The Annals of Statistics
1979, Vol. 7, No. 4, 860–868

A STEPWISE BAYESIAN PROCEDURE

BY FRANCIS C. HSUAN
Temple University

Ordinarily a Bayesian estimation procedure uses one prior distribution to obtain a unique estimation rule (its Bayes rule). From the decision theoretical point of view, this procedure can be regarded as a convenient way to obtain admissible decision rules. However, many intuitively appealing, admissible estimation rules cannot be obtained directly in this way. We propose a new mechanism, called the Stepwise Bayesian Procedure (SBP). When the parameter space contains only finitely-many points and the loss function is strictly convex, this SBP can be used to obtain every admissible estimation rule. A relationship between SBP and the limiting Bayes rules is given.

1. Introduction. There are many ways to classify a prior distribution in decision theory. We introduce the following classifications. (1) A prior distribution is said to be *regular* if it is supported on the entire parameter space (not to be confused with a *regular* measure); otherwise it is said to be *nonregular.* (2) A prior distribution is said to be of *type I* if its corresponding Bayes rule is unique up to the equivalence of risk function (see Ferguson, 1967, for the definition of such equivalence); otherwise it is said to be of *type II.* While a regular prior distribution must be of type I, a nonregular prior may or may not be of type II.

The type I priors (usually the regular ones), coupled with the Bayesian procedure, are widely used by statisticians as a tool to obtain admissible estimation rules (the Bayes estimators). Presumably no type II prior is used for this purpose because it is unable to pinpoint a particular estimation rule. Nevertheless, from the viewpoint of the decision theory a type II prior is just as important as a type I prior in that the Bayes rules with respect to all the type I priors alone do not form a complete class. Often an intuitively appealing, admissible estimator is Bayes only with respect to a type II prior. For example, in the usual binomial estimation problem with the squared error loss function, the maximum likelihood estimator is Bayes with respect to a prior distribution π if and only if π puts positive probability at each of 0 and 1, and none anywhere else (Johnson, 1971). This π can be shown as a type II prior. As a second example, consider the hypergeometric estimation problem discussed in Example 1 of Section 2. The moment estimator, the maximum likelihood estimator, and many nonrandomized admissible estimators share a common property that they are Bayes only with respect to a type II prior.

Of course, the reason that in practice no type II prior is used in Bayesian estimation is that it yields a collection of decision rules (its Bayes rules), rather than

Received May 1975; revised October 1977.
AMS 1970 *subject classifications.* Primary, 62C10; secondary, 62C07, 62F15.
Key words and phrases. Type-II priors, Bayes class, regular priors, sequence of regular priors, strictly convex loss function, minimal complete class.

just a single decision rule. Furthermore, this collection is usually a mixture of admissible as well as inadmissible decision rules. Thus it is desirable to have a mechanism by which the admissible rules can be extracted out of that collection. We now propose such a mechanism, called the Stepwise Bayesian Procedure (SBP). The idea is to apply the Bayesian procedure in a stepwise manner: at each step the Bayesian procedure is used to extract a subcollection from a collection of decision rules which was obtained in the earlier step. The procedure terminates when at a step the SBP results in a collection consisting of a single decision rule. It is shown that such a decision rule must be admissible (Theorem 1).

A general theory on the SBP is yet to be developed. But we shall show that in a decision problem with finite parameter space and strictly convex loss function, every admissible rule can be obtained by SBP in finitely many steps (Theorem 2). In a sense it is a constructive characterization of the minimal complete class (see Example 4).

The ordinary Bayesian procedure utilizes only one (type I) prior distribution in order to obtain a unique decision rule (its Bayes rule). But the SBP calls for a number of (type II) priors in order to obtain a unique decision rule (one prior is used at each step). A philosophical interpretation (in the context of empirical Bayesian approaches) of having more than one prior distribution may not be palpable, but the following mathematical result seems to provide some guidelines toward its usage. Consider a decision problem with a finite parameter space and a strictly convex loss function. Let $\{\pi_n\}$ be a sequence of regular priors. Then there exists a set of finitely many priors (usually of type II) such that the limiting Bayes rule (assume the existence) of $\{\pi_n\}$ is identical with the unique decision rule obtained from SBP by using the finite set of priors stepwisely (Theorem 3). In a sense it means that a set of *finitely* many type II priors is, in effect, equivalent to a set of *countably* many regular priors.

There are three main theorems and seven examples in this paper. Most techniques used in the proof of the theorems are commonly known, therefore only the ideas of proof are given for each theorem. The examples are an integral part of the whole work. Some of them contain new results that are interesting in themselves. A summary of these results follows.

1. We prove that the moment estimator in a hypergeometric distribution is admissible under the squared error loss function (Example 2).

2. A characterization of the minimal complete class, when the parameter space has finitely many points and the loss function is strictly convex, is given (Example 4).

3. We present an example to show that an explicit characterization (i.e., in terms of the values of decision functions) of the minimal complete class is sometimes possible (Example 6).

2. Assumptions and definitions. We follow the same notation as in Ferguson (1967). Consider a statistical decision problem that satisfies the following three conditions:

(A1) The parameter space Θ is a discrete set consisting of finitely many points.

(A2) The action space A is a compact and convex set in the Euclidean space.

(A3) For each $\theta \in \Theta$, the loss function $a \to L(\theta, a)$ is strictly convex over A.

As a practical example where (A1)–(A3) might hold, consider the problem of estimating the proportion of defective items (θ) in a batch of N items coming off an assembly line. Assume that a thorough inspection on every item is economically infeasible. Instead, a sample of n items is randomly drawn and the number of defectives found in the sample, X, is recorded. It is desired to estimate θ based on the observation of X. In this problem the parameter space is discrete, and has finitely many points ($\theta = 0, 1/N, \cdots, (N-1)/N, 1$). Since we are trying to estimate a proportion, it is reasonable to allow an estimator to take values between 0 and 1; and the marginal loss due to an imprecise estimation may be assumed to be an increasing function of the absolute difference between the true proportion and the estimate (e.g., a squared error loss function). Then (A1)–(A3) are satisfied.

We make the following definitions.

DEFINITION 1. Let D be a collection of decision rules. An element $d \in D$ is said to be *Bayes within D* with respect to π if, for the prior distribution π, the decision rule d minimizes the Bayes risk $r(\pi, d')$ over all d' in D.

DEFINITION 2. Let $(\pi_1, \cdots, \pi_k)$ be an ordered set of prior distributions that are mutually singular (i.e., with mutually exclusive supports). The Bayes class with respect to $(\pi_1, \cdots, \pi_k)$, which will be denoted as $D(\pi_1, \cdots, \pi_k)$, is defined inductively as follows: for $k = 1$, $D(\pi_1)$ is the collection of all Bayes rules with respect to π_1; for $k > 1$, $D(\pi_1, \cdots, \pi_k)$ is the set of all Bayes rules within $D(\pi_1, \cdots, \pi_{k-1})$ with respect to π_k.

DEFINITION 3. Let $D(\pi_1, \cdots, \pi_k)$ be the Bayes class with respect to $(\pi_1, \cdots, \pi_k)$. A decision rule in $D(\pi_1, \cdots, \pi_k)$ is called a *stepwise Bayes rule* with respect to $(\pi_1, \cdots, \pi_k)$. If $D(\pi_1, \cdots, \pi_k)$ consists of only one decision rule (up to the equivalence of risk functions), then this rule is called a *unique stepwise Bayes rule*.

Notice that ordinary Bayes rules can be treated as a special class of stepwise Bayes rules where $k = 1$. A few possible misconceptions about the stepwise Bayes rules are listed below. Example 1 serves to clarify these points. It also illustrates the kind of computations involved in deriving stepwise Bayes rules. We shall call this process the Stepwise Bayesian Procedure (SBP).

REMARK 1. A Bayes class is defined in terms of an ordered set of priors. A different ordering of these priors often results in a different Bayes class.

REMARK 2. A stepwise Bayes rule with respect to $(\pi_1, \cdots, \pi_k)$ is necessarily a Bayes rule with respect to π_1. But it need not be Bayes with respect to π_i, $i = 2, \cdots, k$.

REMARK 3. Let the support of π_i be ω_i, $i = 1, \cdots, k$. Then the fact that $\bigcup_{i=1}^{k}\omega_i = \Theta$ is a sufficient, but not necessary, condition to have the stepwise Bayes rule with respect to $(\pi_1, \cdots, \pi_k)$ unique.

EXAMPLE 1. Consider the foregoing estimation problem about the proportion of defective items coming off an assembly line. For illustration purposes we assume that $N = 3$ and $n = 2$. The number of defectives in the sample, X, follows the hypergeometric distribution with population size $N = 3$, subpopulation size $M = 3\theta$ (where θ stands for the proportion of defectives), and sample size $n = 2$. Thus the sample space $\mathcal{X} = \{0, 1, 2\}$, and the parameter space $\Theta = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. Assume that $A = [0, 1]$, and the loss function is $L(\theta, a) = (\theta - a)^2$. A decision rule $d : \mathcal{X} \to A$ is a 3-dimensional vector (d_0, d_1, d_2) in the Euclidean space, where d_i is the estimated value of θ when $x = i$. (We note that in this example it doesn't seem reasonable to allow d_i to take values other than 0, $\frac{1}{3}$, $\frac{2}{3}$ or 1. However, as the size of N gets large the parameter space becomes nearly continuous, and the assignment $A = [0, 1]$ might then be acceptable. Here we let $N = 3$ only for the matter of illustration).

Assume that π_1 is a prior distribution on Θ with probability $\frac{1}{2}$ at $\theta = 0$, and at $\theta = \frac{1}{3}$. We shall denote it by $\pi_1 = (\frac{1}{2}, \frac{1}{2}, 0, 0)$. Then it is straightforward to show that any decision rule d such that $d_0 = \frac{1}{12}$ and $d_1 = \frac{1}{3}$ is Bayes with respect to π_1. Hence $D(\pi_1) = \{d : d_0 = \frac{1}{12}, d_1 = \frac{1}{3}\}$. The risk function of a decision rule d in $D(\pi_1)$ is found to be

$$\begin{aligned} R_d(\theta) &= \frac{1}{144} \quad \text{if} \quad \theta = 0 \\ &= \frac{1}{48} \quad \text{if} \quad \theta = \frac{1}{3} \\ &= \frac{2}{27} + \frac{1}{3}\left(\frac{2}{3} - d_2\right)^2 \quad \text{if} \quad \theta = \frac{2}{3} \\ &= (1 - d_2)^2 \quad \text{if} \quad \theta = 1. \end{aligned}$$

Let $\pi_2 = (0, 0, 1, 0)$. Then the Bayes rule within $D(\pi_1)$ with respect to π_2 is the decision rule in $D(\pi_1)$ that minimizes the Bayes risk $r(\pi_2, d') = \frac{2}{27} + \frac{1}{3}(\frac{2}{3} - d_2)^2$ over all d' in $D(\pi_1)$. Thus the Bayes class $D(\pi_1, \pi_2)$ contains only one decision rule $d^* = (\frac{1}{12}, \frac{1}{3}, \frac{2}{3})$. This d^* is the unique stepwise Bayes rule with respect to (π_1, π_2).

Observe the following facts (and their implications):

(1) The union of the supports of π_1 and π_2 is not Θ (Remark 3).

(2) The decision rule d^* is not Bayes with respect to π_2, thus $D(\pi_1, \pi_2) \neq D(\pi_2, \pi_1)$ (Remarks 1 and 2).

(3) There exists no π such that d^* is the unique Bayes rule with respect to π. (It stands out only if more than one prior is used in a stepwise manner).

(4) Compare d^* with the decision rule $d^\circ = (\frac{1}{12}, \frac{1}{3}, 0)$. Both are Bayes with respect to π_1, but one appears reasonable while the other is not. (The ordinary Bayesian procedure fails to discriminate one from the other).

Consider the moment estimator $\bar{\theta} = (0, \frac{1}{2}, 1)$, the M.L.E. $\hat{\theta}_1 = (0, \frac{1}{3}, 1)$ and $\hat{\theta}_2 = (0, \frac{2}{3}, 1)$, and other reasonable estimators such as $\tilde{\theta} = (\frac{1}{3}, \frac{2}{3}, 1), (0, \frac{1}{3}, \frac{2}{3})$, etc. It is straightforward to show that none of these estimators is a unique Bayes rule in the ordinary sense (i.e., with respect to a type I prior); but all of them can be proved to be unique stepwise Bayes rules, and they are admissible (see Example 6).

3. Unique stepwise Bayes rules and admissible rules. We now state and prove some optimal properties of SBP.

THEOREM 1. *A unique stepwise Bayes rule is admissible.*

PROOF. Let d be the unique decision rule in a Bayes class $D(\pi_1, \cdots, \pi_k)$. If d is not admissible, then there exists d' such that $R(\theta, d') \leqslant R(\theta, d)$ for all $\theta \in \Theta$. This d' must be in $D(\pi_1)$, since $r(\pi_1, d') \leqslant r(\pi_1, d)$. But then this d' is also in $D(\pi_1, \pi_2)$, since $d' \in D(\pi_1)$ and $r(\pi_2, d') \leqslant r(\pi_2, d)$. Repeat this argument. Then this d' is in $D(\pi_1, \cdots, \pi_k)$, which is a contradiction. □

Notice that the validity of Theorem 1 is not contingent on assumptions about the parameter space or about the loss function. It offers a new technique to check the admissibility of a given decision rule. The following Example 2 illustrates this technique. It is interesting to compare this example with the argument given by Johnson (1971) to justify the admissibility of M.L.E. in binomial estimation problem. There is a close resemblance.

EXAMPLE 2. Let X be hypergeometrically distributed with parameters N (the population size), M (the subpopulation size) and n (the sample size). Let $\theta = M/N$, the unknown proportion of the subpopulation. Thus $\Theta = \{0, 1/N, \cdots, (N-1)/N, 1\}$ and $\mathcal{X} = \{0, \cdots, n\}$. Take $A = [0, 1]$ and $L(\theta, a) = (\theta - a)^2$. Then the moment estimator $\bar{\theta}(x) = x/n$ is admissible, since it is the unique stepwise Bayes rule with respect to (π_1, π_2), where

$$\begin{aligned}\pi_1(\theta) &= \frac{1}{2} && \text{if } \theta = 0, 1\\ &= 0 && \text{otherwise;}\end{aligned}$$

and

$$\begin{aligned}\pi_2(\theta) &= 0 && \text{if } \theta = 0, 1\\ &= c/\{\theta(1-\theta)\} && \text{otherwise.}\end{aligned}$$

(Here the value c is so chosen to make π_2 a probability distribution). A sketch of the proof follows.

First, by expanding in binomial series both sides of the following identity

$$(1-x)^{-(h+k+1)} + \frac{x}{h+k+1}\frac{d}{dx}(1-x)^{-(h+k+1)} = (1-x)^{-(h+k+2)}, \tag{3.1}$$

and by collecting all the corresponding coefficients of x^t, we can establish the following equality

$$(3.2)\qquad \Sigma_{i,j}\binom{h+i}{h}\binom{k+j}{k} = \frac{h+k+t+1}{h+k+1}\Sigma_{i,j}\binom{h+i-1}{h-1}\binom{k+j}{k},$$

where the summation is over all nonnegative integers i, j such that $i + j = t$.

Now a straightforward calculation shows that $D(\pi_1) = \{d : d_0 = 0, d_n = 1\}$, where d_i is the decision when $X = i$. The Bayes rule within $D(\pi_1)$ with respect to π_2 can then be shown as d:

$$d_x = \frac{x}{N}\Sigma'_m\binom{m}{x}\binom{N-m-1}{n-x-1}\Big/\Sigma'_m\binom{m-1}{x-1}\binom{N-m-1}{n-x-1},$$

where $x = 1, \cdots, n-1$. Here the summation is over all $m = x, \cdots, (N-n) + x$.

Set $i = m - x, j = N - n - m + x, t = N - n, h = x$ and $k = n - x - 1$ in equation (3.2), we can show that

$$\Sigma'_m\binom{m}{x}\binom{N-m-1}{n-x-1} = \frac{N}{n}\Sigma'_m\binom{m-1}{x-1}\binom{N-m-1}{n-x-1}.$$

Hence the Bayes rule within $D(\pi_1)$ with respect to π_2 is then $d : d_x = x/N \cdot N/n = x/n$, which is the moment estimator. By Theorem 1 it is admissible.

A very desirable property about SBP is stated in the following theorem.

THEOREM 2. *In a decision problem that satisfies* (A1)–(A3), *every admissible rule is unique stepwise Bayes.*

LEMMA 2.1. *Assume* (A1)–(A3). *Let π be a prior distribution and $\omega \subseteq \Theta$ be the support of π. If d_1, d_2 are both Bayes with respect to π, then $R_1(\theta) = R_2(\theta)$ for all $\theta \in \omega$.* (Here $R_i(\cdot)$ is the risk function of d_i).

PROOF. Let $T = \{x : f(x; \theta) > 0$ for some $\theta \in \omega\}$. Then we can write $R_i(\theta)$ as the sum of $R_{i1}(\theta) = E[L(\theta, d_i(X))I_T(X)]$ and $R_{i2}(\theta) = E[L(\theta, d_i(X))I_T^c(X)]$, where $I_T(x)$ is the index function of T and $I_T^c(x) = 1 - I_T(x)$. By definition $f(x; \theta) = 0$ for all $x \in T^c$ and $\theta \in \omega$. Thus $R_{12}(\theta) = R_{22}(\theta)$ for all $\theta \in \omega$. Due to the strict convexity of loss function it follows that $d_1(x) = d_2(x)$ for almost every $x \in T$. Hence $R_{11}(\theta) = R_{21}(\theta)$ for all $\theta \in \Theta$. Therefore $R_1(\theta) = R_2(\theta)$ for all $\theta \in \omega$. □

PROOF OF THEOREM 2. Assume that d is admissible. Then there exists a prior distribution π_1 such that d is Bayes with respect to π_1. Assume, without loss of generality, that the support of π_1 is $\omega_1 = \{\theta_1, \cdots, \theta_m\}$, and that $\Theta = \{\theta_1, \cdots, \theta_n\}$, where $m \leqslant n$. Let $\Gamma \subset R^n$ be the risk set of $D(\pi_1)$ and $\vec{\mathbf{x}} \in \Gamma$ be the risk point corresponding to the decision rule d. Then Lemma 2.1 implies that for every $\vec{\mathbf{y}} \in \Gamma$, we have $y_i = x_i$ for $i = 1, \cdots, m$. Thus it is convenient to consider the reduced risk set $\Gamma^* = \{(z_1, \cdots, z_{n-m})$: For some $\vec{\mathbf{y}} \in \Gamma$, $z_i = y_{m+i}$ for all $i = 1, \cdots, n-m\}$. This Γ^* is closed and convex. Since d is admissible, the reduced risk point z of d must be on the lower boundary of Γ^*. According to the supporting hyperplane theorem there exists a prior distribution π_2, with support $\omega_2 \subseteq \omega_1^c$, such that d is Bayes within $D(\pi_1)$ with respect to π_2. Repeat this argument until d is the only one decision rule within a Bayes class $D(\pi_1, \cdots, \pi_k)$. The value of k must not exceed n since $\pi_1, \cdots, \pi_k$ are mutually singular and the parameter space has only n points. Thus d is a unique stepwise Bayes rule. $\square$

COROLLARY 2.1. *In a decision problem that satisfies* (A1)–(A3), *the set of all unique stepwise Bayes rules is the minimal complete class.*

This corollary is related to a characterization theorem due to Wald and Wolfowitz (1951, Theorem 1). But the assumption of strict convexity of the loss function has made our approach more natural, and our results are more general.

A natural desire at this point is to ask for a sufficient condition on a set of mutually singular priors $\pi_1, \cdots, \pi_k$ so as to yield an unique stepwise Bayes rule. Assume a finite parameter space and a strictly convex loss function. Let the sample space be provided with a σ-finite measure ν with respect to which the density functions $\{f(x; \theta), \theta \in \Theta\}$ exist. Furthermore we assume this ν has the smallest possible support, so that $P_\theta(A) = 0$ for all $\theta \in \Theta$ implies $\nu(A) = 0$, where A is any Borel set in the sample space. (For example, one could take $\nu(A) = \Sigma\lambda_\theta P_\theta(A)$, with $\lambda_\theta > 0$ and $\Sigma\lambda_\theta = 1$). Then the unique stepwise Bayes rule with respect to $(\pi_1, \cdots, \pi_k)$ exists if and only if for almost every (ν) sample point x there exists a posterior distribution with respect to some prior in $\pi_1, \cdots, \pi_k$. Let the supports of π_i be ω_i, $i = 1, \cdots, k$, and $\omega = \cup_i\omega_i$. Then a posterior distribution exists at almost every point x if

$$\nu\{x : f(x; \theta) = 0 \text{ for all } \theta \in \omega\} = 0. \tag{3.3}$$

Notice that this condition (3.3) is imposed on the ω_i's rather than on the π_i's *per se*.

The following two examples illustrate the use of condition (3.3).

EXAMPLE 3. Consider the hypergeometric estimation problem in Example 2 with $N \geqslant 2n$. Let π_1 and π_2 be two priors with supports $\omega_1 = \{0, 1\}$, and $\omega_2 = \{n/N, \cdots, (N-n)/N\}$ respectively. Then there is a unique stepwise Bayes rule with respect to (π_1, π_2), even though $\omega_1 \cup \omega_2 \neq \Theta$.

EXAMPLE 4. Assume (A1), (A2) and $L(\theta, a) = (\theta - a)^2$. Let $(\pi_1, \cdots, \pi_k)$ be a set of mutually singular priors that satisfy condition (3.3). For every $x \in \mathfrak{X}$, let $i(x) = \min\{j : f(x; \theta) \neq 0 \text{ for some } \theta \in \omega_j, 1 \leqslant j \leqslant k\}$. Condition (3.3) assures the existence of such $i(x)$ for almost every x in $\mathfrak{X}$. Let $E(\theta | x, \pi)$ be the mean of the posterior distribution defined at point x and with respect to the prior distribution π, i.e., $E(\theta|x, \pi) = \Sigma_\theta \theta f(x; \theta)\pi(\theta)/\Sigma_\theta f(x; \theta)\pi(\theta)$. Then it can be shown that the unique stepwise Bayes rule with respect to $(\pi_1, \cdots, \pi_k)$ is

$$d : d_x = E(\theta | x, \pi_{i(x)}) \text{ for almost every } (\nu) x \text{ in } \mathfrak{X}. \tag{3.4}$$

Thus the minimal complete class consists of all such decision rules (3.4) with respect to all possible $(\pi_1, \cdots, \pi_k)$'s that satisfy condition (3.3).

4. Unique stepwise Bayes rules and limiting Bayes rules. Let $\{\pi_n\}$ be a sequence of *regular* priors defined on a finite parameter space. From $\{\pi_n\}$ an ordered set of priors $(\pi_{*1}, \cdots, \pi_{*k})$ can be constructed inductively as follows:

(1) When $i = 1$, the prior π_{*1} is an accumulation point of $\{\pi_n\}$.

(2) When $i > 1$, the prior π_{*i} is an accumulation point of a sequence of priors $\{\tau_n\}$, where $\tau_n(\theta_j) = 0$ if $\pi_{*r}(\theta_j) > 0$ for some $r = 1, \cdots, i-1$, and $\tau_n(\theta_j) = c_n \pi_n(\theta_j)$ otherwise. Here the constant c_n is defined so that τ_n is a probability measure on the parameter space Θ.

(3) The process terminates when there is no $\theta_j \in \Theta$ such that $\pi_{*r}(\theta_j) = 0$ for all $1 \leqslant r \leqslant k$.

Since there are only finitely many θ_j's in Θ and all the π_{*i}'s are mutually singular, the number k must not exceed the total number of parameters in Θ. In a sense the prior π_{*i} is an "ith order" accumulation point of $\{\pi_n\}$; as illustrated by the following example.

EXAMPLE 5. Assume that the parameter space consists of four points. Let $\pi_n = c_n(.7n, .3n, .6n^{\frac{1}{2}}, .4n^{\frac{1}{2}})$, where $c_n = 1/(n + n^{\frac{1}{2}})$. Then $\pi_{*1} = (.7, .3, 0, 0)$ and $\pi_{*2} = (0, 0, .6, .4)$.

Clearly the priors $\pi_{*1}, \cdots, \pi_{*k}$ thus defined are mutually singular and satisfy condition (3.3). Therefore the stepwise Bayes rule with respect to $\pi_{*1}, \cdots, \pi_{*k}$ is unique. Let it be d^*. Then d^* is related to $\{d_n\}$, the Bayes rules with respect to $\{\pi_n\}$, as described in the following theorem.

THEOREM 3. *Assume* (A1)–(A3). *Then $d^*(x)$ is an accumulation point of $\{d_n(x)\}$.*

THE IDEA OF PROOF. Let the posterior distribution with respect to π_n be denoted as $\pi_n(\cdot|x)$. Let x be a sample point and $i(x)$ be the smallest integer j such that $f(x, \theta) \neq 0$ for some $\theta \in \omega_j$. Then

$$\Sigma_{\theta \in \Theta} f(x; \theta)\pi_{*j}(\theta) = 0 \quad \text{if } j < i(x);$$
$$> 0 \quad \text{if } j = i(x).$$

Thus, there is a subsequence $\{\pi_m(\cdot|x)\}$ of $\{\pi_n(\cdot|x)\}$ such that $\pi_m(\theta|x) \to \pi_{*i(x)}(\theta|x)$ for every $\theta \in \Theta$.

Let $h_m(a) = \Sigma_\theta L(\theta, a)\pi_m(\theta|x)$ and $h(a) = \Sigma_\theta L(\theta, a)\pi_{*i(x)}(\theta|x)$. Both functions are strictly convex. By the definition of Bayes rules, $d_m(x)$ (or $d^*(x)$) is the unique point at which h_m (or h) attains its minimum. Under assumptions (A1)–(A3), the sequence of functions $\{h_m\}$ converges *uniformly* to the function h. Thus $d_m(x) \to d^*(x)$. $\square$

COROLLARY 3.1. *If the sequence of Bayes rules $\{d_n\}$ is pointwisely convergent, then it converges to d^*.*

COROLLARY 3.2. *Assume* (A1)–(A3). *Let D be the set of all* (*unique*) *Bayes rules with respect to the regular priors. Then the closure* (*under the topology of pointwise convergence*) *of D is the minimal complete class.*

Brown (1973, Theorem 9.3) has shown that under assumptions (A1) and (A2), the set of all risk points of D (as defined in Corollary 3.2) is dense (relative to the weak topology) in the set of all bounded admissible risk points. Our assumption (A3) has made Corollary 3.2 a stronger result.

EXAMPLE 6. Consider the decision problem discussed in Example 1. Let $\pi = (p_0, p_1, p_2, p_3)$ be a regular prior distribution. Then the Bayes rule with respect to π is

$$d : d_0 = \frac{1}{3}p_1/(3p_0 + p_1),$$

$$d_1 = \frac{1}{3}(p_1 + 2p_2)/(p_1 + p_2), \qquad \text{and}$$

$$d_2 = \frac{1}{3}(2p_2 + 9p_3)/(p_2 + 3p_3).$$

It is easy to see that D can be characterized as $\{d : \frac{1}{3}i < d_i < \frac{1}{3}(i + 1), i = 0, 1, 2\}$. Therefore the minimal complete class is $Cl(D) = \{d : \frac{1}{3}i \leqslant d_i \leqslant \frac{1}{3}(i + 1), i = 0, 1, 2\}$.

5. Conclusion. We introduce the idea of stepwise Bayesian procedure. In case of a finite parameter space and a strictly convex loss function, it leads to many fruitful results. In some occasions the condition of finiteness on the parameter space can be relaxed (see Hsuan, 1974, for examples). However, it should be pointed out that the usage of SBP becomes much more intricate when the parameter space is infinite than it would be otherwise. A general theory on this SBP is yet to be developed.

Part of the results in this paper is a reformulation of the work contained in the author's Ph.D. dissertation. The result in Section 4 is new. The author wishes to thank Professor L. D. Brown and Professor Lionel Weiss.

REFERENCES

[1] BROWN, L. D. (1973). *Notes on Decision Theory.* Unpublished class notes.

[2] FERGUSON, T. S. (1967). *Mathematical Statistics—A Decision Theoretical Approach.* Academic Press, New York.

[3] HSUAN, F. C. (1974). Characterization of the minimal complete class in statistical decision theory. Ph.D. dissertation, Cornell Univ.

[4] JOHNSON, B. M. (1971). On the admissible estimators for certain fixed sample binomial problems. *Ann. Math. Statist.* **42** 1579–1587.

[5] WALD, A. and WOLFOWITZ, J. (1951). Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite. *Proc. Second Berkeley Symp. Math. Statist. Probability* 149–157.

DEPARTMENT OF STATISTICS
TEMPLE UNIVERSITY
PHILADELPHIA, PENNSYLVANIA 19122