ON CONSISTENCY AND NULL SETS IN BAYES ESTIMATION

by

LUCIEN ELIE NICOLAS DELBROUCK
B.A., University of British Columbia, 1961

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Mathematics

We accept this thesis as conforming to the

required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September, 1963

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Mathematics.

The University of British Columbia,
Vancouver 8, Canada.

Date September 25, 1963.

Abstract

A basic result of Doob states that, under very weak measurability assumptions, Bayes' estimators are consistent for almost all parameter points. First it is shown that even when this exceptional set is finite, the effect of putting positive prior mass on each point of the set may result in creating a new exceptional set, larger than the original one, rather than in eliminating the lack of consistency. The posterior densities are then studied and it is shown that under fairly strong regularity conditions the corresponding posterior distributions tend, in the limit, to concentrate their mass on a particular point in the parameter set. If in addition, distinct parameter points correspond to distinct probability measures, then it is shown that both the maximum likelihood and the Bayes' estimators are consistent for all parameter values.

I hereby certify that this abstract is satisfactory

Acknowledgement

I wish to express my thanks to Dr. Loraine Schwartz for her help, patience, and encouragement throughout the planning and writing of this thesis, and to Dr. Maurice Sion for his vigorous assistance during its initial stages.

This is also to acknowledge gratefully the financial help received in the form of summer grants from the National Research Council, and teaching assistantships from the Department of Mathematics of the University of British Columbia.

Table of Contents

0. Introduction

As an application of martingale theory to problems of estimation, under fairly mild measurability conditions, and under certain restrictions on the nature of the sample space and of the parameter space, Doob [1] shows that the Bayes' estimates are consistent except on an exceptional set of prior measure 0. In [3] , under Doob's assumptions, and supposing the exceptional set to be finite, Schwartz constructs a new prior measure ascribing positive mass to the elements of the exceptional set; and she conjectures that the Bayes' estimates corresponding to the new measure are consistent for every element of the parameter space. In [4] , under rather stringent continuity conditions on the prior measure as well as on the density functions, Boev exhibits the tendency for the posterior densities to concentrate at a certain point of the parameter space, as the number of observations increases indefinitely.

After recalling Doob's assumptions and results, we examine in some detail the new measure proposed in [3] , and we show that when the Bayes' estimates are not completely specified with respect to the original prior measure, it is sometimes possible to define them in such a way that the exceptional set corresponding to the new measure is non-empty. In the example given, the Bayes' solution which is determined almost everywhere is taken to be the estimator for every sample point. This leads to an estimator which is consistent except at one point for the original prior distribution. The effect of altering the prior distribution is to shift the exceptional set from the set with a single point to one containing several points. The estimation chosen is natural in a mathematical sense: it has the same form throughout the sample space; but it is intuitively unnatural or naive from the point of view of solving the estimation problem. If this latter view is considered, a definition of the estimator on a null set is readily obtained which yields a Bayes' solution consistent at all parameter values.

Subsequently, modifying Boev's assumptions somewhat, we set his conclusions on what we feel to be a firmer basis and we show first that these conclusions provide for the consistency of Bayes' estimates whenever the maximum likelihood estimates are consistent and conversely. Thence, we conclude in proving that under Boev's assumptions, the existence of a consistent procedure is a necessary and sufficient condition for the consistency of maximum likelihood and of Bayes' estimates.

Section 1.

1.0 The Underlying Probability Models.

Let N be the set of natural numbers, and let $\{X_j : j \in N\}$ be a family of completely independent identically distributed random variables, each with range-space $\mathcal{X}$ on which a σ-field $\mathcal{A}$ is defined. For every point θ of a parameter space Θ, a probability measure P_θ is defined on the space $\{\mathcal{X}, \mathcal{A}\}$. If for every $j \in N$, $\mathcal{X}_j$ is a replica of the space $\mathcal{X}$, then for every $n \in N$, and for every $\theta \in \Theta$, we state

Definition 1.0.1

The triplet $\{\mathcal{X}^n, \mathcal{A}^n, P^n_\theta\}$ is a probability space, where $\mathcal{X}^n = \prod_{j=1}^{n} \mathcal{X}_j$, $\mathcal{A}^n$ is the smallest σ-field over all the sets of the form $\prod_{j=1}^{n} A_j$ where $A_j \in \mathcal{A}$ for $j = 1,2,\ldots,n$, and P^n_θ is the product measure on the n-dimensional sets in $\mathcal{A}^n$.

In matters of convergence this definition is unsatisfactory: for our purpose a sample sequence of size n, say, is merely the projection in the space $\mathcal{X}^n$ of some infinite sequence in the space $\prod_{j=1}^{\infty} \mathcal{X}_j$. Accordingly, let $\mathcal{X}^\infty = \prod_{j=1}^{\infty} \mathcal{X}_j$; for every $n \in N$ let $\mathcal{A}_n$ be the smallest σ-field over all the sets of the form $A^n \times \mathcal{X}^\infty$ where $A^n \in \mathcal{A}^n$. Now $\{\mathcal{A}_n\}_{n \in N}$ is an ascending sequence of σ-fields in the space $\mathcal{X}^\infty$; if we let $\mathcal{A}_\infty$ denote the smallest σ-field containing $\bigcup_{j \in N} \mathcal{A}_j$, then it is shown in [2] and [7], among others, that for every $\theta \in \Theta$ there is a unique probability measure P^∞_θ that agrees with P^n_θ on $\mathcal{A}^n$. More precisely

Definition 1.0.2

The triplet $\{\mathcal{X}^\infty, \mathcal{A}_\infty, P^\infty_\theta\}$ is a probability space, where $\mathcal{X}^\infty = \prod_{j=1}^{\infty} \mathcal{X}_j$, $\mathcal{A}_\infty$ is the smallest σ-field containing $\bigcup_{j \in N} \mathcal{A}_j$, and P^∞_θ is a measure such that, for every $n \in N$

$$P_\theta^\infty (A_n) = P_\theta^n (A^n) ,$$

where $A_n = A^n \times \mathcal{X}^\infty$, and $A^n \varepsilon \mathcal{A}^n$.

Now, though for any n ε N, the spaces $\{\mathcal{X}^n, \mathcal{A}^n\}$ and $\{\mathcal{X}^\infty, \mathcal{A}_n\}$ are distinct, there is between these two a one-to-one mapping embodied in the relation

$$A_n = A^n \times \mathcal{X}^\infty$$

for every $A^n \varepsilon \mathcal{A}^n$, and in this sense we regard the measure P_θ^n on $\mathcal{A}^n$ as the contraction of the measure P_θ^∞ to the space $\{\mathcal{X}^\infty, \mathcal{A}_n\}$.

If for every $\theta \varepsilon \Theta$, P_θ is absolutely continuous with respect to a measure ν on $\{\mathcal{X}, \mathcal{A}\}$, by the Radon-Nikodym theorem, there exists an $\mathcal{A}$-measurable function $p_\theta(\cdot)$, unique up to a ν-equivalence, such that, for every $A \varepsilon \mathcal{A}$

$$P_\theta (A) = \int_A p_\theta(x) \; \nu(dx) .$$

Formally

<u>Definition 1.0.3</u>

If P_θ is absolutely continuous with respect to the measure ν on $\{\mathcal{X}, \mathcal{A}\}$, there exists a density function $p_\theta(\cdot)$ for the measure P_θ with respect to the measure ν, unique up to a ν-equivalence, such that

$$P_\theta(\cdot) = \int_{(\cdot)} p_\theta(x) \; \nu(dx) .$$

Moreover, if ν^n is the product measure on the n-dimensional space $\{\mathcal{X}^n, \mathcal{A}^n\}$ obtained from ν on each of its sides, then the function $\prod_{j=1}^n p_\theta(\cdot)$ on $\mathcal{X}^n$, up to ν^n-equivalence, satisfies

$$P_\theta^n (A) = \int_A \prod_{j=1}^n p_\theta(x_j) \; \nu^n(d\{x_1, x_2, \ldots, x_n\})$$

for every $A \varepsilon \mathcal{A}^n$, and for every $n \varepsilon N$.

On the parameter space Θ , sometimes called the space of the possible states of nature, a σ-field $\mathcal{B}$ is given together with a finite measure λ . Without loss of generality, we may assume that

$$\lambda(\Theta) = 1.$$

Furthermore, without regard to any philosophical implication, we shall often refer to λ as the "prior probability" or more simply "the prior". But we do refrain from making this a formal definition.

Turning momentarily to a general situation, when two spaces Y and Z are given together with their associated σ-fields $\mathcal{C}$ and $\mathcal{D}$, respectively, we state

Definition 1.0.4

For every $y \varepsilon Y$, let $f_y(\cdot)$ be a numerical set function on Z ; then if $A \subset Z$, the function $f_{(\cdot)}(A)$ is $\mathcal{C}$-measurable if, and only if, for any real number h,

$$\{ y : f_y(A) \leq h \} \varepsilon \mathcal{C}.$$

By extension, if the range of $f_{(\cdot)}(A)$ is a metric space and $\mathcal{B}$ its Borel field, we say that $f_{(\cdot)}(A)$ is $\mathcal{C}$-measurable if, and only if, for any element $B \varepsilon \mathcal{B}$

$$\{ y : f_y(A) \varepsilon B \} \varepsilon \mathcal{C}.$$

1.1 Preliminary Notions

If $\{X, \rho\}$ is any metric space, for $A \subset X$ such that $A \neq \emptyset$, define

$$d(A) = \sup_{(x,y) \varepsilon A \times A} \rho(x, y).$$

Throughout this paper, both $\mathcal{X}$ and Θ are taken to be complete separable metric spaces, unless otherwise noted; moreover, it is assumed that for every

$A \in \mathcal{A}_n$, $P^n_{(\cdot)}(A)$ is a $\mathcal{B}$-measurable function, where $\mathcal{A}_n$ and $\mathcal{B}$ have the meaning defined in the foregoing subsection. Bearing this in mind, and with the understanding that $\mathcal{A}_n \times \mathcal{B}$ is the smallest σ-field over the sets of the form $A \times B$, where $A \in \mathcal{A}_n$ and $B \in \mathcal{B}$, we define the measure μ_n on $\mathcal{A}_n \times \mathcal{B}$ by the relation

$$(1.1.1) \qquad \mu_n(A \times B) = \int_B P^n_\theta(A)\ \lambda(d\theta)$$

for every $A \in \mathcal{A}_n$, every $B \in \mathcal{B}$, and for every $n \in N$. Letting $B = \Theta$ in (1.1.1), we obtain the so-called <u>marginal probability measure</u>

$$(1.1.2) \qquad P_n(A) = \int_\Theta P^n_\theta(A)\ \lambda(d\theta) = \mu_n(A \times \Theta) \ .$$

Furthermore, when $P_n(A) \neq 0$, the <u>conditional posterior probability of the set $B \in \mathcal{B}$ given $A \in \mathcal{A}_n$</u>, is expressed by the relation

$$(1.1.3) \qquad Q^n(B/A) = \frac{\mu_n(A \times B)}{P_n(A)} \ .$$

If $B \in \mathcal{B}$ is held fixed, then $\mu_n(\cdot \times B)$ is a measure on $\mathcal{A}_n$, absolutely continuous with respect to P_n ; hence there exists, by the Radon-Nikodym theorem, a function $Q^n_{(\cdot)}(B)$, unique up to a P_n -equivalence, and satisfying

$$(1.1.4) \qquad \mu_n(A \times B) = \int_A Q^n_{\underline{x}}(B)\ P_n(d\underline{x})$$

for every $A \in \mathcal{A}_n$.

The function $Q^n_{(\cdot)}(B)$, defined P_n a.e. , is one version of the <u>conditional probability of the set $B \in \mathcal{B}$, given the σ-field $\mathcal{A}_n$.</u>

If the sequence $\underline{x} = (x_1, x_2, \ldots)$ belongs to the $\mathcal{A}_n$ set on which $Q^n_{(\cdot)}(B)$ is defined, the latter may be obtained in the following manner:

select an arbitrary descending sequence of sets $\{A_j\}_{j\varepsilon N}$ such that

$$A_j = \prod_{i=1}^{n} A_{ji} \times \mathcal{X}^{\infty}, \text{ where } A_{ji} \varepsilon \mathcal{Q}, \text{ for } j = 1,2,\cdots,n,$$

$$\underline{x} \varepsilon A_j \text{ for all } j \varepsilon N,$$

$$P_n(A_j) > 0, \text{ for all } j \varepsilon N,$$

$$\lim_{j\varepsilon N} d\left(\prod_{i=1}^{n} A_{ji}\right) = 0,$$

then

(1.1.5)
$$Q^n_{\underline{x}}(B) = \lim_{j\varepsilon N} Q^n(B/A_j).$$

When representation by density functions is available (see Definition 1.0.3), the fact that $P^n_{(\cdot)}(A)$ is $\mathcal{B}$-measurable for all $A \varepsilon \mathcal{Q}_n$ implies that $\prod_{j=1}^{n} p_{(\cdot)}(x_j)$ is $\mathcal{B}$-measurable for ν^n almost every sequence $(x_1, x_2, \ldots, x_n)$, and $\mathcal{Q}^n$-measurable for λ almost all $\theta \varepsilon \Theta$.
By the Fubini theorem then, if $A^n \varepsilon \mathcal{Q}^n$ and $A = A^n \times \mathcal{X}^{\infty}$,

$$P_n(A) = \int_{\Theta}\int_{A^n} \prod_{j=1}^{n} p_\theta(x_j)\, \nu^n(d\{x_1, x_2, \cdots, x_n\})\, \lambda(d\theta)$$
$$= \int_{A^n}\int_{\Theta} \prod_{j=1}^{n} p_\theta(x_j)\, \lambda(d\theta)\, \nu^n(d\{x_1, x_2, \cdots, x_n\}).$$

Hence any $\mathcal{Q}_n$ set, on whose projection in $\mathcal{X}^n$, $\int_{\Theta} \prod_{j=1}^{n} p_\theta(x_j)\, \lambda(d\theta) = 0$ is a P_n-null set; in this case

(1.1.6)
$$Q^n_{\underline{x}}(B) = \frac{\int_B \prod_{j=1}^{n} p_\theta(x_j)\, \lambda(d\theta)}{\int_{\Theta} \prod_{j=1}^{n} p_\theta(x_j)\, \lambda(d\theta)}, \quad \text{a.e. } P_n.$$

The function $\dfrac{\prod_{j=1}^{n} p_{(\cdot)}(x_j)}{\int_{\Theta} \prod_{j=1}^{n} p_\xi(x_j)\, \lambda(d\xi)}$ is commonly called a <u>posterior density function</u>.

It is a trivial matter to show that $Q^n_{\underline{x}}(\cdot)$ is P_n-equivalent to a probability measure on the space $\{\Theta, \mathcal{B}\}$. We now define, P_n $a.e.$, the Bayes' estimate

$$(1.1.7) \qquad \beta_n(\cdot) = \int_{\Theta} \xi \, Q^n_{(\cdot)}(d\xi),$$

and thence:

Definition 1.1.0

The relations (1.1.2), (1.1.4), and (1.1.7) define a Bayes' estimation system $\{\lambda, P, Q, \beta\}$, and this system is said to be

(I) Partially specified, if and only if, $Q^n_{(\cdot)}$ is unspecified on a non-empty P_n-null set, for some $n \varepsilon N$.

(II) Completely specified, if and only if, $Q^n_{(\cdot)}$ is defined everywhere on $\mathcal{X}^n$, for every $n \varepsilon N$.

(III) Consistent, if and only if, it is completely specified and

$$P^\infty_\theta(\{\underline{x}: \beta_n(\underline{x}) \to \theta\}) = 1$$

for all $\theta \varepsilon \Theta$.

It should be clear that Bayes' estimation systems are not unique, in general, and that properly speaking, unless λ be discrete with positive mass at each point, we are dealing with a complex of systems. For, from a fixed completely specified system $\{\lambda, P, Q, \beta\}$, another may be obtained by altering the measures $Q^n_{(\cdot)}$ on P_n-null sets. This will be quite evident in the construction of a counter-example.

1.2 Miscellaneous Conventions.

The effect of identifying the measure P^n_θ on $\mathcal{A}^n$ with the restriction of the measure P^∞_θ to $\mathcal{A}_n$, for any $n \varepsilon N$, is that for practical purposes we do not distinguish between $\mathcal{A}^n$-measurability and $\mathcal{A}_n$-measurability. For this reason, both an infinite sequence in $\mathcal{X}^\infty$ and its projection in the space $\mathcal{X}^n$ are represented by the same symbol $\underline{x}$ when no confusion need arise; for example the subscript n in the Bayes' estimate $\beta_n(\cdot)$ describes the latter as a function on $\mathcal{X}^\infty$ that depends only on the first n coordinates of sequences $\underline{x}$, and for any such sequence the value $\beta_n(\underline{x})$ is obtained by computation in the space $\{\mathcal{X}^n, \mathcal{A}^n\}$. By the same token, for any $\theta \in \Theta$, we view the function with value $\prod_{j=1}^{n} p_\theta(x_j)$, as a mapping on $\mathcal{X}^\infty$ to $\mathcal{X}^n$, by defining

$$\prod_{j=1}^{n} p_\theta(x_j) = \prod_{j=1}^{\infty} \alpha_j ,$$

where

$$\begin{aligned} \alpha_j &= p_\theta(x_j) \quad , \text{ if } j = 1, 2, \ldots, n, \\ &= 1 \quad , \text{ if } j = n+1, n+2, \ldots \end{aligned}$$

for every $\underline{x} = (x_1, x_2, \ldots) \in \mathcal{X}^\infty$.

However, when required, we do represent the projection of the sequence $\underline{x}$ in the space $\mathcal{X}^n$ by the symbol $\underline{x}^{(n)}$.

For simplicity of notation, we let

$$p_n(\xi, \underline{x}) = \prod_{j=1}^{n} p_\xi(x_j) ,$$

$$q_n(\cdot, \underline{x}) = \frac{p_n(\cdot, \underline{x})}{\int_\Theta p_n(\xi, \underline{x})\, \lambda(d\xi)} .$$

By a rectangle in $\mathcal{A}_n$, we mean a set of the form

$$A = A^n \times \mathcal{X}^\infty$$

where $A^n = \prod_{j=1}^{n} A_j$, and $A_j \in \mathcal{A}$ for j = 1,2,...,n.

In general any set $C \in \mathcal{A}_n$ can be represented in the form

$$C = C^n \times \mathcal{X}^\infty$$

where $C^n \in \mathcal{A}^n$. We refer to C^n as the projection or image of C in the space $\mathcal{X}^n$.

A more compact notation for sets is also desirable; for example thus

$$\{\beta_n(\underline{x}) \rightarrow \theta\} \doteq \{\underline{x} : \beta_n(\underline{x}) \rightarrow \theta\} .$$

Now when a system $\{\lambda, P, Q, \beta\}$ is completely specified let $B_0 = \{\theta : P_\theta^\infty(\{\beta_n(x) \to \theta\}) \neq 1\}$. The set B_0 shall be referred to as the exceptional set; often, we shall say that the system $\{\lambda, P, Q, \beta\}$ is consistent except on the set B_0, that the Bayes' estimates are not consistent for $\theta \varepsilon B_0$, or that the Bayes' estimates fail to converge to θ a.e. P_θ on B_0.

The indicator of a set C is a point function χ_C such that

$$\chi_C(t) = 1, \text{ if } t \varepsilon C.$$
$$= 0, \text{ otherwise}.$$

In Section 4, the metric properties of the spaces involved being more in evidence, we shall adhere, without further reference, to the following conventions: <u>open sphere</u>, <u>closed sphere</u>, and <u>neighborhood system</u> of any point x in any metric space $\{X, \rho\}$ are denoted thus

$$I(x,\epsilon) = \{y : y \varepsilon X, \text{and } \rho(x,y) < \epsilon\},$$
$$\bar{I}(x,\epsilon) = \{y : y \varepsilon X, \text{and } \rho(x,y) \leq \epsilon\},$$
$$\mathcal{V}(x) = \{V : V \subset X, \text{ and } \exists\ \epsilon > 0 \ni I(x,\epsilon) \subset V\}.$$

Again, the <u>class of open sets</u>, and the <u>class of closed sets</u>, in the space $\{X, \rho\}$ are, respectively denoted by

$$\mathcal{G}(x) = \{G : G \subset X, \text{and if } x \varepsilon G, \text{ then } G \varepsilon \mathcal{V}(x)\},$$
$$\mathcal{F}(x) = \{F : F \subset X, \text{ and } X \sim F \ \varepsilon\ \mathcal{G}(x)\}.$$

Finally, a set $A \subset X$ is bounded if, and only if,

$$d(A) < \infty,$$

where d(.) has the meaning defined in the previous subsection.

<u>1.3 Doob's Assumptions and Results.</u>

(A.1) $\{\mathfrak{X}, \mathfrak{A}\}$ and $\{\Theta, \mathcal{B}\}$ are both isomorphic to Borel sets in a complete separable metric space.

(A.2) For every $A \varepsilon \mathfrak{A}$, $P_{(\cdot)}(A)$ is a $\mathcal{B}$-measurable function.

(A.3) The measure λ on Θ has finite first and second moments.

(A.4) If $\theta_1 \neq \theta_2$, there exists a set $A \varepsilon \mathcal{A}$ such that $P_{\theta_1}(A) \neq P_{\theta_2}(A)$.

Theorem 1.3.1

Under the assumptions set forth above:

(i) There exists an $\mathcal{A}_\infty$-measurable function f such that

$$P_\theta^\infty(\{f(\underline{x}) = \theta\}) = 1 ,$$

for all $\theta \varepsilon \Theta$.

(ii) $$\lambda\{\theta : P_\theta^\infty(\{Q_{\underline{x}}^n(B) \longrightarrow \chi_B(\theta) ; \text{ for all } B \varepsilon \mathcal{B} \ni \lambda(B) > 0\}) = 1\} = 1.$$

(iii) If the posterior densities exist; i.e. if

$$Q_{\underline{x}}^n(B) = \int_B q_n(\xi, \underline{x})\, \lambda(d\xi), \text{ for all } B \varepsilon \mathcal{B},$$

then

$$\lambda\{\theta : P_\theta^\infty(\{q_n(\alpha, \underline{x}) \rightarrow 0 , \lambda \text{ a.e.}, \alpha \neq \theta\}) = 1\} = 1 .$$

(iv) $$\lambda\{\theta : P_\theta^\infty(\{\beta_n(\underline{x}) \rightarrow \theta\}) = 1\} = 1 .$$

The proof of this theorem is given in [1] and [3] , the assumptions being more sharply delineated in [3] and [5] . To return now to the comment made following Definition 1.1.0, if Doob's assumptions are met, and Θ is discrete with positive λ-mass at each point, a completely specified Bayes' system $\{\lambda, P, Q, \beta\}$ is consistent.

Now for the existence of a consistent procedure, in particular for the existence of a consistent estimation system, it is necessary that the family $\{P_\theta : \theta \varepsilon \Theta\}$ satisfy assumption (A.4); if it does not then no consistent procedure exists, and for this reason, we refer to assumption (A.4) as the "Minimal Consistency Requirement". We shall enlarge on this matter in Section 4.

Section 2.

2.0 The New Prior Measure $\bar{\lambda}$

Doob's assumptions being fulfilled, suppose that

$$B_o = \{\theta_1, \theta_2, \ldots, \theta_k\} \in \mathcal{B}.$$

Define a new prior distribution $\bar{\lambda}$ on Θ by

$$\bar{\lambda}(B) = (1-\epsilon)\lambda(B) + \epsilon\,\varphi(B), \text{ for } B \in \mathcal{B},$$

where $\epsilon \in (0,1)$, $\varphi(B) = 0$ when $B \cap B_o = \emptyset$, and $\varphi(\{\theta_j\}) = a_j > 0$ for $j = 1,2,\ldots,k$, with $\sum_{j=1}^{k} a_j = 1$.

Then

$$\bar{\lambda}(B_o) = \epsilon,$$

$$\bar{\lambda}(\Theta \sim B_o) = (1-\epsilon).$$

In general, for every $B \in B_o$

$$\bar{\lambda}(B) = (1-\epsilon)\lambda(B) + \epsilon \sum_{j=1}^{k} a_j \chi_B(\theta_j).$$

The measure $\bar{\lambda}$ is constructed in [3], and it is conjectured that the system $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$ is consistent, at least when $Q^n_{(\cdot)}$ is defined properly om a P_n-null set. But in Section 3, without disproving this conjecture entirely, we do devise a system $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$ that is not consistent.

2.1 The Bayes' Estimation System $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$

By definition, for any $n \in N$

$$\bar{P}_n(A) = \int_{\Theta} P^n_\theta(A)\,\bar{\lambda}(d\theta)$$

for all $A \in \mathcal{A}_n$.

Therefore

$$\bar{P}_n(A) = (1-\epsilon)\int_{\Theta \sim B_o} P^n_\theta(A)\,\lambda(d\theta) + \epsilon \sum_{j=1}^{k} a_j P^n_{\theta_j}(A).$$

whence

$$(2.1.1) \qquad \overline{P}_n(A) = (1-\epsilon)\, P_n(A) + \epsilon \sum_{j=1}^{k} a_j\, P^n_{\theta_j}(A)$$

for all $A \varepsilon \mathcal{a}_n$.

If $B \varepsilon \mathcal{B}$, and $A \varepsilon \mathcal{a}$, and provided the denominator is non-zero,

$$\overline{Q}^n(B/A) = \frac{\int_{\Theta} \chi_B(\theta)\, P^n_\theta(A)\, \overline{\lambda}(d\theta)}{\int_{\Theta} P^n_\theta(A)\, \overline{\lambda}(d\theta)} .$$

Further, upon dividing both numerator and denominator by the factor $P_n(A)$, if non-zero, in the right-hand side of the above equality, and then expanding in terms of the prior λ , we obtain

$$(2.1.2) \quad \overline{Q}^n(B/A) = \frac{(1-\epsilon)\, Q^n(B/A) + \{P_n(A)\}^{-1}\{\epsilon \sum_{j=1}^{k} a_j\, \chi_B(\theta_j)\, P^n_{\theta_j}(A)\}}{(1-\epsilon)\ \{P_n(A)\}^{-1}\{\epsilon \sum_{j=1}^{k} a_j\, P^n_{\theta_j}(A_j)\}} .$$

A direct computation of posterior probabilities, and Bayes' estimates by means of limiting processes applied to the relations (1.1.5) and (1.1.7) is postponed momentarily; we prefer to assume the existence of density functions, in which case:

$$\overline{Q}^n_{\underline{x}}(B) = \frac{(1-\epsilon)\int_B p_n(\theta,\underline{x})\,\lambda(d\theta) + \epsilon \sum_{j=1}^{k} a_j\, \chi_B(\theta_j)\, p_n(\theta_j,\underline{x})}{(1-\epsilon)\int_{\Theta} p_n(\theta,\underline{x})\,\lambda(d\theta) + \epsilon \sum_{j=1}^{k} a_j\, p_n(\theta_j,\underline{x})} ,$$

for each $\underline{x}$ for which the denominator is non-zero. Dividing both the numerator and denominator in the right-hand side of this equality by $\int_{\Theta} p_n(\theta,\underline{x})\,\lambda(d\theta)$, if non-zero, we obtain

$$(2.1.3) \qquad \bar{Q}^n_{\underline{x}}(B) = \frac{(1-\epsilon) Q^n_{\underline{x}}(B) + \epsilon \sum_{j=1}^{k} a_j q_n(\theta_j, \underline{x}) \chi_B(\theta_j)}{(1-\epsilon) + \epsilon \sum_{j=1}^{k} a_j q_n(\theta_j, \underline{x})} .$$

For the Bayes' estimates, by successive steps, as for the equalities (2.1.2) and (2.1.3), we obtain in turn

$$\bar{\beta}_n(\underline{x}) = \frac{\int_{\Theta} \theta p_n(\theta, \underline{x}) \bar{\lambda}(d\theta)}{\int_{\Theta} p_n(\theta, \underline{x}) \bar{\lambda}(d\theta)} ,$$

$$\bar{\beta}_n(\underline{x}) = \frac{(1-\epsilon) \int_{\Theta} \theta p_n(\theta, \underline{x}) \lambda(d\theta) + \epsilon \sum_{j=1}^{k} \theta_j a_j p_n(\theta_j, \underline{x})}{(1-\epsilon) \int_{\Theta} p_n(\theta, \underline{x}) \lambda(d\theta) + \epsilon \sum_{j=1}^{k} a_j p_n(\theta_j, \underline{x})} .$$

Thence, provided $\int_{\Theta} p_n(\theta, \underline{x}) \lambda(d\theta) \neq 0$,

$$(2.1.4) \qquad \bar{\beta}_n(\underline{x}) = \frac{(1-\epsilon) \beta_n(\underline{x}) + \epsilon \sum_{j=1}^{k} \theta_j a_j q_n(\theta_j, \underline{x})}{(1-\epsilon) + \epsilon \sum_{j=1}^{k} a_j q_n(\theta_j, \underline{x})} .$$

It should be noted again, that the relations (2.1.2), (2.1.3), and (2.1.4) are valid except on a P_n -null set.

Theorem 2.1.0

Suppose that a Bayes' estimation system $\{\lambda, P, Q, \beta\}$ is constructed from the assumptions (A.1), (A.2), (A.3), and (A.4); if B_0 is the exceptional set for the system $\{\lambda, P, Q, \beta\}$ such that

$$B_0 = \{\theta_1, \theta_2, \ldots, \theta_k\} ,$$

if furthermore, for every $\theta \in B_0$

$$P^{\infty}_{\theta}(\{q_n(\theta_j, \underline{x}) \longrightarrow 0, \text{ for } j = 1, 2, \ldots, k\}) = 1,$$

then the system $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$ is consistent.

Proof: By theorem 1.3.1 (iv), $\lambda(B_0)=0$, and the system $\{\lambda, P, Q, \beta\}$ is consistent for $\theta \in \Theta \sim B_0$. Further, if $\alpha = 1,2$:

$$\int_{\Theta} \xi^{\alpha} \bar{\lambda}(d\xi) = \epsilon \sum_{j=1}^{k} \theta_j^{\alpha} a_j + (1-\epsilon) \int_{\Theta} \xi^{\alpha} \lambda(d\xi) ,$$

therefore since λ satisfies (A.3), so does $\bar{\lambda}$.

If $\theta \notin B_0$, then by assumption,

$$P_{\theta}^{\infty}(\{\lim_{n \in N} \sum_{j=1}^{k} \theta_j a_j q_n(\theta_j, \underline{x}) = \lim_{n \in N} \sum_{j=1}^{k} a_j q_n(\theta_j, \underline{x}) = 0\}) = 1 .$$

Hence, by (2.1.4),

$$P_{\theta}^{\infty}(\{\lim_{n \in N} \bar{\beta}_n(\underline{x}) = \lim_{n \in N} \beta_n(\underline{x}) = \theta\}) = 1 .$$

If $\theta \in B_0$, then $\bar{\lambda}(\{\theta\}) > 0$ by construction; hence by the theorem 1.3.1 (iv)

$$P_{\theta}^{\infty}(\{\bar{\beta}_n(\underline{x}) \to \theta\}) = 1 .$$

The question naturally arises whether the convergence to 0, "with probability (P_{θ}^{∞}) 1", of the posterior densities evaluated at any point $\alpha \neq \theta$ insures the convergence of Bayes' estimates to the point θ ; i.e., is it true that

$$\{\underline{x} : q_n(\alpha, \underline{x}) \to 0, \text{ for all } \alpha \neq \theta\} \subseteq \{\underline{x} : \beta_n(\underline{x}) \to \theta\} \quad ?$$

In a later section, we shall undertake to answer this question, at least partially. At any rate, in the counter-example immediately following, the inconsistency of the system $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$ is derived from the non-convergence of the posterior densities for the appropriate values of θ .

Section 3

3.0 Description and Notation for a Particular Probability Model

$\mathcal{X}_j = [0,1]$, for all $j \in N$.

$\Theta = [0,1]$.

λ : the Lebesque measure.

$\mathcal{Q} = \mathcal{B}$: the Borel sets of $[0,1]$.

For $\theta \neq \frac{1}{2}$, $P_\theta(A) = 1$, if $\theta \in A$,

$= 0$, otherwise.

For $\theta = \frac{1}{2}$, $P_\theta(\{0\}) = \frac{1}{2}$,

$P_\theta(\{1\}) = \frac{1}{2}$.

$\underline{x}_D$: any infinite sequence with identical coordinates $x \in [0,1]$.

$D^{(n)}$: the diagonal in $\mathcal{X}^n$; i.e., $\{\underline{x}_D^{(n)} : x \in [0,1]\}$.

$C^{(n)}$: the n-dimensional product of any Borel set $C \subset [0,1]$.

$\tilde{C}^{(n)}$: any n-dimensional cube $\prod_{j=1}^{n} C_j$, such that for every $j \leq n$, C_j is a Borel set in $[0,1]$, $\lambda(C_j)$ is constant, and $D^{(n)} \cap \tilde{C}^{(n)} = \phi$.

For any set $A \in \mathcal{Q}_n \ni A = A^n \times \mathcal{X}^\infty$:

$$D^n(A^n) = D^{(n)} \cap A^n ,$$

$$D(A) = \{ x : \underline{x}_D^{(n)} \in D^n(A^n)\} .$$

Remarks

(a) For any Borel set $C \subset [0,1]$, $C^{(n)} \cap D^{(n)} = \{\underline{x}_D^{(n)} : x \in C\}$.

(b) If $\tilde{C}^{(n)} = \prod_{j=1}^{n} C_j$ then $\bigcap_{j=1}^{n} C_j = \phi$.

(c) The choice of Borel sets rather than the subsets of $[0,1]$ for the σ -field $\mathcal{Q}$ is made to satisfy assumption (A.2).

3.1 Characterization of the Measures P_θ^n

For $\theta \neq \frac{1}{2}$: If $C \in \mathcal{Q}$, then

(3.1.1) $$P_\theta^n(C^{(n)}) = \{P_\theta(C)\}^n = 1, \text{ if } \theta \in C,$$

$= 0$, otherwise.

If $\tilde{C}^{(n)}$ is an "off-diagonal" cube, then

(3.1.2) $$P_\theta^n(C^{(n)}) = 0.$$

In general, if A is any $\mathcal{A}_n$ subset

(3.1.3) $$P_\theta^n(A^n) = 1, \text{ if } \underline{\theta}_D^{(n)} \varepsilon A^n,$$

$$= 0, \text{ otherwise.}$$

Note that

(3.1.4) $$P_\theta^n(\{\mathcal{X}^n \sim D^{(n)}\}) = 0,$$

$$P_\theta^n(D^{(n)}) = 1.$$

<u>For $\theta = \frac{1}{2}$</u>: If $A \varepsilon \mathcal{A}$, then $P_{\frac{1}{2}}(A) = 0, \frac{1}{2}$, or 1, according as A contains respectively none, one, or both of the points $\{0\}$ and $\{1\}$.

If A is a rectangle in $\mathcal{A}_n$, and $A^n = \prod_{j=1}^{n} A_j$, it follows readily that

(3.1.5) $P_{\frac{1}{2}}^n(A^n) = 0$, if at least one of the faces A_j contains neither $\{0\}$, nor $\{1\}$,

$= 2^{-k}$, for $k = 0,1,2,\ldots,n$, if k and (n-k) of the faces A_j contain, respectively, one and both of the points $\{0\}$ and $\{1\}$.

Now if we let $D^\infty = \{\underline{x} : \underline{x}^{(n)} \varepsilon D^{(n)} \text{ for all } n \varepsilon N\}$,

(3.1.6) $$P_\theta^\infty(D^\infty) = 1, \quad \text{if} \quad \theta \neq \tfrac{1}{2},$$

$$P_{\frac{1}{2}}^\infty(\{\mathcal{X}^\infty \sim D^\infty\}) = 1.$$

<u>3.2 The Measures P_n</u>

If $C \varepsilon \mathcal{A}$, then

(3.2.1) $$P_n(C^{(n)}) = \int_\Theta P_\theta^n(C^{(n)})\, \lambda(d\theta) = \int \chi_C(\theta)\, \lambda(d\theta) = \lambda(C)$$

for every $n \varepsilon N$. Moreover, for any cube $\tilde{C}^{(n)}$,

(3.2.2) $$P_n(\tilde{C}^{(n)}) = 0.$$

Observe that

$$P_\theta^n(\underline{\theta}_D^{(n)}) = 1 \ , \ \text{if } \theta \neq \tfrac{1}{2},$$

but

$$P_n(\underline{\theta}_D^{(n)}) = 0 \ , \ \text{for any } \theta \in \Theta .$$

Furthermore, for any "off-diagonal" subset $B \subset \{\mathcal{X}^n \sim D^{(n)}\}$

(3.2.3)
$$P_n(B) = 0 .$$

Be it noted that the relations (3.2.1), (3.2.2) and (3.2.3) are particular instances of the following facts. If A is any rectangle $A^n \times \mathcal{X}^\infty$ in $\mathcal{A}_n$ such that $A^n = \prod_{j=1}^n A_j$, then by (3.1.3),

$$P_n(A^n) = \int \prod_{j=1}^n \chi_{A_j}(\theta)\, \lambda(d\theta) = \int \chi_{\bigcap_{j=1}^n A_j}(\theta)\, \lambda(d\theta) = \lambda(\bigcap_{j=1}^n A_j);$$

clearly

$$\bigcap_{j=1}^n A_j = \{\theta : \underline{\theta}_D^{(n)} \in D^n(A^n)\} = D(A).$$

Therefore

$$P_n(A^n) = \lambda(D(A)).$$

In general if $A \in \mathcal{A}_n$,

$$P_n(A^n) = \int_\Theta P_\theta^n(A^n \cap D^{(n)})\, \lambda(d\theta) + \int_\Theta P_\theta^n(A^n \sim D^{(n)})\, \lambda(d\theta) ;$$

but, by (3.1.4),

$$\int_\Theta P_\theta(A^n \sim D^{(n)})\, \lambda(d\theta) = 0 ;$$

therefore

$$P_n(A^n) = \int \chi_{\{\theta : \underline{\theta}_D^{(n)} \in D^n(A^n)\}}(\theta)\, \lambda(d\theta).$$

Hence, for every $A \varepsilon \mathcal{A}_n$

(3.2.4) $$P_n(A^n) = \lambda(D(A))$$

Observe again, how the dual rôle of the measures P_θ^n is reflected in the measures P_n, we write

$$P_n(A^n) = \int_{\Theta} P_\theta^n(A^n)\,\lambda(d\theta),$$

yet we mean

$$P_n(A) = \int_{\Theta} P_\theta^\infty(A^n \times \mathcal{X}^\infty)\,\lambda(d\theta).$$

But we shun the latter notation on the grounds that it is decidedly too cumbersome.

3.3 The Posterior Probability Measures

For every $B \varepsilon \mathcal{B}$, the equality (1.1.4) defines, almost everywhere P_n, the function $Q_{(\cdot)}^n(B)$ as a version of the Radon-Nikodym derivative of the measure $\mu_n(\,\cdot\,\times B)$ with respect to the measure P_n, for any $n \varepsilon N$. For every $A \varepsilon \mathcal{A}_n$

(3.3.1) $$\mu_n(A \times B) = \int_A Q_{\underline{x}}^n(B)\, P_n(d\underline{x}).$$

By definition, and by (3.2.4),

(3.3.2) $$\mu_n(A \times B) = \int \chi_B(\theta)\, P_\theta^n(A)\, \lambda(d\theta)$$

$$= \int \chi_B(\theta)\, \chi_{D(A)}(\theta)\, \lambda(d\theta)$$

$$= \lambda(B \cap D(A)).$$

Now, by (3.2.4)

$$\int_A \chi_{B^{(n)} \times \mathcal{X}^\infty}(\underline{x}) \, P_n(d\underline{x}) = P_n(B^{(n)} \cap A^n)$$

$$= \lambda(D(B^{(n)} \times \mathcal{X}^\infty \cap A)) .$$

However,

$$D^n(B^{(n)} \cap A^n) = D^n(B^{(n)}) \cap D^n(A^n) ,$$

hence

$$D(B^{(n)} \times \mathcal{X}^\infty \cap A) = B \cap D(A) .$$

Therefore

$$(3.3.3) \qquad \int_A \chi_{B^{(n)} \times \mathcal{X}^\infty}(\underline{x}) \, P_n(d\underline{x}) = \lambda(B \cap D(A)) .$$

It follows, by (3.3.1), (3.3.2) and (3.3.3) that

$$\int_A \chi_{B^{(n)} \times \mathcal{X}^\infty}(\underline{x}) \, P_n(d\underline{x}) = \int_A Q^n_{\underline{x}}(B) \, P_n(d\underline{x})$$

for every $A \in \mathcal{Q}_n$; thence by the Radon-Nikodym theorem,

$$(3.3.4) \qquad Q^n_{\underline{x}}(B) = \chi_{B^{(n)} \times \mathcal{X}^\infty}(\underline{x}) = \chi_{B^{(n)}}(\underline{x}^{(n)}) , \quad a.e. \ P_n ,$$

for any $B \in \mathcal{B}$.

We now specify the system $\{\lambda, P, Q, \beta\}$ completely thus: for every $\underline{x} \in \mathcal{X}^\infty$, for every $B \in \mathcal{B}$, and for every $n \in N$, let

$$Q^n_{\underline{x}}(B) = \chi_{B^{(n)}}(\underline{x}^{(n)}) .$$

This defines $Q^n_{\underline{x}}$ as a proper distribution on $\{\Theta, \mathcal{B}\}$, for every $\underline{x} \in \mathcal{X}^\infty$, and for every $n \in N$. In particular

$$(3.3.5) \qquad Q^n_{\underline{x}_D}(B) = \chi_{B^{(n)}}(\underline{x}_D^{(n)}) = \chi_B(x)$$

for any $x \in [0,1]$. On the other hand, if $\underline{x} \notin D^\infty$, then there exists a $J \in N$ such that for all $n \geq J$, $\underline{x}^{(n)} \notin D^{(n)}$. Consequently, for any fixed $n \geq J$, there exists a number $\epsilon_n > 0$ such that if $B \in \mathcal{B}$, and $d(B) < \epsilon_n$, then $\underline{x}^{(n)} \notin B^{(n)}$, and hence

$$(3.3.6) \qquad Q^n_{\underline{x}}(B) = 0,$$

for all $B \in \mathcal{B}$ such that $d(B) < \epsilon_n$.

3.4 The Exceptional Set of the System $\{\lambda, P, Q, \beta\}$

By (3.3.5) and by (1.1.7)

$$(3.4.1) \qquad \beta_n(\underline{x}_D) = \int_{\Theta} \xi \, Q^n_{\underline{x}_D}(d\xi) = x.$$

For any $\underline{x} \notin D^\infty$, $\exists$ $J(\underline{x}) \in N$ $\ni$ by (3.3.6)

$$(3.4.2) \qquad \beta_n(\underline{x}) = 0,$$

for every $n \geq J(\underline{x})$.

Therefore, if $\theta \neq \frac{1}{2}$, by (3.1.3) and (3.4.1)

$$P^\infty_\theta(\{\beta_n(\underline{x}) \to \theta\}) = P^\infty_\theta(\{\beta_n(\underline{x}_D) \to \theta\}) = 1.$$

If $\theta = \frac{1}{2}$, then by (3.1.6) and (3.4.2)

$$P^n_{\frac{1}{2}}(\{\beta_n(\underline{x}) \to 0\}) = 1.$$

Therefore the system is consistent except on $B_0 = \{\frac{1}{2}\}$ where $\lambda(B_0) = 0$.

3.5 The System $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$

In sub-section 2.1, we avoided an explicit representation of the posterior probabilities and Bayes' estimates by means of limiting processes. We proceed to such a representation for this particular case.

The version of relation 2.1.2 appropriate to the present situation is

$$(3.5.1) \qquad \bar{Q}^n(B/A) = \frac{(1-\epsilon)\, Q^n(B/A) + \{P_n(A)\}^{-1}\{\epsilon\, \chi_B(\frac{1}{2})\, P^n_{\frac{1}{2}}(A)\}}{(1-\epsilon) + \{P_n(A)\}^{-1}\{\epsilon\, P^n_{\frac{1}{2}}(A)\}}.$$

provided $P_n(A) \neq 0$; if in addition $P^n_{\frac{1}{2}}(A) \neq 0$, then

$$(3.5.2) \qquad \bar{Q}^n(B/A) = \frac{(1-\epsilon)Q^n(B/A)\{P_n(A)\}\{P^n_{\frac{1}{2}}(A)\}^{-1} + \epsilon\,\chi_B(\frac{1}{2})}{(1-\epsilon)\{P_n(A)\}\{P^n_{\frac{1}{2}}(A)\}^{-1} + \epsilon} .$$

Now let $C_j = [0, \epsilon_j)$, $0 < \epsilon_{j+1} \leq \epsilon_j < 1$, and $\lim_{n\in N} \epsilon_j = 0$.
Then, by (3.5.2), (3.2.1) and (3.1.5)

$$(3.5.3) \qquad \bar{Q}^n(B/C_j^{(n)} \times \mathcal{X}^\infty) = \frac{(1-\epsilon)Q^n(B/C_j^{(n)} \times \mathcal{X}^\infty)\{2^n\epsilon_j\} + \epsilon\,\chi_B(\frac{1}{2})}{(1-\epsilon)\{2^n\epsilon_j\} + \epsilon} ,$$

for every $n \in N$. Therefore, by (1.1.5)

$$(3.5.4) \qquad \bar{Q}^n_{0_D}(B) = \frac{(1-\epsilon)Q^n_{0_D}(B)\{2^n \lim_{j\in N}\epsilon_j\} + \epsilon\,\chi_B(\frac{1}{2})}{(1-\epsilon)\{2^n \lim_{j\in N}\epsilon_j\} + \epsilon} .$$

But $Q^n_{0_D}$ is well defined for any $n \in N$, and

$$\lim_{j\in N} \epsilon_j = 0 ;$$

hence it is quite clear that

$$(3.5.6) \qquad \bar{Q}^n_{0_D}(B) = \chi_B(\tfrac{1}{2}) ,$$

for every $B \in \mathcal{B}$, and for any $n \in N$.

By a similar argument, it can be shown also that

$$(3.5.7) \qquad \bar{Q}^n_{1_D}(B) = \chi_B(\tfrac{1}{2}) ,$$

for every $B \in \mathcal{B}$, and for any $n \in N$.

Now let $\{P_i\}_{i\in N}$ be a sequence of Borel partitions of the space Θ such that

$$\sup_{B\in P_i} d(B) \leq \frac{1}{i} .$$

Then, for any $n \varepsilon N$, by (3.5.6) ,

$$\frac{1}{2} - \frac{1}{\iota} \leqslant \bar{\beta}_n(\underline{0}_D) = \int_{\Theta} \xi \, Q^n_{\underline{0}_D}(d\xi) \leqslant \frac{1}{2} + \frac{1}{\iota}$$

for all $\iota \varepsilon N$; and hence

$$\bar{\beta}_n(\underline{0}_D) = \frac{1}{2}$$

for every $n \varepsilon N$.

Similarly, by (3.5.7)

$$\bar{\beta}_n(\underline{1}_D) = \frac{1}{2}$$

for every $n \varepsilon N$.

It follows immediately that, by (3.1.3)

$$P^\infty_0(\{\bar{\beta}_n(\underline{x}) \to \tfrac{1}{2}\}) = P^\infty_1(\{\bar{\beta}_n(\underline{x}) \to \tfrac{1}{2}\}) = 1.$$

This last relation shows that the system $\{\bar{\lambda}, \bar{P}, \bar{Q}, \bar{\beta}\}$ is not consistent.

It should be emphasized that the validity of this counter-example rests on a particular choice of the probabilities $Q^n_{\underline{x}}$ for the "off-diagonal" sequences. A more reasonable estimator would be

$$h_n(\underline{x}_D) = x \ , \text{ for any } x \varepsilon [0,1],$$

$$h_n(\underline{x}) = \frac{1}{2} \ , \text{ if } \underline{x}^{(n)} \notin D^{(n)} \text{ for some } n \varepsilon N.$$

Here $h_n(\cdot)$ is a Bayes' estimator that agrees with $\beta_n(\cdot)$, a.e. P_n, namely on the diagonal, but which is consistent for all $\theta \varepsilon \Theta$.

Section 4

4.0 Orientation

Boev's paper, which we now consider is devoted to the study of the asymptotic behaviour of the functions $q_n(\cdot, \underline{x})$, but his results seem to have been gotten without proper foundations. We proceed to reformulate his initial assumptions, thence to describe the behaviour of the functions $q_n(\cdot, \underline{x})$, and finally to examine the behaviour of the Bayes' estimates in the light of Boev's conclusions.

4.1 Assumptions and Basic Lemmas

(B.1) Θ and $\mathfrak{X}$ are two σ-compact subsets of the real line $\{R, \rho\}$, where $\rho(x,y) = |x-y|$, for all $(x,y) \in R \times R$.

(B.2) For every $\theta \in \Theta$, P_θ is absolutely continuous with respect to a fixed measure ν, so that there exists a density function $p_1(\theta, \cdot)$ satisfying

$$P_\theta(\mathfrak{X}) = \int_{\mathfrak{X}} p_1(\theta, x) \; \nu(dx) = 1 .$$

The function $p_1(\cdot, \cdot)$ is continuous and bounded in the product topology of the space $\Theta \times \mathfrak{X}$; moreover, there is a function $\theta_*: \mathfrak{X} \to \Theta$ satisfying

(i) $$\sup_{\xi \in \Theta} p_1(\xi, x) = p_1(\theta_*(x), x) ,$$

(ii) there is a constant $\gamma > 0$, and for every $x \in \mathfrak{X}$, there exists a neighborhood $V_x \in \mathcal{V}(\theta_*(x))$ such that

$$p_1(\xi, x) \geqslant (1 - |\xi - \theta_*(x)|^\gamma) \; p_1(\theta_*(x), x)$$

for all $\xi \in V_x$, and such that

$$\inf_{x \in \mathfrak{X}} d(V_x) = m > 0 .$$

Furthermore, given $\epsilon > 0$, there exists two compact subsets $K_1 \subset \Theta$ and $K_2 \subset \mathcal{X}$ such that

$$p_1(\xi, t) < \epsilon$$

for all pairs $(\xi, t) \notin K_1 \times K_2$.

(B.3) λ has a continuous density p with respect to the Lebesgue measure μ, such that $p(\cdot)$ is bounded on Θ and vanishes only at isolated points. Moreover $\lambda(\Theta) < \infty$, and the first moment of λ is finite also.

(B.4) For ν^n almost every sequence $\underline{x}^{(n)} \varepsilon \mathcal{X}^n$, where ν^n is the n-dimensional ν-measure, and $n \varepsilon N$ is arbitrary, there exists a mapping $Y_n : \Theta \to \mathcal{X}$ satisfying

(i) $$\prod_{j=1}^{n} p_\xi(x_j) = p_n(\xi, \underline{x}) = p_1^n(\xi, y_n(\xi, \underline{x})),$$

(ii) for any $\xi \varepsilon \Theta$, $\lim\limits_{n \varepsilon N} y_n(\xi, \underline{x})$ exists in $\mathcal{X}$.

Throughout this section, unless otherwise noted, it shall be understood that the above assumptions are fulfilled in every statement of proposition.

The function $p_1(\cdot, \cdot)$ being continuous and bounded in the product topology of the space $\Theta \times \mathcal{X}$ and the component spaces of this Cartesian product being σ-compact with the proviso following assumption (B.2 (ii)), for every $\xi \varepsilon \Theta$, and for every $x \varepsilon \mathcal{X}$, the projection mappings $p_1(\xi, \cdot)$ and $p_1(\cdot, x)$ are uniformly continuous in their respective domain. But more ensues, specifically

<u>Lemma 4.1.1</u>

Given any open set $G \varepsilon \mathcal{G}(\Theta)$, empty or non-empty, for every $x \varepsilon \mathcal{X}$, there exists a point $t \varepsilon \Theta \sim G$ such that

$$p_1(t, x) = \sup_{\xi \varepsilon \Theta - G} p_1(\xi, x).$$

Moreover, the function

$$h_\varsigma(\cdot) = \sup_{\xi \in \Theta - \varsigma} p_1(\xi, \cdot)$$

is continuous in $\mathcal{X}$.

Lemma 4.1.2

For every $x \in \mathcal{X}$, and for any $V \in \mathcal{V}(\theta_*(x))$, there exists a number $\epsilon > 0$ such that

$$\sup_{\xi \in \Theta - V} p_1(\xi, x) < (1-\epsilon)\, p_1(\theta_*(x), x) ,$$

and hence the function $\theta_*(\cdot)$ is continuous in $\mathcal{X}$.

Proof: Suppose on the contrary that for some $x \in \mathcal{X}$, $\exists$ $V \in \mathcal{V}(\theta_*(x))$ such that for all $\epsilon \in (0,1)$

(1) $$\sup_{\xi \in \Theta - V} p_1(\xi, x) \geqslant (1-\epsilon)\, p_1(\theta_*(x), x) .$$

Let $t \in \Theta \sim V$ satisfy

(2) $$p_1(t, x) = \sup_{\xi \in \Theta \sim V} p_1(\xi, x) .$$

Choose a decreasing sequence $\{\epsilon_j\}_{j \in N}$ $\ni$

(3) $$\epsilon_j \in (0,1) , \text{ for all } j \in N ,$$

(4) $$\lim_{n \in N} \epsilon_j = 0 .$$

Now by (1) and (2)

$$p_1(t, x) \geqslant (1-\epsilon_j)\, p_1(\theta_*(x), x) ,$$

for every $j \in N$. Thence, by (4)

(5) $$p_1(t, x) \geqslant p_1(\theta_*(x), x) .$$

But this is absurd since $t \notin V$, and $\theta_*(x)$ is unique.

Now let $x_0 \varepsilon \mathcal{X}$ be chosen arbitrarily and let

$$\theta_*(x_0) = \theta_* .$$

Choose any $V \varepsilon \mathcal{V}(\theta_*)$; then by the foregoing argument $\exists \; \epsilon \,\varepsilon\, (0, 1) \ni$

$$h_V(x_0) = \sup_{\xi \,\varepsilon\, \Theta \sim V} p_1(\xi, x_0) < (1-\epsilon)\, p_1(\theta_*, x) .$$

Select two numbers α and β so that

$$h_V(x_0) + \alpha \leqslant (1-\epsilon)\,\{ p_1(\theta_*, x_0) - \beta \} .$$

But by Lemma 4.1.1, $h_V(\cdot)$ and indeed $p_1(\theta_*(\cdot), \cdot)$ being continuous in $\mathcal{X}$, there exists a $\mathcal{U} \varepsilon \mathcal{V}(x_0)$ such that, simultaneously,

$$h_V(x) < h_V(x_0) + \alpha ,$$

$$p_1(\theta_*, x_0) - \beta < p_1(\theta_*(x), x) ,$$

and therefore

$$h_V(x) < p_1(\theta_*(x), x) ,$$

for all $x \,\varepsilon\, \mathcal{U}$. But the last inequality implies clearly that $\theta_*(x) \,\varepsilon\, V$, for all $x \,\varepsilon\, \mathcal{U}$; since V was chosen arbitrarily we conclude that $\theta_*(\cdot)$ is continuous in $\mathcal{X}$.

<u>Lemma 4.1.3</u>

Let $x \,\varepsilon\, \mathcal{X}$, otherwise arbitrary, and suppose that

$$\theta_*(x) = \theta_* .$$

There exists a fixed $\sigma > 0$, and a $\delta > 0$ that may be taken arbitrarily small such that for all $n \,\varepsilon\, N$

$$\int_{\Theta} p_1^n(\xi, x)\, \lambda(d\theta) \geqslant \frac{\sigma\delta}{2}\, p_1^n(\theta_*, x)\, \{1 - \delta^\gamma\}^n .$$

Proof:

Let $p(\theta_*) > 0$; by assumptions (B.2) and (B.3), there exists a $V \varepsilon \mathcal{V}(\theta_*)$ such that

$$p_1(\xi, x) \geqslant (1 - |\xi - \theta_*|^{\gamma})\, p_1(\theta_*, x)\,,$$

$$\inf_{\xi \varepsilon V} p(\xi) = \sigma > 0,$$

for all $\xi \varepsilon V$; and hence

(1) $$\int_{\Theta} p_1^n(\xi, x)\, \lambda(d\xi) \geqslant \sigma p_1^n(\theta_*, x) \int_V (1 - |\xi - \theta_*|^{\gamma})^n \mu(d\xi).$$

Assume that $\beta > 0$ is chosen so that

$$I(\theta_*, \beta) \subset V$$

Therefore, for any $\delta < \beta$

(i) $$[\theta_* - \delta, \theta_* - \tfrac{\delta}{2}] \subset I(\theta_*, \beta)\,,$$

or (ii) $$[\theta_* + \tfrac{\delta}{2}, \theta_* + \delta] \subset I(\theta_*, \beta)\,.$$

The argument being the same in either case, we show it for case (i) only: by a change of variable.

(2) $$\int_V (1 - |\xi - \theta_*|^{\gamma})^n \mu(d\xi) \geqslant \int_{[\theta_* - \delta, \theta_* - \frac{\delta}{2}]} (1 - |\xi - \theta_*|^{\gamma})^n \mu(d\xi)$$

$$= \int_{\frac{\delta}{2}}^{\delta} (1 - t^{\gamma})^n dt \geqslant \frac{\delta}{2}(1 - \delta^{\gamma})^n.$$

Thence, by (1) and (2)

$$\int_{\Theta} p_1^n(\xi, x)\, \lambda(d\xi) \geqslant \frac{\sigma\delta}{2}\, p_1^n(\theta_*, x)\, \{1 - \delta^{\gamma}\}^n.$$

If $p(\theta_*) = 0$, we may by assumption (B.3), choose β such that for any $\xi \in [\theta_* - \delta, \theta_* - \frac{\delta}{2}]$; and for any $\delta < \beta$

$$\inf_{\xi \in [\theta_* - \delta, \theta_* - \frac{\delta}{2}]} p(\xi) > 0 .$$

The argument thence proceeds as before.

<u>Lemma 4.1.4</u>

If $x \in \mathcal{X}$ and

$$\theta_*(x) = \theta ,$$

then for any $\alpha \neq \theta$

$$\lim_{n \in N} \frac{p_1^n(\alpha, x)}{\int_\Theta p_1^n(\xi, x)\, \lambda(d\xi)} = 0 .$$

Moreover, the convergence is uniform on $\{\Theta \sim V\}$, for any $V \in \mathcal{V}(\theta_*)$.

<u>Proof</u>: Choose $V \in \mathcal{V}(\theta_*) \ni \alpha \notin V$; by Lemma 4.1.2 $\exists\ \epsilon \in (0,1) \ni$

(1) $$p_1^n(\alpha; x) \leqslant h_V^n(x) < (1-\epsilon)^n p_1^n(\theta_*, x) .$$

On the other hand, by the previous lemma

(2) $$\int_\Theta p_1^n(\xi, x)\, \lambda(d\xi) \geqslant \sigma p_1^n(\theta_*, x) \left\{ \frac{\delta}{2} (1 - \delta\gamma)^n \right\} ,$$

where $\sigma > 0$ is fixed, and δ may be as small as desired. In particular, if

$$\delta\gamma < \epsilon ,$$

by (1) and (2), we obtain, after simplification,

(3) $$\frac{p_1^n(\alpha, x)}{\int_\Theta p_1^n(\xi, x)\, \lambda(d\xi)} \leqslant \frac{h_V^n(x)}{\int_\Theta p_1^n(\xi, x)\, \lambda(d\xi)} \leqslant \frac{2}{\sigma\delta} \left\{ \frac{1-\epsilon}{1-\delta\gamma} \right\}^n .$$

But, by choice of δ

$$\left\{\frac{1-\epsilon}{1-\delta^{\gamma}}\right\} < 1,$$

therefore

$$\lim_{n\epsilon N}\left\{\frac{1-\epsilon}{1-\delta\gamma}\right\}^{n} = 0 .$$

the conclusion is immediate.

4.2 The Main Convergence Properties

It should be noted that the function $\theta_*(\cdot)$ evaluated at the point $y_n(\xi, \underline{x})$ is independent of ξ ; i.e. if $\underline{x}^{(n)}$ is held fixed then $p_n(\theta_*(y_n(\xi,\underline{x}), \underline{x})$ is constant for all $\xi \epsilon \Theta$, and for any $n \epsilon N$. Recalling the notation in assumption (B.4(i)), we have

$$\prod_{j=1}^{n} p_\xi(x_j) = p_n(\xi, \underline{x}) = p_1^n(\xi, y_n(\xi, \underline{x})) \cdot$$

To use a well-worn statement: $\theta_*(y_n(\xi, \underline{x}))$ is the value of the parameter ξ which maximizes the product $\prod_{j=1}^{n} p_\xi(x_j)$, for any fixed sequence $\underline{x}^{(n)}$; in other words, it is the maximum likelihood estimator of the parameter. In line with the established notational conventions, let

$$\hat{\theta}_n(\underline{x}) = \theta_*(y_n(\cdot, \underline{x})) .$$

Be it noted also that by assumptions (B.2) and (B.4), and by the Intermediate Value theorem, the maximum likelihood estimators do exist for every $\underline{x} \epsilon \mathcal{X}^\infty$, and for every $n \epsilon N$.

<u>Lemma 4.2.1</u>

If $\underline{x} \epsilon \mathcal{X}^\infty$ is such that $\lim_{n\epsilon N} y_n(\xi, \underline{x})$ exists for every $\xi \epsilon \Theta$, then $\lim_{n\epsilon N} \hat{\theta}_n(\underline{x})$ exists; if in addition

$$\lim_{n\epsilon N} \hat{\theta}_n(\underline{x}) = \theta_* ,$$

then for every $\alpha \neq \theta_*$,

$$\lim_{n \in N} q_n(\alpha, \underline{x}) = 0 .$$

Moreover, the convergence is uniform on $\{\Theta - V\}$ for any $V \varepsilon \mathcal{V}(\theta_*)$.

<u>Proof</u>: For simplicity of notation, let

$$\hat{\theta}_n = \hat{\theta}_n(\underline{x}) = \theta_*(y_n(\xi, \underline{x}))$$

for any $\xi \varepsilon \Theta$; by assumption and by lemma 4.1.2,

$$\lim_{n \in N} \hat{\theta}_n = \theta_*(\lim_{n \in N} y_n(\xi, \underline{x})) .$$

Choose $\alpha \neq \theta_*$; there exists a $V \varepsilon \mathcal{V}(\theta_*)$, such that $\alpha \notin V$. Now by lemma 4.1.2, there exists $\epsilon \varepsilon (0,1) \ni$ if

$$\lim_{n \in N} y_n(\alpha, \underline{x}) = y_0 ,$$

then

$$h_V(y_0) < (1-\epsilon)\, p_1(\theta_*, y_0) .$$

Since $h_V(\cdot)$, and $p_1(\cdot,\cdot)$ are continuous, and since $\hat{\theta}_n \to \theta_*$ independently of $\xi \varepsilon \Theta$, $\exists\ J_1 \varepsilon N \ni$ for all $n \geq J_1$

(1) $p_1(\alpha, y_n(\alpha, \underline{x})) \leq \sup\limits_{\xi \varepsilon \Theta - V} p_1(\xi, y_n(\alpha, \underline{x})) \leq (1-\epsilon)\, p_1(\hat{\theta}_n, y_n(\alpha, \underline{x}))$.

On the other hand, by assumption (B.2(ii)), we can find a neighborhood $U_n \varepsilon \mathcal{V}(\hat{\theta}_n)$ and a number $\beta \varepsilon (0,1) \ni$ for all $\xi \in U_n$, and for all $n \varepsilon N$

(2) $$p_1(\xi, y_n(\xi, \underline{x})) \geq (1 - |\xi - \hat{\theta}_n|^{\gamma})\, p_1(\hat{\theta}_n, y_n(\alpha, \underline{x})),$$

and

(3) $$I(\hat{\theta}_n, \beta) \subset U_n .$$

Thus by (1), (2) and (3), and by assumption (B.4), for all $n \geqslant J_1$

$$(4)\quad \frac{p_n(\alpha,\underline{x})}{\int_{\Theta} p_n(\xi,\underline{x})\,\lambda(d\xi)} \leqslant \frac{\sup_{\xi\in\Theta-V} p_n(\xi,\underline{x})}{\int_{\Theta} p_n(\xi,\underline{x})\,\lambda(d\xi)}$$

$$\leqslant \frac{(1-\epsilon)^n\, p_n(\hat\theta_n,\underline{x})}{p_n(\hat\theta,\underline{x})\int_{I(\hat\theta_n,\beta)}(1-|\xi-\hat\theta_n|^\gamma)^n\,\lambda(d\xi)} = \frac{(1-\epsilon)^n}{\int_{I(\hat\theta_n,\beta)}(1-|\xi-\hat\theta_n|^\gamma)^n\,\lambda(d\xi)} .$$

The remainder of the argument parallels the proof of lemma 4.1.3 : if $p(\theta_*) > 0$, assume β to be so chosen that, by assumption (B.3)

$$\inf_{\xi\in I(\theta_*,\beta)} p(\xi) = \sigma > 0 .$$

Next choose $\delta > 0 \ni$

$$\delta < \frac{\beta}{2} ,$$

$$\delta^\gamma < \epsilon .$$

Thus, since $\hat\theta_n \to \theta_*$, $\exists\ J_2 \in N$ such that

$$[\hat\theta_n-\delta,\hat\theta_n-\tfrac{\delta}{2}] \subset [\theta_*-\tfrac{5\delta}{4},\theta_*-\tfrac{\delta}{4}] \subset I(\theta_*,\beta),$$

$$\inf_{\xi\in[\hat\theta_n-\delta,\hat\theta_n-\frac{\delta}{2}]} p(\xi) \geqslant \sigma,$$

for all $n \geqslant J_2$.

Therefore

$$(5)\quad \int_{I(\hat\theta_n,\beta)}(1-|\xi-\hat\theta_n|^\gamma)^n\,\lambda(d\xi) \geqslant \sigma\int_{[\hat\theta_n-\delta,\hat\theta_n-\frac{\delta}{2}]}(1-|\xi-\hat\theta_n|^\gamma)^n\,\mu(d\xi)$$

$$= \sigma\int_{\frac{\delta}{2}}^{\delta}(1-t^\gamma)^n\,dt \geqslant \frac{\sigma\delta}{2}(1-\delta^\gamma)^n ,$$

for all $n \geqslant J_2$.

By (4) and (5), for all $n \geqslant \max(J_1, J_2)$

$$q_n(\alpha, \underline{x}) \leqslant \sup_{\xi \in \Theta - V} q_n(\xi, \underline{x}) \leqslant \frac{2}{\sigma\delta} \left\{\frac{1-\epsilon}{1-\delta^{\gamma}}\right\}^n .$$

The conclusion follows, by the choice of δ .

Again, if $p(\theta_*) = 0$; assume β to be so chosen that

$$\inf_{\xi \in K} p(\xi) > 0$$

for any compact subset $K \subset [\theta_* - \beta, \theta_*)$, then select $\delta < \min(\epsilon^{\frac{1}{\gamma}}, \frac{\beta}{2})$ such that

$$\inf_{\xi \in [\theta_* - \frac{5\delta}{4}, \theta_* - \frac{\delta}{4}]} p(\xi) = \omega > 0,$$

and thence proceed as before.

Theorem 4.2.1

Under the conditions of Lemma 4.2.1, for every neighborhood $V \in \mathcal{V}(\theta_*)$

$$\lim_{n \in N} \int_V q_n(\xi, \underline{x}) \, \lambda(d\xi) = 1 .$$

Proof: By definition, if $V \in \mathcal{V}(\theta_*)$, then

$$\int_V q_n(\xi, \underline{x}) \, \lambda(d\xi) = 1 - \int_{\Theta \sim V} q_n(\xi, \underline{x}) \, \lambda(d\xi).$$

But

$$\int_{\Theta - V} q_n(\xi, \underline{x}) \, \lambda(d\xi) \leqslant \lambda(\Theta - V) \sup_{\xi \in \Theta \sim V} q_n(\xi, \underline{x}) .$$

Since $\lambda(\Theta - V) < \infty$, then by the previous lemma:

$$\lim_{n \in N} \lambda(\Theta - V) \sup_{\xi \in \Theta \sim V} q_n(\xi, \underline{x}) = 0.$$

The desired conclusion follows at once.

Theorem 4.2.2

Under the conditions of Lemma 4.2.1

$$\lim_{n\varepsilon N} \beta_n(\underline{x}) = \lim_{n\varepsilon N} \int_{\Theta} \xi \, q_n(\xi, \underline{x}) \, \lambda(d\xi) = \theta_* .$$

Proof: Note that, for every $n \varepsilon N$,

(1) $$|\beta_n(\underline{x}) - \theta_*| \leqslant \int_{\Theta} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi).$$

Given any $\epsilon > 0$, choose a neighborhood $V \varepsilon \mathcal{V}(\theta_*)$ such that $\mu(V) < \epsilon$. By assumption (B.3), if Θ is unbounded, choose a Borel partition $\Theta = B_1 \cup B_2$, such that B_1 is bounded $V \subset B_1$, and

(2) $$\int_{B_2} |\xi| \, p(\xi) \, \mu(d\xi) \leqslant 1.$$

Now

(3) $$\int_{\Theta} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi) = \int_V |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi)$$
$$+ \int_{\Theta - V - B_2} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi) + \int_{B_2} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi).$$

But since $\mu(V) < \epsilon$, $|\xi - \theta_*| < \epsilon$ on V; hence

(4) $$\int_V |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi) \leqslant \epsilon,$$

for all $n \varepsilon N$; furthermore,

(5) $$\int_{\Theta - V - B_2} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi) \leqslant 2 \sup_{\xi \varepsilon B_1} |\xi| \int_{\Theta - V} q_n(\xi, \underline{x}) \, \lambda(d\xi),$$

(6) $$\int_{B_2} |\xi - \theta_*| \, q_n(\xi, \underline{x}) \, \lambda(d\xi) \leqslant \sup_{\xi \varepsilon B_2} q_n(\xi, \underline{x}) \left\{ \int_{B_2} |\xi| \, \lambda(d\xi) + |\theta_*| \lambda(B_2) \right\}$$
$$\leqslant \sup_{\xi \varepsilon \Theta - V} q_n(\xi, \underline{x}) \{ 1 + |\theta_*| \}.$$

But, by Lemma 4.2.1

$$(7) \qquad \lim_{n\in N} \sup_{\xi\in\Theta\sim V} q_n(\xi,\underline{x}) = \lim_{n\in N} \int_{\Theta\sim V} q_n(\xi,\underline{x})\,\lambda(d\xi) = 0.$$

Therefore, since $2\sup_{\xi\in B_1}|\xi| < \infty$, by (5) and (7)

$$\lim_{n\in N} \int_{\Theta\sim V\sim B_2} |\xi-\theta_*|\,q_n(\xi,\underline{x})\,\lambda(d\xi) = 0\ ;$$

by (6) and (7)

$$\lim_{n\in N} \int_{B_2} |\xi-\theta_*|\,q_n(\xi,\underline{x})\,\lambda(d\xi) = 0\ .$$

Therefore, by (3) and (4),

$$\lim_{n\in N} \int_{\Theta} |\xi-\theta_*|\,q_n(\xi,\underline{x})\,\lambda(d\xi) \leq \epsilon\ .$$

Finally, by (1) and (3)

$$\lim_{n\in N} |\beta_n(\underline{x}) - \theta_*| \leq \epsilon\ .$$

Since ϵ was chosen arbitrarily, it is clear that

$$\lim_{n\in N} \beta_n(\underline{x}) = \theta_*\ .$$

If Θ is bounded, let $B_2 = \emptyset$, and proceed as before.

Remarks:

It is well to note at this point that if we assume boundedness for the set Θ, then we may dispense with the assumption that λ have finite first order moment. Furthermore, as has been shown, there is no need to assume that $\lambda(\Theta) = 1$. In point of fact, it may happen that the functions $p_n(\cdot,\underline{x})$ and $q_n(\cdot,\underline{x})$ meet such exacting integrability conditions that our arguments, except for slight modifications, will yield the same conclusions, even though neither

$\lambda(\Theta)$ nor the first moment of λ is finite. This point shall be illustrated in the example in Section 4.3.

Theorem 4.2.3

If for some $\theta \varepsilon \Theta$,

$$P_\theta^\infty(\{\beta_n(\underline{x}) \longrightarrow \theta\}) = 1,$$

then

$$P_\theta^\infty(\{\hat{\theta}_n(\underline{x}) \longrightarrow \theta\}) = 1,$$

and conversely.

Proof: By assumption (B.4), since P_θ^n is absolutely continuous with respect to γ^n, and if we let

$$X = \{\underline{x} : \lim_{n \varepsilon N} y_n(\xi, \underline{x}) \text{ exists , for any } \xi \varepsilon \Theta\},$$

then

$$P_\theta^\infty(X) = 1.$$

If

$$P_\theta^\infty(\{\beta_n(\underline{x}) \rightarrow \theta\}) = 1,$$

for "probability-one" statements, there is no loss of generality in assuming that

$$X \equiv \{\beta_n(\underline{x}) \longrightarrow \theta\}.$$

Now if $\underline{x} \varepsilon X$, by assumption (B.4) and by theorem 4.1.2, suppose that

$$\lim_{n \varepsilon N} \hat{\theta}_n = \theta_* \ ;$$

clearly, by the previous theorem

$$\lim_{n \in N} \beta_n(\underline{x}) = \theta_* ,$$

and thence

$$\theta_* = \theta .$$

Therefore

$$P_\theta^\infty(\{\hat{\theta}_n(\underline{x}) \to \theta\}) = 1 .$$

The converse follows directly from theorem 4.2.2.

4.3 An Example

Let $p_1(\xi, x) = (2\pi)^{-\frac{1}{2}} \exp[-\frac{1}{2}(x-\xi)^2]$, $x \in (-\infty, \infty)$.

Then:
$$p_n(\xi, \underline{x}) = (2\pi)^{-\frac{n}{2}} \exp[-\tfrac{1}{2} \sum_{j=1}^{n} (x_j - \xi)^2] ,$$

$$p_1^n(\xi, y_n(\xi, \underline{x})) = (2\pi)^{-\frac{n}{2}} \exp[-\tfrac{n}{2}(y_n - \xi)^2] .$$

Solving for y_n, we obtain

$$y_n = \xi \pm \sqrt{\frac{1}{n}\sum_{j=1}^{n} x_j^2 - \frac{2\xi}{n}\sum_{j=1}^{n} x_j + \xi^2} .$$

By substitution

$$p_1(\xi, y_n(\xi, \underline{x})) = (2\pi)^{-\frac{1}{2}} \exp[-\tfrac{1}{2}\{\tfrac{1}{n}\sum_{j=1}^{n} x_j^2 - \tfrac{2\xi}{n}\sum_{j=1}^{n} x_j + \xi^2\}] .$$

Therefore,

$$\log p_1(\xi, y_n(\xi, \underline{x})) = -\tfrac{1}{2}\log 2\pi - \tfrac{1}{2}\{\tfrac{1}{n}\sum_{j=1}^{n} x_j^2 - \tfrac{2\xi}{n}\sum_{j=1}^{n} x_j + \xi^2\} ,$$

$$\frac{\partial}{\partial \xi} \log p_1(\xi, y_n(\xi, \underline{x})) = -\tfrac{1}{2}\{-\tfrac{2}{n}\sum_{j=1}^{n} x_j + 2\xi\} .$$

Hence

$$\hat{\theta}_n(\underline{x}) = \frac{1}{n}\sum_{j=1}^{n} x_j .$$

Case 1.

Let $p(\xi) = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{\xi^2}{2}\right]$, $\xi \in (-\infty, \infty)$,

then :

$$\int_{\Theta} \xi\, q_n(\xi, \underline{x})\, \lambda(d\xi) = \frac{\int_{-\infty}^{\infty} \xi \exp\left[\frac{1}{2}\left\{2\xi\sum_{j=1}^{n} x_j - (n+1)\xi^2\right\}\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[\frac{1}{2}\left\{2\xi\sum_{j=1}^{n} x_j - (n+1)\xi^2\right\}\right] d\xi}$$

$$= \frac{\int_{-\infty}^{\infty} \xi \exp\left[-\frac{n+1}{2}\left\{\frac{1}{n+1}\sum_{j=1}^{n} x_j - \xi\right\}^2\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[-\frac{n+1}{2}\left\{\frac{1}{n+1}\sum_{j=1}^{n} x_j - \xi\right\}^2\right] d\xi}$$

By computation

$$\beta_n(\underline{x}) = \frac{1}{n+1}\sum_{j=1}^{n} x_j .$$

Case 2.

Let $p(\xi) = 1$, $\xi \in (-\infty, \infty)$,

then

$$\int_{\Theta} \xi\, q_n(\xi, \underline{x})\, \lambda(d\xi) = \frac{\int_{-\infty}^{\infty} \xi \exp\left[\frac{1}{2}\left\{2\xi\sum_{j=1}^{n} x_j - n\xi^2\right\}\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[\frac{1}{2}\left\{2\xi\sum_{j=1}^{n} x_j - n\xi^2\right\}\right] d\xi}$$

$$= \frac{\int_{-\infty}^{\infty} \xi \exp\left[-\frac{n}{2}\left\{\frac{1}{n}\sum_{j=1}^{n} x_j - \xi\right\}^2\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[-\frac{n}{2}\left\{\frac{1}{n}\sum_{j=1}^{n} x_j - \xi\right\}^2\right] d\xi} .$$

Hence

$$\beta_n(\underline{x}) = \frac{1}{n}\sum_{j=1}^{n} x_j .$$

Case 3.

Let $p(\xi) = e^{\xi}$, $\xi \in (-\infty, \infty)$,

then

$$\int_{\Theta} \xi \, q_n(\xi, \underline{x})\, \lambda(d\xi) = \frac{\int_{-\infty}^{\infty} \xi \exp\left[\frac{1}{2}\{2\xi(\sum_{j=1}^{n} x_j + 1) - n\xi^2\}\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[\frac{1}{2}\{2\xi(\sum_{j=1}^{n} x_j + 1) - n\xi^2\}\right] d\xi}$$

$$= \frac{\int_{-\infty}^{\infty} \xi \exp\left[-\frac{n}{2}\{\frac{1}{n}(\sum_{j=1}^{n} x_j + 1) - \xi\}^2\right] d\xi}{\int_{-\infty}^{\infty} \exp\left[-\frac{n}{2}\{\frac{1}{n}(\sum_{j=1}^{n} x_j + 1) - \xi\}^2\right] d\xi} .$$

Therefore

$$\beta_n(\underline{x}) = \frac{1}{n}\left(\sum_{j=1}^{n} x_j + 1\right).$$

It is clear, by application of the Law of Large Numbers, that in all three cases

$$P_{\theta}^{\infty}(\{\hat{\theta}_n(\underline{x}) \to \theta\}) = P_{\theta}^{\infty}(\{\beta_n(\underline{x}) \to \theta\}) = 1 .$$

4.4 The Consistency Theorem

Thus far, we have been concerned with cases of convergence at points. Next, we consider a Bayes' estimation system $\{\lambda, P, Q, \beta\}$ constructed under the assumptions (B.1), (B.2), (B.3) and (B.4), and we ask: is such a system consistent? It turns out that restrictive as these assumptions are, they do not necessarily satisfy the minimal consistency requirement. It is indeed quite conceivable that two distinct points θ_1 and θ_2 may be found such that $p_1(\theta_1, \cdot)$ and $p_1(\theta_2, \cdot)$ are ν-equivalent, and hence for all $\mathcal{Q}_{\infty}$-measurable sets A

$$P_{\theta_1}^{\infty}(A) = P_{\theta_2}^{\infty}(A) ,$$

in which case no estimation system, of any type, is consistent. It should be clearly understood that while $\theta_*(x)$ maximizes $p_1(\cdot, x)$ uniquely, it is not assumed that every $\theta \in \Theta$ satisfies

$$p_1(\theta, x) = p_1(\theta_*(x), x),$$

for some $x \in \mathfrak{X}$.

For the purpose of quick reference, we now formalize a few remarks already made, and results obtained in the previous subsections.

<u>Lemma 4.4.0</u>

(i) The maximum likelihood estimates exist, for every $n \in N$, and for every $\underline{x}^{(n)} \in \mathfrak{X}^n$.

(ii) If

$$\lim_{n \in N} x_n = x_0 \in \mathfrak{X},$$

then

$$\lim_{n \in N} p_1(\theta_*(x_n), x_n) = p_1(\theta_*(x_0), x_0).$$

(iii) If $\xi \in \Theta$, and $\theta \in \Theta$, then

$$p_1(\theta_*(y_n(\xi, \underline{x}), y_n(\xi, \underline{x})) = p_1(\theta_*(y_n(\theta, \underline{x}), y_n(\theta, \underline{x})),$$

for all $n \in N$.

(iv) If

$$\lim_{n \in N} x_n = x_0,$$

then

$$\lim_{n \in N} \theta_*(x_n) = \theta_*(x_0).$$

Part (i) and part (iii) of this lemma restate the essence of the opening paragraph in subsection 4.2; part (ii) and part (iv) are direct consequences of the lemmas 4.1.1 and 4.1.2.

Theorem 4.4.0

Under the assumptions (B.1), (B.2), (B.3) and (B.4), a Bayes' estimation system $\{\lambda, P, Q, \beta\}$ is consistent if, and only if, the family $\{P_\theta : \theta \in \Theta\}$ satisfies the minimal consistency requirement; i.e., if, and only if, for any $\theta \in \Theta$, and for every $\xi \neq \theta$, there exists a ν-measurable set A_ξ such that

$$\int_{A_\xi} p_1(\theta, x)\ \nu(dx) \neq \int_{A_\xi} p_1(\xi, x)\ \nu(dx).$$

Proof: Choose $\theta \in \Theta$, and define

(1) $$X_\theta = \left\{ \underline{x} : \frac{p_n(\xi,\underline{x})}{p_n(\theta,\underline{x})} \to 0, \text{ for any } \xi \neq \theta \right\};$$

on the assumption of the minimal consistency requirement Doob showed in [2] that

(2) $$P_\theta^\infty(X_\theta) = 1.$$

As in the proof of theorem 4.2.3, we assume that if $\underline{x} \in X_\theta$, then $\lim\limits_{n \in N} y_n(\xi, \underline{x})$ exists in $\mathfrak{X}$, for any $\xi \in \Theta$.

From (1) and (2), and by assumption (B.4), if $\underline{x} \in X_\theta$, and if $\xi \neq \theta$, then for any $\delta \in (0,1)$ there exists a $J(\xi) \in N \ni$

(3) $$p_n(\xi, \underline{x}) < \delta p_n(\theta, \underline{x}) < p_n(\theta, \underline{x}),$$

and hence

(4) $$p_1(\xi, y_n(\xi, \underline{x})) < \delta^{\frac{1}{n}} p_1(\theta, y_n(\theta, \underline{x})) < p_1(\theta, y_n(\theta, \underline{x}))$$

for every $n \geqslant J(\xi)$.

Let

(5) $$\lim_{n \in N} y_n(\theta, \underline{x}) = y_\theta,$$

and suppose, using Lemmas 4.4.0 (i), (iii) and (iv), that

(6) $$\lim_{n\in N} \hat{\theta}_n(\underline{x}) = \lim_{n\in N} \theta_*(y_n(\cdot, \underline{x})) = \theta_* \neq \theta .$$

By Lemma 4.1.2, $\exists\ V \in \mathcal{V}(\theta_*)$ and $\epsilon \in (0,1)$ such that $\theta \notin V$ and

(7) $$p_1(\theta; y_\theta) < (1-\epsilon)\, p_1(\theta_*, y_\theta) .$$

Choose two positive numbers α and β such that

(8) $$p_1(\theta, y_\theta) + \alpha \leq (1-\epsilon)\{p_1(\theta_*, y_\theta) - \beta\} .$$

By (5) and by assumption (B.2) $\exists\ J_1 \in N \ni$

(9) $$p_1(\theta, y_n(\theta, \underline{x})) < p_1(\theta, y_\theta) + \alpha$$

for all $n \geq J_1$.

By (6) and by Lemmas 4.4.0 (ii) and (iii)

(10) $$\begin{aligned} p_1(\theta_*, y_\theta) &= \lim_{n\in N} p_1(\hat{\theta}_n(\underline{x}), y_n(\theta; \underline{x})) \\ &= \lim_{n\in N} p_1(\hat{\theta}_n(\underline{x}), y_n(\theta_*, \underline{x})) . \end{aligned}$$

But by assumptions (B.2) and (B.4), and by (6) and (10)

(11) $$p_1(\theta_*, y_\theta) = p_1(\theta_*, \lim_{n\in N} y_n(\theta_*, \underline{x})) = \lim_{n\in N} p_1(\theta_*, y_n(\theta_*, \underline{x}) .$$

It follows from this latter equality that $\exists\ J_2 \in N \ni$ for $n \geq J_2$

(12) $$p_1(\theta_*, y_\theta) - \beta < p_1(\theta_*, y_n(\theta_*, \underline{x})) ;$$

and hence, by (8), (9) and (12)

(13) $$p_1(\theta, y_n(\theta, \underline{x})) < p_1(\theta_*, y_n(\theta_*, \underline{x})) ,$$

and therefore

(14) $$p_n(\theta, \underline{x}) < p_n(\theta_*, \underline{x})$$

for all $n \geq \max(J_1, J_2)$.

Clearly, this last inequality contradicts the inequality (3) for all $n \geq \max(J_1, J_2, J(\theta_*))$.

Hence it must be concluded that

$$\theta_* = \theta.$$

Therefore, if $\underline{x} \in X_\theta$,

$$\lim_{n \in N} \hat{\theta}_n(\underline{x}) = \theta.$$

By theorem 4.2.3

$$P_\theta^\infty(\{\beta_n(\underline{x}) \to \theta\}) = 1.$$

Since θ was chosen arbitrarily, the conclusion follows.

The converse is trivially true.

References

1 Doob, J.L. (1949). Applications of the theory of martingales, Le Calcul des Probabilités et ses Applications, Colloques Internationaux du C.N.R.S., Lyon, 23, pp.23-29.

2 ________ (1953). Stochastic Processes, Wiley, New York.

3 Schwartz, L. (1960). On Consistency of Bayes Procedures, Thesis, Univ. of Calif., Berkeley.

4 Boev, G.P. (1959) O Polnom Aposteriornom Raspedelenii, Izv. Vysš. Učebn. Zaved. Matematika, no.5(12), pp.58-66.

5 Le Cam, L. (1958). Les Propriétés Asymptotiques des Solutions de Bayes, Publ. Inst. Stat. Univ. Paris, vol.7, no.3-4, pp.17-35.

6 Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-Existence of Everywhere Proper Conditional Distributions, Ann. Math. Statist., Vol. 28, pp.223-225.

7 Halmos, P. (1950). Measure Theory, Van Nostrand, Toronto-New York-London.

8 Loève, M. (1960). Probability Theory, Van Nostrand, Toronto-New York-London.