APPENDIX TO CHAPTER 4. POINTWISE LIMITS OF BAYES PROCEDURES

This appendix contains a proof of Theorem 4.14, which was used to establish the complete class Theorems 4.16 and 4.24, and will be used again in Chapter 7. As already noted, this theorem has nothing in particular to do with exponential families, but its proof is included here since it is not readily accessible elsewhere. We will state and prove it below in a convenie form which is more general than that stated in Theorem 4.14.

4A.1 Setting

Let $\{p_\theta(x)\colon \theta \in \Theta\}$ be any family of probability densities relativ to a σ-finite measure ν on a measure space $X,\mathcal{B}$. Assume

$$p_\theta(x) > 0 \qquad x \in X, \qquad \theta \in \Theta \qquad . \tag{1}$$

(This assumption is actually used only in Proposition 4A.11 and Theorem 4A.12 Let the action space, A, be a closed convex subset of Euclidean space. The loss function is $L\colon \Theta \times A \to [0, \infty)$. Assume $L(\theta, \cdot)$ is a lower semicontinuous function for each $\theta \in \Theta$. Assume also that

$$\lim_{||a||\to\infty} L(\theta, a) = \infty \quad , \qquad \theta \in \Theta \ . \tag{2}$$

(If A is a bounded set this is trivially satisfied.) If A is bounded let $A^* = A$; if A is unbounded let $A^* = A \cup \{\underset{\sim}{\iota}\}$ denote the one-point compactification of A. Extend the function $L(\theta, \cdot)$ to A^* by defining

$$L(\theta, \underset{\sim}{\iota}) = \infty \quad . \tag{3}$$

A randomized decision procedure on A^* is a kernel $\delta(\cdot|\cdot)$ for which

$$(4) \qquad \begin{aligned} &\delta(\cdot|x) \text{ is a Borel measure on } A^*, \text{ for } x \in X \\ &\delta(A|\cdot) \text{ is } \mathcal{B} \text{ measurable for each measurable set } A \subset A^* . \end{aligned}$$

A nonrandomized procedure is one for which $\delta(\cdot|x)$ is concentrated on a single point, $\delta(x)$, for almost every (ν), $x \in X$. Note we use the symbol δ both for the kernel $\delta(\cdot|\cdot)$ and for the related function $\delta(\cdot)$. Let $\mathcal{D}^*$ denote the collection of all randomized decision procedures. Let $\mathcal{D} \subset \mathcal{D}^*$ denote those giving mass 1 to $A \subset A^*$, and let $\mathcal{D}_n \subset \mathcal{D}$ denote the nonrandomized procedures in $\mathcal{D}$.

As usual, the risk of any procedure is

$$(5) \qquad R(\theta, \delta) = \int_X \int_{A^*} L(\theta, a)\, \delta(da|x)\, p_\theta(x)\nu(dx) \quad .$$

Note that $R(\theta, \delta)$ may take the value ∞. A procedure δ is admissible if

$$(6) \qquad R(\theta, \delta') \le R(\theta, \delta) \;\; \forall\; \theta \in \Theta \Rightarrow R(\theta, \delta') = R(\theta, \delta) \;\; \forall\; \theta \in \Theta \; .$$

The proof of the main result of the appendix is broken down into six main preliminary steps as follows:

(i) $\mathcal{D}^*$ is compact in an appropriate topology;

(ii) $R(\theta, \cdot)$ is lower semicontinuous on $\mathcal{D}^*$;

(iii) $\delta_i \to \delta_0$ with $\delta_i \in \mathcal{D}_n$, $i = 0,1,\ldots$, implies $\delta_i \to \delta_0$ in measure (ν);

(iv) the minimax Theorem for finite Θ;

(v) the closure of the Bayes procedures is a complete class; and

(vi) $\mathcal{D}_n$ is a complete class when $L(\theta, \cdot)$ is strictly convex.

Formal statements of all these results and some corollaries are given below. Complete proofs are also given for all but (i) for which the reader can consult the references cited below.

4A.2 Definitions

We now define the topology on $\mathcal{D}^*$. Let $L_1 = L_1(X, \mathcal{B}_X, \nu)$ denote the Banach space of ν integrable functions. Let C^* denote the (Banach) space of continuous, real-valued functions on the compact set A^*. For every $\delta \in \mathcal{D}^*$, $f \in L_1$, $c \in C^*$ there is a number

$$\beta_\delta(f, c) = \int\int c(a)\delta(da|x)\ f(x)\nu(dx) \quad .$$

Define the topology on $\mathcal{D}^*$ according to the convergence criterion $\delta_\alpha \to \delta$ if

$$(2) \qquad \beta_{\delta_\alpha}(f, c) \to \beta_\delta(f, c) \qquad f \in L_1, \quad c \in C^* \quad .$$

(This is a "weak" topology. The collection of sets of the following form comprise a basis for this topology:

$$\{\delta \in \mathcal{D}^*: |\beta_\delta(f_i, c_j) - \beta_{\delta_0}(f_i, c_j)| < \varepsilon, \quad 1 \le i \le I, \quad 1 \le j \le J,$$
$$\delta_0 \in \mathcal{D}^*, \quad f_i \in L_1, \quad c_j \in C^*, \quad i=1,\ldots,I, \quad j=1,\ldots,J,$$
$$\varepsilon > 0\ \}\ .)$$

<u>4A.3 Theorem</u>

$\mathcal{D}^*$ is compact in the topology defined above.

Proof. This theorem appears in Le Cam (1955) in a form similar to the above. In a somewhat more primitive form the result appears already in Wald (1950). It is interesting to note that this theorem is actually a special case of a result in abstract functional analysis. It follows directly from Theorems V.8.6 and IV.6.3 (the Riesz representation theorem) of the classic treatise of Dunford and Schwartz (1966). For a complete, detailed proof see Farrell (1966, Appendix). ||

As has already been noted in the text, Wald's book, and the paper of Le Cam, both cited above, continue from their versions of Theorem 4A.3 and prove results similar to most of those below; but they do not explicitly state a version of Theorem 4A.12 which is our ultimate goal. See especially Wald (1950, Sections 3.5 and 3.6) and Le Cam (1955, Theorem 3.4).

4A.4 Proposition

The map $R(\theta, \cdot): \mathcal{D}^* \to [0, \infty]$ is lower semi-continuous. In other words, if $\delta_\alpha \to \delta_0$ then

(1) $$\liminf R(\theta, \delta_\alpha) \geq R(\theta, \delta_0), \quad \theta \in \Theta \quad .$$

Proof. Let $\delta_\alpha \to \delta_0$. Let $\theta \in \Theta$ and let $c_B(\cdot) = \min(L(\theta, \cdot), B)$. Then $c_B \in C^*$ and, for any $\delta \in \mathcal{D}^*$, $\beta_\delta(p_\theta, c_B) \uparrow R(\theta, \delta)$ as $B \uparrow \infty$. Thus,

(2) $$\lim_\alpha \inf R(\theta, \delta_\alpha) \geq \lim_\alpha \inf \beta_{\delta_\alpha}(p_\theta, c_B)$$
$$= \beta_{\delta_0}(p_\theta, c_B) \underset{B\uparrow\infty}{\uparrow} R(\theta, \delta_0) \quad .$$

(1) follows directly from (2). ||

We will apply this proposition in roughly the following form:

4A.5 Corollary

Let $\{\theta_1,\ldots,\theta_m\} \subset \Theta$. Let $\Gamma_f \subset R^m$ be the set of available finite risk points -- i.e.

(1) $$\hat{\Gamma}_f = \{r \in R^m: \exists\ \delta \in \mathcal{D}^*,\ R(\theta_j, \delta) = r_j,\ j=1,\ldots,m\} \quad .$$

Let $\Gamma_f \subset R^m$ be the set of points dominated by Γ_f -- i.e.

(2) $$\hat{\Gamma}_f = \{r \in R^m: \exists\ s \in \Gamma_f,\ s \leq r\}$$

where (as usual) $s \leq r$ means $s_j \leq r_j$, $j=1,\ldots,m$. Then $\hat{\Gamma}_f$ is a non-empty, convex, closed subset of R^m.

Notation: In the current context, when $R(\theta_j, \delta) < \infty$, $j=1,\ldots,m$, we write $R(\cdot, \delta)$ to denote the point $r \in R^m$ with $r_j = R(\theta_j, \delta)$, $j=1,\ldots,m$.

Proof. $R(\theta, a_0) = L(\theta, a_0) < \infty$, $\theta \in \Theta$, so $\Gamma_f \neq \phi$, and consequently $\hat{\Gamma}_f \neq \phi$. Γ_f is convex, so also $\hat{\Gamma}_f$ is convex. Suppose $r_i \in \hat{\Gamma}_f$, $i=1,\ldots,$ and $r_i \to r$. Then there exist $\delta_i \in \mathcal{D}^*$ with $R(\cdot, \delta_i) \leq r$, $i=1,\ldots$. Since $\mathcal{D}^*$

is compact there must exist a subsequence $\{i'\}$ such that $\delta_{i'}$ is convergent; $\delta_{i'} \to \delta$. Then

(3) $$(r)_j = \lim(r_{i'})_j \geq \lim\inf R(\theta_j, \delta_{i'})$$
$$\geq R(\theta_j, \delta), \quad j = 1,\ldots,m,$$

by Proposition 4A.4. It follows that $r \in \hat{\Gamma}_f$. This proves that $\hat{\Gamma}_f$ is closed. ||

Here is another useful consequence of Theorem 4A.3 and Proposition 4A.4.

4A.6 Corollary

The set of admissible procedures forms a minimal complete class.

Proof. We give a proof only for the case where $\Theta = \{\theta_1,\ldots,\theta_m\}$ is finite. The corollary will be applied in this form in the proof of Theorem 4A.10. (The proof for general Θ is basically similar, but involves some form of Zorn's lemma. See, e.g. Brown (1977).)

Let $\delta_0 \in \mathcal{D}^*$. To each $\delta \in \mathcal{D}^*$ associate the point $r(\delta) = r \in [0, \infty]^m$ for which $r_j = R(\theta_j, \delta)$, $j = 1,\ldots,m$. Let

(1) $$\bar{\Gamma} = \{r \in [0, \infty]^m\colon \ r = r(\delta) \text{ for some } \delta \in \mathcal{D}^*\} .$$

(This is the same as 4A.5(1), except for the fact that here $r_j = \infty$ is possible so that $r(\delta)$ is defined for all $\delta \in \mathcal{D}^*$, not merely for those δ having finite risk.) Let $\underline{r} \in \bar{\Gamma}$ be a minimal point of $\bar{\Gamma}$ which dominates $r(\delta_0)$. That is,

(2) $$\underline{r} \leq r(\delta_0); \ r \leq \underline{r} \text{ and } r \neq r \Rightarrow r \notin \bar{\Gamma} .$$

(It is shown in Lemma 4A.8 that $\underline{r}_j < \infty$, $j = 1,\ldots,m$, but that fact is not essential here.) Such a point, $\underline{r}$, can be constructed as the limit of a sequence of points $r(\delta_i) \in \Gamma$.

By Theorem 4A.3 the sequence δ_i has an accumulation point in $\mathcal{D}^*$, say $\underset{\sim}{\delta}$. By Proposition 4A.4 $r(\underset{\sim}{\delta}) \le \underset{\sim}{r} \le r(\delta_0)$. Since $\underset{\sim}{r}$ was minimal it follows that $\underset{\sim}{\delta}$ is admissible.

It has thus been shown that any procedure, δ_0, is dominated by an admissible procedure, as asserted by the corollary. ||

4A.7 Proposition

Let $\delta_\alpha \in \mathcal{D}_n$, $\alpha = 1,\ldots,$ and suppose $\delta_\alpha \to \delta_0$ in $\mathcal{D}^*$ with $\delta_0 \in \mathcal{D}_n$. Then $\delta_\alpha(\cdot) \to \delta_0(\cdot)$ in measure (ν). Thus there is a subsequence i' for which $\delta_{i'}(\cdot) \to \delta_0(\cdot)$ a.e.(ν).

Proof. Suppose $\delta_\alpha \to \delta_0$ in $\mathcal{D}^*$ but $\delta_\alpha \not\to \delta_0$ in measure (ν). If $\delta_0 \in \mathcal{D}^*-\mathcal{D}$ then there is an $a_0 \in A$, an $\varepsilon > 0$, and a set S with $\nu(S) > 0$ such that

(1) $$|\delta_0(x) - a_0| < \varepsilon \quad \text{for all} \quad x \in S$$

and

(2) $$\limsup_\alpha \nu(\{x \in S: \ |\delta_\alpha(x) - a_0| > 2\varepsilon\}) > 0 .$$

(To verify (1) and (2) is a standard but nontrivial exercise in measure theory which uses the fact that A is a separable metric space. If $\delta_0 \in \mathcal{D}^*$ then (1) may need to be replaced by $|\delta_0(x)| > 1/\varepsilon$ and, correspondingly, the statement $|\delta_\alpha(x)| < 1/2\varepsilon$ would need to be substituted in (2). Similar substitutions would then need to be made in what follows.) Let $c \in C^*$ satisfy $0 \le c(\cdot) \le 1$, and

$$c(a) = \begin{cases} 1 & |a - a_0| \le \varepsilon \\ 0 & |a - a_0| \ge 2\varepsilon \end{cases},$$

and let $f(\cdot) = \chi_S(\cdot) \in L_1$. Then

(3) $$\beta_{\delta_0}(f, c) = \nu(S),$$

but

$$\beta_{\delta_0}(f, c) \le \nu(\{x: |\delta_\alpha(x) - a_0| < 2\varepsilon\})$$

$$= \nu(S) - \nu(\{x \in S: |\delta(x) - a_0| \ge 2\varepsilon\})$$

so that

$$(4) \qquad \liminf_\alpha \beta_{\delta_\alpha}(f,c) \le \nu(S) - \limsup_\alpha \nu(\{x \in S: |\delta_\alpha(x) - a_0| \ge 2\varepsilon\})$$

$$< \nu(S) .$$

Taken together (3) and (4) contradict the assumption that $\delta_\alpha \to \delta_0$ in $\mathcal{D}^*$. This contradiction shows that $\delta_\alpha \to \delta_0$ in $\mathcal{D}^*$ implies $\delta_\alpha \to \delta_0$ in measure (ν). The second conclusion of the proposition is a standard consequence of this. ||

We now come to the minimax theorem. In preparation for this theorem we prove a simple lemma.

4A.8 Lemma

Let Θ be finite. Then the set of procedures having finite risks is a complete class of procedures in $\mathcal{D}^*$. (In other words, for every $\delta \in \mathcal{D}^*$ there is a procedure $\delta' \in \mathcal{D}^*$ with $R(\theta, \delta') \le R(\theta, \delta)$ and $R(\theta, \delta') < \infty$, $\theta \in \Theta$.)

Proof. Let $a_0 \in A$ and $A_1 = \max\{L(\theta, a_0): \theta \in \Theta\} < \infty$. $B = \{a \in A: \min\{L(\theta, a): \theta \in \Theta\} \le A_1\}$. B is a bounded set because of Assumption 4A.1(2). Hence

$$A_2 = \max\{L(\theta, a): \theta \in \Theta , a \in B\} < \infty .$$

Define δ' to satisfy

$$(1) \qquad \delta'(\{a_0\}|x) = \delta(\{a_0\}|x) + \delta(B^c|x)$$

$$\delta'(A|x) = \delta(A|x) , \quad \{a_0\} \not\subset A, \quad A \subset B .$$

(In words, δ' takes action a_0 whenever δ takes an action outside B.) Then

$\delta'(B|x) \equiv 1$; hence $R(\theta, \delta') \le A_2 < \infty$. Also, by construction,

$$(2) \qquad \int L(\theta, a)\delta'(da|x) \le \int_B L(\theta, a)\delta(da|x) + L(\theta, a_0)\delta(B^c|x)$$

$$\le \int L(\theta, a)\delta(da|x) \quad .$$

Hence $R(\theta, \delta') \le R(\theta, \delta)$. ||

In the language of Corollary 4A.5, used below, the preceding can be interpreted as saying that the set of procedures with risk points in Γ_f is a complete class.

<u>4A.9 Theorem</u>

Let Θ be finite. Let $\delta_0 \in \mathcal{D}^*$ be any procedure for which $R(\cdot, \delta_0) \in \Gamma_f$, and such that

$$(1) \qquad R(\cdot, \delta_0) - \varepsilon \notin \hat{\Gamma}_f$$

for every $\varepsilon > 0$. Then δ_0 is Bayes -- i.e. there exists a prior G giving mass π_j to $\theta_j \in \Theta$, $j=1,\ldots,m$ such that

$$(2) \qquad \sum_{j=1}^{m} \pi_j R(\theta_j, \delta_0) \le \inf_{\delta\in\mathcal{D}^*} \sum_{j=1}^{m} \pi_j R(\theta_j, \delta) \quad .$$

Remark. The minimax risk -- $M = \inf_{\delta\in\mathcal{D}^*} \max \{R(\theta, \delta): \theta \in \Theta\}$ -- must be finite by Lemma 4A.8. (Also, as a consequence of Corollary 4A.5 there must exist a minimax procedure.) If δ_0 is any minimax procedure then it must satisfy (1) and hence must be Bayes. This does not yet prove that the resulting prior G is least favorable -- i.e. $\Sigma \pi_j R(\theta_j, \delta) \ge M$ for all $\delta \in \mathcal{D}^*$. Indeed, this need not be the case. To get a least favorable prior apply the proof of the theorem to the point with coordinates $r_j \equiv M$, $j=1,\ldots,m$. This point need not correspond to any procedure in $\mathcal{D}^*$, but it is in $\hat{\Gamma}$, and the proof of the theorem applies directly to yield $\{\pi_j\}$ such that $M = \sum_{j=1}^{m} \pi_j M \le \inf_{\delta\in\mathcal{D}^*} \sum_{j=1}^{m} \pi_j R(\theta_j, \delta)$. This $\{\pi_j\}$ corresponds to the least favorable distribution.

Proof. $\hat{\Gamma}_f$ is a closed convex subset of R^m by Corollary 4A.5. Condition (1) implies that the point $r_0 = R(\theta, \delta_0)$ lies on the boundary of $\hat{\Gamma}_f$. Hence there exists a nonzero vector $\{\alpha_j\}$ which defines a supporting hyperplane to $\hat{\Gamma}_f$ at r_0 -- i.e.

$$\Sigma\, \alpha_j (r_0)_j = \inf \{\Sigma \alpha_j r_j : \; r \in \hat{\Gamma}_f\} \quad . \tag{3}$$

Since $r_0 \in \hat{\Gamma}_f$, so also is $r_0 + ae_i$ for any unit vector e_i, and $a \geq 0$. Thus (3) yields

$$\alpha_i[(r_0)_i + a] \geq \alpha_i (r_0)_i, \qquad a \geq 0 \; . \tag{4}$$

It follows that $\alpha_i \geq 0$, $i=1,\ldots,m$. Let

$$\pi_i = \frac{\alpha_i}{\sum_{j=1}^{m} \alpha_j} \quad . \tag{5}$$

Then

$$\sum_{j=1}^{m} \pi_j (r_0)_j = \inf \{\Sigma\, \pi_j r_j : \; r \in \hat{\Gamma}_f\} \quad . \tag{6}$$

Furthermore, by Lemma 4A.8, for every $\delta \in \mathcal{D}^*$ there is an $r \in \hat{\Gamma}_f$ such that $r_j \leq R(\theta_j, \delta)$, $j=1,\ldots,m$; so that $\sum_{j=1}^{m} \pi_j r_j \leq \Sigma \pi_j R(\theta_j, \delta)$. The desired result, (2), now follows from (6). ||

<u>4A.10 Theorem</u>

Let B_0 denote the set of Bayes procedures for priors concentrated on finite subsets of Θ. Then $\bar{B}_0$, the closure of B_0 in $\mathcal{D}^*$, is an essentially complete class.

Proof. (Note: the following proof is written in the language of directed sets, nets, and subnets. See, e.g. Dunford and Schwartz (1966). The reader unfamiliar with these concepts, or the equivalent concept of filters and ultrafilters can understand the essence of the proof by considering the case where Θ is countable, for then the nets and subnets can be converted to ordinary sequences and subsequences. If X, $\mathcal{B}$ is Euclidean space -- as in the exponential family situation -- it can be shown by an auxiliary argument that sequences and subsequences also can suffice for the proof, since the topology of $\mathcal{D}^*$ has a countable basis.) Let δ_0 be any procedure.

Let A denote the collection of all finite subsets of Θ formed into a directed set under the obvious partial ordering: $\alpha_1 \leq \alpha_2$ if $\alpha_1 \subset \alpha_2$.

Consider a fixed $\alpha \in A$; $\alpha \subset \Theta$. Consider the statistical problem with parameter space just the finite set α. There must exist a procedure, call it δ_α, which is admissible in this restricted problem and is at least as good as δ_0 -- i.e.

(1) $$R(\theta, \delta_\alpha) \leq R(\theta, \delta_0) \qquad \theta \in \alpha \qquad .$$

Since δ_α is admissible in the restricted problem it satisfies condition 4A.9(1) there. (The existence of δ_α is guaranteed by Corollary 4A.6.) Hence δ_α is Bayes with respect to a prior G_α concentrated on the finite set $\alpha \subset \Theta$.

Let $A' = \{\alpha'\}$ be a (co-final) subnet of A and let $\delta \in \mathcal{D}^*$ be such that $\delta_{\alpha'} \to \delta$. (The existence of A' and δ follows from Theorem 4A.4 by standard topological arguments.) Let $\theta_0 \in \Theta$. Then $\alpha' \supset \{\theta_0\}$ for every α' far enough out in A'. Hence $R(\theta_0, \delta_{\alpha'}) \leq R(\theta_0, \delta_0)$ for any such α' and, by Proposition 4A.5,

$$R(\theta_0, \delta) \leq \liminf_{\alpha'} R(\theta_0, \delta_{\alpha'}) \leq R(\theta_0, \delta_0) \quad .$$

Since $\theta_0 \in \Theta$ is arbitrary, this proves that $\delta \in \bar{B}_0$ is as good as δ_0. Since $\delta_0 \in \mathcal{D}^*$ is also arbitrary this proves $\bar{B}_0$ is an essentially complete class. ||

So far we have not assumed that $L(\theta, \cdot)$ is strictly convex, as is the case in the applications in Chapter 4. We now add this assumption, which is required for the desired complete class theorem.

4A.11 Proposition

Assume

(1) $L(\theta, \cdot)$ is strictly convex on A for each $\theta \in \Theta$.

Let $\delta \in \mathcal{D}^*$, $\delta \notin \mathcal{D}_n$. Then there is a $\delta' \in \mathcal{D}_n$ such that

(2) $$R(\theta, \delta') \leq R(\theta, \delta) , \qquad \theta \in \Theta ,$$

with strict inequality for some $\theta_0 \in \Theta$. In particular, the procedures in $\mathcal{D}_n$ are a complete class.

Proof. If $\delta \in \mathcal{D}^*$ but $\delta \notin \mathcal{D}$ then $\nu(\{x: \delta(\{\underset{\sim}{\iota}\}|x)\}) > 0$. Hence $R(\theta, \delta) \equiv \infty$, $\theta \in \Theta$, by 4A.1(1). Let $a_0 \in A$ and let δ' be defined by $\delta'(x) \equiv a_0$. Then

(3) $$R(\theta, \delta') = L(\theta, a_0) < \infty = R(\theta, \delta) , \qquad \theta \in \Theta .$$

Now, suppose $\delta \in \mathcal{D}$ but $\delta \notin \mathcal{D}_n$. If $R(\theta, \delta) \equiv \infty$ then, again, $\delta'(x) \equiv a_0$ satisfies (3). So, assume $R(\theta_0, \delta') < \infty$ for some $\theta_0 \in \Theta$. Condition (1) and 4A.1(2) guarantees that for some $\varepsilon > 0$, $A_0 \geq 0$

(4) $$L(\theta_0, a) \geq \varepsilon ||a|| - A_0 .$$

(We leave this as an exercise on convex functions. A very similar result is proved in 5.3(3) and (5); see 5.3(4').)

Hence

(5) $$\infty > R(\theta_0, \delta) = \int(\int L(\theta, a)\delta(da|x))p_{\theta_0}(x)\, \nu(dx)$$

$$\geq \varepsilon\int(\int ||a||\, \delta(da|x))p_{\theta_0}(x)\nu(dx) - A_0 .$$

It follows that

(6) $$\int ||a||\, \delta(da|x) < \infty \quad \text{a.e.}(\nu)$$

since $p_{\theta_0}(x) > 0$ a.e.(ν) by 4A.1(1).

Define

(6) $$\delta'(x) = \begin{cases} \int a\delta(da|x) & \text{if} \quad \int ||a||\delta(da|x) < \infty \\ a_0 & \text{otherwise} \end{cases} .$$

Then

$$\int L(\theta, a)\delta(da|x) \leq L(\theta, \delta'(x)) \quad \text{a.e.}(\nu)$$

with strict inequality whenever $\delta(\cdot|x)$ is not concentrated on the point $\delta'(x)$. Since $\delta \notin \mathcal{D}_n$ this occurs with positive probability under ν -- and hence by 4A.1(1) under P_{θ_0}.

Consequently

(7) $$R(\theta, \delta') \leq R(\theta, \delta)$$

with strict inequality for $\theta_0 \in \Theta$. (In fact, there is strict inequality in (7) whenever $R(\theta,\delta) < \infty$.) ||

The desired result now follows as an easy consequence.

<u>4A.12 Theorem</u>

Assume 4A.11(1). Then the set of pointwise limits of sequences of procedures in B_0 is a complete class. (B_0 is defined in 4A.10.)

Proof. As a consequence of 4A.11(1), Jensen's inequality and 4A.1(1) every procedure in B_0 is non-randomized. Also, there cannot be two non-equivalent admissible procedures with equal risk functions, for if $\delta_1 \neq \delta_2$ then

(1) $$(R(\theta, \delta_1) + R(\theta, \delta_2))/2 \geq R(\theta, (\delta_1 + \delta_2)/2)$$

with strict inequality whenever the right-hand side is finite.

The theorem now follows as a direct consequence of Corollary 4A.6, Proposition 4A.11, Theorem 4A.10, and Proposition 4A.7. Here's how:

Because of Corollary 4A.6 there is a unique minimal complete class. It is contained in $\mathcal{D}_n$ by Proposition 4A.11 and in B_0 by Theorem 4A.10 and (1), above. If δ_0 is in this minimal complete class (i.e., if δ_0 is admissible) there is therefore a net $\delta_\alpha \in B_0 = \mathcal{D}_n$ such that $\delta_\alpha \to \delta_0 \in \mathcal{D}_n$ in the topology on $\mathcal{D}^*$. Then, by Proposition 4A.7, $\delta_\alpha(\cdot) \to \delta_0(\cdot)$ a.e.(ν) which is the desired condition. ||

4A.13 Generalizations

(i) Assumption 4A.1(1) and the strict convexity assumption 4A.11(1) are used in the proof of Theorem 4A.12 for only two purposes; namely, to guarantee that

(1) $$\delta \in B_0 \Rightarrow \delta \in \mathcal{D}_n ,$$

and that

(2) $$(\delta_1, \delta_2 \text{ admissible}; \; R(\theta, \delta_1) = R(\theta, \delta_2), \;\; \theta \in \Theta) \Rightarrow \delta_1 = \delta_2 .$$

If (1) and (2) can be established separately, as is the case in some of the applications in Section 7 to the theory of hypotheses tests, then the conclusion of Theorem 4A.12 remains valid without 4A.1(1) and 4A.11(1).

(ii) There is not much hope for something like the conclusion of Theorem 4A.12 unless (1) and (2) are satisfied. However, all the earlier results of this appendix, through Theorem 4A.10, remain valid without the assumptions 4A.1(1) and 4A.11(1) (or (1) and (2)).

(iii) The remaining assumption which can be relaxed without major alterations in the theory is the assumption 4A.1(2) on the loss function. If this assumption is replaced by

(3) $$\lim_{a \to \underset{\sim}{\iota}} L(\theta, a) = \sup \{L(\theta, a) : \; a \in A\}$$

and

(4) $$\sup \{L(\theta, a) : \; a \in A\} < \infty$$

then all results through Theorem 4A.10 remain valid with only a simple modification needed in the statement and proof of Lemma 4.8 to establish that the procedures in $\mathcal{D}$ are a complete class.

(iv) If (3) is valid but not (4) or 4A.1(2), then a peculiar situation may arise. The results through Corollary 4A.7 remain valid, but it is then possible that there may exist admissible procedures having $R(\theta, \delta) = \infty$ for some $\theta \in \Theta$. When this peculiarity occurs the minimax theorem is not valid in the strong form of Theorem 4A.9. (There may exist admissible minimax procedures satisfying 4A.9(1) for which no prior exists satisfying 4A.9(2).) A weaker form of Theorem 4A.9 is, however, valid. Its conclusion is that there exists a sequence of priors defined by $\{\pi^{(\ell)}, \ell=1,\ldots\}$ and corresponding Bayes procedures $\delta^{(\ell)}$ such that $R(\theta, \delta^{(\ell)}) \to R(\theta, \delta_0)$, $\theta \in \Theta$. (The most convenient proof I know of this fact proceeds in a somewhat roundabout fashion using a device found in Wald (1950).)

(v) Brown (1977) contains versions of Theorem 4A.3 and Proposition 4A.4 valid for some situations where it is useful to compactify A in some fashion other than the one point compactification, A*, used above; or where the loss L depends on the observed $x \in X$, as well as on Θ, A; or where the decision rules are restricted *a priori* to lie in some proper subset of $\mathcal{D}$. In many of these situations it is possible to proceed further and also establish the conclusion of Theorem 4A.10.

It is also possible to derive some satisfactory results in the (unusual) situation where A is not a Borel subset of Euclidean space, nor imbeddable as such a subset. Such an extension involves intricacies not present in the preceding treatment of the Euclidean case.

Exercises

4A.2.1

Suppose $A = \{a_0, a_1\}$, corresponding to a hypothesis testing problem. (a_0 = "accept", a_1 = "reject".) For any procedure δ let $\phi(x) = \phi_\delta(x) = \delta(\{a_1|x\})$ denote the critical function of the test. Then, $\delta_i \to \delta$ in the topology on $\mathcal{D}^*$ if and only if $\phi_{\delta_i} \to \phi_\delta$ in the weak* topology on L_∞ (i.e.

$$(1) \qquad \int |\phi_{\delta_i}(x) - \phi_\delta(x)| f(x)\nu(dx) \to 0$$

for every $f \in L_1$. (See e.g. Lehmann (1959, Section A4).)