Robust Inference in Probability under Vague Information

Giuliana Regoli
Dip. Matematica, Università di Perugia
via Vainvitelli 1, 06100 Perugia, Italy.

Abstract

Vague information can be represented as comparison of previsions or comparison of probabilities, and a robust analysis can be done, in order to make inference about some quantity of interest and to measure the imprecision of the answers. In particular, in some decision problems the answer can be unique.

1 Introduction.

In front of problems of description, diagnosis, prevision, forecasting, or decision under uncertainty, the probability theory is the most ancient tool proposed, but for the divination. When the information on the structure of the problem is vague, the use of probability theory -in particular the Bayesian probability theory- is often criticized, because it request an initial precise numerical evaluation. On the other hand nobody refuse the equations of classical mechanics in describing the bullet trajectory, only because it is difficult to calculate exactly the initial position or speed.

The natural answer is: keep the theory, use your imprecise knowledge to give an answer and look for a tool to measure its imprecision: when various probability distributions are compatible with the information, the analysis can be satisfactory even if the value of the quantity of interest is not unique, but "approximatively unique". Of course the degree of approximation has to be chosen by the decision maker, depending on the problem.

So, once abandoned the "dogma of precision", the robust Bayesian analysis (that is the study of the sensitivity of a Bayesian answer to uncertain inputs) has been developed (Berger, 1994 for a review) and theoretical models, such as belief functions and lower probability envelops, have been used in a rigorous probabilistic frame (Wasserman, 1990, Wasserman and Kadane, 1990, Walley, 1991). Those model still need some numerical evaluation, therefore they understand some experience in probability or statistics.

In the real life, if the problem of interest is new, or if the expert of the field is unaware of probability and statistics, any numerical judgement can be misleading. Comparative judgements, while vague, are more reliable, since they are closer to common language and mental categories: a person living in the mountains can be able to compare the quantity of snow of the next two days; a Tour de France organizer can compare how long two stages last, but almost nobody can be able to compute the full probability distributions of the involved random quantities. Those examples suggest that the comparative prevision is a widely used natural concept, therefore it is a good starting point for probability elicitation.

The idea of joining a robust analysis with the comparative probability theory has already been proposed (e.g. Good, 1950, Giron and Rios, 1980), but still it is few exploited, while recently the interest on comparative probability as an effective tool of elicitation is growing up (see also either Coletti, or Garbolino in this same issue).

Here we take up this idea. Note that in order to develop it, we assume as primitive the concept of random variables and their prevision (that is the expectation) rather than events and their probability (e.g. de Finetti 1937).

1.1 Preview.

In next section the concept of comparative prevision is introduced and comparative probability is defined as a particular case of it; the class of all probability distributions compatible with them is described. Particular kinds of elicitation are studied in the following section. Section 4 is devoted to the cases in which the coherence allows to deduce certain comparison from the initial comparative elicitation, while Section 5 considers the problem of making inference about some numerical quantity of interest and the problem of its approximation.

The technical results are deduced from those in Regoli (1994 b), Regoli (1995) and related references.

2 Comparative Probability and Comparative Prevision.

Let $\mathcal{F}$ be a family of random variables (r.v.'s) and let $\tilde{\mathcal{F}}$ the spanned σ-algebra of events. If indicators of events are in $\mathcal{F}$, let $A, B, C, ...$ denote either the events or their indicators.

A *comparative prevision* on $\mathcal{F}$ is a finite or infinite list, $\mathcal{C}$, of comparisons among r.v.'s, equivalently it is a partial binary relation on $\mathcal{F}$. For every $(Y, X) \in \mathcal{C}$ we denote by $Y \preceq X$ the assertion that "*Y is not bigger than X on average*", or "*I expect Y not bigger than X*". The restriction of $\mathcal{C}$ to the set of the events is called *comparative probability* and, for every $(B, A) \in \mathcal{C}$ we denote by $B \preceq A$ the assertion that "*B is not more probable than A*". Possibly a subset $\mathcal{C}'\imath\mathcal{C}$ of strict comparisons is given; in such a case, for every $(Y, X) \in \mathcal{C}'$ we denote by $Y \prec X$ the assertion that "*Y is smaller than X on average*".

A comparative prevision, $\mathcal{C}$, is said to be *coherent on $\mathcal{F}$* if there exists a probability measure P on $\tilde{\mathcal{F}}$ such that

$$Y \preceq X \Rightarrow E^P(Y) \leq E^P(X).$$

being $E^P(X)$ the expected value of X with respect to P; in particular, when only events are considered

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

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

$$Y \prec X \Rightarrow E^P(Y) < E^P(X),$$

we say that $\mathcal{C}$ is *strongly coherent on $\mathcal{F}$*. and that P *strictly represents* (or is *strictly compatible with*) $\mathcal{C}$.

2.1 Coherence.

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 more recently Regoli, 1994 b, for a partial updating, oriented to the present context). Coherence conditions for comparative prevision, such as Buehler's (1974), can be handled by means of linear programming and they only need an arbitrary binary relation on an arbitrary set of r.v.'s. If $\mathcal{C}$ is not coherent, conditions, similar to those quoted above, can check if it is compatible by means of some kind of capacities such as convex capacities or belief functions (Regoli, 1994 a).

Let say that $\mathcal{C} = \{Y_j \preceq X_j, j \in J\}$ is *finite* if J and the algebra $\mathcal{A}_\mathcal{C}$ generated by Y_j, X_j $(j \in J)$ are finite.

Let $\mathcal{C}$ be finite and coherent and let P be a probability on $\mathcal{A}_\mathcal{C}$ representing $\mathcal{C}$; then all σ-additive extensions of P to any σ-algebra containing $\mathcal{A}_\mathcal{C}$ are given by following the mixture procedure used in the next paragraphs. Of course the extension is not unique if $\mathcal{A}_\mathcal{C} \neq \mathcal{F}$ and in general even the probability on $\mathcal{A}_\mathcal{C}$ is not unique.

2.2 The comparative class.

The family, Γ, of probability measures representing $\mathcal{C}$, called *comparative class*, is given by

$$\Gamma = \{P \in Pr(\tilde{\mathcal{F}}),\ E^P(Y_j) \leq E^P(X_j),\ j \in J\}.$$

Family Γ can be described as the close convex hull of its extreme point as follows.

Let $\{C_1, C_2, ...C_n\}$, be the partition generated by $\{X_j, Y_j, j \in J\}$. Pose

$$X_j = \sum \alpha_i^j C_i \text{ and } Y_j = \sum \beta_i^j C_i$$

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

$$\left\{ \sum \alpha_i^j p_i - \sum \beta_i^j p_i \geq 0,\ j \in J \right. \tag{1}$$

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

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 exists a probability measure μ and $b_j \geq 0$, with $\sum\limits_{j=1}^{n} b_j = 1$, such that for every $A \in F$

$$P(A) = \sum_{j=1}^{m} b_j \int_T \sum_{i=1}^{n} q_i^j \delta_{t_i}(A) d\mu(t), \tag{2}$$

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.

Summarizing: Γ is the close convex hull of the probabilities $\sum\limits_{i=1}^{n} q_i^j \delta_{t_i}(A)$.

Obviously, system (**??**) does not have any solution if and only if $\mathcal{C}$ is not representable by any probability measure.

3 Special comparative judgements.

In many problems it is possible to recognize some natural structure useful for describing the given situation. In such a case to guide the probability elicitation trough this structure can be convenient and it can simplify the computation of Γ, as we see in the following.

3.1 Defining a scale.

A particular way of evaluating probabilities is that of looking for a subset of events whose probabilities can be easily or naturally evaluated and use it as a scale (that is compare all other events with them and with their unions). This is the case when the natural comparisons among a subset of events, S, uniquely define a probability measure; for all other events compared with those events, lower and upper bounds of probability are determined. Suppes (1974) proposed this way in order to give a rational foundation to imprecise measurement of beliefs. Depending on the structure of S, the probability imprecision for the events completely compared with those of S, can go from an infimum of $1/2^{n-1}$ to a supremum $1/n$, being n the number of minimal events in S.

The scale proposed by de Finetti (1931) is the simplest: an uniform partition, that is a set of exaustive incompatible events judged equally probable. Usually the reality do not offer an uniform scale, so, one can build an ideal scale (e.g. Good, 1950, and Savage, 1954), for example by imagining an urn containing n balls with different labels and by comparing the events concerning the real problem with those involved in the urn problem. But such comparison is usually not natural and could force the expert judgements and distort his/her evaluation. Moreover even the more general Suppes' scale need some judgement of "perfect equivalence in probability", which is again an idealization: as example when we assert that a coin is fair, we mean that it is roughly well-balanced, that is not exactly physically well-balanced.

3.2 Almost uniform partitions.

In order to avoid the new dogma of "comparative precision", the concept of almost uniform partition can be convenient. It has been defined 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.

Following Savage, a partition, $\{C_i, i = 1, ...n\}$, is said an *n-almost uniform partition* (n-a.u. partition, for short) if

$$\bigcup_{j=1}^{r} C_{i_j} \preceq \bigcup_{j=1}^{r+1} C_{k_j}, \ \forall r < n; \ \forall i_j, k_j \in \{1, ...n\}. \quad (3)$$

Example 1. Let resume the example of the almost fair coin: 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. Let $\theta \in \Theta$ be an unknown parameter. Suppose that in my opinion its density is π_o, and suppose that I am a little bit uncertain about that. Then I can choose a partition, $\{C_i, i = 1, ...n\}$, which is uniform for π_o, assume it as an almost uniform partition and replace π_o by the family of probability distributions representing such an assumption.

This technique can also be used in order to match different opinions of several experts, even if some of them prefer a discrete model.

One can prove (see Regoli, 1995) that a partition is always consistent, provided that it is 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:

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

Therefore given a partition, $\{C_1, C_2, ...C_n\}$, if we only assess that it is an a.u.partition, then it is proper and representable. Moreover the description of the set of all compatible probabilities is very simple: in fact it is determined by the set Δ of all the extreme points of S, given by the $2n$ following points:

$$\left(0, \frac{1}{n-1}, ... \frac{1}{n-1}\right); \quad \left(\frac{1}{n-1}, 0, ... \frac{1}{n-1}\right); \quad ... \left(\frac{1}{n-1}, ... \frac{1}{n-1}, 0\right);$$

$$\left(2, \frac{1}{n+1}, ... \frac{1}{n+1}\right); \quad \left(\frac{1}{n+1}, 2, ... \frac{1}{n+1}\right); \quad ... \left(\frac{1}{n+1}, ... \frac{1}{n+1}, 2\right).$$

On the contrary 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 at least by every probability measure for which $P(C_i) = 1/n$, for all i.

4 Comparative Inference.

For certain problems the goal is a comparison among random variables or events: this is the case either if the goal is to make a decision and choose the best action among some available actions or if it is simply to indicate the most probable event. In such cases, natural whishes arise: is it possible to make some inference from the initial set of comparative previsions? is it sufficient for a conclusion? If the question is about either the most probable event, or about the action giving the best expected value the problem is completely resolved at a theoretical level: it is possible to characterize the pairs of random variables (or pair of events) such that their probabilistic comparison can be deduced from the initial comparative judgements, by the coherence.

Theoretical characterizations (for both weak coherence and strong coherence) can be done (see Regoli, 1994 b, for a review); they implie the following rules for making inference. Note that those rules can be easily implemented in a mechanical support system which only need linear programming.

4.1 Inference rules.

Given a coherent finite set, $\mathcal{C}$, of comparisons among random variables, $\mathcal{C} = \{Y_j \preceq X_j, j \in J\}$, given a pair of random variables, X and Y, which are not already compared, first run the steps i):

i) Check if there exist in $\mathcal{C}$ a finite set of comparisons, $\{Y_j \preceq X_j, j \in F\}$, $F \subset J$, and $\mathbf{y} \in \mathcal{R}^F$, $\mathbf{y} > 0$, such that

$$\sum_{j \in F} y_j (X_j - Y_j) \leq X - Y, \tag{4}$$

if i)

a) if $\mathcal{C}$ is strongly coherent and if in (**??**) there is some j such that $Y_j \prec X_j$, then put $Y \prec X$.

b) if in (**??**) the inequality is a strict inequality, then put $Y \prec X$.

c) if i) and none of a) and b), then put $Y \preceq X$.

ii) Check if there exist in $\mathcal{C}$ a finite set of comparisons, $\{Y'_j \preceq X'_j, j \in F\}$, $F \subset J$, and $\mathbf{y}' \in \mathcal{R}^F$, $\mathbf{y}' > 0$, such that

$$\sum_{j \in F} y'_j (X'_j - Y'_j) \leq Y - X, \tag{5}$$

And follow the analogous rules than in case i).

If both i) and ii) fail, all comparisons between Y and X are allowed by the coherence, then no comparative inference can be done.

Note that, i) and ii) implies that "$Y \sim X$" has to be assumed. Moreover if $\mathcal{C}$ is a strongly coherent prevision, i) and ii) cannot occur together for the pair X, Y, and, if $\mathcal{C}$ is a coherent prevision, i) and ii) cannot occur together if one is a strict inequality. Then if is the case, ii) can be skipped.

Note also that if such rules run for all the pair (A, B), for $A, B \in \mathcal{A}_{\mathcal{C}}$, relation is always a strongly coherent comparative probability: it extends both the two parts of the relation (strong and weak), if the initial comparative prevision/probability is strongly coherent; on the contrary, it forget the initial strong specification and extend the only weak initial comparative prevision, if it is weakly coherent only.

Finally, a set of short rules can be also implemented such as monotonicity, transitivity and additivity.

5 Robust Numerical Inference from Comparative Prevision.

The previous section concerns the case when the goal is some comparison among r.v.'s and characterizes all the comparative inferences which can be done from an initial set of weak and strong comparisons among r.v.'s. Now suppose that our goal is the valuation of a statistical quantity, such as the probability of an event, the (posterior) expected utility of some actions, the moment of a r.v. and so on. This quantity can be of direct interest or can be a way of comparison among r.v.'s, when no sure comparative inference of the previous kind can be done.

Let ψ be the quantity of interest; of course it depends on the probability distribution P, that is $\psi = \psi(P)$. If it is not unique, its range, computed by means of its extreme values gives a measure of the outputs precision.

5.1 Precision of $\psi(\mathbf{P})$.

In all cases quoted above the quantity $\psi(P)$ is a linear or ratio-linear functional. More exactly $\psi(P) = \frac{\phi(P)}{\xi(P)}$, where $\phi(P) = \int_\Theta h(\theta)dP(\theta)$ and $\xi(P) = \int_\Theta k(\theta)dP(\theta)$. In particular let f be the likelihood function which we assume bounded for given data. If ψ is the posterior expectation of a bounded function, h, then

$$\psi(P) = \frac{\int_\Theta h(\theta)f(\theta)dP(\theta)}{\int_\Theta f(\theta)dP(\theta)}. \tag{6}$$

If $P(A) = \int_T P_t(A)d\mu(t) \quad \forall A \in \mathcal{F}$, for some probability measure μ, then $\psi(P)$ is a mixture of $\psi(P_t)$, that is

$$\psi(P) = \int_T \psi(P_t)d\nu(t), \tag{7}$$

where ν is a probability on $\mathcal{T}$.

Then if Γ is a close convex hull of a set $\Lambda\imath\Gamma$ then the range of ψ is determined by

$$\sup \psi(\Gamma) = \sup \psi(\Lambda), \text{ and } \inf \psi(\Gamma) = \inf \psi(\Lambda).$$

Therefore the following theorem holds.

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

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

In particular for a a.u.partition

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

An analogous formula holds for the infimum.

Remark 1. Since the comparative class is a particular moment class (namely a mixture of quantile classes) it can be also treated by all the facilities tuned up for this (Berger, 1994 and Liseo, Moreno and Salinetti, 1995).

Remark 2. It should be noticed that, if a posterior best action has to be chosen, since the denominator in (**??**) and the probability ν in (**??**) do not depend on the actions, then it is sufficient to compare only the values $\sum_{i=1}^{n} q_i^j h(t_i)f(t_i)$ $(t_i \in C_i)$, for different actions h.

6 Conclusions.

Vague information under uncertainty can be technically treated in a rigorous probabilistic framework, in order to make either inference or decisions, even if it is representable just as comparative prevision or comparative probability. Uncertain inputs give uncertain answers, but the degree of imprecision can be computed. Such a process can be summarized in the following steps:

- collect any comparative judgement among uncertain events or quantities.
- represent those judgements by means of a binary relation.
- check the coherence with the probability theory.
    - run the comparative rules, if the goal is either a decision or a comparison.
    - via robust analysis, evaluate the quantity of interest and the degree of its approximation, if the goal is either a quantity or a comparison, but it is not determined by the previous rules.
- if the comparisons are not coherent, or if the accuracy of the analysis is not sufficient, pay more attention to the initial step, correct it or integrate it with additional information and iterate the procedure.

This procedure can be handled by means of an automatic support system, except for the first step.

References

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.

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

de Finetti, B. (1937). La Prevision, ses lois logiques, ses sources subjectives. *Ann. Inst. Poincaré* **7**, 1-68.

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.

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.*

Regoli, G. (1994 a). Rational comparisons and numerical representation. In *Decision Theory and Analysis: Trends and Challenges*, S. Rios ed., 113-126, Kluwer Ac.Pu. Boston

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

Regoli, G. (1995). Comparative probability and robustness *Proceedings of Second International Workshop on Bayesian Robustness, Rimini.*

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

Suppes, P (1974) The measurement of belief. *Proc.Roy. Statist. Soc.*, B, **36**, 160-175.

Walley, P (1991) *Statistical reasoning with imprecise probabilities.* Chapman and Hall, London.

Wasserman, L. (1990). Prior envelops based on belief functions. *Ann. Statist.* **18**, 454-464.

Wasserman, L. and Kadane, J. (1990). Bayes theorem for Choquet capacities. *Ann. Statist.* **18**, 1328-1339.