

Bayesian and Maximum Likelihood Solutions: An Asymptotic Comparison Related to Cost Function

¹Mechakra Hadda, ²Chadli Asssia and ³Tiah Naceur

¹Faculty of Medicine, Badji Mokhtar University, Annaba, Algeria

²Lanos, Badji Mokhtar University, Annaba, Algeria

³Department of Mathematic, Epst Annaba, Algeria

Abstract: Problem statement: Wald showed that the minimax solution is the Bayesian solution with respect to the law a priori the worst. We try to establish a similar result by comparing the Bayesian solution and the solution of maximum likelihood when the parameter space is a compact metrizable group. **Approach:** we take as a priori law Haar measure because we reduce the problem by invariance. We construct a sequence of cost functions for which we obtain a sequence of solutions Bayesian which converges to the solution of the maximum likelihood. **Results:** We show that both solutions are asymptotically equal. **Conclusion/Recommendation:** The generalization when the parameter space is a local compact group.

Key words: Decision theory cost function, haar measure, bayesian solution, maximum likelihood solution, topological group, multivoc function, measurability, a priori law

INTRODUCTION

Problem position: The fundamental problem of statistical decision theory can be summarized as follows: Given the triplet (Θ, D, C) . Where: Θ parameter space. D space of decision rules. C a cost function and $\omega \in \Omega$ a random element on which the law of distribution P_θ depends on a parameter $\theta \in \Theta$.

What rule of decision $\delta(\omega) \in D$ a statistician must choose?). The space Ω cited above is called space of observations, we shall provide it with the σ -algebra \mathfrak{a} . The risk function associated with the rule δ is defined by:

$$R(\theta, \delta) = E(C(\theta, \delta(\omega)))$$

$R(\theta, \delta)$ represents the average cost when θ is estimated by $\delta(\omega)$. The issue is then to choose an optimal statistical decision rule in the sense that $R(\theta, \delta)$ is uniformly minimum. Unknown define q it am and depend of to short-circuit this difficulty, we order the different rules of decision according to d 'other principles such as: The principle of minimax, or the principle of Bayes, or else we consider only rules of decision based on intuitive methods such as the principle of the maximum of likelihood. We shall assume the family (P_θ) , $\theta \in \Theta$ dominated by a likelihood Q on (Ω, \mathfrak{a}) . We shall denote by $L(\theta)$ specific gravity p_θ all over the report in Q . The estimator of the

maximum of likelihood $\hat{\theta}$ is the value of θ which maximizes $L(\theta, \omega)$ at the sight of ω . It is easy to point out that this definition is wholly intuitive. A rule of decision δ_M is called a minimax, if it minimizes (among all δ) the risk maximum that is:

$$\sup_{\theta} (R(\theta, \delta_M)) = \inf_{\delta} \left(\sup_{\theta} (R(\theta, \delta)) \right)$$

In the frame Bayesian, θ define it he am considered to be an unpredictable variable and a law of distribution is μ allocated. This law is called the a priori law. A rule of decision δ_B is called the rule of Bayes, if it minimizes the risk of Bayes, is to say:

$$r(\mu, \delta_B) = \inf_{\delta} (r(\mu, \delta)) = \inf_{\delta} (E_{\mu}(R(\theta, \delta)))$$

where mathematical expectation $E_{\mu}(R(\theta, \delta))$ is calculated in comparison with the law μ . The expression $r(\mu, \delta)$ is called the risk of Bayes. For a more detailed description of these rules, the reader can see (Berger, 1980) or (Fergusson, 1969). The three methods estimation which we have just represented does not give necessarily the same valutors for unknown parameter θ . Even in an asymptotic frame, see (Cam, 1953) for instance for a comparison of performances

Corresponding Author: Mechakra Hadda, Department of Medecin, Badji Mokhtar University, Annaba, Algeria

asymptotic performances of the methods of Bayes and of the methods of the maximum of likelihood. A Wald (1950) however showed that if the rule of minimax decision exists, it is also a rule of Bayesian decision in comparison with the most disadvantageous law. That is to say a rule of Bayes in comparison with the law which would maximize (among all law) the risk of Bayes. The purpose of the present study is to establish a result similar to that of A Wald between valuator of Bayes and valuator of the maximum of likelihood. We shall try to find a frame (Θ, D, C) for which valuator of Bayes and valuator of the maximum of likelihood are equal or asymptotically equal. Before undertaking the building of this frame, let us notice two important things.

We reduce the problem by invariance. This simplification is motivated by the following considerations: In case of equality of Bayes estimator and estimate of the maximum of likelihood. We can say that a priori law μ of the parameter θ has not influence. In that case μ can be interpreted as a priori law "not informative. (Jeffreys, 1998) justifies the lack of information in μ by ownership of invariance See (Florens, 1978).

It is possible to obtain Bayesian resolutions for cost functions more general than those encountered in the model of statistical decision of WALD, where the cost function is expressed as the sum of two positive cost the first one relating to the observation performed and the second one in taken a final decision. We can consider a total cost $C(\theta, z, \omega, t)$ of final decision z when the parameter is worth θ , after observation of realization ω up to time t See (Lanery, 1984). In our case, the noticed system is not dynamics, that is the observation is fixed in lasted and I the information disponible is expressed by the tribu α . We have a cost function:

$$C : (\theta, \delta, \omega) \in \Theta \times D \times \Omega \rightarrow C(\theta, \delta, \omega) \in \mathbb{R}$$

Description of results: In first party: we remind of all preliminary notions which we need such as; the statistical decision theory; topological groups; measure of Haar.

In second party: we compare the Bayesian solution and the solution of the maximum of likelihood when the space parameters are a metrizable compact group. We construct in anticipation a sequence of bounded cost functions. We prove the existence Bayesian solutions and solutions of the maximum of the likelihood under of regularity. We show finally that these two types of solutions are asymptotically identical. The sequence of cost functions is constructed using Uryshon Lemma. The measurability of solutions (Bayesian and of the maximum of likelihood) is

established using the theorem of the measurement section of K. Kuratowski and Ryll-Nardzewski.

Preliminaries: We recall here some preliminary notion such as theory of statistical decision theory, topological groups, Haar measure, which will be used in the next chapters. These notions are elementary, most of the proofs are omitted; they are given in order to the presentation clear.

Recalls of the theory of measure:

Definition 2.1.1: Let (Ω, α) and (Θ, τ) are two measurable spaces. $(P_\theta), \theta \in \Theta$ is a family of likelihood of transition on (Ω, α) defined on (Θ, τ) If (1) $\forall \theta \in \Theta, A \rightarrow p_\theta(A)$ a likelihood on (Ω, α) (2) $\forall A \in \alpha \rightarrow p_\theta(A)$ is τ -measurable

Proposition 2.1.2: If $(P_\theta), \theta \in \Theta$ is a family of likelihood of transition on (Ω, α) parametrized on (Θ, τ) . For any positive function f to $(\Omega \times \Theta, \alpha \times \tau)$ $(\mathbb{R}, \mathcal{B} \mathbb{R})$ The mapping:

$$\theta \rightarrow \int_{\Omega} f(\omega, \theta) P_\theta(d\omega)$$

Is defined and is measurable.

Proof: If $f = X_{A \times B}$ where $A \in \alpha, B \in \tau$:

$$\begin{aligned} \theta \rightarrow \int_{\Omega} X_{A \times B}(\omega, \theta) P_\theta(d\omega) &= \int_A X_B(\theta)(d\omega) = \\ X_B(\theta) P_\theta(A) \end{aligned}$$

Is measurable. The measurability is verified for the characteristic functions and is generalized according to classical technique.

Theorem 2.1.3: Let $(P_\theta), \theta \in \Theta$ a family of likelihood of transition on (Ω, α) defined on (Θ, τ) and μ a likelihood on (Θ, τ) . If there exist a unique likelihood Π on $(\Omega \times \Theta, \alpha \times \tau)$ such as for any function h positive measurable $(\Omega \times \Theta, \alpha \times \tau)$ in $(\mathbb{R}, \mathcal{B} \mathbb{R})$ have:

$$\begin{aligned} \iint_{\Omega \times \Theta} h(\omega, \theta) \Pi(d\omega, d\theta) &= \\ \int_{\Theta} \left[\int_{\Omega} h(\omega, \theta) P_\theta(d\omega) \right] \mu(d\theta) \end{aligned}$$

Proof: Thanks to, the proposition 1. 1, the second member is well defined. We define π by:

$$\begin{aligned} \forall \mu \in \alpha \times \tau, \pi(\mu) &= \\ \int_{\Theta} \left[\int_{\Omega} x u(\theta, \omega) P_\theta(d\omega) \right] \mu(d\theta) \end{aligned}$$

We prove easily that π a likelihood on $(\Omega \times \Theta, \alpha \times \tau)$.

To show the Unicity and the equality given in theorem, we begin with the characteristic functions and is generalized according to classical technique.

Remark 2.1.4: This theorem can be extended in case μ is a positive measure, σ -finie on $(\Omega \times \Theta, \alpha \times \tau)$.

Definition 2.1.5: Let two measurable spaces (Ω, α) and (Θ, τ) and $(P_\theta), \theta \in \Theta$ a family of likelihood of transition on (Ω, α) defined on (Θ, τ) . We call statistical structure (or model) the system:

$$\{\Omega, \alpha, (P_\theta) \in \Theta\}$$

We say that the statistical structure $\{\Omega, \alpha, (P_\theta), \theta \in \Theta\}$ is dominated:

If there is a positive measure, σ -f inite Q in (Ω, α) such as:

$$\forall \theta \in \Theta, P_\theta \ll Q$$

According the theorem of Radon-Nikodym, there is a mapping f_θ of (Ω, α) in $(\mathbb{R}, \mathcal{B} \mathbb{R})$ measurable such as:

$$\forall A \in \alpha, P_\theta(A) = \int_A f_\theta(\omega) Q(d(\omega))$$

f_θ is called probability density of P_θ in comparison with Q ;

We call likelihood function the mapping:

$$L : (\omega, \theta) \in \Omega \times \Theta \rightarrow L(\omega, \theta) = f_\theta(\omega) \in \mathbb{R}^+$$

Remark 2.1.6: 1. The dominating measure is not unique. If:

$$\forall \theta \in \Theta, P_\theta \ll Q$$

where, Q is a positive step, σ -f inite on (Ω, α) and if $Q \ll S$ where v is a measurable positive σ -f inite on (Ω, α)

Then:

$$\forall \theta \in \Theta, P_\theta \ll v$$

We can always and it is often convenient for theoretical calculations to choose as the dominating measure of a probability $(P_\theta), \theta \in \Theta$ (Barra, 1971).

Statistical decision theory: This section reminds certain notions of the theory of decision already mentioned in the previous chapter. We repeat it there with more details. We have:

(Ω, α) a measurable space (space of observations).

(Θ, τ) a measurable space (space of parameters).

(D, \mathcal{D}) a measurable space (space of possible decisions).

$(P_\theta), \theta \in \Theta$ a family of likelihood of transition on (Ω, α) defined on (Θ, τ) . $p_\theta(\cdot)$ governs the observation ω in Ω when, θ is the value of the parameter.

C a function measurable:

$$C : (\theta, \delta, \omega) \in \Theta \times D \times \Omega \rightarrow C(\theta, \delta, \omega) \in \mathbb{R}^+$$

where $C(\theta, \delta, \omega)$ represents the expense of decision δ , when the parameter is θ and when they observed ω . A rule of decision (or strategy) δ a mapping of (Ω, α) in (D, \mathcal{D}) measurable; it consists in deciding $\delta(\omega)$ having noticed ω . We shall denote by $C(\theta, \omega, \delta(\omega))$ the expense of decision $\delta(\omega)$ when we noticed ω and when a parameter θ is and by Δ the set of rules of decision. To solve a problem of statistical decision consists in choosing a rule of decision in Δ according to some criteria acknowledged to be reasonable. We have the risk function defined by:

$$R(\theta, \omega) = \int_\Omega C(\theta, \omega, \delta(\omega)) p_\theta d\omega.$$

As a criterion of choice, we define a preferred relation on Δ . Let δ^* and $\delta' \in \Delta$: δ^* is preferred to:

$$(\delta \leq \delta') \Leftrightarrow [R(\theta, \delta') \leq R(\theta, \delta), \forall \theta \in \Theta]$$

δ^* is equivalent to δ :

$$(\delta \sim \delta') \Leftrightarrow [R(\theta, \delta') \leq R(\theta, \delta), \forall \theta \in \Theta]$$

δ^* is strictly preferred:

$$\delta < \delta' \{ [R(\theta, \delta') \leq R(\theta, \delta), \forall \theta \in \Theta] \\ \text{and } [\exists \theta \in \Theta, R(\theta, \delta') < R(\theta, \delta)] \}.$$

Definition 2.2.3: $\delta \in \Delta$ is called a permissible rulemaking if there is no $\delta' \in \Delta$ such that $\delta < \delta'$. Otherwise, there is no $\delta' \in \Delta$ such that we have:

$$R(\theta, \delta') \leq R(\theta, \delta), \forall \theta \in \Theta \text{ and } \\ \exists \theta \in \Theta, R(\theta, \delta') < R(\theta, \delta)$$

Definition 2.2.4: A decision rule $\delta_0 \in \Delta$ is uniformly optimal if:

$$\forall \delta \in \Delta, \delta \leq \delta_0$$

$$\text{ie } R(\theta, \delta_0) \leq R(\theta, \delta), \forall \delta \in \Delta, \forall \theta \in \Theta$$

$$(x, y) \in G \times G \rightarrow xy \in G$$

It does not exist in general optimal decision rule, since the risk function $R(\theta, \delta)$ depends on a parameter unknown θ , but there are rules of decision-eligible. To choose between them we are led to:

The mapping:

$$(x, y) \in G \times G \rightarrow x^{-1} \in G$$

1) Is to think conservatively = minimax solution. We let:

Is continuous. A given group with a topology compatible with its structure of the group is called topological group.

$$\forall \delta \in \Delta, \rho(\delta) = \sup_{\theta \in \Theta} R(\theta, \delta)$$

Remarks 2.3.2: (1) For any $e \in G$, the translation to the left $x \rightarrow ax$ (respectively the translation to the right $x \rightarrow xa$) is continuous. It is a homeomorphism of d on to itself.

δ_M is called a minimax solution if:

For any $a, b \in G$, the mapping:

$$\rho(\delta_M) \leq \rho(\delta), \forall \delta \in \Delta.$$

$$x \rightarrow axb \text{ (especialy } x \rightarrow axa^{-1} \text{)}$$

That is to say:

Is a homeomorphism of d onto itself.

$$\sup_{\theta \in \Theta} R(\theta, \delta_M) = \inf_{\delta \in \Delta} (\sup_{\theta \in \Theta} R(\theta, \delta))$$

The mapping:

$$x \rightarrow x^{-1}$$

2) Either to give a preconceived idea of the solution: For example, the point estimate of the parameter θ , we choose the value of θ that gives the highest probability of the system studied. This procedure provides a maximum likelihood estimator.

Is bijective and is equal the inverse mapping. It is a homeomorphism of G onto itself.

3) Is to assume that the parameter space Θ , we have a probability μ , that weight between the possible values of the parameter. This is the Bayesian decision theory. In this case, we consider the Bayes risk defined as:

Proposition 2.3.4: Let d a topological group. (i) For any open subset (respectively closed) A of G and for any $x \in G$: groupse xA, Ax and A^{-1} are open (respectively closed). (ii) For any open subset B of G and for any subset A of G : groups BA, AB are open (resp. closed). (iii) If V a neighborhood of e in G and A any non empty subset of G , then GA and AV are neighborhood of A .

$$r(\mu, \delta) = \int_{\Theta} R(\theta, \delta) \mu(d\theta) = \int_{\Theta} \left[\int_{\Omega} C(\theta, \omega, \delta(\omega)) p_{\theta}(d\omega) \right] \mu(d\theta)$$

Remark 2.3.5: If $a \in G$ and V a neighborhood of e then aV and Va are neighborhood of a .

δ_B is called bayesian solution if: $r(\mu, \delta_B) = \inf_{\delta} (r(\mu, \delta))$.

Definition 2.3.6: Let G be a group and E a set. An operation (either operation to the left or action) of G on E is a mapping:

Such situations are common in statistics. For more details, see (Fougereaud and Fuchs, 1967) and (Ulmo and Bernier, 1973)

$$(s, x) \in G \times E \rightarrow s.x \in E$$

Topological groups: Just like that pointed out in the introduction, command us led to study a priori law meeting ownership of invariance; hence the importance to remind of some notions on the topological groups and of the measure of Haar. It is the aim of the present section. Let G a denotes a multiplicative group (not necessarily commutative) and K its neutral element.

such as: 1) If e is the neutral element of G , then we have:

$$e.x = x, \forall x \in E$$

Definition 2.3.1: We say that a topology on G is compatible with the structure of group if both following mappings: a) the mapping:

$\forall s, t \in G$ we have:

$$s.(t.x) = (st).x, \forall x \in E$$

We say that d operate on E .

Definition 2.3.7: Let G be a topological group and E a topological space. We shall say that G operating continuously in E if the mapping $(s, x) \rightarrow s \cdot x$ is continuous.

Lemma 2.3.7: If G is a topological group operating continuously in a topological space E then. for any $s \in G$, the mapping:

$$x \in E \rightarrow s \cdot x \in E$$

Is a homeomorphism.

Remarks 2.3.8: An operation to the right of a group G in a set E is a mapping:

$$(s \cdot x) \in G \times E \rightarrow x \cdot s \in E$$

Such as:

$$x \cdot e = x, \forall x \in E$$

$$x \cdot (st) = (x \cdot s) \cdot t, \forall s, t \in G, \forall x \in E$$

All definitions and ownership enunciated more high remain valid in case of an operation to the right.

Haar measure notations: Let G be a topological group operating continuously to left on a local level compact. We shall denote by: $(\gamma(s))$ the homeomorphism of E in $E(s)$, defined by:

$$\gamma(s)x = s \cdot x, \forall s \in G, \forall x \in E$$

If μ is a measure defined on E , we shall denote by $(\gamma(s)\mu)$ the measure image of μ by $\gamma(s)$. We have:

$$(\gamma(s)\mu)(A) = \mu(s^{-1}A) \forall A.$$

where, A a measurable set of E .

Definition 2.4.1: Let a measure on E . We say that μ is invariant by G if:

$$\gamma(s)\mu = \mu, \forall s \in G$$

Remark 2.4.2: If G is a topological group operating continuously to the right on E . We shall denote by: $\delta(s)$ the homeomorphism of E in E , defined by:

$$\delta(s)x = x \cdot s^{-1}$$

If μ is a measure defined on E , we shall denote by $\delta(s)\mu$ measure picture of by $\delta(s)$

Definition 2.4.3: If μ a measure defined on E . We say that μ is invariant to the right by G if:

$$(\delta(s)\mu) = \mu, \forall s \in G$$

If G is a local level compact group operating on to itself by translation to the left and to the right:

$$\gamma(s)x = s \cdot x \text{ and } \delta(s)x = x \cdot s^{-1}$$

Then we can define G on the notions of invariant measure left and right.

Definition 2.4.4: Let G be a locally compact group Haar measure we call the left (respectively right) of G a positive non-zero, σ - fine G invariant on the left (right respivement).

The existence and Unicity of such a measure is given by the following theorem. The proof of this result is given in (Halmos, 1974)

Theorem 2.4.5: On any locally compact group, there is a Haar measure on the left (respectively right) and a near constant factor, that none exists.

Remark 2.4.6: If G is compact, there exists a Haar measure μ and one on G such that $\mu(G) = 1$. It is called a normalized Haar measure

Multivoc mappings definitions and notations: We use multivocs mappings to establish certain results It seems to us make a small paragraph. For more details the reader can consult (Berge, 1966).

Let X and Y two set and for any $x \in X$ we match a subset $\Phi(x)$ of the Y . We say that correspondence:

$$x \in X \rightarrow \phi(x) \subset Y$$

Is multivocal mapping of X in Y :

$$X^* = \{x \in X \text{ such as } \phi(x) \neq \emptyset\}$$

Is called set of definition of ϕ . $Y^* = \bigcup_{x \in X} \phi(x)$ is

called set of values of Φ . If $\Phi(x)$ is composed of only one element, we say that Φ is a valued mapping of X in Y . Upper reverses Φ^+ :

$$\forall B \subset Y : \phi^+(B) = \{x \in X / \phi(x) \subset B\}$$

Inferior reverse Φ^- :

$$\forall B \subset Y : \phi^-(B) = \{x \in X / \phi(x) \cap B \neq \emptyset\}$$

Theorem 2.5.1: We have:

$$\phi^-\left(\bigcup_i A_i\right) = \bigcup_i \phi^-(A_i)$$

The proof is immediate. To end this party, we quote a theorem of Kuratowski and Ryll-Nardzewski, which will be used. Before let us give the following definition.

Definition 2.5.2: A Polish space is a metrisable, separable space, on which there is a metric which is compatible with the topology for which the space is complete.

Theorem 2.5.3: Let X a Polish space and (U, \mathfrak{u}) a measurable space. We consider this multifocal mapping of U in X such as: $\forall u \in U, \Phi(u)$ is closed and non empty in X . We assume that:

$$\phi^-(G) = \{u / \phi(u) \cap G \neq \emptyset\} \in \mathfrak{u}$$

For all G open in X . Then exists ϕ mapping measurable of (U, \mathfrak{u}) in (X, \mathcal{B}_X) such as:

$$\forall u \in U, \phi(\phi(u)) \in \phi(u)$$

For proof, the reader can consult (Kuratowski and Ryll-Nardzewski, 1965).

Bayesian solutions and maximum likelihood solutions when Θ is compact: In this chapter, we give a mathematical framework that allows us to assert that the solutions of Bayesian and maximum likelihood solutions are asymptotically equals. We construct a sequence of cost functions $C_n(\theta, z, \omega)$ bounded, translation invariant, continuous in θ and z and measurable in ω (see Lemma 2.2.1). For this type of cost function, we obtain a Bayesian sequence $(\tilde{\theta}_n)_n$. Solutions which converge pointwise to the solution of maximum likelihood $\hat{\theta}$ (see Theorem.2.4.1). We generalize the Theorem 2.4.1. In fact we remove the restrictive assumption of Unicity of the solution of maximum likelihood.

Assumptions, ratings and results: H1: Let be (Ω, \mathfrak{a}) a measurable space (space of observations).

H2: Θ is the parameter space, we assume that Θ is a group (written multiplicatively) compact and metrizable. We denote by e its unit element, by the d distance of translation invariant and by τ the tribu of Borel on Θ .

Let $(\varepsilon_n)_n$ be a sequence of positive real numbers such that:

$$(\varepsilon_n)_n \text{ decreasing to } 0$$

$$2\varepsilon_{n+1} < \varepsilon_n$$

$$\varepsilon_0 < (\text{diameter}\Theta \wedge 1)$$

The sequence $(v_n = B(e, \varepsilon_n))_{n \in \mathbb{N}}$ is a neighborhood base of e such that:

$$B^*(e, \varepsilon_n).B^*(e, \varepsilon_n) \subset B(e, \varepsilon_{n-1})$$

where, $B^*(e, \varepsilon_n)$ is the closed ball with center e and radius ε_n .

H3: μ is the Haar probability on (Θ, τ) (Halmos, 1974).

H4: (p_θ) θ in Θ is a family of transition probability parameter on (Ω, \mathfrak{a}) set to (Θ, τ) . We assume that there is a probability Q on (Ω, \mathfrak{a}) such that:

$$\forall \theta \in \Theta, p_\theta \ll Q$$

We denote by $L, (\theta)$ the probability density p_θ par rapport to Q .

H5: We assume that:

$$\forall \omega \in \Omega, \theta \rightarrow L(\theta, \omega)$$

Is continuous on Θ .

H6: When θ is the unknown value of the parameter, it is estimated by d called final decision and belongs to D . We assume that $D = \Theta$ (current condition of regularity).

H7: We assume the uniqueness of the solution of maximum likelihood ie:

$$\phi(\omega) = \theta : L(\theta, \omega) = \sup(L(x, \omega), x \in \Theta = \hat{\theta}(\omega))$$

Theorem 2.1.1: For any likelihood function L satisfies H5 and under hypotheses H1, H2, ..., H7, there is a sequence of cost functions $(C_n)_{n \in \mathbb{N}}$ defined on $\Theta \times \Theta \times \Omega$ with values in \mathbb{R} , satisfying the following conditions:

- Bounded and invariant by translation.
- Continuous on $\Theta \times \Theta$
- Measurable on Ω And such as:

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \hat{\theta} \text{ point wisely}$$

Lemmas:

Lemma 3.2.1: Gives the existence and measurability of the solution of maximum likelihood.

Lemma 3.2.1: Under the hypotheses H1, H2, H4 and H5, there exists a mapping $\hat{\theta}$ of (Ω, \mathcal{a}) in (Θ, τ) measurable such as:

$$L(\hat{\theta}(\omega), \omega) \geq L(\theta, \omega), \forall \omega \in \Omega, \forall \theta \in \Theta$$

Proof:

$$\forall \omega \in \Omega, \exists \hat{z}_\omega \in \Theta \text{ such as } L(\hat{z}_\omega, \omega) \geq L(\theta, \omega), \forall \theta \in \Theta$$

Because:

$$\forall \omega \in \Omega, \theta \mapsto L(\theta, \omega)$$

Is continuous on the compact set Θ . It remains to prove that we can choose a \hat{z}_ω measurable in ω .

Consider the following multivoc mapping

$$\phi : \omega \in \Omega \rightarrow \phi(\omega) = \{\theta : L(\theta, \omega) = M(\omega)\} \in \Theta$$

Where

$$M(\omega) = \sup(L(x, \omega), x \in \Theta)$$

For any $\omega \in \Omega, \phi(\omega)$ is non empty and closed in Θ . Let G be an open in Θ ; in a metric space every open set is a countable union of closed $G = \bigcup_k F_k$ where F_k is a closed in Θ .

To proof that:

$$\phi^-(G) = \{\omega / \phi(\omega) \cap G \neq \emptyset\} \in \mathcal{a}$$

It suffices to proof that:

$$\phi^-(F) = \{\omega / \phi(\omega) \cap F \neq \emptyset\} \in \mathcal{a}$$

For any a closed in Θ because $\phi^-(G) = \bigcup_k \phi^-(F_k)$ (see theorem 2.5.1).

F a closed in Θ , is itself compact:

$$\begin{aligned} \phi^-(F) &= \{\omega / \phi(\omega) \cap F \neq \emptyset\} = \{\omega / \sup_{\theta \in F} L(\theta, \omega) \\ &= M(\omega)\} = \{\omega / M_F(\omega) = M(\omega)\} \end{aligned}$$

Where $M_F(\omega) = \sup_{\theta \in F} L(\theta, \omega)$

Let $D = \{\theta_i / i \in \mathbb{N}\}$ a countable dense subset in Θ :

$$\sup_{\theta \in F} L(\theta, \omega) = \sup_{i \in \mathbb{N}} L(\theta_i, \omega)$$

Donc:

$$\omega \rightarrow M(\omega) = \sup_{\theta \in \Theta} L(\theta, \omega) = \sup_{i \in \mathbb{N}} L(\theta_i, \omega)$$

Is measurable (it is a countable supremum of measurable functions).

It is the same for:

$$M_F(\omega) = \sup_{\theta \in F} L(\theta, \omega) = \sup_{\theta \in D_F} \{L(\theta, \omega), \theta \in D_F\}$$

There D_F is a countable dense subset in F . Therefore:

$$\phi^-(F) = \{\omega / M_F(\omega) = M(\omega)\} \in \mathcal{a}$$

Because M and M_F are two random variables defined on (Ω, \mathcal{a}) with values in separable metric space \mathbb{R}^+ . We have then $\phi^-(G) \in \mathcal{a}$ for every open d in Θ by the.

Theorem 2.5.3: It exist a mapping $\hat{\theta}$ of (Ω, \mathcal{a}) in (Θ, τ) measurable such as:

$$\forall \omega \in \Omega, \hat{\theta}(\omega) \in \phi(\omega)$$

Ie:

$$\forall \omega \in \Omega, L(\hat{\theta}(\omega), \omega) \geq L(\theta, \omega), \forall \theta \in \Theta$$

Lemma 3.2.2: Is crucial for the proof of Theorem 3.1.1. It gives the construction of the cost function.

Lemma 3.2.2: Under the hypotheses H1, H2, H4, H5, H6 and H7, it exists a sequence of cost functions $(C_n)_{n \in \mathbb{N}}$ defined on $\Theta \times \Theta \times \Omega$ with values in \mathbb{R} , such

that $\forall J$: (i) C_n is bounded and invariant by translation. (ii) C_n is continuous on $\Theta \times \Theta$ (iii) C_n is measurable on Ω .

Proof: Thanks to the hypothesis H7 we have:

$$\sup\{L(x, \omega), x \notin \hat{\theta}(\omega). B(e, \epsilon_n)\} < L(\hat{\theta}(\omega), \omega)$$

This supremum is reached because $\Theta - \hat{\theta}(\omega)B(e, \epsilon_n)$ is a nonempty compact. Let:

$$\forall \omega \in \Omega, e_n(\omega) = L(\hat{\theta}(\omega), \omega) - \sup\{L(x, \omega), x \notin \hat{\theta}(\omega). B(e, \epsilon_n)\}$$

$$\forall \omega \in \Omega, e_n(\omega) > 0$$

There exists a neighborhood $U_{n,\omega}$ of e , contained in $B(e, \epsilon_n)$ such that:

$$\forall \theta \in \hat{\theta}(\omega). U_{n,\omega}, \\ L(\theta, \omega) > [L(\hat{\theta}(\omega), \omega) - e_n(\omega) / 2]$$

$\hat{\theta}(\omega). U_{n,\omega}$ can be chosen as the open ball with center:

$\hat{\theta}(\omega)$ and radius $\rho_n(\omega)$ where:

$$\rho_n(\omega) = \sup\{r > 0 : d(\theta, \hat{\theta}(\omega)) < r \Rightarrow \\ L(\theta, \omega) > [L(\hat{\theta}(\omega), \omega) - e_n(\omega) / 2]\} \\ = \sup\{r > 0 : \inf_{\theta \in B(\hat{\theta}(\omega), r)} \\ \left[L(\hat{\theta}(\omega), \omega) - \frac{e_n(\omega)}{2} \right]\}$$

There exists a mapping $\gamma_{n,\omega}$ of Θ in $[0,1]$ continuous, with support in $U_{n,\omega}$ and is equal to 1 in e . In other words we have:

$$\gamma_{n,\omega} : \Theta \rightarrow [0,1]$$

Such as:

$$\theta \rightarrow \gamma_{n,\omega}(\theta) = \begin{cases} 1 & \text{if } \theta = e \\ 0 & \text{if } \theta \notin U_{n,\omega} \end{cases}$$

Where:

$$U_{n,\omega} = B(e, \rho_n(\omega))$$

Consider:

$$\gamma_{n,\omega} : \Theta \rightarrow [0,1] \text{ as } \theta \rightarrow \gamma_{n,\omega}(\theta) = \\ \frac{\rho_n(\omega) - d(\theta, e)}{\rho_n(\omega)} \vee 0$$

This mapping suggests the following cost function:

$$C_n : \Theta \times \Theta \times \Omega \rightarrow [0,1] \text{ such as } (\theta, z, \omega) \rightarrow \\ C_n(\theta, z, \omega) = 1 - \gamma_{n,\omega}(z^{-1}\theta)$$

I.e.:

$$C_n(\theta, z, \omega) = 1 - \frac{\rho_n(\omega) - d(\theta, z)}{\rho_n(\omega)} \vee 0$$

It is clear that C_n is continuous in θ and z , bounded and invariant by translation i.e.:

$$C_n(x\theta, xz, \omega) = C_n(\theta, z, \omega), \forall x \in \Theta$$

Since we have to evaluate quantities such as:

$$\int C_n(\theta, z, \omega) L(\theta, \omega) \xi(\theta) d\omega$$

It is important that $C_n(.,., \omega)$ be measurable in. To show that $C_n(.,., \omega)$ is measurable in ω , is to prove the measurability of $\rho_n(\omega)$ par rapport ω , since $C_n(\theta, z, \omega)$ is written into the form (1.1). Prove then that the mapping ρ_n of (Ω, \mathcal{a}) in $(\mathbb{R}_+^*, \mathcal{B}_{\mathbb{R}_+}^*)$ defined as follows:

$$\rho_n(\omega) = \sup\{r > 0 : \inf_{\theta \in B(\hat{\theta}(\omega), r)} \\ L(\theta, \omega) > [L(\hat{\theta}(\omega), \omega) - e_n(\omega) / 2]\}$$

is measurable. Let:

$$X(r, \omega) = \inf_{\theta \in B(\hat{\theta}(\omega), r)} L(L(\theta, \omega))$$

To show that:

$$\omega \rightarrow \rho_n(\omega)$$

Is measurable. We must prove that:

$$\forall s \in \mathbb{R}_+^*, \text{ the set } \{\omega : \rho_n(\omega) > s\} \text{ is in } \mathcal{a}$$

We note that:

$$\begin{aligned} \{\omega : \rho_n(\omega) > s\} &= \bigcap_{s_i < s} \\ \{\omega : X(s_i, \omega) [L(\hat{\theta}(\omega), \omega) - e_n(\omega) / 2] \\ s_i \in Q\} \end{aligned}$$

The problem amounts to proving that the two following mappings:

$$\begin{aligned} \omega &\rightarrow e_n(\omega) = \\ L(\hat{\theta}(\omega), \omega) - \sup\{L(x, \omega), x \notin \theta(\omega).B(e, \varepsilon_n)\} \\ \omega &\rightarrow X(s, \omega) = \inf_{\theta \in B(\hat{\theta}(\omega), s)} (L(\theta, \omega)) \end{aligned}$$

are measurable. The proof of the measurability of the first mapping is established. To show that for s fixed:

$$\omega \rightarrow X(s, \omega)$$

is measurable. Let θ^0 be a countable dense subset of $B(e, s)$ the set $\{(\hat{\theta}(\omega)\theta_i / \theta_i \in \theta^0)\}$ is dense in $\hat{\theta}(\omega)B(e, s) = B(\hat{\theta}(\omega), s)$ the following mapping:

$\phi_i(\omega)B(e, s) = B(\hat{\theta}(\omega)\theta_i, \omega) \rightarrow L(\hat{\theta}(\omega)\theta_i, \omega)$ is measurable. Therefore:

$$\omega \rightarrow \inf_i \phi_i(\omega) = \inf_{\theta \in B(\hat{\theta}(\omega), s)} (L(\theta, \omega))$$

Lemma 3.2.3 Under the hypotheses H1, H2, ..., H6, there is a measurable mapping $\tilde{\theta}_n$ of (Ω, a) in (Θ, τ) such as:

$$\begin{aligned} \forall \omega \in \Omega, \forall d \in \odot \int_{\odot} C_n(\theta, \tilde{\theta}_n(\omega)\omega)L(\theta, \omega) \\ \mu(d\theta) \leq \int_{\odot} C_n(\theta, d, \omega)\mu(d\theta) \end{aligned}$$

Proof: It is clear that minimizing:

$$\int_{\odot} C_n(\theta, d, \omega)L(\theta, \omega)\mu(d\theta)$$

Require, by definition of C_n , to maximize:

$$\int_{\odot} \gamma_{n, \omega}(d^{-1}\theta)L(\theta, \omega)\mu(d\theta)$$

The maximum of this last integral is reached because:

$$\forall \omega \in \Omega, d \rightarrow \int_{\odot} \gamma_{n, \omega}(d^{-1}\theta)L(\theta, \omega)\mu(d\theta)$$

Is continuous on the compact set Θ . Therefore:

$$\begin{aligned} \forall \omega \in \Omega, \exists \hat{d}_\omega^n \in \Theta \text{ such as} \\ \int_{\odot} \gamma_{n, \omega}(\hat{d}_\omega^n L(\theta, \omega)\mu(d\theta) \\ = \int_{\odot} \gamma_{n, \omega}(x^{-1}\theta)L(\theta, \omega)\mu(d\theta), \forall x \in \Theta \end{aligned}$$

It remains to prove that we can choose a \hat{d}_ω^n measurable in ω . For this consider the following multivoc mapping:

$$\begin{aligned} \phi_n : \omega \in \Omega \rightarrow \phi_n(\omega) = \\ \{d : \int_{\odot} \gamma_{n, \omega}(d^{-1}\theta)L(\theta, \omega)\mu(d\theta) = g_n(\omega)\} \in \Theta \end{aligned}$$

Where :

$$g_n(\omega) = \sup_{x \in \Theta} \int_{\odot} \gamma_{n, \omega}(x^{-1}\theta)L(\theta, \omega)\mu(d\theta)$$

We check the conditions of Theorem K. Kuratowski and R. Nardowski same way as for the mapping Φ (Lemma 3.2.1). We can say that there is a measurable mapping $\tilde{\theta}_n$ of (Ω, a) in (Θ, τ) such as:

$$\forall \omega \in \Omega, \tilde{\theta}_n(\omega) \in \phi_n(\omega)$$

Proof of Theorem 3.1.1:

The existence of the following cost functions $(C_n)_{n \in \mathbb{N}}$ is established in the Lemma 3.2.2. It remains to show that:

$$\lim_{x \rightarrow \infty} \tilde{\theta}_n = \hat{\theta} \text{ point wisely}$$

Recall the definitions of $\Phi_n(\omega)$ and $\Phi(\omega)$:

$$\begin{aligned} \phi_n(\omega) &= \{d : \int_{\odot} \gamma_{n, \omega}(d^{-1}\theta)L(\theta, \omega)\mu(d\theta) = \\ &\sup_{x \in \Theta} \int_{\odot} \gamma_{n, \omega}(x^{-1}\theta)L(\theta, \omega)\mu(d\theta)\} \end{aligned}$$

$$\begin{aligned} \phi(\omega) &= \theta : L(\theta, \omega) = \sup \\ &\{L(x, \omega), x \in \Theta\} = \{\hat{\theta}(\omega)\} \end{aligned}$$

And suppose that there is:

$$\begin{aligned} g_n \in \phi_n(\omega) \text{ such as } g_n B(e, \rho_n(\omega)) \cap \\ \hat{\theta}(\omega)B(e, \varepsilon_n) = \emptyset \end{aligned}$$

For these considerations we have:

a) On the one hand, we have:

$$\int_{\odot} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)(d\theta) = \int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)\mu(d\theta)$$

Because the function $\gamma_{n,\omega}$ has support in $B(e,\rho_n(\omega))$ and we have:

$$\int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)(\theta,\omega)\mu(d\theta) \leq L(\hat{\theta}(\omega),\omega) - e_n(\omega) \int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)\mu(d\theta)$$

Because :

$$\sup\{L(\theta,\omega), \theta \notin (\hat{\theta}(\omega).B(e,\epsilon_n))\} = [L(\hat{\theta}(\omega),\omega) - e_n(\omega)]$$

b) On the other hand we have:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)L(\theta,\omega)\mu(d\theta) > L(\hat{\theta}(\omega),\omega) - e_n(\omega) / 2 \int_{\hat{\theta}(\omega)B(e,\rho_n(\omega))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)\mu(d\theta)$$

Because:

$$\forall \theta \in (\hat{\theta}(\omega)B(e,\rho_n(\omega))) \\ L(\theta,\omega) > L(\hat{\theta}(\omega),\omega) - \frac{e_n(\omega)}{2}$$

Then:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)L(\theta,\omega)\mu(d\theta) > [L(\hat{\theta}(\omega),\omega) - e_n(\omega)] \int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)\mu(d\theta)$$

That:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)\mu(d\theta) = \int_{\odot} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)\mu(d\theta) \int_{\odot} \gamma_{n,\omega}(\theta)\mu(d\theta)$$

And also:

$$\int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)\mu(d\theta) = \int_{\odot} \gamma_{n,\omega}(g_n^{-1}\theta)\mu(d\theta) = \int_{\odot} \gamma_{n,\omega}(\theta)\mu(d\theta)$$

Then we have:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}(((\hat{\theta}(\omega))^{-1}\theta))L(\theta,\omega)\mu(d\theta) > [L(\hat{\theta}(\omega),\omega) - e_n(\omega)] \int_{\odot} \gamma_{n,\omega}(\theta)\mu(d\theta)$$

And:

$$\int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)\mu(d\theta) \leq [L(\hat{\theta}(\omega),\omega) - e_n(\omega)] \int_{\odot} \gamma_{n,\omega}(\theta)\mu(d\theta)$$

Then:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)L(\theta,\omega)\mu(d\theta) > \int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)\mu(d\theta)$$

$$\int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)\mu(d\theta) \leq [L(\hat{\theta}(\omega),\omega) - e_n(\omega)] \int_{\odot} \gamma_{n,\omega}(\theta)\mu(d\theta)$$

Then:

$$\int_{(\hat{\theta}(\omega)B(e,\rho_n(\omega)))} \gamma_{n,\omega}((\hat{\theta}(\omega))^{-1}\theta)L(\theta,\omega)\mu(d\theta) > \int_{g_n B(e,\rho_n(\omega))} \gamma_{n,\omega}(g_n^{-1}\theta)L(\theta,\omega)\mu(d\theta)$$

This contradicts the definition of g_n we have necessarily:

$$\forall g_n \in \phi_n(\omega) : g_n B(e,\rho_n(\omega)) \cap \hat{\theta}(\omega)B(e,\epsilon_n) \neq \emptyset \\ \forall g_n \in \phi_n(\omega) : (\hat{\theta}(\omega))^{-1}g_n \in B(e,\epsilon_n) \\ B(e,\rho_n)^{-1} \subset B(e,\epsilon_n), B(e,\epsilon_n) \subset B(e,\epsilon_{n-1})$$

Then: $\forall g_n \in \phi_n(\omega) : (\hat{\theta}(\omega))^{-1}g_n \in B(e,\epsilon_n)$.And:

$$\forall g_n \in \phi_n(\omega) : \lim_{n \rightarrow \infty} g_n = \hat{\theta}(\omega)$$

As:

$$\tilde{\theta}_n(\omega) \in \phi_n(\omega)$$

We have:

$$\lim_{n \rightarrow \infty} \tilde{\theta}_n(\omega) = \hat{\theta}(\omega)$$

Therefore:

$$\lim_{n \rightarrow \infty} \tilde{\theta}_n(\omega) = \hat{\theta} \text{ pointwisely}$$

$$\theta \rightarrow \gamma_{n,\omega}(\theta) = \begin{cases} 1 & \text{if } \theta = e \\ 0 & \text{if } \theta \notin B(e, \rho_n(\omega)) \end{cases}$$

Remarks 1): The result holds for any sequence of function $\tilde{\theta}_n$ of Ω in Θ non necessarily measurable such that:

$$\tilde{\theta}_n(\omega) \in \phi_n(\omega)$$

2) We prove that $\hat{\theta}$ is measurable, because it is the pointwise limit of measurable functions.

Asymptotic comparison between $\tilde{\theta}_n$ and $\hat{\theta}$ when the estimator $\hat{\theta}$ is not unique: At this party, we generalize the Theorem 3.1.1, in the sense that we are considering more hypotheses H7.

Theorem 3.5.1 for all the likelihood function L satisfying H5 and under the hypotheses H1... H6, there exists: a sequence of cost functions $(C_n)_{n \in \mathbb{N}}$ defined on $\Theta \times \Omega$ with values in \mathbb{R} , such that $\forall n C_n$:

- i) is bounded and invariant by translation.
- ii) is continuous on $\Theta \times \Theta$
- iii) is measurable on Ω and such that:

$\lim_{n \rightarrow \infty} \tilde{\theta}_n$ (if it exists) is a maximum likelihood estimator.

Proof: The construction of the following cost function is established using the same technique as before (Lemmas 3.2). We ask:

$$\begin{aligned} \phi(\omega) &= \{\theta : L(\theta, \omega) = M(\omega)\} \text{ where } M(\omega) \\ &= \sup\{L(x, \omega), x \in \Theta\} \\ e_n(\omega) &= M(\omega) - \sup\{L(\theta, \omega), \theta \notin \phi(\omega)B(\varepsilon_n)\} > 0. \forall \omega \in \Omega \\ Z_n(\omega) &= \inf_{m > 1} \sup \{L(\theta, \omega), \theta \notin \phi(\omega)B^*(e, \varepsilon_n(1-1/m))\} \end{aligned}$$

$$\rho_n(\omega) = \sup \left\{ \begin{array}{l} r > 0 : \inf_{\theta \in \phi(\omega)B(e,r)} L(\theta, \omega) \\ > M(\omega) - e_n(\omega) / 2 \end{array} \right\} \wedge \varepsilon_n$$

There exists a $\gamma_{n,\omega}$ of Θ in $[0,1]$ continuous, to stand (à support) in $B(e, \rho_n(\omega))$ and is equal to 1 in e . In other word we have:

$$\gamma_{n,\omega} : \Theta \rightarrow [0,1]$$

Consider:

$$\gamma_{n,\omega}(\theta) = \frac{\rho_n(\omega) - d(\theta, e)}{\rho_n(\omega)} \vee 0$$

This mapping suggests the following cost function:

$$\begin{aligned} C_n : \Theta \times \Theta \times \Omega &\rightarrow [0,1] \\ (\theta, z, \omega) &\rightarrow C_n(\theta, z, \omega) = 1 - \gamma_{n,\omega}(z^{-1}\theta) \end{aligned}$$

Ie:

$$C_n(\theta, z, \omega) = 1 - \frac{\rho_n(\omega) - d(\theta, e)}{\rho_n(\omega)} \vee 0$$

It is clear that $\forall n \in \mathbb{N} C_n$ is continuous in θ and z . C_n .bounded, invariant by translation, ie:

$$C_n(x\theta, xz, \omega) = C_n(\theta, z, \omega), \forall \theta \in \Omega$$

The measurability of C_n in ω is deduced from that of the two following mappings:

$$\omega \rightarrow e_n(\omega) = M(\omega) - \sup\{L(\theta, \omega), \theta \notin \phi(\omega)B(e, \varepsilon_n)\}$$

$$\omega \rightarrow \inf\{L(\theta, \omega), \theta \in \phi(\omega)B(e, r)\}$$

Problem1: We must prove that $\omega \rightarrow e_n$ is measurable. We start by showing that:

$$\omega \rightarrow M(\omega) = \sup\{L(\theta, \omega), \theta \in \Theta\}$$

is measurable. Let $D = \{\theta_i, i \in \mathbb{N}\}$ be a countable dense subset in Θ :

$$\omega \rightarrow M(\omega) = \sup_i L(\theta_i, \omega)$$

Is then measurable. We now show that:

$$\omega \rightarrow \sup\{L(\theta, \omega), \theta \notin \phi(\omega)B(e, \varepsilon_n)\}$$

Is measurable.

We have firstly, for ω fixed and n fixed:

$$\sup\{L(\theta, \omega), \theta \in \phi(\omega)B(e, \epsilon_n)\} = Z_n(\omega) \quad (3.5.1)$$

The proof of (3.5.1) is in the appendix.
Secondly:

$$\psi : (\theta, \omega) \in \Theta \times \Omega \rightarrow d(\theta, \phi(\omega)) \in \mathbb{R}$$

Is measurable. The proof of the measurability of Ψ is in the appendix. Then:

$$\forall m > 1, B_m = \psi^{-1}(0, \epsilon_n(1 - 1/m)) = \{(\theta, \omega) : d(\theta, \phi(\omega)) \leq \epsilon_n(1 - 1/m)\} \in \tau \otimes a$$

Modify L in B_m :

$$\tilde{L}_m(\theta, \omega) = \begin{cases} L(\theta, \omega) & \text{if } (\theta, \omega) \in B_m \\ -\infty & \text{otherwise} \end{cases}$$

$\forall m > 1, \tilde{L}_m$ is measurable on $\Theta \times \Omega$. Let:

$$A_m^\omega = \Theta - \phi(\omega)B^*(e, \epsilon_n(1 - 1/m))$$

A_m^ω is open in $\Theta A_m^\omega \cap D$ is dense in A_m^ω . Therefore:

$$\begin{aligned} \omega \rightarrow Z_n(\omega) &= \inf(\sup_{m>1, \theta \in A_m^\omega} L(\theta, \omega)) = \\ \inf(\sup_{m>1, \theta \in A_m^\omega \cap D} L(\theta, \omega)) &= \inf(\sup_{m>1, \theta \in D} \tilde{L}(\theta, \omega)) \end{aligned}$$

is measurable on Ω . So:

$$\omega \rightarrow e_n(\omega) = M(\omega) - Z_n(\omega)$$

Is measurable on Ω . Thus the problem 1 is solved.

Problem2: Show that: With r fixed:

$$\omega \rightarrow \inf\{L(\theta, \omega), \theta \in \phi(\omega)B(e, r)\}$$

Is measurable. Let B:

$$\begin{aligned} B = \psi^{-1}(0, r] &= \{(\theta, \omega) : \\ d(\theta, \phi(\omega)) < r\} &\in \tau \otimes a \end{aligned}$$

Modify L on B:

$$\hat{L}(\theta, \omega) = \begin{cases} L(\theta, \omega) & \text{if } (\theta, \omega) \in B \\ +\infty & \text{otherwise} \end{cases}$$

\hat{L} is measurable on $\Theta \times \Omega$.

And:

$$\begin{aligned} \omega \rightarrow \inf\{L(\theta, \omega), \theta \in \phi(\omega)B(e, r)\} &= \\ \inf\{L(\theta, \omega), \theta \in \phi(\omega)B(e, r) \cap D\} &= \\ = \inf\{\hat{L}(\theta, \omega), \theta \in D\} \end{aligned}$$

is also measurable. We have constructed a sequence of cost functions, which satisfies the conditions of the theorem. To end show that for this type of cost functions, $\lim_{n \rightarrow \infty} \hat{\theta}_n$ (if it exists) is a maximum likelihood estimator. Suppose that, there is $g_n \in \Phi_n(\omega)$ such that:

$$g_n B(e, \rho_n(\omega)) \cap \phi(\omega)B(e, \epsilon_n) = \phi$$

For these considerations we have:

$$\int_{g_n B(e, \rho_n(\omega)) \cap \phi(\omega)B(e, \epsilon_n)} \gamma_{n, \omega}(g_n^{-1}\theta)(\theta, \omega) \mu(d\theta) \leq M(\omega) - e_n(\omega) \int_{\Theta} \gamma_{n, \omega}(\theta) \mu(d\theta)$$

And $\forall f \in \Phi(\omega)$:

$$\begin{aligned} \int_{f B(e, \rho_n(\omega)) \cap \phi(\omega)B(e, \epsilon_n)} \gamma_{n, \omega}(f^{-1}\theta)L(\theta, \omega) \mu(d\theta) &M(\omega) - \\ e_n(\omega) / 2 \int_{\Theta} \gamma_{n, \omega}(\theta) \mu(d\theta) &> [M(\omega) - e_n(\omega)] \\ \int_{\Theta} \gamma_{n, \omega}(\theta) \mu(d\theta) \end{aligned}$$

This gives:

$$\begin{aligned} \forall f \in \Phi(\omega), \int_{f B(e, \rho_n(\omega)) \cap \phi(\omega)B(e, \epsilon_n)} \gamma_{n, \omega}(f^{-1}\theta)L(\theta, \omega) \mu(d\theta) &> \\ \int_{g_n B(e, \rho_n(\omega)) \cap \phi(\omega)B(e, \epsilon_n)} \gamma_{n, \omega}(g_n^{-1}\theta)L(\theta, \omega) \mu(d\theta) \end{aligned}$$

This contradicts the definition of g_n . ($g_n \in \Phi_n(\omega)$) We obviously have:

$$\begin{aligned} \forall g_n \in \Phi_n(\omega) : g_n B(e, \rho_n(\omega)) \cap \hat{\theta}(\omega)B(e, \epsilon_n) &\neq \emptyset \\ \Rightarrow \forall g_n \in \Phi_n(\omega) : (\hat{\theta}(\omega))^{-1}g_n \in B(e, \epsilon_n) \cdot B(e, \rho_n)^{-1} &\subset \\ B(e, \epsilon_n) \cdot B(e, \epsilon_n) \subset B(e, \epsilon_{n-1}) \end{aligned}$$

We know that there is $\tilde{\theta}_n$ of (Ω, a) in (Θ, τ) measurable such that $\tilde{\theta}_n(\omega) \in \Phi_n(\omega)$ From the foregoing $\lim_{n \rightarrow \infty} \tilde{\theta}_n(\omega)$ (if it exists) $\in \Phi(\omega)$, as $\lim_{n \rightarrow \infty} \tilde{\theta}_n(\omega)$ is still a measurable mapping from (Ω, a) in (Θ, τ) . So $\lim_{n \rightarrow \infty} \tilde{\theta}_n(\omega)$ (if it exists) is a maximum likelihood estimator.

CONCLUSION

Our study deals with the maximum likelihood estimator as a limit Bayesian estimators. We think conceptually it is not one of the first times, leads a result concerning a sequence of cost. Usually the cost is an intrinsic data of the decision problem and within a Bayesian framework focuses attention on the dependence of the optimal solution to the law a priori. For us this law is fixed and what is the cost function that varies.

Appendix: Lemma (Proof of (3.5.1)) Let Θ be a compact metric space, A be a compact subset of Θ , f be a continuous mapping of Θ in \mathbb{R} , $n, m \in \mathbb{N}$ and $(\epsilon_n)_n$ a sequence of real numbers. Let:

$$F = \{\theta : d(\theta, A) \geq \epsilon_n\}$$

$$G_m = \{\theta : d(\theta, A) > \epsilon_n(1 - 1/m)\}$$

$$\tilde{G}_m = \{\theta : d(\theta, A) \geq \epsilon_n(1 - 1/m)\}$$

Then:

$$\sup_{x \in F} f(x) = \inf_{m > 1} (\sup_{x \in G_m} f(x)) = \inf_{m > 1} (\sup_{x \in \tilde{G}_m} f(x))$$

Proof:

$$\inf_{m > 1} (\sup_{x \in G_m} f(x)) = \inf_{m > 1} (\sup_{x \in \tilde{G}_m} f(x)) \text{ because } G_{m+1} \subset \tilde{G}_{m+1} \subset G_m :$$

if $F \neq \emptyset$, then

$$F = \bigcap_m G_m = \bigcap_m \tilde{G}_m$$

We have: A):

$$\forall m > 1, \sup_{x \in \tilde{G}_m} f(x) \geq \sup_{x \in F} f(x)$$

$$\Rightarrow \forall m > 1, \inf_{m > 1} \left(\sup_{x \in \tilde{G}_m} f(x) \right) \geq \sup_{x \in F} f(x) \tag{1}$$

Since:

$$\sup_{\tilde{G}_m} f(x) = f(x_m), x_m \in \tilde{G}_m$$

We can suppose that the sequence $(x_m)_m$ converges to x :

$$d(x_m, A) \geq \epsilon_n(1 - 1/m)$$

$$d(x_m, A) \rightarrow d(x, A), m \rightarrow \infty$$

Then:

$$d(x, A) \geq \epsilon_n \Rightarrow x \in F$$

$$\sup_{\tilde{G}_m} f(x) = f(x_m) \rightarrow f(x), m \rightarrow \infty$$

Since f is continuous:

$$\inf_{m > 1} (\sup_{\tilde{G}_m} f(x)) = \inf_{m > 1} (f(x_m)) = f(x)$$

Because $(f(x_m))_m$ is decreasing. But:

$$f(x) \leq (\sup_F f(x), x \in F)$$

$$(1) \text{ and } (2) \Rightarrow \inf_{m > 1} (\sup_{\tilde{G}_m} f(x)) = \sup_F f(x)$$

2) If $F = \emptyset$, there exists necessarily m such that $G_m = \emptyset$. Indeed: $(\tilde{G}_m)_m$ is a decreasing sequence of compact. If \tilde{G}_m is not empty, then $\bigcap \tilde{G}_m \equiv \emptyset$ (compactness property).

But for:

$$x_m \in \tilde{G}_m, \forall m > 1$$

We have:

$$d(x, A) \geq \epsilon_n(1 - 1/m), \forall m > 1$$

Therefore:

$$d(x, A) \epsilon_n \Rightarrow x \in F$$

Which is nonsense. Therefore:

$$F = \bigcap_m G_m = \bigcap_m \tilde{G}_m = \emptyset$$

And:

$$\sup_F f(x) = \inf_{m > 1} (\sup_{G_m} f(x)) = \inf_{m > 1} (\sup_{\tilde{G}_m} f(x)) = -\infty$$

Proof of mesurability of ψ :

$$\Psi : (\theta, \omega) \in \Theta \times \Omega \mapsto d(\theta, \Phi(\omega)) \in \mathbb{R}$$

Consider $k(\theta)$ the family of compacts for θ , equipped with the Hausdorff topology, which a sub-base is $g \cup g'$. where g and g' are defined by:

$$\begin{aligned}
 g \text{ open} &\stackrel{\text{def}}{=} \{K \in K(\Theta) / K \cap G, \text{Gopenin}\Theta\} && \leq d(\theta', x_1) - d(\theta', x_2) \\
 & && \leq d(x_1, x_2) = d(x_2, K) \\
 g' \text{ open} &\stackrel{\text{def}}{=} \{K \in K(\Theta) / K \subset G', G' \text{openie}\Theta\} && \leq \sup_{x \in K} d(x, K) \leq \delta(k, k')
 \end{aligned}$$

Let A and B be two nonempty closed in Θ :

$$\delta(A, B) = \max \left\{ \sup_{x \in A} d(x, B), \sup_{y \in B} d(y, A) \right\}$$

Defines a distance in $k(\theta)$ called the Hausdorff distance. The topology induced by δ is as defined above Θ is a metric space sees (Christensen, 1974). Show that the mapping defined below is measurable:

$$\begin{aligned}
 \phi : \omega \in \Omega &\rightarrow \phi(\omega) = \{\theta : L(\theta, \omega) = M(\omega)\} \in K(\Theta) \\
 \phi^{-1}\{k / K \cap F \neq \emptyset\} &= \{\theta : L(\theta, \omega) \cap F \neq \emptyset\} \in a
 \end{aligned}$$

(See the proof of Lemma 3.2.1:

$$\begin{aligned}
 \phi^{-1}\{K / K \subset F\} &= \{\omega : \phi(\omega) \subset F\} = A \\
 A^c &= \{\omega : \phi(\omega) \cap F^c \neq \emptyset\} = \bigcup_n \{\omega : \phi(\omega) \cap F_n \neq \emptyset\}
 \end{aligned}$$

Because $F^c = \bigcup_n F_n$ where F_n is closed in $\Theta \Rightarrow$:

$$\begin{aligned}
 A^c \in a, &\text{ because } \{\omega : \phi(\omega) \cap F \neq \emptyset\} \in a, \forall n \\
 \Rightarrow A &= \phi^{-1}\{K / K \subset F\} \in a
 \end{aligned}$$

Therefore Φ is measurable. We now show that the following mapping:

$$d : (K, \theta) \in k(\Theta) \times \Theta \rightarrow d(K, \theta) \in \mathbb{R}$$

Is continuous (uniformly). Define on $k(\Theta) \times \Theta$ the following distance:

$$\begin{aligned}
 d'((k, \theta), (k', \theta')) &= d(\theta, \theta') + \delta(k, k') \\
 |d(\theta, k) - d(k', \theta')| &\leq |d(\theta, k) - d(\theta', k)| + \\
 &|d(\theta', k) - d(k', \theta)| \\
 |d(\theta, k) - d(\theta', k)| &\leq d(\theta, \theta')
 \end{aligned}$$

Let $x_1 \in k, x_2 \in k$ such that $d(\theta', x_1) = d(\theta', k)$ and $d(\theta', x_2) = d(\theta', k')$.

Let $x_1' \in k$ such that $d(x_2, x_1') = d(x_2, x_1)$. we have:

$$d(\theta', k) - d(\theta', k) = d(\theta', x_1) - d(\theta', x_2)$$

It is the sum of:

$$\begin{aligned}
 d(\theta', k') - d(\theta', k) &\leq \delta(k, k') \\
 \Rightarrow |d(\theta', k) - d(\theta', k')| &\leq \delta(k, k')
 \end{aligned}$$

There fore:

$$\begin{aligned}
 |d(\theta, k) - d(k', \theta')| &\leq d(\theta, \theta') + \delta(k, k') = \\
 &= d'((k, \theta), (k', \theta'))
 \end{aligned}$$

This proves the uniform continuity for:

$$(K, \theta) \mapsto d(K, \theta)$$

So the following mapping:

$$\begin{aligned}
 \Psi : (\theta, \omega) \in \Theta \times \Omega &\rightarrow ((\theta, \phi(\omega))) \\
 \in \Theta \times k(\Theta) &\rightarrow D(\theta, \phi(\omega)) \in \mathbb{R}
 \end{aligned}$$

Is measurable, because it is composed of two measurable mappings.

REFERENCES

- Barra, J.R., 1971. Notions Fondamentales De Statistique mathématique: Maitrise De Mathematiques et Applications Fondamentales. 1st Edn., Dunod, Paris, pp: 558.
- Berge, C., 1966. Espaces Topologiques, Fonctions Multivoques. 2nd Edn., Dunod, Paris, pp : 283.
- Berger, J.O., 1980. Statistical Decision Theory, Foundations, Concepts and Methods. Springer-Verlag, New York, ISBN-10: 3540904719 pp: 425.
- Christensen, J.P.R., 1974. Topology and Borel Structure: Descriptive Topology and Set Theory with Applications to Functional Analysis and Measure Theory, Issue 10. 1st Edn., Elisver, North Holland, ISBN-10: 0444106081, pp: 133.
- Fergusson, T.S., 1969. Mathematical Statistics: A Decision Theoretic Approach. 1st Edn., Academic Press, New-York, pp: 396.
- Florens, J.P., 1978. Measures a priori et invariance dans une experience Bayesian, Pub. I.S.U.P., 13: 29-55.
- Fougereaud, C. and A. Fuchs, 1967. Statistique. 1st Edn., Dunod, Paris.

- Halmos, P.R., 1974. *Measure Theory*. 1st Edn., Narosa Publishing House, New Delhi, ISBN-10: 0387900888 pp: 304.
- Jeffreys, H., 1998. *Theory of Probability*. 3Edn., Oxford University Press, Oxford, ISBN-10: 0198503687 pp: 459.
- Lanery, E., 1984. *Solutions Bayésiennes Sans Désintégration*. 1st Edn., Université Paris, Dauphine, Paris.
- Cam, L.L.M., 1953. *On some Asymptotic Properties of Maximum Likelihood Estimates and Related Estimators*. 1st Edn., University of California Press, Berkeley, pp: 53.
- Kuratowski, K. and C. Ryll-Nardewski, 1965, A general theorem on selectors. *Bull. Acad. Sci. Polon.*, 13: 397-403.
- Wald, A., 1971. *A Statical Decision Function*. 2nd Edn., Chelsea Pub., New-York, ISBN-10: 0828402434, 179.
- Ulmo, J. and J. Bernier, 1973. *Presses Universitaires de France*,. 1st Edn., Presses Universitaires de France, Paris, pp: 330.