Ann. Inst. Statist. Math.
Vol. 45, No. 2, 303–315 (1993)

UNBIASED BAYES ESTIMATES AND IMPROPER PRIORS*

GUIDO CONSONNI[1] AND PIERO VERONESE[2]

[1]*Dipartimento di Economia, Politica e Metodi Quantitativi, Università di Pavia, Via S. Felice, 5, I-27100 Pavia, Italy*
[2]*Dipartimento di Economia Politica, Istituto di Metodi Quantitativi, Università Commerciale Luigi Bocconi, Via U. Gobbi, 5, I-20136 Milano, Italy*

(Received March 5, 1991; revised June 30, 1992)

Abstract. Given two random variables (X, Y) the condition of unbiasedness states that: $E(X \mid Y = y) = y$ and $E(Y \mid X = x) = x$ both almost surely (a.s.). If the prior on Y is proper and has finite expectation or non-negative support, unbiasedness implies $X = Y$ a.s. This paper examines the implications of unbiasedness when the prior on Y is improper. Since the improper case can be meaningfully analysed in a finitely additive framework, we revisit the whole issue of unbiasedness from this perspective. First we argue that a notion weaker than equality a.s., named coincidence, is more appropriate in a finitely additive setting. Next we discuss the meaning of unbiasedness from a Bayesian and fiducial perspective. We then show that unbiasedness and finite expectation of Y imply coincidence between X and Y, while a weaker conclusion follows if the improper prior on Y is only assumed to have positive support. We illustrate our approach throughout the paper by revisiting some examples discussed in the recent literature.

Key words and phrases: Coincidence, dF-coherence, equality almost surely, finite additivity, improper prior, unbiasedness.

1. Introduction

This paper was motivated by Bickel and Mallows (1988) (henceforth B&M).

Assuming the standard Kolomogorovian setup of probability theory, let (X, Y) be two random variables satisfying the following condition:

$$\text{(U) unbiasedness} \qquad \begin{aligned} E(X \mid Y = y) &= y \qquad \text{(a.s.)} \\ E(Y \mid X = x) &= x \qquad \text{(a.s.)}, \end{aligned}$$

* This work was partially supported by C.N.R. grant N.80.02970.10 (G.C.) and by C.N.R. grant "altri interventi" (P.V.). A preliminary draft was written while the Authors were visiting the Department of Statistics at Carnegie Mellon University.

The paper is the result of close cooperation between the two authors. However subsections 3.1 and 3.3 are mainly due to G.C. while subsection 3.2 and section 4 are mainly due to P.V.

where a.s. means *almost surely.*

It is shown, in the above paper, that (U) together with either of the following conditions

(FE) finite expectation	$E(\|Y\|) < +\infty$
(NN) non-negativity	$Y \geq 0 \quad \text{(a.s.)}$

leads to $X = Y$ a.s. Of course, one obtains the same conclusion if (FE) and (NN) are referred to X instead of Y.

In a Bayesian framework, the above conditions may be read as follows: given a probability distribution of X conditional on the mean parameter Y, if the prior on Y satisfies (FE) or (NN) and induces a posterior distribution such that $E(Y \mid X = x) = x$, then $X = Y$ a.s. On the other hand this result need no longer be true, according to B&M, when the prior on Y is improper. However, their analysis of this case appears somewhat incomplete, since it does not address explicitly the issue of evaluating the joint probability distribution of (X, Y). We try to fill this gap by adopting a finitely additive approach which allows a probabilistic treatment of improper priors.

The finitely additive nature of the resulting joint probability distribution suggests to reexamine the relevance of the concept of equality a.s. This leads to the adoption of a weaker notion, named coincidence by de Finetti (1937). These issues are discussed in Section 2 which contains also relevant notation and definitions used in the paper.

Section 3 deals with three topics. In Subsection 3.1 we show that, for a fixed statistical model of X given Y, there do not always exist priors on Y inducing coincidence between X and Y: this happens, for example, for location families. On the other hand we also show that, for scale families, such priors do exist. Since unbiasedness appears to be a natural condition to achieve coincidence, we offer some statistical interpretations of condition (U) in Subsection 3.2 from a Bayesian and fiducial perspective.

A critical discussion of B&M's results relative to the improper case is contained in Subsection 3.3. In particular B&M's examples are revisited, from a finitely additive viewpoint, in order to show that, even if equality a.s. fails, coincidence may hold.

Some general results relating to improper priors and unbiasedness are contained in Section 4. In particular we show that (U) and (FE) imply coincidence between X and Y, while a weaker conclusion follows if (FE) is replaced by (NN).

Three short Appendices summarize the main concepts for finitely additive probability distributions that we use in the paper and review a useful technique to evaluate probabilities when the prior is improper.

2. Notation and definitions

Given a set $\mathcal{Z}$ and a σ-field of subsets of $\mathcal{Z}$, $\mathcal{Q}_{\mathcal{Z}}$, a *probability measure* on $\mathcal{Q}_{\mathcal{Z}}$ is a non-negative σ-additive function defined on $\mathcal{Q}_{\mathcal{Z}}$ and taking value 1 on $\mathcal{Z}$.

If the assumption of σ-additivity is replaced by the weaker condition of finite additivity, then the corresponding set function is called *probability.* The natural domain of a probability is any algebra of events. Nevertheless we shall henceforth consider only probabilities defined on σ-algebras.

Let W be a random variable (r.v.) on $(\mathcal{Z}, \mathcal{Q}_{\mathcal{Z}}, \nu)$ where ν is a probability. If B is a Borel set of $\boldsymbol{R}$, the set function $\nu_W(B) = \nu\{z : W(z) \in B\}$ is called the *probability distribution* of W. If $B = (-\infty, w]$ then the function $F_W(w) = \nu_W\{(-\infty, w]\}$ is called the *probability distribution function* (p.d.f.) of W induced by ν_W.

We point out, at this stage, that we shall consider inferential procedures which satisfy the condition of *dF-coherence*, introduced by Regazzini (1987).

We refer the reader to this paper for a thorough treatment of this concept and, in particular, to its Section 3 for a discussion of inference from improper priors within this framework. Notice that for *improper prior* we mean a non-negative, σ-additive, non-finite, σ-finite measure ρ. If, for a given σ-finite measure τ, there exists a density g such that $\rho(B) = \int_B g(z)\tau(dz)$, for all Borel sets B, then g is named improper prior too; indeed the latter interpretation is standard in much of the current statistical literature. Finally a simple and intuitive method to construct a probability from g is illustrated in Appendix 1.

The terms *conditional probability*, *conditional probability distribution* and *conditional p.d.f.* will always be interpreted, in this paper, according to the usual Kolmogorovian setup, i.e. in terms of a Radon-Nikodym derivative, see e.g. Loéve ((1978), Section 30), provided they are dF-coherent.

Assume that the conditions for applying the formal Bayes theorem with an improper prior g are satisfied. Then one may find a probability, π say, such that the standard posterior generated from g through the above rule is dF-coherent relative to the prior π.

Consider two r.v.'s X and Y taking values, respectively, in $\mathcal{X}$ and $\mathcal{Y}$, with $\mathcal{X}$ and $\mathcal{Y}$ Borel subsets of $\boldsymbol{R}$. Let P be a probability defined on the σ-algebra of Borel subsets of $\boldsymbol{R}^2$, generated by $\mathcal{B} \cap \mathcal{X} \times \mathcal{B} \cap \mathcal{Y}$, where $\mathcal{B}$ is the Borel σ-algebra of $\boldsymbol{R}$. If A belongs to this σ-field we shall often write, with a slight abuse of notation, $P\{(X,Y) \in A\}$ instead of $P(A)$. In particular if $A = A_x \times \mathcal{Y}$ with $A_x \subseteq \mathcal{B} \cap \mathcal{X}$, then we shall also write $P(A) = P(X \in A_x)$.

DEFINITION 2.1. X and Y are said to be *a.s. equal* with respect to P if

$$P\{X = Y\} = 1 \tag{2.1}$$

written $X = Y$ a.s. $[P]$.

In the sequel we shall omit the symbol $[P]$ whenever it is clear from the context.

It is well known that, if P is a probability measure, then (2.1) is equivalent to

$$P\{|X - Y| < \epsilon\} = 1 \qquad \forall \epsilon > 0. \tag{2.2}$$

When σ-additivity does not hold, (2.1) implies (2.2), while the converse is not generally true. This happens because the probability associated to the r.v. $|X - Y|$ might present an adherent mass to zero; see Appendix 3 for the definition of adherent mass.

On the other hand (2.2) is often perfectly adequate to describe the notion of equality in practical terms (e.g. if X and Y are two measurements, it is experimentally impossible to ascertain their value without error).

Furthermore current mathematical literature on charges (e.g. Bhaskara Rao and Baskara Rao ((1983), p. 88)) adopts (2.2) as the standard definition of equality a.s., since results involving (2.1) in a σ-additive context typically have a natural counterpart in a finitely additive setting if (2.1) is replaced with (2.2).

However to emphasize the distinction between (2.1) and (2.2) we adopt the following

DEFINITION 2.2. (de Finetti (1937)). X and Y are said to be *coincident* with respect to P if

$$P\{|X - Y| < \epsilon\} = 1, \qquad \forall \epsilon > 0,$$

written $X \approx Y$ $[P]$.

In the sequel we shall assume that a statistical model, i.e. a conditional probability distribution for X, given $Y = y$, is given for every $y \in \mathcal{Y}$, and that an improper prior g is assigned to Y.

Having established that a probability π can be associated to g, we can use π, in conjunction with the statistical model, to generate a probability P for (X, Y); see Appendix 2. Since P need not be σ-additive, we shall examine the implications of g on P in terms of the notion of coincidence (2.2).

3. Unbiasedness and improper priors

3.1 *Prior probabilities and coincidence*

As recalled in the Introduction, B&M's results may be seen as providing conditions, on the prior for the mean parameter Y, which imply $X = Y$ a.s.

A preliminary question which appears natural to ask concerns the *existence* of such priors. We shall however replace the condition of equality a.s. with that of coincidence because of the previous remarks.

The following two propositions provide an answer for two important classes of statistical models.

PROPOSITION 3.1. *Let F be a continuous p.d.f., induced by a probability measure, with support the real line. For $y \in (-\infty, +\infty)$, let the conditional p.d.f. of X, given $Y = y$, belong to the location family $F(x - y)$. Then there exists no prior for Y, for which $X \approx Y$.*

PROOF. Let π be a prior for Y. Then

$$\begin{aligned} P\{|X - Y| < \epsilon\} &= \int_{-\infty}^{+\infty} (F(\epsilon) - F(-\epsilon))\pi(dy) \\ &= F(\epsilon) - F(-\epsilon) < 1 \qquad \forall \epsilon > 0. \end{aligned}$$

Here, and in the sequel, integrals are of Stieltjes type; see Bhaskara Rao and Bhaskara Rao ((1983), p. 115). □

PROPOSITION 3.2. *Let F be a continuous p.d.f., induced by a probability measure, with support $(0, +\infty)$. Given any $y \in (0, +\infty)$ let the conditional p.d.f. of X, given $Y = y$, belong to the scale family $F(x/y)$. Then, if π denotes the prior of Y, a necessary and sufficient condition for $X \approx Y$ is $\pi\{(0, \epsilon]\} = 1$ $\forall \epsilon > 0$, i.e. π assigns unitary adherent mass to the right of $y = 0$.*

PROOF. i) Necessity.

$$\begin{aligned} &P\{|X - Y| \leq \epsilon\} \\ &\quad = \int_0^\epsilon F(1 + \epsilon/y)\pi(dy) + \int_\epsilon^{+\infty} (F(1 + \epsilon/y) - F(1 - \epsilon/y))\pi(dy) \\ &\quad \leq \pi\{(0, \epsilon]\} + F(2)(1 - \pi\{(0, \epsilon)\}). \end{aligned}$$

Since $F(2) < 1$, (2.2) implies $\pi\{(0, \epsilon]\} = 1$ $\forall \epsilon > 0$.

ii) Sufficiency. Assume now $\pi\{(0, \epsilon]\} = 1$ for all $\epsilon > 0$. Then

$$\begin{aligned} P\{|X - Y| < \epsilon\} &= \int_0^\epsilon F(1 + \epsilon/y)\pi(dy) \\ &= \lim_{y \to 0^+} F(1 + \epsilon/y) = 1. \end{aligned} \qquad \square$$

Remark 1. In the above proof we have used the standard procedure to evaluate the joint probability P relative to (X, Y), for a fixed conditional probability distribution of X given Y and prior for Y. Notice however that this procedure, while legitimate, is not compulsory in the setup of dF-coherent probabilities.

Remark 2. A prior π with $\pi\{(0, \epsilon]\} = 1$ $\forall \epsilon > 0$, is said to assign unitary adherent mass to the right of 0. Finitely additive priors with total mass adherent either to $\inf \mathcal{Y}$ or $\sup \mathcal{Y}$ (as π in Proposition 3.2) typically correspond to improper priors according to the limiting procedure described in Appendix 1.

Remark 3. A non-intuitive corollary of Proposition 3.2 is that $X \approx Y$ may hold even without unbiasedness being satisfied as in the following example: let the conditional p.d.f. of X, given $Y = y$, be negative exponential with scale parameter $1/y$ and let π be generated by the improper prior $g(y) = 1/y^3$ $(y > 0)$. In this case $E(Y \mid X = x) = x/2$, as a standard prior to posterior computation shows, and $\pi\{(0, \epsilon]\} = 1$ for all $\epsilon > 0$.

Remark 4. Notice that Propositions 3.1 and 3.2 still hold if the condition of coincidence between X and Y is replaced with that between X and Y^*, where $Y^* = E(X \mid Y)$, i.e. $Y^* = Y + K$ in Proposition 3.1 and $Y^* = KY$ in Proposition 3.2, with $K = \int z dF(z)$. In statistical terms changing Y to Y^* is equivalent to reparametrizing the model in terms of the mean-parameter.

3.2 *On the notion of unbiasedness*

The previous subsection has shown that, given a conditional probability distribution for X given Y (i.e. a statistical model such as the scale family), there exist priors for Y such that $X \approx Y$.

Going back to B&M's results, we now discuss the unbiasedness condition (U) which appears to be quite natural in order to realize coincidence between X and Y.

Notice first that (U) is stated for one observation X and generalizes to a sample $(X_1, \ldots, X_n)$ $(n > 1)$ only when a real valued sufficient statistic is available.

Secondly we remark that condition (U) may be interpreted as saying that our prior information on Y is so poor that, whatever the experimental outcome X, this will be taken as an estimate of Y. Hence (U) can be regarded as a condition of "non-informativity" for the prior on Y.

Suppose for example that $(X_1, \ldots, X_n)$ is a sample from a real regular natural exponential family (see Brown (1986)) so that (U) may be rewritten in terms of the sufficient statistic $\bar{X}_n = n^{-1} \sum X_i$ as

$$E(\bar{X}_n \mid \mu) = \mu, \quad \text{a.s.}, \qquad E(\mu \mid \bar{X}_n) = \bar{X}_n, \quad \text{a.s.},$$

where $\mu = \mu(\theta) = E(X_i \mid \theta)$, and θ is the natural parameter.

Using Theorems 1 and 2 of Diaconis and Ylvisaker (1979) and Theorem 3.3 of Cifarelli and Regazzini (1987) it is straightforward to show that, under some regularity conditions, the improper prior $g(\theta) = c$, with $c > 0$, is the only one that satisfies (U).

This result is consistent with the standard practice of viewing a constant prior as non-informative.

Finally we point out that condition (U) is always satisfied by a fiducial probability distribution whenever it can be regarded as an actual Bayesian posterior distribution. To see this recall a result by Lindley (1958) according to which a fiducial probability distribution for the real parameter Y^* is, under some regularity conditions on the model, Bayesian posterior if and only if the p.d.f. of X^* given $Y^* = y^*$ (where X^* is a sufficient statistic) can be written as $H(u(x^*) - v(y^*))$ for some p.d.f. H (induced by a probability measure), and increasing functions $u(\cdot)$ and $v(\cdot)$ with a, possibly improper, uniform prior on $v(Y^*)$. Since H can be assumed, without loss of generality, to have expectation zero, setting $X = u(X^*)$ and $Y = v(Y^*)$, it follows that $E(X \mid Y = y) = y$ and $E(Y \mid X = x) = \int yh(x-y)dy / \int h(x-y)dy = x$ (where h is the density corresponding to H), so that (U) is satisfied. We conclude that, when a fiducial distribution is also a posterior distribution, the unbiasedness condition (U) is always satisfied by suitable X and Y.

3.3 *A discussion of Bickel and Mallows' results for the improper case*

The implications of (U) when the prior is improper have been investigated by B&M. They claim that (U) and (NN) can fail to imply $X = Y$ a.s. and offer two examples to support this conclusion.

If however, besides (U) and (NN), the further condition that "the supports of X and Y are discrete with X having no point of accumulation" is added, then $X = Y$ a.s. still holds (Theorem 4 of B&M).

Condition (FE) is not mentioned by B&M presumably because it is deemed to be irrelevant, from their viewpoint, in the presence of improper priors.

To better appreciate their argument and illustrate our point of view consider the following example.

Example 3.1. (Example 2 of B&M) Let W and Y be independent r.v.'s, with W having a continuous p.d.f. H, induced by a probability measure, having support $(0,+\infty)$ and such that $E(W)=1$. Set $X=WY$; then $E(X \mid Y=y)=y$. If $g(y)=1/y^2$ $(y>0)$ is the improper prior for Y (so that (NN) holds), then standard calculations lead to $E(Y \mid X=x)=x$, so that (U) holds. On the other hand (2.1) fails since $X=WY$ with W not degenerate.

This conclusion is correct and misleading at the same time, since it hides an essential feature of the joint probability distribution of (X,Y) which is possible to explore if a finitely additive approach is embraced. We will now show (see the continuation of Example 3.1 below) that indeed X and Y are coincident, i.e. (2.2) holds. Therefore we do not see this example as providing convincing evidence that improper priors behave differently from proper priors (which are actually probability measures), since we regard definition (2.2) more meaningful in this context than (2.1).

Example 3.1. (continued) Following the procedure outlined in Appendix 1 and 2, we first derive the marginal p.d.f. of Y, G say. Recalling that $\mathcal{Y}=(0,+\infty)$, take the sequence of intervals $\{(\alpha,\beta);\alpha\to 0^+,\beta\to+\infty\}$. Then

$$\begin{aligned} G(y) &= \lim_{\substack{\alpha\to 0^+\\ \beta\to+\infty}} \left(\int_{(-\infty,y]\cap(\alpha,\beta)} 1/t^2dt\right)\Big/\left(\int_\alpha^\beta 1/y^2dy\right)\\ &= \lim_{\substack{\alpha\to 0^+\\ \beta\to+\infty}} \begin{cases} 0 & y\le\alpha\\ \dfrac{\int_\alpha^y 1/t^2dt}{\int_\alpha^\beta 1/y^2dy} & \alpha<y<\beta\\ 1 & y\ge\beta \end{cases}\\ &= \begin{cases} 0 & y\le 0\\ 1 & y>0. \end{cases} \end{aligned}$$

Notice that G is not right-continuous. The quantity $G(0^+)-G(0)=1$ represents the probability adherent (to the right) of zero. This is a typical example of a non-σ-additive p.d.f.

Since, for $y>0$, $G(y)=1$, it follows that $P\{X\le x, Y\le y\}=P\{X\le x\}$. Now, for $x>0$,

$$\begin{aligned} P\{X\le x\} = P\{WY\le x\} &= \int_0^{+\infty}\int_0^{x/y} dH(w)dG(y)\\ &= \int_0^{+\infty} H(x/y)dG(y) = \lim_{y\to 0^+} H(x/y) = 1. \end{aligned}$$

Therefore

$$P\{X \leq x, Y \leq y\} = 1 \qquad x, y > 0$$

whence

$$P\{|X - Y| > \epsilon\} = 0 \qquad \forall \epsilon > 0$$

i.e. $X \approx Y$.

The above discussion applies also to

Example 3.2. (Example 3 of B&M) Let W and Y be independent r.v.'s each taking values in the set $\{2^j : j = 0, \pm 1, \pm 2, \ldots\}$. Assume that the p.d.f. of W is induced by a probability measure, while the prior on Y is improper with $g(y) = 1/y$, $y = 2^j$; $j = 0, \pm 1, \pm 2, \ldots$.

As in Example 3.1 take $E(W) = 1$ and $X = WY$. It is easy to show that (NN) and (U) are satisfied.

Taking the sequence $\{\{2^{-m}, \ldots, 2^n\}, m, n, = 1, 2, \ldots\}$ we conclude, through the usual limiting procedure, that the p.d.f. of Y presents again unitary adherent mass at the point zero. Arguing as before we therefore derive that $X \approx Y$.

As a consequence, B&M's remark that the presence of a point of accumulation (at 0) in the support of Y (and X) implies failure of equality a.s. appears to be irrelevant as far as coincidence is concerned.

This issue is further pursued in the next example which shows that even if X has no point of accumulation, then (U) and (NN) do not imply $X \approx Y$.

Example 3.3. Let W be a non degenerate r.v. taking values in the set of integers lying between $-K$ and K (K a positive integer), and having a p.d.f., induced by a probability measure, with density h such that $E(W) = 0$. Let Y take values in the set $\mathcal{Y} = \{0, \pm 1, \pm 2, \ldots\}$ and let $g(y) = c$ with $c > 0$, be the improper prior for Y. Assume Y independent of W and set $X = W + Y$.

Choosing the sequence of sets $\{-m, \ldots, 2^m\}$ $(m = 0, 1, \ldots)$ which converges to $\mathcal{Y}$, it is straightforward to show that Y presents unitary adherent mass at $+\infty$, so that (NN) holds. Since $X = W + Y$, X will also be a.s. non-negative. It is immediate to check that $E(X \mid Y = y) = y$ since $E(W) = 0$.

Adopting the rule employed by B&M to compute $E(Y \mid X = x)$, which could be formally justified using our limiting procedure, we have:

$$E(Y \mid X = x) = \frac{\sum_{y=-\infty}^{+\infty} y h(x - y) g(y)}{\sum_{y=-\infty}^{+\infty} h(x - y) g(y)} = \frac{\sum_{z=-\infty}^{+\infty} (x - z) h(z)}{\sum_{z=-\infty}^{+\infty} h(z)} = x.$$

Consequently (U) is also satisfied. Nevertheless, since W is neither degenerate on zero, nor admits adherent mass to zero, it follows that $X \approx Y$ cannot hold.

4. Some general results

The present section discusses, in a general setting, the implications of unbiasedness when the prior is improper.

Having chosen to work with dF-coherent inferences, we shall evaluate probabilities, in the presence of an improper prior g, as follows:

A1: the prior probability π is computed from g according to the limiting procedure outlined in Appendix 1.

A2: given a conditional probability distribution for X given Y admitting density, the probability relative to (X, Y) and the posterior distribution of Y given X (derived through a formal application of Bayes theorem) are computed as in Appendix 2.

THEOREM 4.1. *Let X and Y be r.v.'s taking values respectively in $\mathcal{X} \subseteq (-\infty, +\infty)$ and $\mathcal{Y} = (A, B)$ with $-\infty \leq A < B \leq +\infty$. Assume a conditional probability distribution for X, given Y, is assigned and that Y has an improper prior g with g continuous.*

Then, if the probability assessments are made according to A1 *and* A2, (U) *and* (FE) *imply $X \approx Y$.*

PROOF. Since g is continuous, A1 implies that the prior π will present an adherent mass a, say, to A and $(1-a)$ to B $(0 \leq a \leq 1)$. If $0 < a < 1$, then (FE) implies that A and B must be both finite. Indeed if either A or B is infinite then $E(|Y|) = +\infty$, because of the presence of positive adherent masses at A and B, violating (FE). To see this assume for simplicity $A > -\infty$ and $B = +\infty$. Then

$$E(|Y|) \geq \int_A^n |y| \pi(dy) + n\pi\{(n, +\infty)\} \qquad \text{for all } n > 0,$$

so that $E(|Y|) = +\infty$, since $\pi\{(n, +\infty)\} = 1 - a > 0$ for all $n > 0$.

Next we show that $\inf \mathcal{X} = \inf \mathcal{Y} = A$ and $\sup \mathcal{X} = \sup \mathcal{Y} = B$. Suppose $\inf \mathcal{X} < A$; then there exists $x \in \mathcal{X}$ with $\inf \mathcal{X} \leq x < A$ so that, by (U), $E(Y \mid X = x) = x < A$, which is impossible. Next suppose $\inf \mathcal{X} > A$; then there exists $y \in \mathcal{Y}$ with $A \leq y < \inf \mathcal{X}$ so that, by (U), $E(X \mid Y = y) = y < \inf \mathcal{X}$, which is impossible. Hence $\inf \mathcal{X} = \inf \mathcal{Y} = A$. Similarly one shows that $\sup \mathcal{X} = \sup \mathcal{Y} = B$.

Assume now $A < 0 < B$ and denote with F_y the conditional p.d.f. of X, given $Y = y$. From condition (U) we have

$$A = \lim_{y \to A^+} E(X \mid Y = y) = \lim_{y \to A^+} \left\{ \int_0^B (1 - F_y(x))dx - \int_A^0 F_y(x)dx \right\}$$

whence

$$B - A = \lim_{y \to A^+} \int_A^B F_y(x)dx.$$

Since $F_y(x) \leq 1 \ \forall x$, we can apply Fatou's lemma and obtain

$$(B-A) \leq \int_A^B \limsup_{y \to A^+} F_y(x)dx \leq (B-A);$$

whence, $\limsup F_y(x) = 1$ for $x > A$.

Recall now that, letting p_y and ν represent the Lebesgue-Stieltjes measures corresponding to F_y and to a p.d.f. F, respectively, the condition $\limsup p_y(C) \leq \nu(C)$ for all closed sets $C \subseteq (-\infty, +\infty)$ is equivalent to $F_y \to F$ weakly (see Rao ((1984), p. 222)). Since $\limsup p_y(C) = 1$ for all $C = [x_0, x]$ ($x_0 \leq A$, $x > A$) the above condition implies $\nu(C) = 1$, so that:

$$F(x) = \begin{cases} 0 & x < A \\ 1 & x \geq A. \end{cases} \tag{4.1}$$

The same results hold if A and B ($B > A$) are both non-positive or non-negative.

Denote, as usual, with P the probability relative to (X, Y). Then

$$\begin{aligned} P\{X \leq x, Y \leq y\} &= \int_A^y F_t(x)\pi(dt) \\ &= a \lim_{t \to A^+} F_t(x) \qquad \text{because of the structure of } \pi \\ &= aF(x) = a \qquad \text{because of (4.1)} \end{aligned} \tag{4.2}$$

for $x > A$ and $A < y < B$.

Furthermore it is immediate to verify that

$$P\{X > x, Y > y\} = 1 - a \qquad x < B, \quad y < B. \tag{4.3}$$

Hence P presents, when X and Y are both continuous, adherent mass a at the point (A, A) and $(1-a)$ at (B, B), whereas, if X and Y are both discrete, the masses will be concentrated, with natural modifications for the mixed case.

We can therefore conclude that $X \approx Y$.

It remains to examine the case in which $a = 0$ or $a = 1$. When $a = 0$ condition (FE) only implies that $B < +\infty$. Furthermore (4.2) implies $P\{X \leq x, Y \leq y\} = 0$ ($x > A$, $A < y < B$) whence the whole mass is adherent (or concentrated) to (B, B), so that $X \approx Y$ again.

When $a = 1$, a similar argument leads to the conclusion that $A > -\infty$ and $P\{X > x, Y > y\} = 0$ ($x < B$, $y < B$), so that $X \approx Y$ again. □

From Example 3.3 we know that (NN) and (U) do not necessarily imply $X \approx Y$, the existence of a point of accumulation in the support of Y being irrelevant.

Yet in both examples the set $\mathcal{Y}$ is unbounded and it can be checked that the following result holds:

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\} = P\{Y \leq y\} \qquad -\infty < x, \quad y < +\infty,$$

so that, in particular,

$$P\{X \leq x, Y \leq x\} = P\{X \leq x\} = P\{Y \leq x\} \qquad -+\infty < x < +\infty. \tag{4.4}$$

As remarked in Cifarelli and Regazzini (1987), condition (4.4), within the context of σ-additivity, is equivalent to $X = Y$ a.s. whereas if, as it happens in our case, $\lim_{y \to -\infty} P\{Y \leq y\} > 0$ or/and $\lim_{y \to +\infty} P\{Y \leq y\} < 1$, (4.4) only states that

$$P\{|X - Y| > \epsilon \text{ and } |Y| < C\} = 0 \qquad \forall \epsilon > 0 \quad \text{and} \quad 0 < C < +\infty,$$

so that, from (4.4), in general we cannot even conclude that $X \approx Y$. Consequently (4.4) will be taken to identify only a condition of *perfect linear association* between X and Y.

The essential difference between perfect linear association and coincidence is that, under the former, the adherent mass at $(+\infty, +\infty)$ or/and $(-\infty, -\infty)$ is allowed to lay outside the region $|x - y| \leq \epsilon$.

The following theorem shows that a slight modification of the assumptions of unbiasedness and positivity actually induces a condition of perfect linear association.

THEOREM 4.2. *Under the same conditions of Theorem* 4.1, (U) *and* $\inf \mathcal{Y} > -\infty$ (*or* $\sup \mathcal{Y} < +\infty$) *imply that* X *and* Y *are perfectly linearly associated.*

PROOF. Assume $\inf \mathcal{Y} = A > -\infty$. Condition (U) implies that $\inf \mathcal{X} = \inf \mathcal{Y}$ and $\sup \mathcal{X} = \sup \mathcal{Y} = B \leq +\infty$.

If $B < +\infty$, following the proof of Theorem 4.1, deduce (4.3) which implies perfect linear association.

Assume now $B = +\infty$ and, for simplicity, take $A = 0$. Then, arguing as in Theorem 4.1, from $\lim E(X \mid Y = y) = 0$, $(y \to 0^+)$ deduce

$$0 \geq \int_0^{+\infty} \liminf_{y \to 0^+} (1 - F_y(x))dx \geq 0,$$

whence $F(x) = \lim_{y \to 0^+} F_y(x) = 1$ $(x > 0)$ and the conclusion follows from (4.2). If $B = +\infty$ and A is an arbitrary constant (or $B < +\infty$ and $A \geq -\infty$), a similar argument leads to the conclusion. □

Acknowledgements

We wish to thank Patrizia Berti, David Lane and Pietro Rigo for helpful comments. We also benefited from discussions of this paper at seminars held at Carnegie Mellon University and University of Minneapolis. The suggestions of a referee are also gratefully acknowledged.

Appendix

1. *Improper priors in a finitely additive context*

Let Y be a r.v., taking values in $\mathcal{Y}$, having an improper prior g so that, for a given σ-finite measure τ,

$$\int_{\mathcal{Y}} g(y)\tau(dy) = +\infty.$$

We outline, omitting measurability details, a method to associate a probability to g; see Regazzini ((1987), Section 3).

Let $\{\mathcal{Y}_n\}$ be a sequence of subsets of $\mathcal{Y}$ converging to $\mathcal{Y}$ from below such that

$$0 < \rho(\mathcal{Y}_n) = \int_{\mathcal{Y}_n} g(y)\tau(dy) < +\infty \qquad \text{for all } \mathcal{Y}_n. \tag{A.1}$$

Then the probability that $Y \in B \subseteq \mathcal{Y}$, conditional on $Y \in \mathcal{Y}_n$, is given by

$$\pi_n(B) = \rho(B \cap \mathcal{Y}_n)/\rho(\mathcal{Y}_n).$$

Finally the (non-σ-additive) prior π induced by g is computed as $\pi(B) = \lim_{n\to+\infty} \pi_n(B)$, whenever the limit exists.

2. *Evaluation of joint probabilities and dF-coherent posteriors for improper priors*

Assume, as in standard statistical applications, that a conditional probability for X given Y is assigned having density f_y, and that Y has an improper prior g. The previous procedure of first performing the analysis conditional on $Y \in \mathcal{Y}_n$ and then passing to the limit may be applied also to compute probabilities relative to (X, Y) and to obtain a dF-coherent posterior for Y.

Notice that, if $0 < \int_{\mathcal{Y}} f_y(x)g(y)\tau(dy) < +\infty$, such a posterior, q_x say, may be also obtained using a standard formal Bayes calculation, i.e.

$$q_x(dy) = \frac{f_y(x)g(y)\tau(dy)}{\int_{\mathcal{Y}} f_y(x)g(y)\tau(dy)};$$

see Regazzini ((1987), p. 856).

3. *Adherent masses*

Let π be a probability. Define $G(y) = \pi\{(-\infty, y]\}$ and $G_l(y) = \pi\{(-\infty, y)\}$.

If $(G_l(y) - \lim_{t\to y^-} G(t)) = p_l(y) > 0$, then $p_l(y)$ is said to be the probability adherent to the left of y. Similarly if $(\lim_{t\to y^+} G(t) - G(y)) = p_r(y) > 0$, then $p_r(y)$ is said to be the probability adherent to the right of y. If π is also σ-additive then $p_l(y) = p_r(y) = 0$ for all y.

Furthermore if $G(y) = p_0$ for all $y \le y_0$, say, G presents right adherent mass p_0 to $-\infty$; similarly if $1 - G(y) = p_0^*$ for all $y > y_0^*$ then G presents left adherent mass p_0^* to $+\infty$.

References

Bhaskara Rao, K. P. S. and Bhaskara Rao, M. (1983). *Theory of Charges*, Academic Press, London.

Bickel, P. J. and Mallows, C. L. (1988). A note on unbiased Bayes estimates, *Amer. Statist.*, **42**, 132–134.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families*, Institute of Mathematical Statistics, Hayward, California.

Cifarelli, D. M. and Regazzini, E. (1987). Priors for exponential families which maximize the association between past and future observations, *Probability and Bayesian Statistics* (ed. R. Viertl), 83–95, Plenum Press, New York.

de Finetti, B. (1937). A proposito di "correlazione", *Supplemento Statistico ai Nuovi Problemi di Politica, Storia ed Economia*, **3**, 41–57.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families, *Ann. Statist.*, **7**, 269–281.

Lindley, D. V. (1958). Fiducial distributions and Bayes theorem, *J. Roy. Statist. Soc. Ser. B*, **20**, 102–107.

Loéve, M. (1978). *Probability Theory II*, 4th ed., Springer, New York.

Rao, M. M. (1984). *Probability Theory with Applications*, Academic Press, Orlando.

Regazzini, E. (1987). de Finetti's coherence and statistical inference, *Ann. Statist.*, **15**, 845–864.