TRABAJOS DE ESTADISTICA
Vol. 5, núm. 1, 1990, pp. 93 a 105

EQUIVALENCE OF DECISION PROBLEMS[1]

GARCÍA-NOGALES, A.
Dpto. de Matemáticas
Universidad de Extremadura
06071 Badajoz (Spain)

SUMMARY

A general and strong notion of equivalence of decision problems is given. Some results and examples are given to show that this natural notion is well adapted to the methodology of statistics.

Key words and phrases: decisión problem, experiment, equivalence.
AMS Subject classification: 62C05.

RESUMEN

En este trabajo se introduce una definición de equivalencia de problemas de decisión. Los resultados y ejemplos que presentamos muestran que esta definición de equivalencia se adapta bien a la metodología de la estadística.

Título: Equivalencia de problemas de decisión.
Palabras y frases claves: problema de decisión, experimento, equivalencia.

Clasificación AMS: 62C05.

1. INTRODUCTION

In this paper a strong notion of equivalence of statistical decision problems is sistematically studied. The problem of comparing experiments is a classical one in statistics. A definition of a experiment being

Recibido septiembre 1989. Revisado febrero 1990.

[1] Part of this paper has been presented in the XVIII Reunión Nacional de Estadística, Investigación Operativa e Informática on September 1989 at Santiago University (Spain).

more informative than another was given by Blackwell in 1951; this gives in a natural way a definition of equivalence of experiments: two experiments will be said equivalent (in this sense) if each one is more informative than the other. A brief survey on these topics could be find in Lehmann (1988). Our definition of equivalence is stronger enough so that the risk sets for the two problems coincide as it is shown in theorem 1. Furthermore, the introduced notion of equivalence constitutes itself a ststistical procedure to obtain solution for a statistical decision problem equivalent to another one of wich a solution is known, and the optimality properties (for many and usual points of view) of this solution are preserved by equivalence as we will show in the following. Example 1 below illustrates the results obtained in the paper. An interesting application of our definition in normal linear models will be briefly exposed in Example 2. A mainly theoretical application is also announced in Example 3.

The results will be presented in a general setting to avoid unnecessary duplication of proofs for discrete and continuous models. Let us fixe the notations to be used.

A experiment is a triplet $\mathscr{E} = (\Omega, \mathscr{U}, \{P_\theta/\theta \in \Theta\})$ where $(\Omega, \mathscr{U})$ is a measurable space and $\{P_\theta/\theta \in \Theta\}$ a family of probability measures on $\mathscr{U}$. Usually, the parameter space Θ is endowed with a σ-field $\mathscr{I}$. In this case we shall write

$$\mathscr{E} = (\Omega, \mathscr{U}, \{P_\theta/\theta \in (\Theta, \mathscr{I})\}) \tag{1}$$

A decision problem is a triplet

$$\mathscr{D} = [\mathscr{E}, (\Delta, \mathbf{D}), W] \tag{2}$$

where $\mathscr{E}$ is a experiment, $(\Delta, \mathbf{D})$ is a measurable space (the decision space) and W is a loss function, i.e., a nonnegative random variable (r.v.) on $(\Theta, \mathscr{I}) \times (\Delta, \mathbf{D})$. We shall say that $\mathscr{E}$ is the experiment support of $\mathscr{D}$. An experiment or a decision problem is said to be n-dimensional (or real, if $n = 1$) if Ω is a Borel subset of $\mathbb{R}^n$ and $\mathscr{U}$ is the Borel σ-field of Ω.

The solutions for a decision problem $\mathscr{D}$ are called strategies; we recall that a (random) strategy for $\mathscr{D}$ is a transition probability S on $\Omega \times \mathbf{D}$, i.e., a map $S : \Omega \times \mathbf{D} \to [0, 1]$ such that for every $\omega \in \Omega$, $S(\omega, \cdot)$ is a probability measure on $\mathbf{D}$, and for every $D \in \mathbf{D}$, $S(\cdot, D)$ is a real random variable (r.r.v.) on $(\Omega, \mathscr{U})$. A nonrandom strategy S is determined by a r.v. $s : (\Omega, \mathscr{U}) \to (\Delta, \mathbf{D})$ by means of $S(\omega, D) = \varepsilon_{s(\omega)}(D)$ where, for $\delta \in \Delta$, $\varepsilon_\delta(D) = 1$ if $\delta \in D$, $=0$ otherwise.

The loss function is used as a measure of the goodness of solutions for the decision problem. The mean loss of a strategy S is the map W_S defined on $\Theta \times \Omega$ by $W_S(\theta, \omega) = \int_\Delta W(\theta, \delta) S(\omega, d\delta)$ and the risk function is defined as $R_{W,S}(\theta) = \int_\Omega W_S(\theta, \omega) P_\theta(d\omega)$. In particular, if S is the nonrandom strategy determined by s, we have $R_{W,S}(\theta) = \int_\Omega W(\theta, s(\omega)) P_\theta(d\omega)$. If S_1 and S_2 are two strategies for $\mathscr{D}$, S_1 is said to be preferable to S_2 if the risk function of S_1 is less or equal to the one of S_2 at every point of Θ.

Let $(\Omega, \mathscr{U}, P)$ be a probability space and let $X : (\Omega, \mathscr{B})$ be a r.v. We shall denote by P^X the probability law of X on $\mathscr{B}$, i.e., $P^X(B) = P(X^{-1}(B))$ for all $B \in \mathscr{B}$. We shall denote by $\mathscr{R}$ the Borel σ-field of $\mathbb{R}$ and, if $A \in \mathbb{R}$, by $\mathscr{R}(A)$ the Borel σ-field of A; however, $\mathscr{R}^+$ shall means the Borel σ-field of $\mathbb{R}^+$. Analogous notations shall be used for $\mathbb{R}^n$.

2. EQUIVALENCE OF DECISION PROBLEMS

Definition 1: Two decision problems

$$\mathscr{D} = [(\Omega, \mathscr{U}, \{P_\theta / \theta \in (\Theta, \mathscr{I})\}), (\Delta, \mathbf{D}), W]$$

and

$$\mathscr{D}^* = [(\Omega^*, \mathscr{U}^*, \{P^*_{\theta^*} / \theta^* \in (\Theta^*, \mathscr{I}^*)\}), (\Delta^*, \mathbf{D}^*), W^*]$$

are said to be equivalent if there are three one-to-one and onto r.v.

$$\Psi : (\Omega, \mathscr{U}) \to (\Omega^*, \mathscr{U}^*) \quad , \quad \Gamma : (\Theta, \mathscr{I}) \to (\Theta^*, \mathscr{I}^*) \quad \text{and}$$
$$\Lambda : (\Delta, \mathbf{D}) \to (\Delta^*, \mathbf{D}^*)$$

whose inverses are also r.v. (i.e., bimeasurable) and such that

i) $P_\theta^\Psi = P^*_{\Gamma(\theta)}$ for every $\theta \in \Theta$

ii) $W(\theta, \delta) = W^*(\Gamma(\theta), \Lambda(\delta))$ for every $\theta \in \Theta$ and every $\delta \in \Delta$.

If we want to be more precise, we shall say that $\mathscr{D}$ and $\mathscr{D}^*$ are (Ψ, Γ, Λ)-equivalents.

Remark: Obviously, if a decision problem is equivalent to another one, the second is equivalent to the first, and if two decision problems are equivalents to a third problem then they are equivalents.

Definition 2: Two experiments

$$\mathscr{E} = (\Omega, \mathscr{U}, \{P_\theta/\theta \in (\Theta, \mathscr{I})\}) \quad \text{and} \quad \mathscr{E}^* = (\Omega^*, \mathscr{U}^*, \{P^*_{\theta^*}/\theta^* \in (\Theta^*, \mathscr{I}^*)\})$$

are said to be equivalent if there exist two maps Ψ and Γ as those in definition 1 such that i) holds. Briefly, we shall say that they are (Ψ, Γ)-equivalent.

Remark: In the case $\Theta = \Theta^*$ and $\Gamma(\theta) = \theta$ it is easy to see that this notion of equivalence of experiments is stronger than the usual one; see, for ex, Lehmann (1988) for this definition.

Definition 3: Let $\mathscr{D}$ and $\mathscr{D}^*$ be two (Ψ, Γ, Λ)-equivalent decision problems and let S and S^* be strategies for $\mathscr{D}$ and $\mathscr{D}^*$, respectively. We shall say that this strategies are associate if

$$S(\omega, D) = S^*(\Psi(\omega), \Lambda(D))$$

for every ω and D.

Remark: It is easy to see that, if S and S^* are associate strategies, then S is nonrandom iff S^* is nonrandom; in fact, if $S(\omega, \cdot) = \varepsilon_{s(\omega)}$ then $S^*(\omega^*, \cdot) = \varepsilon_{s^*(\omega^*)}$, where $s^*(\omega^*) = \Lambda(s(\Psi^{-1}(\omega^*)))$.

The following theorem states the relationship between associate strategies.

Theorem 1: *Let $\mathscr{D}$, $\mathscr{D}^*$, S and S^* be as in definition 3; then*

$$W^*_{S^*}(\Gamma(\theta), \Psi(\omega)) = W_S(\theta, \omega)$$

and

$$R_{W,S}(\theta) = R_{W^*,S^*}(\Gamma(\theta))$$

for every θ and ω.

Proof: For any given $\omega \in \Omega$, $S(\omega, \cdot)$ is a probability measure on $\mathbf{D}$ and the probability law $S(\omega, \cdot)^\Lambda$ of Λ is a probability measure on $\mathscr{D}^*$ such that

$$S(\omega, \cdot)^\Lambda(D^*) = S(\omega, \Lambda^{-1}(D^*)) = S^*(\Psi(\omega), D^*)$$

for all $D^* \in \mathscr{D}^*$. Hence $S(\omega, \cdot)^\Lambda = S^*(\Psi(\omega), \cdot)$. Then, if $\theta \in \Theta$ and $\omega \in \Omega$ and if we define $\theta^* = \Gamma(\theta)$ and $\omega^* = \Psi(\omega)$, we have

$$W^*_{S^*}(\Gamma(\theta), \Psi(\omega)) = W^*_{S^*}(\theta^*, \omega^*) = \int_{\Delta^*} W^*(\theta^*, \delta^*) S^*(\omega^*, d\delta^*) =$$
$$= \int_{\Delta^*} W^*(\theta^*, \delta^*)(S(\omega, \cdot)^\Lambda)(d\delta^*) = \int_\Delta W^*(\theta^*, \Lambda(\delta)) S(\omega, d\delta) =$$
$$= \int_\Delta W^*(\Gamma(\theta), \Lambda(\delta)) S(\omega, d\delta) = \int_\Delta W(\theta, \delta) S(\omega, d\delta) = W_S(\theta, \delta).$$

Furthermore

$$R_{W^*, S^*}(\Gamma(\theta)) = R_{W^*, S^*}(\theta^*) = \int_{\Omega^*} W^*_{S^*}(\theta^*, \omega^*) P^*_{\theta^*}(d\omega^*) =$$
$$= \int_{\Omega^*} W^*_{S^*}(\theta^*, \omega^*) P^\Psi_\theta(d\omega^*) = \int_\Omega W^*_{S^*}(\theta^*, \Psi(\omega)) P_\theta(d\omega) =$$
$$= \int_\Omega W_S(\theta, \omega) P_\theta(d\omega) = R_{W, S}(\theta)$$

wich completes the proof. □

Remark: In particular, if we have two equivalent decision problems and we know an optimun strategy for one of them (in the sense that the risk in uniformly minimized), we also know an optimum strategy for the other one and it can be construct from the first.

If $\mathscr{E}$ is an experiment as in (1), we shall denote by $\mathscr{E}^n$, $n = 1, 2, \ldots$, the product experiment

$$\mathscr{E}^n = (\Omega^n, \mathscr{U}^n, \{P^n_\theta / \theta \in (\Theta, \mathscr{I})\}).$$

If $\mathscr{D}$ is a decision problem as in (2), we shall denote by $\mathscr{D}^n$ the product decision problem

$$\mathscr{D}^n = [\mathscr{E}^n, (\Delta, \mathbf{D}), W]$$

We have them the following result.

Theorem 2: *If $\mathscr{D}$ and $\mathscr{D}^*$ are two decision problems (Ψ, Γ, Λ)-equivalent then $\mathscr{D}^n$ is $(\Psi^n, \Gamma, \Lambda)$-equivalent to $\mathscr{D}^{*n}$, $n = 1, 2, \ldots$, where*

$$\Psi^n(\omega_1, \ldots, \omega_n) = (\Psi(\omega_1), \ldots, \Psi(\omega_n)).$$

Proof: Ψ^n is a one-to-one, onto and bimeasurable map. To prove that the distribution law of Ψ^n under P^n_θ is equal to $P^*_{\Gamma(\theta)}$ it will b enough to show that this is the case for the measurable rectangles in the product σ-field $\mathscr{U}^{*n}$; but this is an easy consequence of i) in Definition 1. □

Now, we shall consider two equivalent experiments. The following results suggest that the statistics on one and other can be identified. The first is an immediate consequence of the definition of equivalence.

Theorem 3: *Let $\mathscr{E}$ and $\mathscr{E}^*$ be two (Ψ, Γ)-equivalent experiments as in Definition 2 and let T^* a real statistic on $(\Omega^*, \mathscr{U}^*)$. If $T = T^* \circ \Psi$ then*

$$E_\theta(T) = E_{\Gamma(\theta)}(T^*),\ \theta \in \Theta$$

in the sense that if one of these integrals exists then the other also exists and both coincide.

Theorem 4: *Let $\mathscr{E}$ and $\mathscr{E}^*$ be equivalent experiments as in Definition 2, $(\mathscr{X}, \mathscr{C})$ a measurable space, $T^* : (\Omega^*, \mathscr{U}^*) \to (\mathscr{X}, \mathscr{C})$ a statistic and $T = T^* \circ \Psi$. Then T^* is a sufficient (resp., complete) statistic for $\mathscr{E}^*$ iff so is T for $\mathscr{E}$.*

Proof: Let $\mathscr{B}$ (resp., $\mathscr{B}^*$) be the σ-field $T^{-1}(\mathscr{C})$ (resp., $T^{*-1}(\mathscr{C})$). Then $\mathscr{B} = \Psi^{-1}(\mathscr{B}^*)$. Let us suppose that T^* is sufficient, i.e., $\mathscr{B}^*$ is a sufficient sub-σ-field of $\mathscr{U}^*$. We must to show that for any $A \in \mathscr{U}$ there is a r.r.v. g_A on $(\Omega, \mathscr{B})$ such that $P_\theta(A \cap B) = \int_B g_A(\omega) P_\theta(d\omega)$, for all $\theta \in \Theta$ and all $B \in \mathscr{B}$. It will be enough to show that

$$g_A(\omega) = P^*_{\Gamma(\theta)}(\Psi(A) \mid \mathscr{B}^*)(\Psi(\omega))$$

works since the conditional probability in the second term not depends in fact on θ, $\mathscr{B}^*$ being sufficient.

But, for all B in $\mathscr{B}$ and all θ in Θ, we have

$$\begin{aligned} &\int_B P^*_{\Gamma(\theta)}(\Psi(A) \mid \mathscr{B}^*)(\Psi(\omega)) P_\theta(d\omega) = \\ &= \int_{\Psi(B)} P^*_{\Gamma(\theta)}(\Psi(A) \mid \mathscr{B}^*)(\omega^*) P^\Psi_\theta(d\omega^*) = \\ &= \int_{\Psi(B)} P^*_{\Gamma(\theta)}(\Psi(A) \mid \mathscr{B}^*)(\omega^*) P^*_{\Gamma(\theta)}(d\omega^*) = \\ &= P^*_{\Gamma(\theta)}(\Psi(A \cap B)) = P^\Psi_\theta(\Psi(A \cap B)) = P_\theta(A \cap B). \end{aligned}$$

This shows that T is sufficient if so is T^*.

The assertion relative to completeness is a consequence of theorem 3. □

Next, we give some examples.

Example 1: Let us consider the (exponential) experiment

$$\mathscr{E} = (\mathbb{R}^+, \mathscr{R}^+, \{P_\theta/\theta > 0\})$$

where

$$dP_\theta(x) = \theta^{-1}\cdot\exp(-\theta^{-1}x)\,dx,\ x > 0.$$

The map $\Psi : x \in \mathbb{R}^+ \to e^{-x} \in]0,1[$ is one-to-one onto and bimeasurable, and $\Psi^{-1}(y) = -\log y,\ 0 < y < 1$. The map $\Gamma : \theta \in]0, +\infty[\to \theta^{-1} - 1 \in]-1, +\infty[$ is also one-to-one onto and bimeasurable, and $\Gamma^{-1}(\theta^*) = (1 + \theta^*)^{-1}$ for $\theta^* > -1$. Now, let us also consider the experiment

$$\mathscr{E}^* = (]0,1[,\ \mathscr{R}(]0,1[),\ \{P^*_{\theta^*}/\theta^* > -1\})$$

where

$$dP^*_{\theta^*}(y) = (1 + \theta^*)y^{\theta^*}\,dy,\ 0 < y < 1.$$

The experiment $\mathscr{E}$ and $\mathscr{E}^*$ are (Ψ, Γ)-equivalents.

Now, let $g : \mathbb{R}^+ \to \mathbb{R}$ be a r.v. and let us consider the problem of estimating g in $\mathscr{E}$ when we use the loss function

$$W(\theta, x) = (x - g(\theta))^2 \quad , \quad x \in \mathbb{R}^+,\ \theta > 0.$$

If we make

$$W^*(\theta^*, x) = \left(x - g\left(\frac{1}{1 + \theta^*}\right)\right)^2 \quad , \quad x \in \mathbb{R},\ \theta^* > -1$$

the decision problems (estimation problems, in this case) $\mathscr{D} = [\mathscr{E}, (\mathbb{R}, \mathscr{R}), W]$ and $\mathscr{D}^* = [\mathscr{E}^*, (\mathbb{R}, \mathscr{R}), W^*]$ are (Ψ, Γ, Λ)-equivalent where $\Lambda(x) = x,\ x \in \mathbb{R}$.

The product decision problems $\mathscr{D}^n$ and $\mathscr{D}^{*n}$ are also equivalent. Since

$$T_n(x_1, ..., x_n) = n^{-1} \sum_{i=1}^{n} x_i \quad , \quad x_1, ..., x_n > 0$$

is a complete sufficient statistic for $\mathscr{E}^n$, then

$$T^*_n(y_1, ..., y_n) = -n^{-1} \sum_{i=1}^{n} \log y_i \quad , \quad 0 < y_1, ..., y_n < 1$$

is a complete sufficient statistic for $\mathscr{E}^{*n}$.

Now, let us consider the problem of testing a hypotheses $\Theta_0 \subset \mathbb{R}^+$ in $\mathscr{E}$ against $\Theta_1 = \Theta_0^c$. If $\phi : \Omega \to [0,1]$ is a level α UMP test for this problem, then $\phi^* = \phi \circ \Psi^{-1}$ is a level α UMP test for the problem of testing $\Gamma(\Theta_0)$ against $\Gamma(\Theta_1)$ in $\mathscr{E}^*$.

Example 2: The experiment associate to a statistical inference problem on a (normal) linear model has the form

$$\left(\mathbb{R}^n, \mathscr{R}^n, \left\{\prod_{i=1}^{n} P_{(\mu_i,\sigma)}/(\mu_1, ..., \mu_n) \in V, \sigma > 0\right\}\right)$$

where V is an s-dimensional subspace of $\mathbb{R}^n$ with $s < n$ and $P_{(\mu,\sigma)}$ is the distribution law of r.r.v. $N(\mu, \sigma^2)$.

Let C be an $n \times n$ orthogonal matrix such that its s first columns constitutes a orthonormal basis of V and let us consider the transformation Ψ of $\mathbb{R}^n$ onto itself defined by $\Psi(x) = y$, where $(y_1, ..., y_n) = (x_1, ..., x_n)C$. This reduces the problem to the named canonical form; in fact, this is an experiment equivalent to the first one in which the work is more simple. The transformation Ψ induces, in a natural way, a transformation Γ between the parameter spaces; namely,

$$\Gamma : (\mu_1, ..., \mu_n, \sigma) \in V \times \mathbb{R}^+ \to (v_1, ..., v_n, \sigma) \in W \times \mathbb{R}^+$$

where

$W = \{(y_1, ..., y_n) \in \mathbb{R}^n/y_{s+1} = \cdots = y_n = 0\}$ and $(v_1, ..., v_n) = (\mu_1, ..., \mu_n)\cdot C$.

We refer to Lehmann (1983) and Lehmann (1986) for more details on this type of problems.

Example 3: It is well known that for any complete and separable metric space there is a one-to-one map from in onto $(\mathbb{R}, \mathscr{R})$ that is also bimeasurable; see, for ex., Ash (1972, p. 195). In particular, every n-dimensional decision problem is equivalent to a real decision problem.

3. EQUIVALENCE AND SOME ASYMPTOTIC PROPERTIES

We have shown that the above notion of equivalence is well adapted to the decision problem stated in the introduction. Next, we shall give some results to justify that this notion of equivalence preserves also some other usual properties of solutions of a statistical inference problem.

Theorem 5: *Let $\mathscr{E}$ and $\mathscr{E}^*$ be two experiments (Ψ, Γ)-equivalent, $f : \Theta \to \mathbb{R}$ a r.v. and $T_n, n = 1, 2, ...,$ a estimate of f. Let $f^*(\theta^*)$*

$=f(\Gamma^{-1}(\theta^*))$, $\theta^* \in \Theta^*$, *and* $T_n^* = T_n \circ (\Psi^n)^{-1}$. *Then* T_n *is a consistent estimate of* f *iff so it* T_n^* *of* f^*.

Proof: Let $\varepsilon > 0$ and $\theta \in \Theta$. For $\theta^* = \Gamma(\theta)$, we have

$$\{\omega^* = (\omega_1^* . \ldots, \omega_n^*) \in \Omega^{*n} / |T_n^*(\omega^*) - f^*(\theta^*)| > \varepsilon\} = \\ = \Psi^n(\{\omega \in \Omega^n / |T_n(\omega) - f(\theta)| > \varepsilon\}).$$

Hence

$$P_{\theta^*}^{*n}(\{\omega^* / |T_n^*(\omega^*) - f^*(\theta^*)| > \varepsilon\}) = P_\theta^n(\{\omega / |T_n(\omega) - f(\theta)| > \varepsilon\})$$

and this finishes the proof. □

Example 1: (continued) The statistic $T_n(x_1, \ldots, x_n) = n^{-1} \sum x_i$ is a consistent estimate for θ in $\mathscr{E}$. Then $T_n^*(y_1, \ldots, y_n) = -n^{-1} \sum \log y_i$ is a consistent estimate for $(1 + \theta^*)^{-1}$ in $\mathscr{E}^*$.

Let $\mathscr{E}$ be an experiment and let $g : \Theta \to \mathbb{R}$ be measurable. For $n = 1, 2, \ldots$, let T_n be an estimate of g on Ω^n. It is said that T_n is asymptotically normal if for all x in $\mathbb{R}$ and all θ in Θ

$$P_\theta^n(n^{\frac{1}{2}}(T_n - g(\theta)) \leqslant x) \underset{n \to \infty}{\longrightarrow} \Phi_\theta(x)$$

where Φ_θ denotes the distribution function of a r.r.v. $N(0, v(\theta))$. We shall denote this fact by $n^{\frac{1}{2}}(T_n - g(\theta)) \xrightarrow{d_\theta} N(0, v(\theta))$, $\forall\, \theta$.

Then, we have the following results.

Theorem 6: *Let* $\mathscr{E}$ *and* $\mathscr{E}^*$ *be two experiments* (Ψ, Γ)*-equivalent,* g *a r.r.v. on* Θ *and* T_n *an asymptotically normal estimate of* g *on* Ω^n *as defined above if* $g^*(\theta^*) = g(\Gamma^{-1}(\theta^*))$ *and* $T_n^* = T_n \circ (\Psi^n)^{-1}$, *then* $n^{\frac{1}{2}}(T_n^* - g^*(\theta^*)) \xrightarrow{d_{\theta^*}} N(0, v^*(\theta^*))$, $\theta^* \in \Theta^*$, *where* $v^*(\theta^*) = v(\Gamma^{-1}(\theta^*))$.

Proof: For $x \in \mathbb{R}$, $\theta \in \Theta$ and $n = 1, 2, \ldots$ we shall denote

$$A_{n,x,\theta} = \{\omega \in \Omega^n / n^{\frac{1}{2}}(T_n - g(\theta)) \leqslant x\}$$

and

$$A_{n,x,\theta^*}^* = \{\omega^* \in \Omega^{*n} / n^{\frac{1}{2}}(T_n^* - g^*(\theta^*)) \leqslant x\}$$

Therefore $A^*_{n,x,\theta^*} = \Psi^n(A_{n,x,\theta})$ if $\theta^* = \Gamma(\theta)$ and then

$$P^{*n}_{\theta_*}(A^*_{n,x,\theta^*}) = P^n_\theta(A_{n,x,\theta})$$

which gives the proof. □

From now on, the framework will be the following. $\mathscr{E}$ and $\mathscr{E}^*$ will denote (Ψ, Γ)-equivalent as above, Θ and Θ^* are open subsets of $\mathbb{R}$ and Γ a one-to-one and everywhere differentiable map from Θ onto Θ^* whose derivative is never nulle. We shall suppose that $\mathscr{E}$ is dominate by a σ-finite measure μ on $\mathscr{U}$ and we shall denote by p_θ the density of P_θ and $L(\omega, \theta) = L_\omega(\theta) = \log p_\theta(\omega)$. Finally, we shall consider that the Fisher information $I(\theta)$ is well defined; we shall denote by $V_\theta(\omega)$ the derivative of L_ω at the point θ and, then, $I(\theta) = E_\theta(V_\theta^2)$. Analogous assumptions shall be made for $\mathscr{E}^*$.

With obvious notations, we have the following lemma.

Lemma 1: *It the preceding setting, if μ is a probability*

a) *$\mathscr{E}^*$ is dominated by $\mu^* = \mu^\Psi$ and $p^*_{\theta^*}(\Psi(\omega)) = p_\theta(\omega)$ if $\theta^* = \Gamma(\theta)$, where $\mu^\Psi(A^*) = \mu(\Psi^{-1}(A^*))$ and $p^*_{\theta^*} = dP^*_{\theta^*}/d\mu^*$.*

b) *$V_\theta(\omega) = \Gamma'(\theta) \cdot V^*_{\Gamma(\theta)}(\Psi(\omega))$ and $I(\theta) = (\Gamma'(\theta))^2 \cdot I^*(\Gamma(\theta))$ for all ω and θ.*

Proof: a) It is immediate.

b) If $L^*(\omega^*, \theta^*) = L^*_{\omega^*}(\theta^*) = \log p^*_{\theta^*}(\omega^*)$, it happens that $L(\omega, \theta) = L^*(\Psi(\omega, \Gamma(\theta))$, for all ω and θ, and then, if $\theta^* = \Gamma(\theta)$ and $\omega^* = \Psi(\omega)$,

$$V_\theta(\omega) = \frac{dL^*_{\omega^*}}{d\theta}(\Gamma(\theta)) = \frac{dL^*_{\omega^*}}{d\theta^*}(\theta^*) \cdot \Gamma'(\theta) = \Gamma'(\theta) \cdot V^*_{\theta^*}(\omega^*)$$

where $V^*_{\theta^*}(\omega^*)$ is the derivative of $L^*_{\omega^*}$ at the point θ^*.

The rest follows easily from this. □

A estimate T_n of $g : \Theta \to \mathbb{R}$ on Ω^n said to be asymptotically efficient if $n^{\frac{1}{2}}(T_n \to g(\theta)) \xrightarrow{d_\theta} N(0, v(\theta))$ for all θ, where $v(\theta) = (g'(\theta))^2 \cdot (I(\theta))^{-1}$. We assume, obviously, that g is differentiable in Θ. The following result holds in the preceding framework.

Theorem 7: *Let $g^*(\theta^*) = g(\Gamma^{-1}(\theta^*))$ and $T^*_n = T_n \circ (\Psi^n)^{-1}$. Then T^*_n is an asymptotically efficient estimates of g^* if so is T_n of g.*

Proof: It is shown in theorem 6 that $v^*(\theta^*) = v(\Gamma^{-1}(\theta^*))$ (here, we use the notations used there). Hence, if T_n is asymptotically efficient, it follows from this and the previous lemma that

$$\begin{aligned} v^*(\theta^*) &= I(\Gamma^{-1}(\theta^*))^{-1} \cdot (g'(\Gamma^{-1}(\theta^*)))^2 = \\ &= I^*(\theta^*)^{-1} \cdot (\Gamma'(\Gamma^{-1}(\theta^*)))^{-2} \cdot (g'(\Gamma^{-1}(\theta^*)))^2 = \\ &= I^*(\theta^*)^{-1} \cdot (g^{*-1}(\theta^*))^2. \ \square \end{aligned}$$

Example 1: (again) The CLT shows that T_n is asymptotically normal $N(0, \theta^2)$, for all $\theta > 0$. Since $I(\theta) = \theta^{-2}$, $\theta > 0$, T_n is an asymptotically efficient estimate of θ. Hence, T_n^* is an asymptotically efficiente estimate of $(1 + \theta^*)^{-1}$.

4. EQUIVALENCE AND THE BAYESIAN APPROACH

The bayesian treatment of a decision problem requires the presence of a prior distribution Q on the parameter space $(\Theta, \mathscr{I})$. We shall call bayesian the experiment (1) and the decision problem (2) when this is the case. A Bayes solution for such a problem is a strategy S wich minimizes the Bayes risk

$$R_{W,S}^{Q} = \int_{\Theta} R_{W,S}(\theta)\, dQ(\theta).$$

In addition to those of Def. 1, an assumption on the prior distributions Q and Q^* must be made in order to obtain a right notion of equivalence of two bayesian decision problems $\mathscr{D}$ and $\mathscr{D}^*$ or two bayesian experiments $\mathscr{E}$ and $\mathscr{E}^*$; namely

iii) $Q^{\Gamma} = Q^*$.

The following result is a consequence of this and theorem 1.

Theorem 8: *Let $\mathscr{D}$ and $\mathscr{D}^*$ be equivalent bayesian decision problems. If S and S^* are associate strategies then*

$$R_{W,S}^{Q} = R_{W^*,S^*}^{Q^*}$$

Next, we analyse the relationship between the posterior distributions. First we fix some notations. $\mathscr{E}$ and $\mathscr{E}^*$ will be (Ψ, Γ)-equivalent bayesian experiments. With obvious notations, we shall suppose that $\{P_\theta/\theta \in \Theta\}$ is dominated by a σ-finite measure μ on $\mathscr{U}$. Then the $P^*_{\theta^*}$ are also dominated by $\mu^* = \mu^\Psi$. The report between the respectives densities is given by the lemma in the precedent paragraph. We shall suppose that $P_\theta(A)$ is a transition probability on $\Theta \times \mathscr{U}$; then so is also $P^*_{\theta^*}(A^*)$ on $\Theta^* \times \mathscr{U}^*$. This in the case if the likelihood function $\mathscr{L}(\omega, \theta) = p_\theta(\omega)$ is a r.r.v. on $(\Omega \times \Theta, \mathscr{U} \times \mathscr{I})$ as it is shown in Barra (1971). Two probability measures β and β^* are defined on $\mathscr{U}$ and $\mathscr{U}^*$, resp., by

$$\beta(A) = \int_\Theta P_\theta(A)\, dQ(\theta) \quad \text{and} \quad \beta^*(A^*) = \int_{\Theta^*} P^*_{\theta^*}(A^*)\, dQ^*(\theta^*).$$

A posterior distribution for the bayesian experiment $\mathscr{E}$ is (if there exist) a transition probability $\pi_\omega(T)$ on $\Omega \times \mathscr{I}$ such that $\int_T P_\theta(A)\, dQ(\theta) = \int_A \pi_\omega(T)\, d\beta(\omega)$ for every $A \in \mathscr{U}$ and $T \in \mathscr{I}$. Under the made assump tions, it can be shown that there is always a posterior distribution for $\mathscr{E}$ if Θ is a Borel subset of some euclidean space (or, more generally, if there is a one-to-one, onto and bimeasurable map from Θ onto a Borel subset of some euclidean space). One has then

$$\frac{d\pi_\omega}{dQ}(\theta) = \frac{p_\theta(\omega)}{\int_\Theta p_\theta(\omega)\, dQ(\theta)} \quad \mu\text{-a.e.} \tag{3}$$

In the precedent set up, we have the following result.

Theorem 9: *Let $\mathscr{E}$ and $\mathscr{E}^*$ be (Ψ, Γ)-equivalent bayesian experiments as above. Suppose that there exists a posterior distribution $\pi_\omega(T)$, $\omega \in \Omega$, for $\mathscr{E}$. Then*

$$\pi^*_{\omega^*} = \pi^\Gamma_\omega \quad \textit{if} \quad \Psi(\omega) = \omega^*$$

is a posterior distribution for $\mathscr{E}^*$ and its density with regard to Q^*, wich is given by the shattered analogous of (3), satisfy

$$\frac{d\pi^*_{\Psi(\omega)}}{dQ^*}(\Gamma(\theta)) = \frac{d\pi_\omega}{dQ}(\theta) \quad \textit{for all } \omega \textit{ and } \theta.$$

Proof: It is easy to see that $\beta^*(\Psi(A)) = \beta(A)$ for all $A \in \mathscr{U}$. Obviously $\pi^*_{\omega^*}$ is a transition probability on $\Omega^* \times \mathscr{I}^*$. Furthermore, since $Q^* = Q^\Gamma$ and $\beta^* = \beta^\Psi$, we have for $A^* \in \mathscr{U}^*$ and $T^* \in \mathscr{I}^*$ that

$$\int_{T^*} P^*_{\theta^*}(A^*)\, dQ^*(\theta^*) = \int_T P^*_{\Gamma(\theta)}(A^*)\, dQ(\theta) = \int_T P_\theta(A)\, dQ(\theta) =$$
$$= \int_A \pi_\omega(T)\, d\beta(\omega) = \int_A \pi^*_{\Psi(\omega)}(T^*)\, d\beta(\omega) = \int_{A^*} \pi^*_{\omega^*}(T^*)\, d\beta^*(\omega^*)$$

where $\Psi(A) = A^*$ and $\Gamma(T) = T^*$. This shows that π_ω is a posterior distribution for $\mathscr{E}^*$. The rest follows from the definition of $\pi^*_{\omega^*}$ and the lemma 1. □

Remark: The Bayes solution for the problem of estimating a r.r.v. f on Θ when we use the squared error loss is $T_B(\omega) = \int_\Theta f(\theta)\pi_\omega(d\theta)$. It follows from the precedent results that $T^*_B(\omega^*) = T_B(\Psi^{-1}(\omega^*))$ is the Bayes solution for the shattered equivalent decision problem.

Example 1: (concluded) Let us suppose the parameter space $(\Theta, \mathscr{I})$ considered in this example endowed with the prior distribution Q whose density with regard to the Lebesgue measure on $\mathbb{R}^+$ is $q(\theta) = \theta^{-2}$ if $\theta > 1$, $=0$ otherwise. The posterior distribution $\pi_x(x > 0)$ is then given by its density with regard to Q

$$\frac{d\pi_x}{dQ}(\theta) = \frac{x^2\theta^{-1}e^{-x}/\theta}{1 - x - e^{-x}} \quad , \quad \theta > 0$$

and hence

$$d\pi_x(\theta) = \frac{x^{-2}\theta^{-3}e^{-x}/\theta}{1 - x - e^{-x}} d\theta \quad , \quad \theta > 1.$$

The prior distribution Q^* on Θ^* is the uniform distribution on $]-1, 0[$, and then for $0 < y < 1$

$$d\pi^*_y(\theta^*) = \frac{(1+\theta^*)y^{(1+\theta^*)}(\log y)^2}{1 - y\log y} d\theta^* \quad \text{if } -1 < \theta^* < 0, \ =0 \text{ otherwise.}$$

The Bayes estimate of θ is $T_B(x) = \dfrac{x(1 - e^{-x})}{1 - x - e^{-x}}$ for $x > 0$ and then the Bayes estimate of $(1 + \theta^*)^{-1}$ is

$$T^*_B(y) = \frac{(y - 1)\log y}{1 - y + \log y} \quad \text{for} \quad 0 < y < 1.$$

REFERENCES

ASH, R. B. (1972): *Real analysis and probability*, Academic Press, New York.

BARRA, J. R. (1971): *Notions fondamentales de statistique mathematique*, Dunod, Paris.

LEHMANN, E. L. (1986): *Testing statistical hypotheses,* 2nd ed., Wiley, New York.

LEHMANN, E. L. (1983): *Theory of point estimation,* Wiley, New York.

LEHMANN, E. L. (1988): «Comparing location experiments», *Ann. Statist.,* **16**, 521-533.