THE USE OF PREVIOUS EXPERIENCE IN REACHING STATISTICAL DECISIONS[1]

By J. L. Hodges, Jr., and E. L. Lehmann

University of California, Berkeley, and University of Chicago and Stanford University

1. Summary. Instead of minimizing the maximum risk it is proposed to restrict attention to decision procedures whose maximum risk does not exceed the minimax risk by more than a given amount. Subject to this restriction one may wish to minimize the average risk with respect to some guessed a priori distribution suggested by previous experience. It is shown how Wald's minimax theory can be modified to yield analogous results concerning such restricted Bayes solutions. A number of examples are discussed, and some extensions of the above criterion are briefly considered.

2. Introduction. Among various possible approaches to the problem of defining a best decision procedure we may mention the following two extremes.

(i) *The Bayes solution.* If the unknown parameter θ is a random variable distributed according to a known probability distribution λ, and if $R_\delta(\theta)$ denotes the risk function of the decision procedure δ, we simply minimize with respect to δ the average risk $\int R_\delta(\theta)\, d\lambda(\theta)$.

(ii) *The minimax principle.* Here one focuses attention on the maximum of the risk function and wishes to minimize $\sup_\theta R_\delta(\theta)$. The reader may consult Wald [1] for definitions and examples of these terms.

Of these two methods of treating the problem the first one assumes complete knowledge of the a priori distribution, an assumption that is usually not satisfied in practice. Even if extensive past experience is available, it will in most cases be difficult to exclude the possibility of some change in conditions. On the other hand, the minimax principle forces us to act as if θ were following a particular a priori distribution, the one least favorable to us, even though we may feel pretty sure that actually θ is distributed in quite a different manner. Thus it would seem that the minimax principle is suitable, if at all, only in situations characterized by a complete absence of past experience or other sources of knowledge concerning θ.

The situations occurring in practice usually lie between the two extremes just described. On the one hand, one does frequently have a good idea as to the range of θ, and as to which values in this range are more or less likely. On the other hand, such information cannot be expected to be either sufficiently precise or sufficiently reliable to justify complete trust in the Bayes approach.

The purpose of the present paper is to discuss an approach to the problem of

[1] Work sponsored in part by the Office of Naval Research.

optimal decisions that utilizes the available information but at the same time provides a safeguard in case this information is not correct.

Suppose that the maximum risk of the minimax procedure is $\underline{C}$. This is the smallest possible value for the maximum of a risk function. But we may be willing to tolerate a somewhat bigger maximum risk $C_0 > \underline{C}$ if, in case the guess at λ has been a good one, there results a substantial decrease in the average risk. This leads to the

DEFINITION. The procedure δ_0 is said to be a *restricted Bayes solution* with respect to the a priori distribution λ and subject to the restriction

$$R_\delta(\theta) \leqq C_0 \text{ for all } \theta, \tag{2.1}$$

if it minimizes $\int R_\delta(\theta)\, d\lambda(\theta)$ among all procedures satisfying (2.1).

This definition takes into account two aspects of the risk function, its supremum and its average with result to the distribution λ. Within a certain range of values each of these can be improved at the expense of the other. The proper balance between the two, that is, the value we select for C_0, depends on the confidence we have in θ's actually following a distribution close to λ and on the decrease in $\int R_\delta\, d\lambda$ that can be achieved by further increasing C_0.

To the above approach can also be given the following slightly different form. Instead of setting an upper bound for the risk, we specify a constant $0 \leqq \rho_0 \leqq 1$ and minimize

$$\rho_0 \int R_\delta(\theta)\, d\lambda(\theta) + (1 - \rho_0) \sup_\theta R_\delta(\theta). \tag{2.2}$$

Here it is ρ_0 that indicates the confidence we have in λ. The two principles are clearly equivalent; for if δ_0 minimizes (2.2) and if $\sup R_{\delta_0}(\theta) = C_0$, then δ_0 is a restricted Bayes solution, and the converse also holds.

The formulations given here may be applicable also to games played against an opponent rather than against Nature. This would be the case if one believed from past experience that the opponent is likely to make certain mistakes. One could then take advantage of these and still protect oneself in case the opponent has improved.

3. Restricted Bayes solutions. The principal aim of the present section is to obtain sufficient conditions for a decision procedure to be a restricted Bayes solution. For this purpose the modified problem mentioned at the end of the last section turns out to be the more natural one to consider; the results concerning restricted Bayes solutions follow as immediate corollaries. All of the theorems of this section will be simple generalizations of the corresponding results in Wald's minimax theory.

THEOREM 1. *Let ν_0 be a distribution for which there exists a constant* $0 < \rho_0 \leqq 1$

and a distribution μ_0 *such that*

$$\nu_0 = \rho_0\lambda + (1 - \rho_0)\mu_0 . \tag{3.1}$$

Then if the Bayes solution δ_0 *of* ν_0 *satisfies*

$$\int R_{\delta_0}(\theta)\, d\mu_0(\theta) = \sup_\theta R_{\delta_0}(\theta), \tag{3.2}$$

the procedure δ_0 *minimizes*

$$\rho_0 \int R_\delta(\theta)\, d\lambda(\theta) + (1 - \rho_0) \sup R_\delta(\theta). \tag{3.3}$$

PROOF. Let δ be any procedure. Then

$$\rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \sup R_\delta(\theta) \geqq \rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \int R_\delta\, d\mu_0$$

$$\geqq \rho_0 \int R_{\delta_0}\, d\lambda + (1 - \rho_0) \int R_{\delta_0}\, d\mu_0 = \rho_0 \int R_{\delta_0}\, d\lambda + (1 - \rho_0) \sup R_{\delta_0}(\theta).$$

COROLLARY. *Under the assumptions of Theorem* 1 *suppose that* $\sup R_{\delta_0}(\theta) = C_0$. *Then* δ_0 *minimizes* $\int R_\delta\, d\lambda$ *among all procedures satisfying* $R_\delta(\theta) \leqq C_0$ *for all* θ.

Since a distribution ν_0 with the required properties does not always exist, we state the following generalization of Theorem 1. *If* ν_i *is a sequence of a priori distributions with* $\nu_i = \rho_0\lambda + (1 - \rho_0)\mu_i$, *and if* δ_i *are the associated Bayes solutions, then*

$$\lim_{i\to\infty} \int R_{\delta_i}\, d\mu_i = \sup R_{\delta_0}(\theta) \tag{3.4}$$

is a sufficient condition for δ_0 *to minimize* (3.3).

Before proceeding with this development let us discuss briefly the decomposition (3.1). For a given pair of distributions λ, ν consider the totality of numbers $0 \leqq \rho \leqq 1$ such that $\nu = \rho\lambda + (1 - \rho)\mu$ for some μ. It is easily seen that this set is a closed interval $0 \leqq \rho \leqq \rho_\lambda$, and we shall call ρ_λ the λ-*component* of ν. It is of interest to note that under the conditions of Theorem 1 we have $\rho_0 = \rho_\lambda$ unless C_0 is the minimax risk. For if ρ_0 is not the λ-component of ν_0 it follows that μ_0 must have a positive λ-component. Thus any point of increase of λ is also one of μ_0 and we have $\int R_{\delta_0}\, d\lambda = C_0$. But by Theorem 1 this is possible only if C_0 is the minimax risk.

As in the minimax theory the distribution ν_0 of Theorem 1 is "least favorable" for the statistician, in a sense made precise in

THEOREM 2. *Under the assumptions of Theorem* 1, *the distribution* ν_0 *maximizes the Bayes risk among all a priori distributions* ν *that permit a representation*

$$\nu = \rho\lambda + (1 - \rho)\mu \text{ with } \rho \geqq \rho_0 .$$

PROOF. Let $\nu = \rho\lambda + (1 - \rho)\mu$ be any distribution with $\rho \geqq \rho_0$. Then

$$\int R_{\delta_0}\, d\nu_0 = \rho_0 \int R_{\delta_0}\, d\lambda + (1 - \rho_0) \sup R_{\delta_0}(\theta) \geqq \rho \int R_{\delta_0}\, d\lambda$$

$$+ (1 - \rho) \sup R_{\delta_0}(\theta) \geqq \rho \int R_{\delta_0}\, d\lambda + (1 - \rho) \int R_{\delta_0}\, d\mu$$

$$\geqq \int R_{\delta_\nu}(\theta)\, d\nu(\theta).$$

Thus δ_0 is a maximum solution for Nature (in Wald's interpretation of a decision problem as a two-person zero sum game whose players are Nature and the statistician), if she is restricted to a priori distributions whose λ-component is at least ρ_0.

From the proof of Theorem 2 it is seen that ν_0 not only maximizes the Bayes risk but also—and this is a slightly stronger result—the restricted Bayes risk, that is, the quantity $\inf \int R_\delta\, d\lambda$ when δ is restricted by the condition $R_\delta(\theta) \leqq C_0$ for all θ.

In Theorem 1 we gave sufficient conditions for a procedure δ_0 to be a restricted Bayes solution, and the distribution ν_0 in terms of which these conditions were formulated was further characterized in Theorem 2. We must still prove the existence of a distribution with the desired properties, at least for some class of decision problems. This is easily done, along the lines of Wald's proof of Theorem 3.10 of [1], under the following assumptions.

ASSUMPTION 1.

$$\inf_\delta \sup_\mu \left[\rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \int R_\delta\, d\mu\right]$$

$$= \sup_\mu \inf_\delta \left[\rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \int R_\delta\, d\mu\right].$$

This states that the decision problem is strictly determined when Nature is restricted to distributions with λ-components $\geqq \rho_0$.

ASSUMPTION 2. *There exists a least favorable distribution* μ_0, *that is a distribution that maximizes*

$$\inf_\delta \left[\rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \int R_\delta\, d\mu\right].$$

ASSUMPTION 3. *There exists a decision procedure* δ_0 *that minimizes*

$$\sup_\mu \left[\rho_0 \int R_\delta\, d\lambda + (1 - \rho_0) \int R_\delta\, d\mu\right].$$

It follows from Wald's work that 1 and 3 hold, for example, when Conditions 3.1–3.6 of [1] are satisfied, while in general 2 requires the stronger Condition

(3.7). All these conditions are satisfied, for example, when we are dealing with discrete or absolutely continuous distributions, the problem is nonsequential, the loss function is bounded and the parameter space is compact.

To prove our result, we note that

$$\sup_{\mu}\left[\rho_0 \int R_\delta \, d\lambda + (1 - \rho_0) \int R_\delta \, d\mu\right] = \rho_0 \int R_\delta \, d\lambda + (1 - \rho_0) \sup_{\theta} R_\delta (\theta).$$

Therefore, we have by assumption

$$\rho_0 \int R_{\delta_0} \, d\lambda + (1 - \rho_0) \sup R_{\delta_0}(\theta) = \inf_{\delta}\left[\rho_0 \int R_\delta \, d\lambda + (1 - \rho_0) \int R_\delta \, d\mu_0\right],$$

and hence

$$\rho_0 \int R_{\delta_0} \, d\lambda + (1 - \rho_0) \sup R_{\delta_0}(\theta) = \rho_0 \int R_{\delta_0} \, d\lambda + (1 - \rho_0) \int R_{\delta_0} \, d\mu_0.$$

But this implies, when $\rho_0 < 1$, that

$$\sup R_{\delta_0} (\theta) = \int R_{\delta_0} \, d\mu_0.$$

On the other hand, the result is true vacuously when $\rho_0 = 1$.

4. A continuity theorem. We shall next consider the dependence of the restricted Bayes solution on the restricting quantity C_0. The main result will be a continuity theorem that permits us, at least in some cases, to characterize the topological structure of the restricted Bayes solutions.

Throughout this section we shall make the following two assumptions whose validity was proved by Wald (Theorems 3.1 and 3.2a of [1]) under very general conditions.

ASSUMPTION A. *The space of decision functions is compact, that is, every sequence δ_i of decision procedures possesses a subsequence δ_{i_j} that converges to some decision function δ^*. Here convergence means what Wald calls "regular convergence"* (see pp. 65–66 of [1]).

ASSUMPTION B. *If $\delta_i \to \delta^*$, then for every distribution ν*

$$\int R_{\delta_i} \, d\nu \to \int R_{\delta^*} \, d\nu.$$

This convergence of integrals implies the pointwise convergence of the risk functions, as may be seen by letting the distributions ν degenerate at single points θ. In some cases (for example, if the loss function is bounded) the two notions of convergence are equivalent.

Let $\bar{C} = \sup R_{\delta_\lambda}(\theta)$ be the maximum risk of the unrestricted Bayes solution (for simplicity assume that δ_λ is unique; otherwise we would put $\bar{C} = \inf_{\delta_\lambda} \sup_\theta R_{\delta_\lambda}(\theta)$, the inf taken with respect to all Bayes solutions), and let $\underline{C}$ be the maximum risk of the minimax procedure. Then $\underline{C} \leqq \bar{C}$ and we may exclude the case $\underline{C} = \bar{C}$ as trivial. For any $\underline{C} \leqq C \leqq \bar{C}$ let $\delta_\lambda(C)$ be an associated restricted Bayes solution, and $r_\lambda(C) = \int R_{\delta_\lambda(C)}\, d\lambda$ the corresponding restricted Bayes risk. Concerning $r_\lambda(C)$ we have the

LEMMA. *The function $r_\lambda(C)$ is convex, continuous, and strictly decreasing.*

PROOF.

(i) $r_\lambda(C)$ is obviously nonincreasing. If $C' < C''$, $0 < \gamma < 1$ and

$$\delta = \gamma\delta_\lambda(C') + (1 - \gamma)\delta_\lambda(C''),$$

then

$$\int R_\delta\, d\lambda = \gamma r_\lambda(C') + (1 - \gamma)\, r_\lambda(C'')$$

and

$$\sup R_\delta(\theta) \leqq \gamma C' + (1 - \gamma)C'',$$

from which the convexity of $r_\lambda(C)$ follows.

(ii) From well known properties of convex functions, the continuity of $r_\lambda(C)$ is now obvious except at the point $C = \underline{C}$. But let $C_i \downarrow \underline{C}$, let δ_i be the corresponding restricted Bayes solutions and δ_{i_j} the convergent subsequence guaranteed by Assumption A. If $\delta_{i_j} \to \delta^*$, we have $r_\lambda(C_i) = \int R_{\delta_{i_j}}\, d\lambda \to \int R_{\delta^*}\, d\lambda$. By monotonicity $r_\lambda(C) \geqq \int R_{\delta^*}\, d\lambda$. Also, since $R_{\delta_{i_j}}(\theta) \to R_{\delta^*}(\theta)$ for each θ, we have that $\sup R_{\delta^*}(\theta) \leqq \lim C_{i_j} = \underline{C}$, so that $\sup R_{\delta^*}(\theta) = \underline{C}$. But this implies $r_\lambda(C) \leqq \int R_{\delta^*}\, d\lambda$ and hence $r_\lambda(C) = \int R_{\delta^*}\, d\lambda = \lim r_\lambda(C_{i_j})$.

(iii) Before proving that $r_\lambda(C)$ is strictly decreasing we shall now show that for any $\underline{C} < C < \bar{C}$ we have

$$\sup_\theta R_{\delta_\lambda(C)}(\theta) = C.$$

Suppose that $\sup R_{\delta_\lambda}(\theta) < C$. If $\bar{C} < \infty$ there exists $0 < \gamma < 1$ so that $\delta = \gamma\delta_\gamma + (1 - \gamma)\delta_\lambda(C)$ still satisfies $\sup R_\delta(\theta) \leqq C$, but we would have $\int R_\delta\, d_\lambda < \int R_{\delta_\lambda(C)}\, d\lambda$. If $\bar{C} = \infty$, then $C_i \to \bar{C}$ implies that $\int R_{\delta_\lambda(C_i)}\, d\lambda \to \int R_{\delta_\lambda(\bar{C})}\, d\lambda$ and the same argument applies.

(iv) Strict monotonicity of $r_\lambda(C)$ is an obvious consequence of (iii). For let $C' < C''$ and suppose that $r_\lambda(C') = r_\lambda(C'')$. Then $\delta_\lambda(C')$ would be a solution not only corresponding to C' but also to C'' in contradiction to (iii).

We can now state a closure and continuity theorem.

THEOREM 3. *If $\{\delta_i\}$ is a sequence of restricted Bayes solutions converging regularly to δ^*, then δ^* is a restricted Bayes solution, and $\sup_\theta R_{\delta^*}(\theta) = \lim_{i\to\infty} \sup_\theta R_{\delta_i}(\theta)$.*

PROOF. Let r_i, r^* be the Bayes risks of δ_i, δ^* respectively. By Assumption

B, $r_i \to r^*$. By the lemma, we can conclude that the sequence $C_i = \sup_\theta R_{\delta_i}(\theta)$ is convergent; denote $\lim C_i$ by C_0. Clearly $C^* = \sup R_{\delta^*}(\theta) \geqq C_0$, and since for each θ $R_{\delta_i^*}(\theta) = \lim R_{\delta_i}(\theta) \leqq C_0$, we conclude $C^* \leqq C_0 = C^*$. The lemma now assures δ^* to be a restricted Bayes solution.

COROLLARY. *The set of risk functions corresponding to restricted Bayes solutions is closed with respect to the convergence of Assumption B.*

PROOF. Let R_i be the risk functions corresponding to restricted Bayes solutions δ_i, and let $R_i \to R^*$. By the compactness Assumption A, we may extract a subsequence δ_{i_j} which converges regularly to some δ^*. By Theorem 3, δ^* is a restricted Bayes solution, whose risk function R^* is the limit of $R_{\delta_{i_j}}$, and hence of R_{δ_i}.

5. Extensions. We shall now mention briefly some extensions of the notion of a restricted Bayes solution. To give a first simple example, it may happen that we have ideas concerning the range of the parameter but not concerning a possible distribution over this range. Thus it may be known that $\theta \varepsilon \Omega$ and may further be indicated by other considerations that actually $\theta \varepsilon \omega$ where $\omega \subset \Omega$. We may then, subject to the condition $R_\delta(\theta) \leqq C_0$ for all $\theta \varepsilon \Omega$, wish to minimize $\sup_{\theta \varepsilon \omega} R_\delta(\theta)$. For example, when testing the hypothesis that three means θ_1, θ_2, θ_3 are equal we may believe that the most likely alternatives are such that $\theta_1 < \theta_2 < \theta_3$ without however definitely being able to exclude the other possibilities.

We can get further refinements as follows. Let $\Lambda_0 \subset \Lambda_1 \subset \cdots \subset \Lambda_{r-1} \subset \Lambda_r$ be a nested set of families of distributions. We may be certain that the true distribution of θ is an element of Λ_r (Λ_r may of course contain all distributions that assign probability 1 to a single point), fairly sure that it lies in Λ_{r-1}, still believe that is in Λ_{r-2}, etc. Accordingly we could select a decreasing sequence of constants $C_r > C_{r-1} > \cdots > C_1$, each C_i greater than the restricted minimax risk at this stage, and minimize $\sup_{\lambda \varepsilon \Lambda_i} \int R_\delta \, d\lambda$ subject to the condition $\sup_{\lambda \varepsilon \Lambda_i} \int R_\delta \, d\lambda \leqq C_i$ for $i = 1, \cdots, r$.

There is an extension of Theorem 1 to the present case that we give as an example of how the theory of the earlier sections generalizes.

THEOREM 4. *Suppose there exists $\rho_0 > 0$ and a distribution*

$$\nu_0 = \rho_0 \lambda_0 + \sum_{i=1}^{r} \rho_i \lambda_i, \qquad \rho_i \geqq 0, \qquad \sum_{i=0}^{r} \rho_i = 1,$$

with $\lambda_i \varepsilon \Lambda_i$ and such that the Bayes solution δ_0 of ν_0 satisfies

$$\sup_{\lambda \varepsilon \Lambda_i} \int R_{\delta_0} \, d\lambda = C_i = \int R_{\delta_0} \, d\lambda_i \qquad \textit{for } i = 1, \cdots, r \tag{1}$$

and

$$\sup_{\lambda \varepsilon \Lambda_0} \int R_{\delta_0} \, d\lambda = \int R_{\delta_0} \, d_{\lambda_0}. \tag{2}$$

Then δ_0 *minimizes* $\sup_{\lambda\varepsilon\Lambda_0} \int R_\delta \, d\lambda$ *subject to*

$$\sup_{\lambda\varepsilon\Lambda_i} \int R_\delta \, d\lambda \leqq C_i \qquad \textit{for } i = 1, \cdots, r. \tag{3}$$

Proof. Let δ be any procedure satisfying (3). Then

$$\sum_{i=0}^{r} \rho_i \sup_{\lambda\varepsilon\Lambda_i} \int R_\delta \, d\lambda \geqq \sum \rho_i \int R_\delta \, d\lambda_i \geqq \sum \rho_i \int R_{\delta_0} \, d\lambda_i = \sum \rho_i \sup_{\lambda\varepsilon\Lambda_i} \int R_\delta \, d\lambda.$$

But for $i = 1, \cdots, r$ we have

$$\sup_{\lambda\varepsilon\Lambda_i} \int R_\delta \, d\lambda \leqq C_i = \sup_{\lambda\varepsilon\Lambda_i} \int R_{\delta_0} \, d\lambda$$

and the conclusion follows.

Of the conditions, (2) of course becomes vacuous in case Λ_0 contains only a single distribution.

6. Examples. We conclude the paper by discussing a number of examples which serve to illustrate the ideas and theorems. As might be expected, it is more difficult to obtain explicit results with the restricted Bayes approach than with the logically simpler minimax or Bayes principles. In fact, the minimax principle has been successfully applied, in most cases, by guessing the answer and then verifying it through use of the specialized form of Theorem 1. It appears that the restricted Bayes solutions are often much less simple in their mathematical structure than the minimax solutions, and will accordingly be much harder to guess. We suspect that the widespread application of the restricted Bayes approach would require numerical methods in combination with theoretical results of the kind given above.

Example 1. Let $X = (X_1, X_2, \cdots, X_n)$ be random variables having the joint density function $p_\theta(x) = c(\theta)b(x)e^{t(x)q(\theta)}$, where θ is an unknown real parameter and $q(\theta)$ is a monotonely increasing function of θ. This form includes many of the distributions frequently used in statistics, such as the Poisson, binomial, and normal with known mean or variance. Consider the problem of testing, at a given level of significance, the hypothesis $H:\theta = \theta_0$ against the alternatives $\theta \neq \theta_0$. We are indifferent to alternatives $\theta_1 \leqq \theta \leqq \theta_2$, where $\theta_1 < \theta_0 < \theta_2$. As risk function, we shall use the probability of false acceptance. Thus, the risk is $1 - \beta(\theta)$, where $\beta(\theta)$ is the power function. Suppose finally that past experience or other considerations suggest that, should the hypothesis $\theta = \theta_0$ be false, then one of the alternatives $\theta \geqq \theta_2$ is true.

It is known [2] that an essentially complete class of tests consists in those of the form $w(k)$: reject H if $t(x) \leqq k$ or $t(x) \geqq f(k)$, where $f(k)$ is determined by the given level of significance. We may restrict attention to these tests. Denote the power of $w(k)$ at alternative θ by $\beta_k(\theta)$. If we had complete confidence in the presumption that either $\theta = \theta_0$ or $\theta \geqq \theta_2$, we should seek to maximize the power against the latter alternatives, with no regard to the power against the alternatives $\theta \leqq \theta_1$.

This leads to the known uniformly most powerful one-sided test $w(-\infty)$. This test is obtainable as the Bayes test corresponding to any a priori distribution for θ which assigns all of the probability to values $\theta \geqq \theta_2$. However, the power of $w(-\infty)$ tends to 0 as θ tends to its lower limit, so we may get very poor performance if our presumption is not correct. At the other extreme, if we placed no reliance on the presumption, we might seek the minimax test. It is easy to see that the power functions all have unique minima, and are continuous in k for fixed θ, whence the minimax test will be $w(\underline{k})$, where $\underline{k}$ is determined by the condition $\beta_{\underline{k}}(\theta_1) = \beta_{\underline{k}}(\theta_2)$. This test then gives no better protection against large alternatives than against small, and thus makes no use at all of our presumption that large alternatives are the ones to fear.

The restricted Bayes approach suggests a compromise, under which we would seek to maximize $\inf_{\theta \geq \theta_2} \beta(\theta)$ subject to $\inf_\theta \beta(\theta) \geqq C$. It can be shown that, for $-\infty < k \leqq \underline{k}$, $\inf_{\theta \geq \theta_2} \beta_k(\theta) = \beta_k(\theta_2)$, while $\inf_\theta \beta_k(\theta)$ is attained at a finite value of $\theta \leqq \theta_1$. From these facts the Lemma of [2] enables us to conclude, for $k' < k < k''$, $k \leqq \underline{k}$, that $\inf_{\theta \geq \theta_2} \beta_k(\theta) > \inf_{\theta \geq \theta_2} \beta_{k''}(\theta)$, while $\inf_\theta \beta_k(\theta) > \inf_\theta \beta_{k'}(\theta)$. It follows that for every $-\infty < k \leqq \underline{k}$, $w(k)$ is a restricted Bayes solution in the sense given. The same result holds if we assign to θ, any a priori distribution under which $P(\theta \geqq \theta_2) = 1$, and apply the restricted Bayes principle in the narrower sense of Section 2.

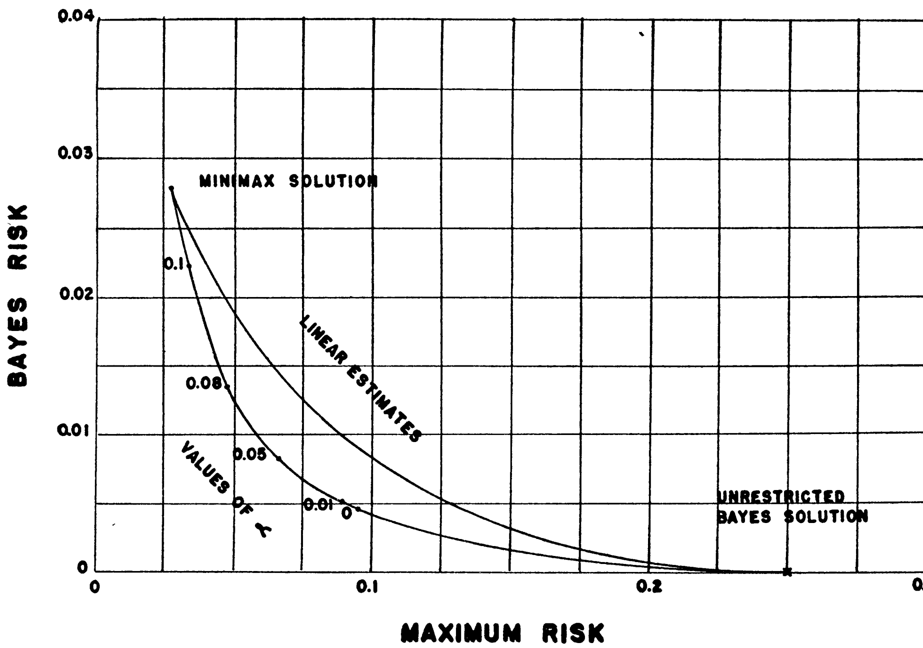
It is easily seen that $C(k) = \inf_\theta \beta_k(\theta)$ is a continuous, monotonely increasing function of k in the interval $-\infty \leqq k \leqq \underline{k}$. The tests $w(k)$, $k \leqq \underline{k}$, thus provide essentially unique restricted Bayes solutions for all possible values of C. In this example, our principle provides nothing essentially new, since the admissible tests already form only a one-parameter family.

EXAMPLE 2. Consider next the binomial random variable X = number of successes on n independent trials, each having the probability p of success, and the problem of estimating p from X. Take as loss function the square of the error of estimate. Suppose that a theoretical examination of the experimental situation reveals considerations of symmetry which suggest that $p = \frac{1}{2}$. We wish to design an estimate which will take advantage of this theory, in the sense of providing small risk at $p = \frac{1}{2}$, but will not place complete reliance in the theory, in the sense of giving some control over the risk for all values of p. This problem falls within the framework of the restricted Bayes approach, if we give to p the a priori distribution λ which assigns all of its probability to the value $\frac{1}{2}$. Our objective is then to minimize the risk at $\frac{1}{2}$, subject to a given maximum C for the risk over $0 \leqq p \leqq 1$.

Denote the estimate corresponding to a value x of X by $\delta(x)$, and the risk involved in using this estimate by $R_\delta(p) = E[\delta(X) - p]^2$. The ordinary Bayes estimate is $\bar{\delta}(x) = \frac{1}{2}$ for every x, since this estimate reduces the risk at $\frac{1}{2}$ to 0, while every other estimate has positive risk at $\frac{1}{2}$. The maximum risk of $\bar{\delta}$ is $\bar{C} = \frac{1}{4}$, attained at $p = 0$ and $p = 1$. The minimax estimate for this problem is known (see [3]) to be $\underline{\delta}(x) = \dfrac{\sqrt{n}}{\sqrt{n}+1}\dfrac{X}{n} + \dfrac{1}{2(\sqrt{n}+1)}$. This estimate has the constant risk $\underline{C} = \frac{1}{4}(\sqrt{n}+1)^2$.

In getting the restricted Bayes estimates, we first observe that we need only consider those estimates which possess the symmetry property $\delta(x) + \delta(n - x) = 1$. For, if $\delta(x)$ is any estimate, the estimate $\delta'(x) = 1 - \delta(n - x)$ will have the same maximum risk and the same Bayes risk as δ, for any a priori distribution which is symmetric about $p = 1/2$. Thus, the estimate $\frac{1}{2}(\delta(x) + \delta'(x))$, which possesses the symmetry property cited, will have the same Bayes risk as δ and no larger maximum risk.

Consider the distribution μ_1 which assigns probability 1/2 to each of the points 0 and 1. The Bayes solution δ_1 corresponding to the a priori distribution $\rho\lambda + (1 - \rho)\mu_1$ is easily found to be of the form: $\delta_1(0) = 1 - \delta_1(n)$, $\delta_1(x) = 1/2$ for $0 < x < n$. By virtue of Theorem 2, δ_1 is a restricted Bayes estimate if it achieves



its maximum risk only at 0 and 1. But this is true for $\bar{R}(p) = (p - \frac{1}{2})^2$, and hence by continuity (Theorem 3) must hold true for some interval $(C_1, \bar{C})$ of values of C. In fact, C_1 may be calculated from the condition $R'_1(0) = 0$. Again using the continuity of the variation of the risk function with C, we see that for some interval (C_2, C_1), the distribution μ_2 will consist of probability 1/2 at points α and $1 - \alpha$, the value of α varying continuously with C. At $C = C_2$, the risk function will develop further maxima, requiring a modification of the form of the distribution μ, and so forth. But in any case, the distribution μ must have only finitely many points of positive probability, since the risk function must achieve its maximum at all such points, is a polynomial in p, and is not constant except for $C = \underline{C}$.

The computing program outlined in the preceding paragraph was carried out numerically for the case $n = 4$, leading to the relation between C and $R(1/2)$

shown in the graph. The restricted Bayes estimates are (except for the extremes of minimax and Bayes estimates) not linear functions of X. For comparison, the graph also shows the minimum risk at $p = 1/2$ attainable for given maximum risk using only linear estimates. We see that, over a wide range of values of C, the use of nonlinear estimates permits nearly a 50 per cent reduction in the risk at $p = 1/2$.

EXAMPLE 3. As an illustration of the application of our principle to a finite decision problem, consider the following genetic situation. We are concerned with the inheritance of a simple Mendelian trait. Two individuals F and G are crossed to produce n progeny, each of whom is then crossed with a known hybrid to produce a single offspring. Of the n third-generation individuals, R are found to be recessive, while the remaining $n - R$ are either dominant or hybrid. Suppose it is known from other evidence that F and G are neither both dominant nor both recessive. Then the number k of recessive genes which they possess between them is either 1, 2, or 3. Our problem is to infer the value of k from the observed value of R. As our risk, we take the probability of wrong inference.

It is easily seen that R has the binomial distribution corresponding to $p = k/8$. The complete class of inference procedures is obtainable as the class of Bayes procedures corresponding to a priori distributions λ over the three possible values of p. If $P(p = i/8) = \lambda_i$, $i = 1, 2, 3$, then the Bayes solution, when R is observed to have the value r, is to decide for that value of k for which

$$\lambda_k \binom{n}{k} (k/8)^r (1 - k/8)^{n-r}$$

is greatest. These solutions have the following structure. There exist two numbers a and b, such that our decision is for $k = 1$ if $R < a$, for $k = 2$ if $a < R < b$, for $k = 3$ if $b < R$, while if $R = a$, we choose between $k = 1$ and $k = 2$, and if $R = b$, we choose between $k = 2$ and $k = 3$ according to certain probabilities. Let $P(\text{choosing } k = 1 \mid R = a) = \pi_1$, and $P(\text{choosing } k = 2 \mid R = b) = \pi_2$.

Suppose the proportions of dominant and recessive genes in the population at large are known to be μ and $1 - \mu$, respectively. Then, if F and G are bred under panmixia (i.e., chosen independently and randomly from the population), we should have

$$P(k = 1):P(k = 2):P(k = 3) = 2(1 - \mu)^2:3\mu(1 - \mu):2\mu^2.$$

The Bayes solution corresponding to this distribution for k would be the reasonable decision procedure to use if we were sure that F and G had been bred under panmixia. If we placed no reliance at all in the panmixia hypothesis, we might prefer to employ the minimax solution, which is characterized by equal probabilities of error corresponding to the three possible values of k. If we placed some reliance in the panmixia hypothesis, but not complete reliance, we might impose a limit on the permissible probability of error for any value of k, and subject to this limit seek to minimize the average probability of error under panmixia.

The minimax, Bayes, and restricted Bayes solutions are easy to obtain numerically. For illustration, suppose $n = 20$ and $\mu = 0.8$. The three solutions are as follows.

Solution	a	b	π_1	π_2
Minimax	3	7	0.482	0.142
Bayes	0.5	4.5	···	···
Restricted Bayes, maximum probability of error = 1/2	2	5	0.868	0.828

The performance characteristic properties of the three solutions may be conveniently compared in tabular form.

Solution	*Maximum probability of error*	*Average probability of error under panmixia*
Minimax	0.354	0.354
Restricted Bayes	0.500	0.239
Bayes	0.931	0.234

The restricted Bayes solution loses little efficiency as compared with the Bayes solution if panmixia holds, but gives considerably better protection if it does not hold.

It may be remarked that, in any finite decision problem, the restricted Bayes solution may be found by means of a finite number of applications of the Neyman-Pearson fundamental lemma. Suppose there are m decisions, so that the risks associated with any decision procedure may be represented as a vector $(r_1, r_2, \cdots, r_m)$ in m-dimensional Euclidean space. Except in the trivial situation in which the Bayes solution itself satisfies the restrictive condition that $r_i \leqq C$ for every $i = 1, 2, \cdots, m$, we may conclude from part (iii) of the proof of the Lemma of Section 3, that the restricted Bayes solution will have $r_i = C$ for at least one i. Let $I(C)$ be the set of i such that $r_i = C$. The restricted Bayes solution corresponding to C may then be obtained by minimizing $\sum \lambda_j r_j$, subject to $r_i = C$ for all i in $I(C)$. If we compute these minimizing solutions for all sets I, using the Neyman-Pearson lemma, we may select the restricted Bayes solution from among them by inspection.

REFERENCES

[1] A. Wald, *Statistical Decision Functions*, John Wiley and Sons, 1950.

[2] E. L. Lehmann, "On families of admissible tests," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 97–104.

[3] J. L. Hodges, Jr. and E. L. Lehman, "Some problems in minimax point estimation," *Annals* of *Math. Stat.*, Vol. 21 (1950), pp. 182–197.