The "statistical experiment"-equivalence for prior distributions

Claudio Macci

Abstract

Two prior distributions are said to be $(P^a : a \in A)$-equivalent when they have in common all the families of posterior distributions (with respect to a fixed statistical experiment $(P^a : a \in A)$).
It is shown that two $(P^a : a \in A)$-equivalent prior distributions are necessarily mutually absolutely continuous and two cases of statistical experiment in some sense opposite are presented.
Furthermore a partial order for statistical experiments can be defined in a natural way by comparing the quotient sets of prior distributions w.r.t. the $(P^a : a \in A)$-equivalences.
Finally a result about the ϵ-contaminations is presented.

1 Introduction and preliminaries

In this paper we shall refer to the frame of *Bayesian experiments* (see e.g. [5]). Throughout this paper we shall denote the *parameter space* by $(A, \mathcal{A})$ and the *sample space* by $(S, \mathcal{S})$ and we shall assume they are two Polish spaces. Then, given a Markov kernel $(P^a : a \in A)$ from $(A, \mathcal{A})$ to $(S, \mathcal{S})$ and a probability measure μ on $\mathcal{A}$, we can consider the probability measure $\Pi_{\mu,(P^a:a\in A)}$ on $\mathcal{A} \otimes \mathcal{S}$ such that

$$\Pi_{\mu,(P^a:a\in A)}(E \times X) = \int_E P^a(X) d\mu(a), \quad \forall E \in \mathcal{A} \ \textit{and} \ \forall X \in \mathcal{S}. \tag{1}$$

Received by the editors January 1997 – In revised form August 1997.
Communicated by M. Hallin.
1991 *Mathematics Subject Classification :* 62A15, 04A05, 06A06.
Key words and phrases : Bayesian experiment, Markov kernel, equivalence, partial order, ϵ-contaminations.

Then, for the *Bayesian experiment* $\mathcal{E} = (A\times S, \mathcal{A}\otimes\mathcal{S}, \Pi_{\mu,(P^a:a\in A)})$, we can say that μ is the *prior distribution* and $(P^a : a \in A)$ are the *sampling distributions*; $(P^a : a \in A)$ is also called *statistical experiment*. Furthermore we can also say that $\mathcal{E}$ is *regular* (see e.g. Remark (i) in [5], page 31), i.e. we have a Markov kernel $(\mu^s : s \in S)$ from $(S, \mathcal{S})$ to $(A, \mathcal{A})$ such that

$$\Pi_{\mu,(P^a:a\in A)}(E \times X) = \int_X \mu^s(E)dP_\mu(s) \ \ (\forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}), \tag{2}$$

where P_μ is the *predictive distribution*, i.e. the probability measure on $\mathcal{S}$ such that

$$X \in \mathcal{S} \mapsto P_\mu(X) = \Pi_{\mu,(P^a:a\in A)}(A \times X). \tag{3}$$

A Markov kernel $(\mu^s : s \in S)$ satisfying (2) is called family of *posterior distributions* and, when we need, we shall denote condition (2) as follows:

$$\Pi_{\mu,(P^a:a\in A)}(da, ds) = \mu^s(da)P_\mu(ds).$$

The aim of this paper is to define and to study the $(P^a : a \in A)$-*equivalence* between prior distributions (where $(P^a : a \in A)$ is a fixed statistical experiment); as we shall see in Definition 1, two prior distributions are said to be $(P^a : a \in A)$-equivalent when they have in common all the families of posterior distributions with respect to the statistical experiment $(P^a : a \in A)$.

In Section 2 we shall define the $(P^a : a \in A)$-equivalence and we shall show that, for any statistical experiment $(P^a : a \in A)$, two $(P^a : a \in A)$-equivalent prior distributions are necessarily mutually absolutely continuous.

In Section 3 we shall consider two different cases concerning the statistical experiment $(P^a : a \in A)$ which are in some sense opposite. Indeed, for the first one, two prior distributions are $(P^a : a \in A)$-equivalent if and only if they coincide while, for the second one, the mutual absolute continuity between two prior distributions is sufficient for their $(P^a : a \in A)$-equivalence. Furthermore in Section 4 we shall present an example which represents an intermediate case.

In Section 5, we shall study some properties of a partial order for the family of the statistical experiments; this partial order can be defined in a natural way by comparing the quotient sets of prior distributions w.r.t. the $(P^a : a \in A)$-equivalences.

Finally we shall present a result about the ϵ-*contaminations* (Section 6) and some concluding remarks.

At the end of this Section it is useful to introduce some notation.
The family of Markov kernels from $(A, \mathcal{A})$ to $(S, \mathcal{S})$ (i.e. the family of statistical experiments) will be denoted by $\kappa(\mathcal{A}, \mathcal{S})$;
the family of Markov kernels from $(S, \mathcal{S})$ to $(A, \mathcal{A})$ will be denoted by $\kappa(\mathcal{S}, \mathcal{A})$;
the family of probability measures on $\mathcal{A}$ will be denoted by $\Phi(\mathcal{A})$ (in this paper it will be seen as the family of prior distributions);
the symbol $\equiv$ will denote the mutual absolute continuity between two positive measures;
given a set $C \in \mathcal{A} \otimes \mathcal{S}$, we shall consider the notation

$$C(a, .) = \{s \in S : (a, s) \in C\}, \ \ \forall a \in A$$

and

$$C(.,s) = \{a \in A : (a,s) \in C\}, \quad \forall s \in S.$$

Finally let us recall two propedeutic results: the first one follows from Lemma 7.4 in [6] (page 287), for the second one see [4] (Theorem, page 57).

Proposition 1. *Let* $(P^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ *and* $\mu \in \Phi(\mathcal{A})$ *be such that*

$$\mu(\{a \in A : P^a \ll \lambda\}) = 1$$

for a suitable σ*-finite measure* λ.
Then we have a jointly measurable function f_λ *such that*

$$\mu(\{a \in A : X \in \mathcal{S} \mapsto P^a(X) = \int_X f_\lambda(a,s)d\lambda(s)\}) = 1 \tag{4}$$

and

$$P_\mu(\{s \in S : E \in \mathcal{A} \mapsto \mu^s(E) = \frac{\int_E f_\lambda(a,s)d\mu(a)}{\int_A f_\lambda(a,s)d\mu(a)}\}) = 1. \tag{5}$$

Proposition 2. *Let* $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ *be such that*

$$Q^a \ll P^a, \quad \forall a \in A. \tag{6}$$

Then there exists a jointly measurable function h *such that*

$$X \in \mathcal{S} \mapsto Q^a(X) = \int_X h(a,s)dP^a(s), \quad \forall a \in A.$$

2 The $(P^a : a \in A)$-equivalence for prior distributions

Let us start with the definition of $(P^a : a \in A)$-equivalence on $\Phi(\mathcal{A})$ (where $(P^a : a \in A)$ is a fixed statistical experiment).

Definition 1. μ_1 *and* μ_2 *are said to be* $(P^a : a \in A)$-equivalent *if and only if the two following conditions hold:*

$$\exists(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A}) : \quad \Pi_{\mu_k,(P^a:a\in A)}(da, ds) = \mu^s(da)P_{\mu_k}(ds), \quad (k = 1, 2); \tag{7}$$

$$P_{\mu_1} \equiv P_{\mu_2}. \tag{8}$$

In this case we shall write $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$.

In other words we have $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$ if and only if the Bayesian experiments

$$\mathcal{E}_k = (A \times S, \mathcal{A} \otimes \mathcal{S}, \Pi_{\mu_k,(P^a:a\in A)}) \quad (k = 1, 2)$$

have in common all the families of posterior distributions; indeed, almost surely w.r.t. the predictive distribution, any Bayesian experiment has a unique family of posterior distributions.

The next proposition allows to express the $(P^a : a \in A)$-equivalence between two prior distributions by a more easily-handled equivalent condition.

Proposition 3. *We have* $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$ *if and only if (7) and*

$$\mu_1 \equiv \mu_2. \tag{9}$$

contemporary hold.

Proof. Assume that $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$. Then we have (7) while (9) follows from (8); indeed, by taking into account (1), (2) and (7), we obtain

$$\mu_1(E) = \int_S \mu^s(E) dP_{\mu_1}(s) = 0 \Leftrightarrow \mu_2(E) = \int_S \mu^s(E) dP_{\mu_2}(s) = 0.$$

Assume that (7) and (9) contemporary hold. Then we obtain $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$ deriving (8); indeed, by taking into account (3) and (1), we have

$$P_{\mu_1}(X) = \int_A P^a(X) d\mu_1(a) = 0 \Leftrightarrow P_{\mu_2}(X) = \int_A P^a(X) d\mu_2(a) = 0.$$

■

In what follows we shall use the symbol $[\mu]_{(P^a:a\in A)}$ to denote the $(P^a : a \in A)$-equivalence class of μ while the symbol $[\mu]_\equiv$ will be used to denote the class of all the prior distributions mutually absolutely continuous w.r.t. μ. In other words we set

$$[\mu]_{(P^a:a\in A)} = \{\nu \in \Phi(\mathcal{A}) : \nu \wp \mu (mod(P^a : a \in A))\}$$

and

$$[\mu]_\equiv = \{\nu \in \Phi(\mathcal{A}) : \nu \equiv \mu\};$$

thus, by Proposition 3, we can say that

$$\{\mu\} \subset [\mu]_{(P^a:a\in A)} \subset [\mu]_\equiv. \tag{10}$$

Both the inclusions in (10) are equality if and only if μ is a degenerating probability measure (i.e. a probability measure concentrated on a singleton).

Proposition 4. *Let $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ and $\mu \in \Phi(\mathcal{A})$ be such that*

$$\mu(\{a \in A : P^a = Q^a\}) = 1.$$

Then we have

$$[\mu]_{(P^a:a\in A)} = [\mu]_{(Q^a:a\in A)}.$$

Proof. We have to show that $\nu \in [\mu]_{(P^a:a\in A)}$ if and only if $\nu \in [\mu]_{(Q^a:a\in A)}$.
Given $\nu \in [\mu]_{(P^a:a\in A)}$, for a suitable $(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ we have

$$\Pi_{\mu,(P^a:a\in A)}(da, ds) = \mu^s(da) P_\mu(ds)$$

and

$$\Pi_{\nu,(P^a:a\in A)}(da, ds) = \mu^s(da) P_\nu(ds);$$

furthermore, by Proposition 3, we also have $\nu \equiv \mu$ and, consequently, we can say that

$$\nu(\{a \in A : P^a = Q^a\}) = 1.$$

Thus $\Pi_{\mu,(Q^a:a\in A)} = \Pi_{\mu,(P^a:a\in A)}$ and $\Pi_{\nu,(Q^a:a\in A)} = \Pi_{\nu,(P^a:a\in A)}$.
In conclusion $\nu \in [\mu]_{(Q^a:a\in A)}$ follows from Proposition 3; indeed $\nu \equiv \mu$ and there exists $(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ such that

$$\Pi_{\mu,(Q^a:a\in A)}(da, ds) = \mu^s(da) P_\mu(ds)$$

and

$$\Pi_{\nu,(Q^a:a\in A)}(da, ds) = \mu^s(da) P_\nu(ds).$$

Similarly, given $\nu \in [\mu]_{(Q^a:a\in A)}$, we obtain $\nu \in [\mu]_{(P^a:a\in A)}$. ■

3 Two opposite cases of statistical experiments

First of all let us recall the following definition (see e.g. [1]).

Definition 2. *The statistical experiment* $(P^a : a \in A)$ *is said to be* dominated *if there exists a* σ*-finite measure* λ *such that*

$$P^a \ll \lambda, \quad \forall a \in A.$$

Example 1. *Consider* $(A, \mathcal{A}) = (S, \mathcal{S}) = (]0, \infty[, \mathcal{B})$ *where* $\mathcal{B}$ *is the usual Borel* σ*-algebra and let* $(P^a : a \in A)$ *be such that*

$$X \in \mathcal{S} \mapsto P^a(X) = a^{-1} m(X \cap]0, a[), \quad \forall a \in A,$$

where m *is the usual Lebesgue measure on* $\mathcal{B}$*. Then* $(P^a : a \in A)$ *is a dominated statistical experiment with* $\lambda = m$*.*

The assumption of dominated statistical experiment is common in mathematical statistics. Several examples are presented in [3] (see Section 9) for illustrating the construction of *conjugate families* of prior distributions; these examples shall refer the two following cases: $(S, \mathcal{S})$ is a *discrete* measurable space (i.e. S is at most countable and $\mathcal{S}$ is the power set of S); $(S, \mathcal{S})$ is an interval of the real line equipped with the usual Borel σ-algebra and λ is the usual Lebesgue measure.

Now we present two other definitions which allow to introduce the two opposite cases of statistical experiments cited in the Introduction; the first one is a slight modification of Definition 2, for the second one see Definition 1.1 in [9]. Each definition will be followed from an example.

Definition 3. *The statistical experiment* $(P^a : a \in A)$ *is said to be* strongly dominated *if there exists a* σ*-finite measure* λ *such that*

$$P^a \equiv \lambda, \quad \forall a \in A. \tag{11}$$

Example 2. *Consider* $(A, \mathcal{A}) = (S, \mathcal{S}) = (]-\infty, \infty[, \mathcal{B})$ *where* $\mathcal{B}$ *is the usual Borel* σ*-algebra and let* $(P^a : a \in A)$ *be such that*

$$X \in \mathcal{S} \mapsto P^a(X) = \int_X \frac{\exp(-\frac{(s-a)^2}{2})}{\sqrt{2\pi}} ds, \quad \forall a \in A.$$

Then $(P^a : a \in A)$ *is strongly dominated with* λ *as the usual Lebesgue measure.*

Definition 4. *The statistical experiment* $(P^a : a \in A)$ *is said to be* completely orthogonal *if there exists a set* $K \in \mathcal{A} \otimes \mathcal{S}$ *such that*

$$P^a(K(a, .)) = 1, \quad \forall a \in A \tag{12}$$

and

$$a \neq b \Rightarrow K(a, .) \cap K(b, .) = \emptyset. \tag{13}$$

Example 3. *Let* $(G, \mathcal{G})$ *be an arbitrary Polish space and consider* $(A, \mathcal{A}) = (S, \mathcal{S}) = (G, \mathcal{G})$*. Moreover let* $(P^a : a \in A)$ *be such that*

$$X \in \mathcal{S} \mapsto P^a(X) = 1_X(a), \quad \forall a \in A.$$

Then $(P^a : a \in A)$ *is a completely orthogonal statistical experiment with*

$$K = \{(a, s) \in G \times G : a = s\}.$$

For a completely orthogonal statistical experiment $(P^a : a \in A)$ we have

$$a \neq b \Rightarrow P^a \bot P^b;$$

then $(P^a : a \in A)$ is *totally informative* (see definition in [7], page 235).

Remark. *Obviously a strongly dominated statistical experiment is dominated but we can have dominated statistical experiments which are not strongly dominated (see Example 1). We also remark that a statistical experiment cannot be completely orthogonal and strongly dominated contemporary but we can easily build completely orthogonal statistical experiments which are dominated; for instance consider Example 3 with* $(G, \mathcal{G})$ *discrete.*

Now we shall prove the first result which can be seen as a generalization of Proposition 2.1 in [8].

Proposition 5. *Let* $(P^a : a \in A)$ *be a strongly dominated statistical experiment. Then*

$$\mu_1 \wp \mu_2 (mod(P^a : a \in A)) \Leftrightarrow \mu_1 = \mu_2.$$

Proof. Obviously we have to show that

$$\mu_1 \wp \mu_2 (mod(P^a : a \in A)) \Rightarrow \mu_1 = \mu_2;$$

indeed the inverse implication holds because $\cdot \wp \cdot (mod(P^a : a \in A))$ is an equivalence. By the hypothesis, (11) holds for a suitable σ-finite measure λ and we have $P_{\mu_1}, P_{\mu_2} \equiv \lambda$ for any pair $\mu_1, \mu_2 \in \Phi(\mathcal{A})$. Moreover $(P^a : a \in A)$ is dominated and there exists a function f_λ satisfying (4) by Proposition 1; more precisely we can choose a version of such f_λ which is positive and finite.

Now let us assume that $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$. Then, by (5), λ a.e. we have

$$\int_E f_\lambda(a, s) d\mu_1(a, s) = k(s) \int_E f_\lambda(a, s) d\mu_2(a), \quad \forall E \in \mathcal{A} \tag{14}$$

with

$$k(s) = \frac{\int_A f_\lambda(a, s) d\mu_1(a)}{\int_A f_\lambda(a, s) d\mu_2(a)}.$$

Moreover, by Proposition 3, we have $\mu_1 \equiv \mu_2$. Thus, $\mu_2 \otimes \lambda$ a.e., we obtain

$$f_\lambda(a, s) \frac{d\mu_1}{d\mu_2}(a) = k(s) f_\lambda(a, s)$$

that is equivalent to

$$\frac{d\mu_1}{d\mu_2}(a) = k(s)$$

because f_λ is positive and finite.

Then, by taking the integral over A w.r.t. μ_2, λ a.e. we have

$$k(s) = 1; \tag{15}$$

thus, by putting (15) in (14) and by taking in (14) the integral over S w.r.t. λ, we obtain

$$\mu_1(E) = \mu_2(E), \quad \forall E \in \mathcal{A}$$

as a consequence of Fubini theorem and (4). In conclusion we have $\mu_1 = \mu_2$. ■

Before presenting the opposite case, let us consider a propedeutic result.

Proposition 6. *Let* $(P^a : a \in A)$ *be a completely orthogonal statistical experiment and put*

$$R = \{s \in S : K(., s) \neq \emptyset\}. \tag{16}$$

Then

$$\#(K(., s)) = 1, \ \ \forall s \in R, \tag{17}$$

where $\#(K(., s))$ *denotes the cardinality of* $K(., s)$*.*
Moreover, for any $\mu \in \Phi(\mathcal{A})$*, we have*

$$\Pi_{\mu,(P^a:a\in A)}(K) = 1 \tag{18}$$

and consequently

$$P_\mu(\{s \in S : \mu^s(K(., s)) = 1\}) = P_\mu(R) = 1. \tag{19}$$

Proof. To obtain (17) we reason by contradiction. Let us suppose there exists $s \in R$ such that we have $a, b \in K(., s)$ with $a \neq b$. Then we obtain $s \in K(a, .) \cap K(b, .)$ that is impossible because (13) holds.
Moreover, by (12), we obtain (18). Indeed

$$\Pi_{\mu,(P^a:a\in A)}(K) = \int_A P^a(K(a, .))d\mu(a) = \mu(A) = 1.$$

Finally we also have

$$\int_S \mu^s(K(., s))dP_\mu(s) = \Pi_{\mu,(P^a:a\in A)}(K) = 1.$$

Thus we obtain (19); indeed

$$P_\mu(\{s \in S : \mu^s(K(., s)) = 1\}) = 1$$

and consequently $P_\mu(R^c) = 0$. ∎

Then we can prove the following

Proposition 7. *Let* $(P^a : a \in A)$ *be a completely orthogonal statistical experiment. Then*

$$\mu_1 \equiv \mu_2 \Leftrightarrow \mu_1 \wp \mu_2 (mod(P^a : a \in A)).$$

Proof. The proof consists to show that

$$\mu_1 \equiv \mu_2 \Rightarrow \mu_1 \wp \mu_2 (mod(P^a : a \in A));$$

indeed, by Proposition 3, the inverse implication holds in general.
By Proposition 6 we know that $K(., s)$ has only one element when it is not empty. Then it is useful to put

$$K(., s) = \{a^*(s)\}, \ \ \forall s \in R,$$

where R is the set defined by (16).
Now let $\mu \in \Phi(\mathcal{A})$ be arbitrarily fixed. By (18) we have

$$\Pi_{\mu,(P^a:a\in A)}(E \times X) = \Pi_{\mu,(P^a:a\in A)}((E \times X) \cap K) =$$

$$= \int_X \mu^s(E \cap K(.,s))dP_\mu(s), \quad \forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}$$

and, by (19), we obtain

$$\Pi_{\mu,(P^a:a\in A)}(E \times X) = \int_{X\cap R} \mu^s(E \cap K(.,s))dP_\mu(s) =$$

$$= \int_{X\cap R} \mu^s(E \cap \{a^*(s)\})dP_\mu(s), \quad \forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}.$$

Thus a Markov kernel $(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ such that

$$E \in \mathcal{A} \mapsto \mu^s(E) = 1_E(a^*(s)), \quad \forall s \in R$$

represents a family of posterior distributions for $\mathcal{E} = (A \times S, \mathcal{A} \otimes \mathcal{S}, \Pi_{\mu,(P^a:a\in A)})$ which does not depend on μ. Then, given $\mu_1, \mu_2 \in \Phi(\mathcal{A})$ satisfying (9), we can say that (7) holds. Thus the proof is complete by Proposition 3. ■

In conclusion, by Proposition 5 and Proposition 7 respectively, for any $\mu \in \Phi(\mathcal{A})$ we have the following situations:
when $(P^a : a \in A)$ is strongly dominated

$$[\mu]_{(P^a:a\in A)} = \{\mu\}; \tag{20}$$

when $(P^a : a \in A)$ is completely orthogonal

$$[\mu]_{(P^a:a\in A)} = [\mu]_{\equiv}. \tag{21}$$

The author thinks that, even if $(P^a : a \in A)$ in Example 1 is not strongly dominated, (20) holds for any $\mu \in \Phi(\mathcal{A})$. Finally, by taking into account Proposition 4, as immediate consequences of Proposition 5 and Proposition 7 respectively, we have the ensuing results.

Proposition 8. *Let $(P^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ and $\mu \in \Phi(\mathcal{A})$ be such that*

$$\mu(\{a \in A : P^a \equiv \lambda\}) = 1$$

for a suitable σ-finite measure λ.
Then (20) holds.

Proposition 9. *Let $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ and $\mu \in \Phi(\mathcal{A})$ as in Proposition 4. Moreover assume that $(Q^a : a \in A)$ is completely orthogonal.*
Then (21) holds.

4 An intermediate case

In this Section we shall consider an example which can be considered as an intermediate case. Indeed, as we shall see, we can find a prior distribution $\mu \in \Phi(\mathcal{A})$ such that both the inclusions in (10) are strict or, equivalently, conditions (20) and (21) are both false.

Before presenting this example, it is useful to consider some further notation. In this Section the Lebesgue measure on the real line will be denoted by m and, for $x \in]-\infty, \infty[$, we shall set

$$I_x =]x - \frac{1}{2}, x + \frac{1}{2}[.$$

Example 4. *Consider $(A, \mathcal{A}) = (S, \mathcal{S}) = (]-\infty, \infty[, \mathcal{B})$ where $\mathcal{B}$ is the usual Borel σ-algebra and let $(P^a : a \in A)$ be such that*

$$X \in \mathcal{S} \mapsto P^a(X) = m(X \cap I_a), \quad \forall a \in A.$$

Then in particular, according to Definitions 2 and 3, $(P^a : a \in A)$ is dominated with $\lambda = m$ but it is not strongly dominated.

As an immediate consequence we can say that, for any $\mu \in \Phi(\mathcal{A})$, the set

$$S_\mu = \{s \in S : \mu(I_s) = 0\}$$

has measure zero w.r.t. P_μ. Moreover, by considering (5) with $\lambda = m$, μ_1 and μ_2 are $(P^a : a \in A)$-equivalent if and only if they are mutually absolutely continuous and the ensuing condition holds:

$$\mu_1(I_s), \mu_2(I_s) > 0 \Rightarrow \mu_1(\cdot|I_s) = \mu_2(\cdot|I_s).$$

Then, under the assumption of $\mu_1(I_s)$ and $\mu_2(I_s)$ positive (note that they are both positive or both zero because of the mutual absolute continuity between μ_1 and μ_2) we shall consider the condition $\mu_1(\cdot|I_s) = \mu_2(\cdot|I_s)$ whence we obtain

$$\mu_1(\{a \in A : \frac{1_{I_s}(a)}{\mu_1(I_s)} = \frac{1_{I_s}(a)}{\mu_2(I_s)} \frac{d\mu_2}{d\mu_1}(a)\}) = 1$$

and

$$\mu_1(E_s|I_s) = 1,$$

where

$$E_s = \{a \in A : \frac{d\mu_2}{d\mu_1}(a) = \frac{\mu_2(I_s)}{\mu_1(I_s)}\};$$

thus, if $\mu_1(I_s)$ and $\mu_2(I_s)$ are positive, we can say that, almost surely w.r.t. $\mu_1(\cdot|I_s)$, the density $\frac{d\mu_2}{d\mu_1}(a)$ is equal to a constant (depending on s).

Now let $s, t \in S$ be arbitrarily fixed and assume that

$$\mu_1(I_s \cap I_t) > 0;$$

in this case $|s - t| < 1$ and we also have $\mu_2(I_s \cap I_t) > 0$ because μ_1 and μ_2 are mutually absolutely continuous. Then we obtain

$$\mu_1(E_s \cap E_t|I_s \cap I_t) = 1.$$

Indeed we have

$$\mu_1(E_s \cap E_t | I_s \cap I_t) = \frac{\mu_1(E_s \cap E_t \cap I_s \cap I_t)}{\mu_1(I_s \cap I_t)} = \frac{\mu_1(E_s \cap E_t \cap I_s | I_t)\mu_1(I_t)}{\mu_1(I_s \cap I_t)} =$$

$$= \frac{\mu_1(E_s \cap I_s | I_t)\mu_1(I_t)}{\mu_1(I_s \cap I_t)} = \frac{\mu_1(E_s \cap I_s \cap I_t)}{\mu_1(I_s \cap I_t)} = \frac{\mu_1(E_s \cap I_t | I_s)\mu_1(I_s)}{\mu_1(I_s \cap I_t)} =$$

$$= \frac{\mu_1(I_t | I_s)\mu_1(I_s)}{\mu_1(I_s \cap I_t)} = \frac{\mu_1(I_t \cap I_s)}{\mu_1(I_s \cap I_t)} = 1$$

and then

$$\mu_1(I_s \cap I_t) > 0 \Rightarrow \frac{\mu_2(I_s)}{\mu_1(I_s)} = \frac{\mu_2(I_t)}{\mu_1(I_t)}.$$

Thus, if I is an interval of the real line having positive measure w.r.t. μ_1 (and μ_2), there exists a positive constant c_I such that

$$\mu_1(\{a \in A : \frac{d\mu_2}{d\mu_1}(a) = c_I\} | I) = 1$$

when one of the two next conditions holds:
$m(I) < 1$;
$m(I) \geq 1$ but we cannot find open subintervals J of I such that $m(J) = 1$ and $\mu_1(J) = 0$.

In conclusion, given two mutually absolutely continuous prior distributions μ_1 and μ_2, we can say that they are $(P^a : a \in A)$-equivalent if and only if one can find disjoint and closed intervals (eventually reduced to a single point) such that the distance between two different intervals is not less than 1 and, almost surely w.r.t. μ_1, in each one of those intervals the density $\frac{d\mu_2}{d\mu_1}(a)$ is constant.

Before concluding this Section we shall present two not trivial examples of $[\mu]_{(P^a:a\in A)}$ (i.e. two examples in which the inclusions in (10) are strict).
The first one is quite simple. Let μ be such that

$$E \in \mathcal{A} \mapsto \mu(E) = \frac{1}{2}\{m(E \cap [0,1]) + m(E \cap [2,3])\};$$

then (20) and (21) fail because

$$[\mu]_{(P^a:a\in A)} = \{\alpha m(\cdot \cap [0,1]) + (1-\alpha)m(\cdot \cap [2,3]) : \alpha \in]0,1[\}.$$

The following slight modification of the previous example is more illustrative. For $\beta \in]0,1[$ set

$$E \in \mathcal{A} \mapsto \mu_\beta(E) = \frac{\beta}{2}[m(E \cap [0,1]) + 1_E(2)] + (1-\beta)m(E \cap [2,3]);$$

thus we have

$$E \in \mathcal{A} \mapsto \mu_\beta(E | [0,1]) = m(E \cap [0,1])$$

and

$$E \in \mathcal{A} \mapsto \mu_\beta(E | [2,3]) = \frac{\frac{\beta}{2}1_E(2) + (1-\beta)m(E \cap [2,3])}{1 - \frac{\beta}{2}}$$

with $\mu_\beta([0,1]) = \frac{\beta}{2}$ and $\mu_\beta([2,3]) = 1 - \frac{\beta}{2}$. Then

$$\mu_\beta = \frac{\beta}{2}\mu_\beta(\cdot|[0,1]) + (1-\frac{\beta}{2})\mu_\beta(\cdot|[2,3]);$$

thus (20) and (21) are false because

$$[\mu_\beta]_{(P^a:a\in A)} = \{\alpha\mu_\beta(\cdot|[0,1]) + (1-\alpha)\mu_\beta(\cdot|[2,3]) : \alpha \in]0,1[\}.$$

5 A partial order for the statistical experiments

In this Section we shall consider a partial order between statistical experiments. This partial order will be denoted by $\succeq$ and in some sense its definition follows in a natural way from the $(P^a : a \in A)$-equivalence defined in Section 2. Indeed we shall set

$$(P^a : a \in A) \succeq (Q^a : a \in A) \tag{22}$$

if and only if we have

$$\mu_1 \wp \mu_2 (mod(P^a : a \in A)) \Rightarrow \mu_1 \wp \mu_2 (mod(Q^a : a \in A)).$$

Thus (22) holds if and only if any $(P^a : a \in A)$-equivalence class is a subset of a $(Q^a : a \in A)$-equivalence class.

First of all, by Proposition 5 and Proposition 7 respectively, we can say that the strongly dominated statistical experiments are maximal and the completely orthogonal statistical experiments are minimal.

Furthermore we can consider another partial order between statistical experiments. It will be denoted by $\gg$ and it is defined as follows; we have

$$(P^a : a \in A) \gg (Q^a : a \in A) \tag{23}$$

if and only if condition (6) holds.

Before presenting the next result, it is useful to consider the following notation. When we consider $(Q^a : a \in A)$ as statistical experiment and μ as prior distribution, the predictive distribution will be denoted by Q_μ; in other words we set

$$X \in \mathcal{S} \mapsto Q_\mu(X) = \Pi_{\mu,(Q^a:a\in A)}(A \times X). \tag{24}$$

The first result of this Section shows that (23) is stronger than (22).

Proposition 10. *Let $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ be such that (23) holds.*
Then we have (22).

Proof. Let $\mu_1, \mu_2 \in \Phi(\mathcal{A})$ be such that $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$. Then, by Proposition 3, (9) holds and we shall complete the proof by showing that

$$\Pi_{\mu_k,(Q^a:a\in A)}(da, ds) = \nu^s(da)Q_{\mu_k}(ds) \quad (k = 1, 2), \tag{25}$$

for a suitable $(\nu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$.
It is useful to remark that, for h as in Proposition 2, we have

$$\Pi_{\mu_k,(Q^a:a\in A)}(E \times X) = \int_E Q^a(X)d\mu_k(a) = \int_E [\int_X h(a,s)dP^a(s)]d\mu_k(a) =$$

$$= \int_{E\times X} h(a,s)d\Pi_{\mu_k,(P^a:a\in A)}(a,s), \ \ \forall E \in \mathcal{A} \ \ \textit{and} \ \ \forall X \in \mathcal{S} \ \ (k = 1,2)$$

and, by (24), we obtain

$$X \in \mathcal{S} \mapsto Q_{\mu_k}(X) = \int_{A\times X} h(a,s)d\Pi_{\mu_k,(P^a:a\in A)}(a,s) =$$
$$\int_X [\int_A h(a,s)d\mu^s(a)]dP_{\mu_k}(s) \ \ (k = 1,2), \quad (26)$$

with $(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ satisfying (7). Now let us consider $(\nu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ such that

$$P_{\mu_k}(\{s \in S : E \in \mathcal{A} \mapsto \nu^s(E) = \frac{\int_E h(a,s)d\mu^s(a)}{\int_A h(a,s)d\mu^s(a)}\}) = 1, \ \ (k = 1,2).$$

Then, by (26), we have

$$\int_X \nu^s(E)dQ_{\mu_k}(a,s) = \int_X \frac{\int_E h(a,s)d\mu^s(a)}{\int_A h(a,s)d\mu^s(a)} \int_A h(a,s)d\mu^s(a)dP_{\mu_k}(s) =$$

$$= \int_X [\int_E h(a,s)d\mu^s(a)]dP_{\mu_k}(s) = \int_{E\times X} h(a,s)d\Pi_{\mu_k,(P^a:a\in A)}(a,s) =$$
$$= \Pi_{\mu_k,(Q^a:a\in A)}(E \times X), \ \ \forall E \in \mathcal{A} \ \ \textit{and} \ \ \forall X \in \mathcal{S} \ \ (k = 1,2).$$

Thus (25) holds and the proof is complete. ■

Remark. *Condition (23) is not necessary to have (22). Indeed let us consider the following example:* $(A, \mathcal{A}) = (S, \mathcal{S}) = (]-\infty, \infty[, \mathcal{B})$ *with* $\mathcal{B}$ *as the usual Borel* σ*-algebra,* $(P^a : a \in A)$ *is as in Example 2 and set*

$$X \in \mathcal{S} \mapsto Q^a(X) = 1_X(a) \ \ \forall a \in A.$$

In this case (23) is false; indeed we have

$$Q^a \perp P^a, \ \ \forall a \in A.$$

But $(P^a : a \in A)$ *is strongly dominated and* $(Q^a : a \in A)$ *is completely orthogonal; then, by Proposition 5 and Proposition 7, we have (22).*

In conclusion we can state the following immediate consequences of Proposition 10 and Proposition 4.

Proposition 11. *Let* $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ *be such that*

$$Q^a \equiv P^a, \ \ \forall a \in A.$$

Then, for any $\mu \in \Phi(\mathcal{A})$, *we have*

$$[\mu]_{(P^a:a\in A)} = [\mu]_{(Q^a:a\in A)}. \tag{27}$$

Proposition 12. *Let* $(P^a : a \in A), (Q^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ *and* $\mu \in \Phi(\mathcal{A})$ *be such that*

$$\mu(\{a \in A : Q^a \equiv P^a\}) = 1.$$

Then (27) holds.

Proposition 13. *Let* $(P^a : a \in A), (P_1^a : a \in A), (P_2^a : a \in A) \in \kappa(\mathcal{A}, \mathcal{S})$ *be such that*

$$P^a = \alpha_a P_1^a + (1 - \alpha_a) P_2^a, \ \ \forall a \in A$$

where

$$a \in A \mapsto \alpha_a \in]0, 1[$$

is a measurable mapping w.r.t. $\mathcal{A}$.
Then we have

$$\mu_1 \wp \mu_2 (mod(P^a : a \in A)) \Rightarrow \mu_1 \wp \mu_2 (mod(P_k^a : a \in A)) \ \ (k = 1, 2).$$

6 A result about the "ϵ-contaminations"

Given a prior distribution μ and $\epsilon \in]0, 1[$, we can consider the ϵ-contamination class of prior distributions (see e.g. [2])

$$\Gamma_\epsilon^\mu = \{(1 - \epsilon)\mu + \epsilon\nu : \nu \in \Phi\}$$

where Φ determines the allowed contaminations which are mixed with μ.

Throughout this paper we shall consider $\Phi = [\mu]_\equiv$. Hence, by taking into account (10) and the properties of mutual absolute contuinuity between positive measures, given a statistical experiment $(P^a : a \in A)$ we have

$$\{\mu\} \subset \Gamma_\epsilon^\mu, [\mu]_{(P^a:a\in A)} \subset [\mu]_\equiv.$$

It is easy to check that, when we have one of the two extreme cases in (10) (i.e. $[\mu]_{(P^a:a\in A)} = \{\mu\}$ or $[\mu]_{(P^a:a\in A)} = [\mu]_\equiv$), we have

$$\Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)} = \{(1 - \epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\}. \tag{28}$$

Indeed, when $[\mu]_{(P^a:a\in A)} = \{\mu\}$, we have

$$\Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)} = \{(1 - \epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\} = \{\mu\}$$

while, when $[\mu]_{(P^a:a\in A)} = [\mu]_\equiv$, we have

$$\Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)} = \{(1 - \epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\} = \Gamma_\epsilon^\mu.$$

However, as we shall see, (28) holds in general, i.e. it also holds when both the inclusions in (10) are strict. Before proving this fact, we need a propedeutic result.

Proposition 14. *Let us consider* $\mu_1, \mu_2 \in [\mu]_{(P^a:a\in A)}$ *and* $\epsilon \in]0,1[$. *Then*

$$\nu = \epsilon\mu_1 + (1-\epsilon)\mu_2 \in [\mu]_{(P^a:a\in A)}.$$

Proof. By the hypothesis we have $\mu_1 \wp \mu_2 (mod(P^a : a \in A))$ and then

$$[\mu]_{(P^a:a\in A)} = [\mu_1]_{(P^a:a\in A)} = [\mu_2]_{(P^a:a\in A)}. \tag{29}$$

Thus (7) holds and, for $(\mu^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ as in (7), we have

$$\Pi_{\nu,(P^a:a\in A)}(da, ds) = \epsilon\Pi_{\mu_1,(P^a:a\in A)}(da, ds) + (1-\epsilon)\Pi_{\mu_2,(P^a:a\in A)}(da, ds) =$$

$$= \epsilon\mu^s(da)P_{\mu_1}(ds) + (1-\epsilon)\mu^s(da)P_{\mu_2}(ds) = \mu^s(da)[\epsilon P_{\mu_1}(ds) + (1-\epsilon)P_{\mu_2}(ds)] =$$
$$= \mu^s(da)P_\nu(ds).$$

Moreover, by Proposition 3, we have $\mu_1, \mu_2 \equiv \mu$; then $\nu \equiv \mu_1, \mu_2$.
In conclusion, by Proposition 3, we can say that $\nu \in [\mu_1]_{(P^a:a\in A)}$ (or equivalently $\nu \in [\mu_2]_{(P^a:a\in A)}$) and, by (29), $\nu \in [\mu]_{(P^a:a\in A)}$. ■

Now we can prove the final result.

Proposition 15. *Condition (28) holds for any prior distribution* μ.

Proof. By (10) we can say that

$$\{(1-\epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\} \subset \Gamma_\epsilon^\mu$$

and, by Proposition 14,

$$\{(1-\epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\} \subset [\mu]_{(P^a:a\in A)}.$$

Thus we have

$$\{(1-\epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\} \subset \Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)}$$

and we complete the proof by showing that

$$\Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)} \subset \{(1-\epsilon)\mu + \epsilon\nu : \nu \in [\mu]_{(P^a:a\in A)}\}. \tag{30}$$

To this aim, let us consider $\rho \in \Gamma_\epsilon^\mu \cap [\mu]_{(P^a:a\in A)}$ arbitrarily fixed.
Then there exists $\nu_\rho \in [\mu]_\equiv$ such that

$$\rho = (1-\epsilon)\mu + \epsilon\nu_\rho$$

and we have

$$P_\mu(\{s \in S : \mu^s = \rho^s\}) = 1 \tag{31}$$

for a suitable $(\rho^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ such that $\Pi_{\rho,(P^a:a\in A)}(da, ds) = \rho^s(da)P_\rho(ds)$.
For deriving (30) we have to show that $\nu_\rho \in [\mu]_{(P^a:a\in A)}$.
It is easy to check that (31) can be rewritten as follows

$$P_\mu(\{s \in S : \mu^s = (1-\epsilon)\frac{dP_\mu}{dP_\rho}(s)\mu^s + \epsilon\frac{dP_{\nu_\rho}}{dP_\rho}(s)(\nu_\rho)^s\}) = 1$$

for $((\nu_\rho)^s : s \in S) \in \kappa(\mathcal{S}, \mathcal{A})$ such that $\Pi_{\nu_\rho,(P^a:a\in A)}(da, ds) = (\nu_\rho)^s(da)P_{\nu_\rho}(ds)$; then

$$P_\mu(\{s \in S : (1-(1-\epsilon)\frac{dP_\mu}{dP_\rho}(s))\mu^s = \epsilon\frac{dP_{\nu_\rho}}{dP_\rho}(s)(\nu_\rho)^s\}) = 1.$$

Thus we obtain

$$P_\mu(\{s \in S : \epsilon\frac{dP_{\nu_\rho}}{dP_\rho}(s)\mu^s = \epsilon\frac{dP_{\nu_\rho}}{dP_\rho}(s)(\nu_\rho)^s\}) = 1$$

whence it follows

$$P_\mu(\{s \in S : \mu^s = (\nu_\rho)^s\}) = 1; \tag{32}$$

indeed we have

$$P_\mu(\{s \in S : 1 = (1-\epsilon)\frac{dP_\mu}{dP_\rho}(s) + \epsilon\frac{dP_{\nu_\rho}}{dP_\rho}(s)\}) = 1.$$

In conclusion, by (32), $\nu_\rho \in [\mu]_{(P^a:a\in A)}$ because $\nu_\rho \equiv \mu$. ∎

7 Concluding remarks

The results presented in this paper show that the $(P^a : a \in A)$-equivalence classes are mostly reduced to a single prior distribution. As already stated, the author thinks that, for $(P^a : a \in A)$ as in the Example 1, (20) holds for any prior distributions; in other words it should not be necessary to consider a strongly dominated statistical experiment to have all the corresponding $(P^a : a \in A)$-equivalence classes reduced to a single prior distribution. In general, by taking into account the example presented in Section 4, the idea is that we can have $(P^a : a \in A)$-equivalence classes with more than one prior distribution only when we can find pairs of mutually singular sampling distributions.

In some sense it could have been more interesting to come up with an analogous equivalence weaker than the previous one so that it would be easier to have equivalence classes not reduced to a single prior distribution, even for strongly dominated statistical experiments. A way to do that it would be to consider a similar theory for a given observation; in other words, for a given $s \in S$, the equivalence $\wp(s)$ on $\Phi(\mathcal{A})$ defined as follows

$$\mu_1\wp(s)\mu_2 \Leftrightarrow \mu_1^s = \mu_2^s$$

seems to be more interesting.
The trouble is that this equivalence is not well-defined because, in general, the family of posterior distributions $(\mu^s : s \in S)$ is almost surely unique w.r.t. the predictive distribution P_μ; thus, for a given $s \in S$ such that $P_\mu(\{s\}) = 0$, we can set μ^s in an arbitrary way. The author does not know how to solve this problem.

Acknowledgements. This work has been supported by MURST 60% funds. The author thanks the referee. His comments led to an improvement in the content of the paper.

References

[1] R.R. Bahadur, Sufficiency and Statistical Decision Functions, Ann. Math. Stat., vol. 25 (1954), p. 423-462.

[2] J.O. Berger, Robust Bayesian Analysis: Sensitivity to the Prior, J. Stat. Plann. Inference, vol. 25 (1990), p. 303-328.

[3] M.H. DeGroot, Optimal Statistical Decisions (McGraw-Hill, New York, 1970).

[4] C. Dellacherie, P. Meyer, Probabilités et Potentiel (Chap. V-VIII) Théorie des Martingales (Hermann, Paris, 1980).

[5] J. Florens, M. Mouchart, J. Rolin, Elements of Bayesian Statistics (Marcel Dekker Inc., New York, 1990).

[6] R.S. Liptser, A.N. Shiryiayev, Statistics of Random Processes I, General Theory (Springer-Verlag, New York, 1977).

[7] H. Heyer, Theory of Statistical Experiments (Springer-Verlag, New York, 1982).

[8] C. Macci, Some results about Bayesian experiments with discrete sample space and positive sampling probabilities, Statistica (Bologna), vol. 55 (1995), p. 95-100.

[9] R.D. Mauldin, D. Preiss, H. v. Weizsacker, Orthogonal Transition Kernels, Ann. Probab., vol. 11 (1983), p. 970-988.

Claudio Macci
Dipartimento di Matematica
Università degli Studi di Roma "Tor Vergata"
Viale della Ricerca Scientifica
I-00133 Rome
Italy.