$$\sum_{C_i \subset A_j} p_i - \sum_{C_i \subset B_j} p_i \geq 0,\ j = 1, 2, ...N \tag{4}$$

Then $\Gamma = \{P \in Pr(\Theta),\ P(C_i) = p_i,\ (p_i,\ i = 1,2,...,n) \in S\}$.

If $\Delta = \{Q^1, Q^2, ...Q^m\}$ denotes the finite set of the extreme points of S, and $T = \prod_{i=1}^{n} C_i$, then $P \in \Gamma$ if and only if there exist $\mu \in Pr(T)$ and $b_j \geq 0$, with $\sum_{j=1}^{n} b_j = 1$, such that for every $A \in F$

$$P(A) = \sum_{j=1}^{m} b_j \int_T \sum_{i=1}^{n} q_i^j \delta_{t_i}(A) d\mu(t), \tag{5}$$

where $Q^j = (q_1^j, ...q_n^j) \in \Delta$, $(j = 1,2,...m)$, $t_i \in C_i$ and where δ_t denotes the probability measure degenerate at t.

Obviously, the system above does not have any solution if and only if $\mathcal{C}$ is not representable by any probability measure.

Now, let f be the likelihood function which we assume bounded for the given data. Suppose the quantity of interest, ψ, is the posterior expectation of a bounded function, h, that is

$$\psi(P) = \frac{\int_\Theta h(\theta) f(\theta) dP(\theta)}{\int_\Theta f(\theta) dP(\theta)}.$$

Lemma 1 and (5) yield

$$\psi(P) = \sum_{j=1}^{m} a_j \int_T \frac{\sum_{i=1}^{n} q_i^j h(t_i) f(t_i)}{\sum_{i=1}^{n} q_i^j f(t_i)} d\nu(t),$$

where ν is defined as in the proof of Lemma 1 and the a_j's are defined in a similar way, being P a convex combination of probability measures. Therefore the next Theorem holds.

THEOREM 1 *If Γ is a comparative class, the supremum of $\psi(P)$, as P ranges over Γ, is given by*

$$\sup_{P\in\Gamma}\{\psi(P)\} = \sup_{Q^j\in\Delta,\ t_i\in C_i} \left\{ \frac{\sum_{i=1}^{n} q_i^j h(t_i) f(t_i)}{\sum_{i=1}^{n} q_i^j f(t_i)} \right\}.$$

An analogous formula holds for the infimum.

REMARK 1. In the general case, the computation of Δ is not polinomial (e.g. Chvàtal, 1983 and Dyer, 1983). In some special cases, simplified algorithms or particular kinds of assessments can play a key role to reduce the complexity, as we shall see in the next section. Once computed, Δ can be used without loosing any information on Γ, at least in many kind of problems, such as sequential updating or comparison of different (posterior) expected values.

REMARK 2. When $\mathcal{C}$ is directly elicited and its cardinality is reasonably small, an alternative way to deal with Γ is via Generalized Moment Theory (Kemperman 1987, Salinetti 1994, Liseo, Moreno and Salinetti 1995): this allows to avoid the computation of Δ. In fact, rewriting Γ as

$$\Gamma = \{P \in Pr(\Theta),\ \int_\Theta \{I_{B_j}(\theta) - I_{A_j}(\theta)\} dP(\theta) \leq 0,\ j = 1, 2, \ldots N\},$$

the infimum of $\psi(P)$, as P ranges over Γ, is given by the solution in λ of the equation

$$\sup_{d_j \geq 0,\ j=1,2,\ldots N} \left\{ \inf_\theta \{(h(\theta) - \lambda) f(\theta) - \sum d_j (I_{B_j}(\theta) - I_{A_j}(\theta))\} \right\} = 0.$$

An analogous expression gives the supremum of $\psi(P)$.

REMARK 3. It is possible to join $\mathcal{C}$ with some quantitative judgments of the kind $P(D_i) \leq \alpha_i$ or $P(D_i) \geq \alpha_i$, $i = 1, \ldots M$, just by using these constraints to define Γ and including them either in system (4), which has to be written down considering the related partition of Θ, or in the equation of Remark 2.

4. Almost uniform partitions. This section is devoted to almost uniform partitions which are a special kind of comparative assessment; we will see how the analysis of the previous section can be conveniently used in this case.

The concept of almost uniform partition has been introduced by Savage (1954), as a technical tool to deduce an "unambiguous assignment of a numerical probability" from a comparative probability, avoiding the de Finetti's postulate of uniform partitions (i.e. there exist partitions of Θ in arbitrary many equivalent events). The idea of uniform partition is easy to understand, but only very rarely can be applied to real problems. This idea can be replaced by the more ductile and realistic idea of almost uniform partition. Following Savage, a partition of Θ, $\{C_i, i = 1, \ldots n\}$, is said an *n-almost uniform partition* (n-a.u. partition, for short) if

$$\bigcup_{j=1}^{r} C_{i_j} \preceq \bigcup_{j=1}^{r+1} C_{k_j},\ \forall r < n;\ \forall i_j, k_j \in \{1, \ldots n\}. \tag{6}$$

EXAMPLE 1. Even if I am not convinced that the coin in my wallet is totally fair, I can claim that, for a quite great particular n, all sequences of heads and tails of length n form a 2^n-almost uniform partition.

EXAMPLE 2. Suppose that several experts have similar, yet different opinions about the shape of the prior density on Θ. Nevertheless they may find a common almost uniform partition, for example by selecting an uniform partition from one of the elicited densities. Since the comparative approach does not require the choice between discrete and continuous models, this solution can be accepted also by those experts who prefer a discrete model.

In fact, under some algebraic conditions, including completeness, either de Finetti's or Savage's postulate implies the existence of a unique finitely additive probability representing a comparative probability.

We can add that, if $\mathcal{C}_\alpha$, $\alpha \in I$, is a net of comparative probabilities, such that $\mathcal{C} = \bigcup_\alpha \mathcal{C}_\alpha$ verifies the de Finetti's or Savage's postulate, the net of the corresponding families, Γ_α, converges to the unique P which represents $\mathcal{C}$. Moreover if $I = \mathbf{N}$, and $\mathcal{C}_\alpha$ contains an α-uniform or an α-almost uniform partition, for every $\alpha \in I$, then it can be shown that the imprecision of $\mathcal{C}_\alpha$ converges to 0, where *imprecision* of $\mathcal{C}$ -or of Γ- is defined as

$$\mathcal{I}(\Gamma) = \sup_{A \in \mathcal{A}_C} \left\{ \sup_{P \in \Gamma} P(A) - \inf_{P \in \Gamma} P(A) \right\}.$$

This encourages the use of partial -rather than complete- judgments in an actual elicitation process, but the problem of the consistency of $\mathcal{C}$ arises again.

For this reason, we will consider only a.u.partitions which are proper, where we say that a partition, $\{C_i, i = 1, ...n\}$, is *proper* in $\mathcal{C}$, if $\mathcal{C}$ does not contain any comparison of the following type:

$$\bigcup_{j=1}^{s} C_{i_j} \preceq \bigcup_{j=1}^{r} C_{k_j}, \text{ with } r < s.$$

Actually, proper partitions are always consistent, as the following Theorem proves, while if an a.u. partition is not proper, its consistency has to be directly checked. Of course, if $\mathcal{C}$ is non-contradictory, a proper n-a.u. partition is representable by means of every probability measure for which $P(C_i) = 1/n$, for all i.

THEOREM **2** *Given an arbitrary set of comparisons $\mathcal{C}$, if it contains a proper partition $\{C_i, i = 1, ...n\}$, then it is consistent on the algebra generated by such partition.*

PROOF: Suppose that $\mathcal{C}$ is not consistent on such algebra. Then, condition (1) holds. Without loss of generality, we can suppose that y_j are natural numbers and write, for some finite set, F, $B_j \preceq A_j$, $j \in F$,

$$\sup \sum_{j \in F} (I_{A_j} - I_{B_j}) < 0.$$

This is equivalent to

$$\text{card}\{j \in F : C_i \subset A_j\} < \text{card}\{j \in F : C_i \subset B_j\} \qquad \forall i = 1...n$$

Therefore, for some j_o,

$$r = \text{card}\{i : C_i \subset A_{j_o}\} < \text{card}\{i : C_i \subset B_{j_o}\} = s.$$

Since A_{j_o} and B_{j_o} are logically dependent on $\{C_i, i = 1, ...n\}$, then

$$A_{j_o} = \bigcup_{j=1}^{r} C_{i_j} \text{ and } B_{j_o} = \bigcup_{j=1}^{s} C_{k_j}.$$

Since $r < s$, and the comparison "$B_{j_o} \preceq A_{j_o}$" is contained in $\mathcal{C}$, then the given partition is not a proper partition in $\mathcal{C}$. □

Now, we use the same notation as in Section 3, and prove a theorem which describes the family Γ_n representing a general almost uniform partition. To do so, we need a lemma which can also be used to describe the family Γ_n^o representing an almost uniform partition with particular additional comparisons between the C_i's.

LEMMA 2 *Let Γ_n^o be the set of all probabilities compatible with the two assessments that $\{C_1, C_2, ...C_n\}$ is an a.u. partition of Θ and that $C_1 \preceq C_2 \preceq ... \preceq C_n$. Then the set $\Delta \subset \mathcal{R}^n$ of all the extreme points of S is given by*

$$\Delta = \{Q^{i,i+1},\ 0 \leq 2i \leq n-1,\ Q^{j,n-j},\ 0 \leq 2j \leq n-1\},$$

where

$$Q^{i,k} = \left(\overbrace{\frac{i}{n(i+1)-k}, ... \frac{i}{n(i+1)-k}}^{k},\ \overbrace{\frac{i+1}{n(i+1)-k}, ... \frac{i+1}{n(i+1)-k}}^{n-k} \right).$$

PROOF: Given the constraints $p_1 \leq p_2 \leq ...p_n$, the system derived from (6) reduces to

$$\sum_{i=n-k+1}^{n} p_i \leq \sum_{i=1}^{k+1} p_i; \text{ with } 1 \leq k < n-k$$

To complete the proof, the method given by Classens et al. (1991) can be adapted. □

THEOREM 3 *Let Γ_n be the set of all probabilities compatible with the only assessment that $\{C_1, C_2, ...C_n\}$ is an a.u. partition of Θ. Then the set $\Delta \subset \mathcal{R}^n$ of all the extreme points of S is given by*

$$\Delta = \{V^{(-1)}, V^{(-2)}, ...V^{(-n)}, V^{(1)}, V^{(2)}, ...V^{(n)}\},$$

where

$$V^{(-j)} = (q_i^j), \text{ with } q_j^j = 0, \text{ and } q_i^j = 1/(n-1), \text{ if } i \neq j$$

and

$$V^{(j)} = (q_i^j), \text{ with } q_j^j = 2/(n+1), \text{ and } q_i^j = 1/(n+1), \text{ if } i \neq j.$$

Therefore

$$\sup_{P \in \Gamma}\{\psi(P)\} = \sup_{t_i, \in C_i} \{ \sup_{j=1,...n} \left\{ \frac{\sum_{i=1}^n h(t_i)f(t_i) \pm h(t_j)f(t_j)}{\sum_{i=1}^n f(t_i) \pm f(t_j)} \right\} \}.$$

PROOF: In order to determine Δ, consider the result of Lemma 2 for all possible orderings on $\{C_i\}$. The union of the relative families Γ_n^o's gives Γ_n.□

The next Example shows how the previous results can be applied.

EXAMPLE 3. Given $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$, suppose we can choose among three actions, g, h and k, each of which costs \$ 3 and that can bear respectively $g = (5,2,4,2)$; $h = (3,2,4,3)$; $k = (2,3,4,3)$. Asking several experts in the field, their opinions agree just on assessing that $\{\theta_i\}$ is an almost uniform partition and that $\theta_1 \preceq \theta_2 \preceq \theta_3 \preceq \theta_4$. Let Γ be the set of probabilities representing their common opinions, that is the convex envelop of Δ, which (by Lemma 2) is given by
$\Delta = \{(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}); (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}); (\frac{1}{6}, \frac{1}{6}, \frac{2}{6}, \frac{2}{6}); (\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{2}{5})\}$.

An experiment is performed whose result, y, yields likelihood $(3,4,6,5)$. It is easy to compute the posterior means w.r. to each of the points of Δ, and to deduce that k is the best action for each probability in Γ. Moreover the ranges of posterior means are respectively: $r(g) = (2.80, 3.1\overline{6})$; $r(h) = (3.09, 3.28)$; $r(k) = (3.13, 3.4)$. Note that the posterior means of h are always greater than the cost while those of g can be smaller. If we had replaced Γ with only the uniform probability distribution instead, we could have chosen indifferently action g or k, as, in that case: $E(k|y) = E(g|y) = 3.1\overline{6}$ versus $E(h|y) = 3.\overline{1}$.

REMARK 4. As already noticed in Remark 1, the direct specific treatment by means of extreme points of a comparative class, could be, in general too hard, because of the dimension of the problem. In an almost uniform partition instead, the cardinality of Δ is $2n$ (respectively n if the comparisons among the sets of the partition are known as in Lemma 2) and the extreme points of the family Γ are explicitely computed in Theorem 3. Notice that, in this case, the cardinality of $\mathcal{C}$ is fairly large (see (6)).

REMARK 5. When the elicitation is too rough for a robust conclusion, any additional comparison (possibly numerical) gives a refinement of the initial family: of course it is important to look for some kind of new assessment which robustifies the conclusion and avoids an excessive increase in the dimension of the problem. Unfortunately, so far, there is no natural procedure to refine an n-a.u.partition to an m-a.u.partition, with $n < m$. It seems to be convenient to pick out a C_i and elicit an a.u.partition of it. If the most probable C_i is chosen, after comparing the events of the algebra generated by the new partition, a much more precise family is obtained and it may make the original partition "more uniform". On the other hand, it could be more convenient to choose, each time, the C_i which makes the quantity of interest more sensitive; in fact, this should increase the robustness more rapidly.

Acknowledgements I would like to thank Prof. Walter Racugno who suggested me the question which gave rise to the present research. I am also grateful to Prof. Elias Moreno for his valuable suggestions on an earlier version of this paper. Finally, I wish to thank the referees for their useful comments and advices.

REFERENCES

BERGER, J. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer Verlag N.Y.

BERGER, J. (1994). An overview of Robust Bayesian Analysis. (with discussion) *Test* **3**, 5-124.

BUEHLER, R.J. (1976). Coherent preferences. *Ann. Statist.* **4**, 1051-1064.

CHVÁTAL, V. (1983). *Linear Programming.* Freeman and Co. N.Y.

CLAESSENS, M.N.A.J. LOOTSMA, F.A. and VOGT, F.J. (1991). An elementary proof of Paelinck's theorem on the convex hull of ranked criterion weights. *Eur. J. Oper. Res.* **52**, 255-258.

DYER, M.E. (1983). The complexity of vertex enumeration methods . *Math. Oper. Res.* **8**, 381-402.

DE FINETTI, B. (1931). Sul significato soggettivo della probabilità. *Fund. Math.* **17**, 298-329.

FISHBURN, P.C. (1986). The axiom of Subjective Probability. (with discussion) *Statist. Sci.* **1**, 335-358.

GIRON, F.J. and RIOS, S (1980). Quasi-Bayesian behaviour: a more realistic approach to decision making? In *Bayesian Statistics* J.M.Bernardo, M.H.De Groot, D.V.Lindley, A.F.M. Smith eds. 17-38. Univ Press. Valencia.

GOOD, I.J. (1950). *Probability and weighing of evidence.* Griffin London.

KEMPERMAN, J.K.B. (1987) Geometry of the moment problem. *Proc. of Symposia in Appl. Math.* **37**, 93-122.

LISEO, B. MORENO, E. and SALINETTI, G (1995). Bayesian Robustness for classes of Bidimensional Priors with Given Marginals. *Proceedings of Second International Workshop on Bayesian Robustness, Rimini.*

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

REGOLI, G. (1994). Qualitative probabilities in an elicitation process. *Atti XXXVII Riunione scientifica SIS* San Remo April 1994, 153- 165.

SALINETTI, G. (1994). Discussion on "An overview of Robust Bayesian Analysis" by J.O. Berger *Test* **3**, 5-124.

SAVAGE, L.J. (1954). *The Foundation of Statistics.* Wiley N.Y.

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

DIP. MATEMATICA
VIA VAINVITELLI 1
06100 PERUGIA, ITALY