On prior distributions which give rise to a dominated Bayesian experiment

Claudio Macci

Abstract

When the statistical experiment is dominated (i.e. when all the sampling distributions are absolutely continuous w.r.t. a σ-finite measure), all the probability measures on the parameter space are prior distributions which give rise to a dominated Bayesian experiment.
In this paper we shall consider the family $\mathbb{D}$ of prior distributions which give rise to a dominated Bayesian experiment (w.r.t. a fixed statistical experiment not necessarily dominated) and we shall think the set of all the probability measures on the parameter space endowed by the total variation metric d.
Then we shall illustrate the relationship between $d(\mu, \mathbb{D})$ (where μ is the prior distribution) and the probability to have sampling distributions absolutely continuous w.r.t. the predictive distribution.
Finally we shall study some properties of $\mathbb{D}$ in terms of convexity and extremality and we shall illustrate the relationship between $d(\mu, \mathbb{D})$ and the probability to have posteriors and prior mutually singular.

1 Introduction.

In this paper we shall consider the terminology used in [5]. Let $(S, \mathcal{S})$ (*sample space*) and $(A, \mathcal{A})$ (*parameter space*) be two Polish Spaces and denote by $\mathbb{P}(\mathcal{A})$ and by $\mathbb{P}(\mathcal{S})$ the sets of all the probability measures on $\mathcal{A}$ and $\mathcal{S}$ respectively.
Furthermore let $(P^a : a \in A)$ be a fixed family of probability measures on $\mathcal{S}$ (*sampling distributions*) such that $(a \mapsto P^a(X) : X \in \mathcal{S})$ are measurable mappings w.r.t. $\mathcal{A}$.

Received by the editors December 1995 – In revised form in August 1996.
Communicated by M. Hallin.
1991 *Mathematics Subject Classification :* 60A10, 62A15, 62B15, 52A07.
Key words and phrases : Bayesian experiment, Lebesgue decomposition, distance between a point and a set, extremal subset, extremal point.

Then, for any $\mu \in \mathbb{P}(\mathcal{A})$ (*prior distribution*), we can consider the probability space $\mathcal{E}_\mu = (A \times S, \mathcal{A} \otimes \mathcal{S}, \Pi_\mu)$ (*Bayesian experiment*) such that

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

Moreover we shall denote by P_μ the *predictive distribution*, i.e. the probability measure on $\mathcal{S}$ such that

$$P_\mu(X) = \Pi_\mu(A \times X), \quad \forall X \in \mathcal{S}. \tag{2}$$

Finally we can say that $\mathcal{E}_\mu$ is *regular* because $(S, \mathcal{S})$ and $(A, \mathcal{A})$ are Polish Spaces, (see e.g. [5], Remark (i), page 31); in other words we have a family $(\mu^s : s \in S)$ of probability measures on $\mathcal{A}$ (*posterior distributions*) such that

$$\Pi_\mu(E \times X) = \int_X \mu^s(E) dP_\mu(s), \quad \forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}. \tag{3}$$

We stress that the family $(\mu^s : s \in S)$ satisfying (3) is P_μ a.e. unique; moreover $\mathcal{E}_\mu$ is said to be *dominated* if $\Pi_\mu << \mu \otimes P_\mu$.

Before stating the next result it is useful to introduce the following notation. Let g_μ be a version of the density of the absolutely continuous part of Π_μ w.r.t. $\mu \otimes P_\mu$ and assume that the singular part of Π_μ w.r.t. $\mu \otimes P_\mu$ is concentrated on a set $D_\mu \in \mathcal{A} \otimes \mathcal{S}$ having null measure w.r.t. $\mu \otimes P_\mu$; in other words the Lebesgue decomposition of Π_μ w.r.t. $\mu \otimes P_\mu$ is

$$\Pi_\mu(C) = \int_C g_\mu d[\mu \otimes P_\mu] + \Pi_\mu(C \cap D_\mu), \quad \forall C \in \mathcal{A} \otimes \mathcal{S}.$$

Furthermore put

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

and

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

Now we can recall the following result (see [7], Proposition 1).

Proposition 1. *μ a.e. the Lebesgue decomposition of P^a w.r.t. P_μ is*

$$P^a(X) = \int_X g_\mu(a, s) dP_\mu(s) + P^a(X \cap D_\mu(a, .)), \quad \forall X \in \mathcal{S}. \tag{4}$$

P_μ a.e. the Lebesgue decomposition of μ^s w.r.t. μ is

$$\mu^s(E) = \int_E g_\mu(a, s) d\mu(a) + \mu^s(E \cap D_\mu(., s)), \quad \forall E \in \mathcal{A}.$$

As an immediate consequence we obtain the next

Corollary 2. *The following statements are equivalent:*

$$\mathcal{E}_\mu \ \ dominated; \tag{5}$$

$$\mu(\{a \in A : P^a << P_\mu\}) = 1; \tag{6}$$

$$P_\mu(\{s \in S : \mu^s << \mu\}) = 1. \tag{7}$$

Corollary 3. *The following statements are equivalent:*

$$\Pi_\mu \perp \mu \otimes P_\mu;$$

$$\mu(\{a \in A : P^a \perp P_\mu\}) = 1;$$
$$P_\mu(\{s \in S : \mu^s \perp \mu\}) = 1.$$

From now on we shall use the following notation; for any $\mu \in \mathbb{P}(\mathcal{A})$, we put

$$B_\mu^{(ac)} = \{a \in A : P^a << P_\mu\},$$

$$B_\mu^{(sg)} = \{a \in A : P^a \perp P_\mu\}$$

and, for a given family $(\mu^s : s \in S)$ of posterior distributions,

$$T_\mu^{(ac)} = \{s \in S : \mu^s << \mu\},$$

$$T_\mu^{(sg)} = \{s \in S : \mu^s \perp \mu\}.$$

Remark. *For any $Q \in \mathbb{P}(\mathcal{S})$ we can say that*

$$\{a \in A : P^a << Q\}, \{a \in A : P^a \perp Q\} \in \mathcal{A}.$$

Indeed (see e.g. [3], Remark, page 58) we can consider a jointly measurable function f such that $f(a, \cdot)$ is a version of the density of the absolutely continuous part of P^a w.r.t. Q and, consequently, we have

$$\{a \in A : P^a << Q\} = \{a \in A : \int_S f(a,s)dQ(s) = 1\}$$

and

$$\{a \in A : P^a \perp Q\} = \{a \in A : \int_S f(a,s)dQ(s) = 0\}.$$

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

$$B_\mu^{(ac)}, B_\mu^{(sg)} \in \mathcal{A}$$

and, by reasoning in a similar way, we can also say that

$$T_\mu^{(ac)}, T_\mu^{(sg)} \in \mathcal{S}$$

for any given family $(\mu^s : s \in S)$ of posterior distributions.

Remark. *In general $T_\mu^{(ac)}$ and $T_\mu^{(sg)}$ depend on the choice of the family $(\mu^s : s \in S)$ satisfying (3) we consider. On the contrary, by the P_μ a.e. uniqueness of $(\mu^s : s \in S)$, the probabilities $P_\mu(T_\mu^{(ac)})$ and $P_\mu(T_\mu^{(ac)})$ do not depend on that choice.*

In this paper we shall concentrate the attention on the set

$$\mathbb{D} = \{\mu \in \mathbb{P}(\mathcal{A}) : \ (5) \ \ holds\}.$$

We remark that when $(P^a : a \in A)$ is a *dominated statistical experiment* (see e.g. [1]), i.e. when each P^a is absolutely continuous w.r.t. a fixed σ-finite measure, we have $\mathbb{D} = \mathbb{P}(\mathcal{A})$.
However we can say that $\mathbb{D}$ is always not empty; indeed we have the following

Proposition 4. *$\mathbb{D}$ contains all the* discrete *probability measures on $\mathcal{A}$ (i.e. all the probability measures in $\mathbb{P}(\mathcal{A})$ concentrated on a set at most countable).*

Proof. Let $\mu \in \mathbb{P}(\mathcal{A})$ be concentrated on a set C_μ at most countable. Then, by noting that

$$P_\mu(X) = \sum_{a \in C_\mu} P^a(X)\mu(\{a\}) \ \ (\forall X \in \mathcal{S}),$$

(6) holds and, by Corollary 2, $\mu \in \mathbb{D}$. ∎

Remark. *It is known (see [2], Theorem 4, page 237) that each $\mu \in \mathbb{P}(\mathcal{A})$ is the weak limit of a sequence of discrete probability measures. Then, if we consider $\mathbb{P}(\mathcal{A})$ as a topological space with the weak topology, $\mathbb{D}$ is dense in $\mathbb{P}(\mathcal{A})$ by Proposition 4.*

In Section 2 we shall consider $\mathbb{P}(\mathcal{A})$ endowed with the *total variation metric d* defined as follows:

$$(\mu, \nu) \in \mathbb{P}(\mathcal{A}) \times \mathbb{P}(\mathcal{A}) \mapsto d(\mu, \nu) = \sup\{|\mu(E) - \nu(E)| : \ \ E \in \mathcal{A}\}. \tag{8}$$

Then we shall prove that

$$\mu(B_\mu^{(ac)}) + d(\mu, \mathbb{D}) = 1, \ \ \forall \mu \in \mathbb{P}(\mathcal{A}) \tag{9}$$

where $d(\mu, \mathbb{D})$ is the distance between μ and $\mathbb{D}$, i.e.

$$d(\mu, \mathbb{D}) = \inf\{d(\mu, \nu) : \nu \in \mathbb{D}\}. \tag{10}$$

Hence $\mu(B_\mu^{(ac)})$ increases when $d(\mu, \mathbb{D})$ decreases.

In Section 3 we shall consider $\mathbb{D}$ and $\mathbb{P}(\mathcal{A})$ as subsets of $\mathbb{M}(\mathcal{A})$ (i.e. the vector space of the signed measures on $\mathcal{A}$) and we shall study some properties $\mathbb{D}$ in terms of convexity and extremality.

In Section 4 we shall prove an inequality concerning $d(\mu, \mathbb{D})$ and the probability (w.r.t. P_μ) to have posterior distributions and prior distribution mutually singular and, successively, we shall present two examples.

2 The proof of (9).

In this Section we shall prove the formula (9).
To this aim we need some further notation. Put

$$\mathcal{A}^* = \{E \in \mathcal{A} : \exists Q_E \in \mathbb{P}(\mathcal{S}) \ \ such \ \ that \ \ P^a << Q_E, \ \ \forall a \in E\}$$

and

$$F(\mu) = \sup\{\mu(E): \quad E \in \mathcal{A}^*\}. \tag{11}$$

$F(\mu)$ defined in (11) has big importance in what follows; indeed we shall prove (9) showing that, for any $\mu \in \mathbb{P}(\mathcal{A})$, $F(\mu)$ is equal to $1 - d(\mu, \mathbb{D})$ and $\mu(B_\mu^{(ac)})$. Before doing this, we need some propedeutic results.

Lemma 5. *Let $\mu \in \mathbb{P}(\mathcal{A})$ be such that $\mu(\{a \in A : P^a << Q\}) = 1$ for some $Q \in \mathbb{P}(\mathcal{S})$.*
Then $\mu(B_\mu^{(ac)}) = 1$.

Proof. By the hypothesis we can say that (see e.g. [6], Lemma 7.4, page 287)

$$P_\mu(\{s \in S: \quad \mu^s(E) = \frac{\int_E f_Q(a,s)d\mu(a)}{\int_A f_Q(a,s)d\mu(a)}, \quad \forall E \in \mathcal{A}\}) = 1$$

where f_Q is a jointly measurable function such that

$$\mu(\{a \in A: \quad P^a(X) = \int_X f_Q(a,s)dQ(s), \quad \forall X \in \mathcal{S}\}) = 1.$$

Hence we have $P_\mu(T_\mu^{(ac)}) = 1$ and, by Corollary 2, $\mu(B_\mu^{(ac)}) = 1$. ∎

Lemma 6. *For any $\mu \in \mathbb{P}(\mathcal{A})$ there exists a set $A_\mu \in \mathcal{A}^*$ such that $F(\mu) = \mu(A_\mu)$.*

Proof. The statement is obvious when $F(\mu) = 0$; indeed we have $\mu(E) = 0$ for any $E \in \mathcal{A}^*$.
Thus let us consider the case $F(\mu) > 0$.
Then, for any $n \in \mathbb{N}$, we have a set $A_n \in \mathcal{A}^*$ such that $\mu(A_n) > F(\mu) - \frac{1}{n}$ and we can say that

$$\mu(\cup_{n\in\mathbb{N}} A_n) > F(\mu) - \frac{1}{n}, \quad \forall n \in \mathbb{N};$$

thus

$$\mu(\cup_{n\in\mathbb{N}} A_n) \geq F(\mu).$$

Furthermore the probability measure Q defined as follows

$$Q = \sum_{n\in\mathbb{N}} \frac{Q_{A_n}}{2^n}$$

is such that

$$P^a << Q, \quad \forall a \in \cup_{n\in\mathbb{N}} A_n.$$

Thus $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}^*$ and $\mu(\cup_{n\in\mathbb{N}} A_n) = F(\mu)$.
In other words we can put $A_\mu = \cup_{n\in\mathbb{N}} A_n$. ∎

Lemma 7. *Let $\mu \in \mathbb{P}(\mathcal{A})$ be such that $F(\mu) = 1$. Then $\mu \in \mathbb{D}$.*

Proof. By Lemma 6 we have a set $A_\mu \in \mathcal{A}^*$ such that $\mu(A_\mu) = 1$; in other words there exists $Q \in \mathbb{P}(\mathcal{S})$ such that $\mu(\{a \in A: \quad P^a << Q\}) = 1$.
Then, by Lemma 5, $\mu(B_\mu^{(ac)}) = 1$ and $\mu \in \mathbb{D}$ follows from Corollary 2. ∎

Lemma 8. *Let $\mu \in \mathbb{P}(\mathcal{A})$ be such that $F(\mu) = 0$. Then*

$$\mathbb{D} \subset \{\nu \in \mathbb{P}(\mathcal{A}) : \quad \mu \perp \nu\}.$$

Proof. Let $\nu \in \mathbb{D}$ be arbitrarily fixed. Then $\nu(B_\nu^{(ac)}) = 1$ immediately follows.
Moreover we have $\mu(B_\nu^{(ac)}) = 0$; indeed $F(\mu) = 0$.
Then $\mu \perp \nu$ and the proof is complete. ■

In this Section, when $F(\mu) \in]0,1[$, we put $\mu_1 = \mu(\cdot|A_\mu)$ and $\mu_2 = \mu(\cdot|A_\mu^c)$.

Lemma 9. *Let $\mu \in \mathbb{P}(\mathcal{A})$ be such that $F(\mu) \in]0,1[$. Then*

$$F(\mu_1) = 1 \tag{12}$$

and

$$F(\mu_2) = 0. \tag{13}$$

Proof. By construction we have $F(\mu_1) \leq 1$. Then (12) holds; indeed we have $\mu_1(A_\mu) = 1$ with $A_\mu \in \mathcal{A}^*$.
To prove (13) we reason by contradiction.
Assume that $F(\mu_2) > 0$ and let $Q \in \mathbb{P}(\mathcal{S})$ be defined as follows

$$Q = \frac{Q_{A_\mu} + Q_{A_{\mu_2}}}{2};$$

then we can say that

$$P^a << Q, \quad \forall a \in A_\mu \cup A_{\mu_2}. \tag{14}$$

Now, since we have

$$\mu = F(\mu)\mu_1 + (1 - F(\mu))\mu_2,$$

we obtain

$$\mu(A_\mu \cup A_{\mu_2}) = F(\mu)\mu_1(A_\mu \cup A_{\mu_2}) + (1 - F(\mu))\mu_2(A_\mu \cup A_{\mu_2}) =$$

$$= F(\mu) + (1 - F(\mu))\mu_2(A_{\mu_2}) > F(\mu).$$

But this is a contradiction; indeed, by (14), we have $A_\mu \cup A_{\mu_2} \in \mathcal{A}^*$ and consequently

$$\mu(A_\mu \cup A_{\mu_2}) \leq F(\mu).$$

■

The identity (9) will immediately follow from the two next Propositions.

Proposition 10. *For any $\mu \in \mathbb{P}(\mathcal{A})$ we have*

$$F(\mu) = 1 - d(\mu, \mathbb{D}).$$

Proof. If $F(\mu) = 1$ we have $\mu \in \mathbb{D}$ by Lemma 7 and $d(\mu, \mathbb{D}) = 0$.
If $F(\mu) = 0$ we have $\mathbb{D} \subset \{\nu \in \mathbb{P}(\mathcal{A}) : \ \mu \perp \nu\}$ by Lemma 8 and, by (8),

$$\mathbb{D} \subset \{\nu \in \mathbb{P}(\mathcal{A}) : \ d(\mu, \nu) = 1\}.$$

Thus, by (10), we have $d(\mu, \mathbb{D}) = 1$.
Then let us consider the case $F(\mu) \in]0, 1[$.
By (12) and by Lemma 7, $\mu_1 \in \mathbb{D}$. Moreover, by construction, we have $\mu_1 \perp \mu_2$; thus, by (8),

$$d(\mu_1, \mu_2) = 1.$$

Then, for any $\nu \in \mathbb{D}$, we put

$$E_\nu = A_\mu \cup B_\nu^{(ac)}$$

and we obtain

$$d(\mu, \nu) \geq |\mu(E_\nu) - \nu(E_\nu)| = |F(\mu)\mu_1(E_\nu) + (1 - F(\mu))\mu_2(E_\nu) - 1| =$$

$$= |F(\mu)1 + (1 - F(\mu))0 - 1| = 1 - F(\mu);$$

indeed, by (13), $\mu_2(B_\nu^{(ac)}) = 0$.
Then the proof is complete; indeed $\mu_1 \in \mathbb{D}$ and we have

$$\begin{aligned} d(\mu, \mu_1) &= \sup\{|\mu(E) - \mu_1(E)| : E \in \mathcal{A}\} = \\ &= \sup\{|F(\mu)\mu_1(E) + (1 - F(\mu))\mu_2(E) - \mu_1(E)| : E \in \mathcal{A}\} = \\ &\qquad (1 - F(\mu))d(\mu_1, \mu_2) = (1 - F(\mu)). \end{aligned}$$

■

Proposition 11. *For any $\mu \in \mathbb{P}(\mathcal{A})$ we have*

$$F(\mu) = \mu(B_\mu^{(ac)}).$$

Proof. If $F(\mu) = 1$ we have $\mu \in \mathbb{D}$ by Lemma 7; then, by Corollary 2, we have $\mu(B_\mu^{(ac)}) = 1$.
If $F(\mu) = 0$ we have necessarily $\mu(B_\mu^{(ac)}) = 0$.
Then let us consider the case $F(\mu) \in]0, 1[$.
By taking into account that

$$\mu = F(\mu)\mu_1 + (1 - F(\mu))\mu_2,$$

we have $P_{\mu_1} << P_\mu$; indeed, by (2),

$$P_\mu = F(\mu)P_{\mu_1} + (1 - F(\mu))P_{\mu_2}.$$

Thus $B^{(ac)}_{\mu_1} \subset B^{(ac)}_{\mu}$ and, consequently,

$$1 = \mu_1(B^{(ac)}_{\mu_1}) = \mu_1(B^{(ac)}_{\mu});$$

indeed $\mu_1 \in \mathbb{D}$ by (12) and Lemma 7.
Then we obtain the following inequality:

$$\mu(B^{(ac)}_\mu) \geq \mu(A_\mu \cap B^{(ac)}_\mu) = F(\mu)\mu_1(A_\mu \cap B^{(ac)}_\mu)+$$

$$+(1 - F(\mu))\mu_2(A_\mu \cap B^{(ac)}_\mu) = F(\mu)1 + (1 - F(\mu))0 = F(\mu).$$

Now put $Q = \frac{Q_{A_\mu} + P_\mu}{2}$; then

$$P^a << Q, \ \ \forall a \in A_\mu \cup B^{(ac)}_\mu.$$

Thus $A_\mu \cup B^{(ac)}_\mu \in \mathcal{A}^*$ and, consequently, $F(\mu) = \mu(A_\mu \cup B^{(ac)}_\mu)$.
Then

$$F(\mu) = \mu(A_\mu \cup B^{(ac)}_\mu) = \mu(A_\mu) + \mu(B^{(ac)}_\mu - A_\mu)$$

whence $\mu(B^{(ac)}_\mu - A_\mu) = 0$ and we obtain the following inequality:

$$\mu(B^{(ac)}_\mu) = \mu(B^{(ac)}_\mu \cap A_\mu) + \mu(B^{(ac)}_\mu - A_\mu) = \mu(B^{(ac)}_\mu \cap A_\mu) \leq \mu(A_\mu) = F(\mu).$$

This completes the proof; indeed we have $\mu(B^{(ac)}_\mu) \geq F(\mu)$ and $\mu(B^{(ac)}_\mu) \leq F(\mu)$. ■

Remark. *By (9) and Corollary 2 we have $d(\mu, \mathbb{D}) = 0$ if and only if $\mu \in \mathbb{D}$. Thus we can say that, if we consider $\mathbb{P}(\mathcal{A})$ as a topological space with the topology induced by d, $\mathbb{D}$ is a closed set.*

3 Convexity and extremality properties.

The first result in this Section shows that $\mathbb{D}$ is a convex set.

Proposition 12. $\mathbb{D}$ *is a* convex *set (see e.g. [8], page 100), i.e.*

$$\mu_1, \mu_2 \in \mathbb{D}, \ \ \mu_1 \neq \mu_2 \ \ \Rightarrow \ \ t\mu_1 + (1 - t)\mu_2 \in \mathbb{D}, \ \ \forall t \in [0, 1].$$

Proof. Let $\mu_1, \mu_2 \in \mathbb{D}$ (with $\mu_1 \neq \mu_2$) and $t \in [0, 1]$ be arbitrarily fixed and put

$$\mu = t\mu_1 + (1 - t)\mu_2. \tag{15}$$

Thus we have $\mu_1, \mu_2 << \mu$ and, moreover, $P_{\mu_1}, P_{\mu_2} << P_\mu$; indeed, by (15), we obtain

$$\Pi_\mu = t\Pi_{\mu_1} + (1 - t)\Pi_{\mu_2}, \tag{16}$$

whence

$$P_\mu = tP_{\mu_1} + (1-t)P_{\mu_2}.$$

Then $\mu \in \mathbb{D}$. Indeed, by taking into account that $\mu_1, \mu_2 \in \mathbb{D}$, (16) can be rewritten as follows

$$\Pi_\mu(C) = t\int_C g_{\mu_1} d[\mu_1 \otimes P_{\mu_1}] + (1-t)\int_C g_{\mu_2} d[\mu_2 \otimes P_{\mu_2}] =$$

$$= \int_C [t g_{\mu_1}(a,s)\frac{d\mu_1}{d\mu}(a)\frac{dP_{\mu_1}}{dP_\mu}(s) +$$

$$+ (1-t) g_{\mu_2}(a,s)\frac{d\mu_2}{d\mu}(a)\frac{dP_{\mu_2}}{dP_\mu}(s)] d[\mu \otimes P_\mu](a,s), \quad \forall C \in \mathcal{A}\otimes\mathcal{S}.$$

■

In the following we need the next

Lemma 13. *Let $\mu \in \mathbb{D}$ be such that $\nu << \mu$. Then $\nu \in \mathbb{D}$ and*

$$P_\nu(X) = \int_X [\int_A g_\mu(a,s) d\nu(a)] dP_\mu(s), \quad \forall X \in \mathcal{S}. \tag{17}$$

Proof. By Corollary 2 and Proposition 1 we have

$$\mu(\{a \in A: \quad P^a(X) = \int_X g_\mu(a,s) dP_\mu(s), \quad \forall X \in \mathcal{S}\}) = 1$$

whence

$$\nu(\{a \in A: \quad P^a(X) = \int_X g_\mu(a,s) dP_\mu(s), \quad \forall X \in \mathcal{S}\}) = 1;$$

indeed $\nu << \mu$.
Then

$$\Pi_\nu(E \times X) = \int_E P^a(X) d\nu(a) = \int_E [\int_X g_\mu(a,s) dP_\mu(s)] d\nu(a) =$$

$$= \int_X [\int_E g_\mu(a,s) d\nu(a)] dP_\mu(s), \quad \forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}$$

and (17) follows from (2) (with ν in place of μ). Furthermore we have

$$\Pi_\nu(E \times X) = \int_X [\frac{\int_E g_\mu(a,s) d\nu(a)}{\int_A g_\mu(a,s) d\nu(a)} \int_A g_\mu(a,s) d\nu(a)] dP_\mu(s) =$$

$$= \int_X [\frac{\int_E g_\mu(a,s) d\nu(a)}{\int_A g_\nu(a,s) d\nu(a)}] dP_\nu(s), \quad \forall E \in \mathcal{A} \ \ and \ \ \forall X \in \mathcal{S}.$$

Thus (7) holds for $\mathcal{E}_\nu$ and $\nu \in \mathbb{D}$ by Corollary 2. ■

The next result is an immediate consequence of Lemma 13.

Proposition 14. $\mathbb{D}$ *is* extremal *for* $\mathbb{P}(\mathcal{A})$ *(see e.g. [8], page 181), i.e.*

$$t\mu_1 + (1-t)\mu_2 \in \mathbb{D} \ \ with \ \ t \in]0,1[\ \ and$$

$$\mu_1, \mu_2 \in \mathbb{P}(\mathcal{A}) \quad \Rightarrow \quad \mu_1, \mu_2 \in \mathbb{D}.$$

Proof. Let $\mu \in \mathbb{D}$ be such that $\mu = t\mu_1 + (1-t)\mu_2$ with $t \in]0,1[$ and $\mu_1, \mu_2 \in \mathbb{P}(\mathcal{A})$. Then $\mu_1, \mu_2 \in \mathbb{D}$ by Lemma 13; indeed, by construction, we have $\mu_1, \mu_2 << \mu$. ■

Before proving the next Propositions, it is useful to denote by $EX(\mathbb{D})$ the set of the *extremal points* of $\mathbb{D}$ (see e.g. [8], page 181); thus we put

$$EX(\mathbb{D}) = \{\mu \in \mathbb{D} : \quad \mu = t\mu_1 + (1-t)\mu_2 \ \ with \ \ t \in]0,1[$$

$$and \ \ \mu_1, \mu_2 \in \mathbb{D} \ \ \Rightarrow \ \ \mu_1 = \mu_2 = \mu\}.$$

Thus we can prove the next results.

Proposition 15. *If $\mu \in \mathbb{D}$ is not concentrated on a singleton, then $\mu \notin EX(\mathbb{D})$.*

Proof. If $\mu \in \mathbb{D}$ is not concentrated on a singleton, there exists a set $B \in \mathcal{A}$ such that $\mu(B) \in]0,1[$ and we can say that

$$\mu = \mu(B)\mu(\cdot|B) + (1-\mu(B))\mu(\cdot|B^c).$$

Then $\mu(\cdot|B), \mu(\cdot|B^c) \in \mathbb{D}$ by Lemma 13 and $\mu(\cdot|B)$ and $\mu(\cdot|B^c)$ are both different from μ; indeed $\mu(B) \in]0,1[$. Thus we can say that $\mu \notin EX(\mathbb{D})$. ∎

Proposition 16. *If $\mu \in \mathbb{D}$ is concentrated on a singleton, then $\mu \in EX(\mathbb{D})$.*

Proof. Assume that $\mu \in \mathbb{D}$ is concentrated on a singleton; in other words there exists $b \in A$ such that

$$\mu(E) = 1_E(b), \quad \forall E \in \mathcal{A}.$$

Then, if we have

$$\mu = t\mu_1 + (1-t)\mu_2 \ \ with \ \ t \in]0,1[\ \ and \ \ \mu_1, \mu_2 \in \mathbb{D},$$

we obtain

$$1 = t\mu_1(\{b\}) + (1-t)\mu_2(\{b\}).$$

Then we have necessarily $\mu_1(\{b\}) = \mu_2(\{b\}) = 1$; thus $\mu_1 = \mu_2 = \mu$. ∎

Proposition 17.

$$EX(\mathbb{D}) = \{\mu \in \mathbb{P}(\mathcal{A}) : \quad \mu \ \ is \ \ concentrated \ \ on \ \ a \ \ singleton\}$$

Proof. By Proposition 15 and Proposition 16 we have

$$EX(\mathbb{D}) = \{\mu \in \mathbb{D} : \quad \mu \ \ is \ \ concentrated \ \ on \ \ a \ \ singleton\}.$$

Then the proof is complete; indeed, by Proposition 4, all the probability measures concentrated on a singleton belong to $\mathbb{D}$. ∎

4 A consequence about Posteriors and two examples.

In Section 2 we proved equation (9). From a statistical point of view it is more interesting a relationship between $d(\mu, \mathbb{D})$ and the probability to have a particular Lebesgue decomposition between posteriors distributions and prior distribution.

Then, in the first part of this Section, we shall prove that

$$P_\mu(T_\mu^{(sg)}) \leq d(\mu, \mathbb{D}), \quad \forall \mu \in \mathbb{P}(\mathcal{A}). \tag{18}$$

We stress that $T_\mu^{(sg)}$ can be seen as the set of samples which give rise to posterior distributions concentrated on a set of probability zero w.r.t. the prior distribution μ.

Equation (18) immediately follows from (9) and from the next

Proposition 18. *We have*

$$P_\mu(T_\mu^{(sg)}) \leq 1 - \mu(B_\mu^{(ac)}), \quad \forall \mu \in \mathbb{P}(\mathcal{A}).$$

Proof. By (1), (2) and (4) we have

$$P_\mu(T_\mu^{(sg)}) = \int_A P^a(T_\mu^{(sg)}) d\mu(a) = \int_A [\int_{T_\mu^{(sg)}} g_\mu(a,s) dP_\mu(s) + P^a(T_\mu^{(sg)} \cap D_\mu(a,.))] d\mu(a)$$

whence it follows

$$P_\mu(T_\mu^{(sg)}) = \int_{T_\mu^{(sg)}} [\int_A g_\mu(a,s) d\mu(a)] dP_\mu(s) + \int_A P^a(T_\mu^{(sg)} \cap D_\mu(a,.)) d\mu(a);$$

thus, by Proposition 1, we obtain

$$P_\mu(T_\mu^{(sg)}) = \int_A P^a(T_\mu^{(sg)} \cap D_\mu(a,.)) d\mu(a).$$

Then we can conclude that

$$P_\mu(T_\mu^{(sg)}) = \int_{(B_\mu^{(ac)})^c} P^a(T_\mu^{(sg)} \cap D_\mu(a,.)) d\mu(a) \leq \mu((B_\mu^{(ac)})^c) = 1 - \mu(B_\mu^{(ac)});$$

indeed, as a consequence of (4), we have

$$\int_{B_\mu^{(ac)}} P^a(D_\mu(a,.)) d\mu(a) = 0.$$

■

In conclusion we can say that $P_\mu(T_\mu^{(sg)})$ cannot be too big when μ is near $\mathbb{D}$ (w.r.t. the distance d). More precisely, when $\mu \notin \mathbb{D}$, we can have $P_\mu(T_\mu^{(sg)}) = 0$ (see the example in [7], Section 4) or $P_\mu(T_\mu^{(sg)}) > 0$ but, in any case, $P_\mu(T_\mu^{(sg)})$ cannot be greater than the d-distance between μ and $\mathbb{D}$.

Now we shall consider two examples. For the first one we shall derive $\mathbb{D}$ by using the results in Section 2 and in Section 3 while, for the second one, we shall present the different cases concerning (9) and (18) for some particular choices of prior distributions.

In the first example we shall consider $(A, \mathcal{A})$ and $(S, \mathcal{S})$ both equal to $([0,1], \mathcal{B})$, where $\mathcal{B}$ denotes the usual Borel σ-algebra. Moreover we shall put

$$X \in \mathcal{S} \mapsto P^a(X) = \frac{1}{2}[1_X(a) + \lambda(X)], \quad \forall a \in B = [0, \frac{1}{4}] \cup [\frac{3}{4}, 1] \tag{19}$$

and

$$X \in \mathcal{S} \mapsto P^a(X) = a1_X(\frac{1}{2}) + (1-a)\lambda(X), \quad \forall a \in A - B =]\frac{1}{4}, \frac{3}{4}[\tag{20}$$

where λ is the Lebesgue measure.
We stress that the statistical experiment $(P^a : a \in A)$ defined by (19) and (20) is not dominated because, for any $a \in B$, $\{a\}$ is an atom of P^a.
As we shall see, the set B has a big importance to say when a prior distribution μ belongs to $\mathbb{D}$.

For doing this let us consider the following notation; given a a prior distribution μ, we put

$$I(\mu) = \int_{A-B} a d\mu(a);$$

then we obtain

$$X \in \mathcal{S} \mapsto P_\mu(X) = \frac{1}{2}\int_B [1_X(a) + \lambda(X)]d\mu(a) + \int_{A-B}[a1_X(\frac{1}{2}) + (1-a)\lambda(X)]d\mu(a) =$$

$$= \frac{1}{2}\mu(B \cap X) + \frac{1}{2}\mu(B)\lambda(X) + I(\mu)1_X(\frac{1}{2}) + (1 - \mu(B) - I(\mu))\lambda(X) =$$

$$= \frac{1}{2}\mu(B \cap X) + (1 - \frac{\mu(B)}{2} - I(\mu))\lambda(X) + I(\mu)1_X(\frac{1}{2}).$$

For our aim, let us consider the following

Lemma 19. *Assume μ is* diffuse *(i.e. μ assigns probability zero to each singleton). Then*

$$d(\mu, \mathbb{D}) = \mu(B). \tag{21}$$

Proof. We have three cases: $\mu(B) = 1$, $\mu(B) = 0$ and $\mu(B) \in]0,1[$.
If $\mu(B) = 1$, we have $I(\mu) = 0$ and

$$X \in \mathcal{S} \mapsto P_\mu(X) = \frac{1}{2}[\mu(X) + \lambda(X)];$$

then $\mu(B_\mu^{(ac)}) = \mu(\emptyset) = 0$ and (21) follows from (9).
If $\mu(B) = 0$, we have $I(\mu) \in]\frac{1}{4}, \frac{3}{4}[$ and

$$X \in \mathcal{S} \mapsto P_\mu(X) = (1 - I(\mu))\lambda(X) + I(\mu)1_X(\frac{1}{2});$$

then $\mu(B_\mu^{(ac)}) = \mu(A - B) = 1 - \mu(B)$ and (21) follows from (9).
Finally, if $\mu(B) \in]0,1[$, we have $I(\mu) \in]\frac{1}{4}(1-\mu(B)), \frac{3}{4}(1-\mu(B))[$ and we can say that P_μ has $\{\frac{1}{2}\}$ as a unique atom and its diffuse part is absolutely continuous w.r.t. λ; then

$$\mu(B_\mu^{(ac)}) = \mu(A - B) = 1 - \mu(B)$$

and (21) follows from (9). ∎

Now we can prove the next

Proposition 20. *$\mu \in \mathbb{D}$ if and only if*

$$\mu = p\mu_{(ds)} + (1-p)\mu_{(df)} \tag{22}$$

where $p \in [0,1]$, $\mu_{(ds)}$ is a discrete probability measure on $\mathcal{A}$, $\mu_{(df)}$ is a diffuse probability measure on $\mathcal{A}$ such that

$$\mu_{(df)}(B) = 0. \tag{23}$$

Proof. Let us start by noting that, for any $\mu \in \mathbb{P}(\mathcal{A})$, (22) holds in general (always with $p \in [0,1]$, $\mu_{(ds)}$ discrete probability measure on $\mathcal{A}$ and $\mu_{(df)}$ diffuse probability measure on $\mathcal{A}$).
If $p = 1$, we have $\mu \in \mathbb{D}$ by Proposition 4.
If $p = 0$, by Lemma 19 we have $\mu \in \mathbb{D}$ if and only if (23) holds.
Finally, if $p \in]0,1[$, we have two cases: when (23) holds, $\mu \in \mathbb{D}$ by Proposition 12 (i.e. by the convexity of $\mathbb{D}$); when (23) fails, $\mu \notin \mathbb{D}$ by Proposition 14 (i.e. because $\mathbb{D}$ is extremal w.r.t. $\mathbb{P}(\mathcal{A})$). Indeed, by taking into account that $\mathbb{D}$ is an extremal subset, when we have

$$\mu = t\mu_1 + (1-t)\mu_2$$

with $t \in]0,1[$, $\mu_1 \in \mathbb{D}$ and $\mu_2 \notin \mathbb{D}$, we can say that $\mu \notin \mathbb{D}$. ∎

The second example refers to a nonparametric problem (see example 4 in [5], page 45).
The results in Section 2 and in Section 4 will be used for a class of prior distributions called *Dirichlet Processes* (see the references cited therein).
For simplicity let $(S, \mathcal{S})$ be the real line equipped with the usual Borel σ-algebra, put

$$A = \{a : \mathcal{S} \to [0,1]\} = [0,1]^{\mathcal{S}}$$

and, for $\mathcal{A}$, we take the product σ-algebra (i.e. the σ-algebra generated by all the cylinders based on a Borel set of $[0,1]$ for a finite number of coordinates).
Furthermore let $(P^a : a \in A)$ be such that $P^a = a$ when a is a probability measure on $\mathcal{S}$ and let μ be the Dirichlet Process with parameter α, where α is an arbitrary finite measure on $\mathcal{S}$; thus it will be denoted by μ_α.
In what follows we shall refer to the results shown by Ferguson (see [4]).
First of all we can say that, μ_α almost surely, a is a discrete probability measure on $\mathcal{S}$ and

$$P_{\mu_\alpha} = \frac{\alpha(\cdot)}{\alpha(S)}.$$

Moreover we can say that each addendum in (9) assumes the values 0 and 1 only; more precisely:
$\mu_\alpha(B^{(ac)}_{\mu_\alpha}) = 1$ (and $d(\mu_\alpha, \mathbb{D}) = 0$, i.e. $\mu_\alpha \in \mathbb{D}$) when α is discrete;
$\mu_\alpha(B^{(ac)}_{\mu_\alpha}) = 0$ (and $d(\mu_\alpha, \mathbb{D}) = 1$), when α is not discrete.
Consequently, by Corollary 2, when α is discrete we obtain

$$P_{\mu_\alpha}(T^{(ac)}_{\mu_\alpha}) = 1;$$

thus equation (18) gives $0 \leq 0$.
On the contrary, when α is diffuse, we have $\mu_\alpha(B^{(sg)}_{\mu_\alpha}) = 1$ and

$$P_{\mu_\alpha}(T^{(sg)}_{\mu_\alpha}) = 1$$

follows from Corollary 3; thus equation (18) gives $1 \leq 1$.
Finally let us consider α neither discrete nor diffuse.
It is known that (see [4], Theorem 1) that

$$P_{\mu_\alpha}(\{s \in S : (\mu_\alpha)^s = \mu_{\alpha+\delta_s}\}) = 1$$

where δ_s denotes the probability measure concentrated on s.
Then, if we put

$$K_\alpha = \{s \in S : \alpha(\{s\}) > 0\} = \{s \in S : P_{\mu_\alpha}(\{s\}) > 0\},$$

we have $P_{\mu_\alpha}(T^{(ac)}_{\mu_\alpha}) = P_{\mu_\alpha}(K_\alpha)$ and $P_{\mu_\alpha}(T^{(sg)}_{\mu_\alpha}) = P_{\mu_\alpha}((K_\alpha)^c)$; thus, in this case, equation (18) gives the strict inequality $P_{\mu_\alpha}((K_\alpha)^c) < 1$.

Acknowledgements. This work has been supported by CNR funds.
I thank the referees. Their suggestions led to an improvement in both the content and readability of the paper: the proof of Proposition 11 is a simplified version suggested by an anonymous referee and Professor C. P. Robert has suggested the Dirichlet Process as a possible example.

References

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

[2] P. Billingsley, Convergence of probability measures (John Wiley and Sons, New York, 1968).

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

[4] T. S. Ferguson, A Bayesian Analysis of Some Nonparametric Problems, Ann. Statist., vol. 1 (1973), p. 209-230.

[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] C. Macci, On the Lebesgue Decomposition of the Posterior Distribution with respect to the Prior in Regular Bayesian Experiments, Statist. Probab. Lett., vol. 26 (1996), p. 147-152.

[8] A. E. Taylor, D. C. Lay, Introduction to Functional Analysis (Second edition, John Wiley and Sons, New York, 1980).

Dipartimento di Matematica,
Università degli Studi di Roma "Tor Vergata",
Viale della Ricerca Scientifica,
00133 Rome,
Italy.