CHAPTER

8 decision problems

8.1 ELEMENTS OF A DECISION PROBLEM

Problems of the type considered in the preceding chapter are examples of decision problems in which the statistician must choose a most preferred distribution from a given class of probability distributions on a set R of rewards. In this chapter we shall specify in more detail the structure of a broad class of decision problems of this type.

Consider a particular experiment whose possible outcomes w belong to a space Ω. Suppose that the statistician, without knowing the outcome of the experiment, must make a decision the consequences of which will depend on the outcome of the experiment. Let D be the space of all possible decisions d which might be made by the statistician, and let R be the space of all possible rewards r which might be received as a result of the statistician's decision d and the outcome w of the experiment. Specifically, we shall denote by $\sigma(w, d)$ the reward in R that would be received if the statistician makes the decision d and the outcome of the experiment is w.

We shall suppose that all the assumptions in Chap. 6 are satisfied. Hence, we shall assume that there exists a probability distribution P on the space Ω of outcomes whose value $P(A)$ is specified for each event A

belonging to an appropriate σ-field $\mathcal{A}$ of subsets of Ω. We shall let W denote the unknown outcome. Then, for any event $A \in \mathcal{A}$, it may be said that $\Pr(W \in A) = P(A)$.

Furthermore, we shall suppose that the statistician's preferences among the rewards in R satisfy all the assumptions in Chap. 7. Hence, we shall assume that there exists a utility function U on the set R. It is assumed that U is a measurable function with respect to an appropriate σ-field $\mathcal{B}$ of subsets of R.

For each fixed decision $d \in D$, the function σ induces a probability distribution P_d on the set R of rewards. For any subset $B \in \mathcal{B}$, the value $P_d(B)$ is specified as follows:

$$P_d(B) = \Pr[\sigma(W, d) \in B] = P\{w\colon \sigma(w, d) \in B\}. \tag{1}$$

In order that the distribution P_d defined by Eq. (1) will be meaningful, the following requirement must be met. For each subset $B \in \mathcal{B}$, the set $\{w\colon \sigma(w, d) \in B\}$ must belong to the σ-field $\mathcal{A}$. We shall assume that this requirement is met for every decision $d \in D$. Then, for each probability distribution P_d for which the utility function U can be integrated, the expected utility $E(U|P_d)$ can be computed as follows:

$$E(U|P_d) = \int_R U(r)\, dP_d(r) = \int_\Omega U[\sigma(w, d)]\, dP(w). \tag{2}$$

The statistician should choose, if possible, a decision d which maximizes $E(U|P_d)$.

In a context like the present one, where a decision must be made without knowledge of the outcome W of some experiment, W is called a *parameter* and the set Ω of all possible values of W is called the *parameter space*. Furthermore, in these decision problems, it has become standard to specify for each reward $r \in R$ the negative of its utility, rather than its utility, and to call this number the *loss*. Hence, for each outcome $w \in \Omega$ and each decision $d \in D$, the loss $L(w, d)$ is defined by the equation

$$L(w, d) = -U[\sigma(w, d)]. \tag{3}$$

It follows from this discussion that the elements of a decision problem may be regarded as a parameter space Ω, a decision space D, and a real-valued loss function L which is defined on the product space $\Omega \times D$. For any point $(w, d) \in \Omega \times D$, the number $L(w, d)$ represents the loss when the value of the parameter W is w and the statistician chooses decision d. It is assumed that for each $d \in D$, the loss $L(\cdot, d)$ is an $\mathcal{A}$-measurable function on the space Ω.

Let P be any given probability distribution of the parameter W. For any decision $d \in D$, the expected loss, or *risk*, $\rho(P, d)$ is specified by the equation

$$\rho(P, d) = \int_\Omega L(w, d)\, dP(w). \tag{4}$$

It will be assumed that the integral in Eq. (4) is finite for every $d \in D$. Any decision d for which this assumption is not true can usually be eliminated from the set D. It follows from Eqs. (2) and (3) that the statistician should choose, if possible, a decision d which minimizes the risk $\rho(P, d)$.

8.2 BAYES RISK AND BAYES DECISIONS

Consider now a decision problem defined by a parameter space Ω, a decision space D, and a loss function L. For any distribution P of the parameter W, the *Bayes risk* $\rho^*(P)$ is defined to be the greatest lower bound for the risks $\rho(P, d)$ for all the decisions $d \in D$. Hence, $\rho^*(P)$ is specified by the equation

$$\rho^*(P) = \inf_{d \in D} \rho(P, d). \tag{1}$$

Any decision d^* whose risk is equal to the Bayes risk is called a *Bayes decision against the distribution* P. Hence, a decision d^* is a Bayes decision against the distribution P if, and only if, $\rho(P, d^*) = \rho^*(P)$.

If the distribution of the parameter W is P, any Bayes decision against P will be an optimal decision for the statistician because the risk cannot be smaller for any other decision. It is possible, however, that no decision in the space D is a Bayes decision. This situation occurs when the greatest lower bound in Eq. (1) is not actually attained for any decision $d \in D$. In such a situation, the statistician should select any decision $d \in D$ for which the risk $\rho(P, d)$ is sufficiently close to the Bayes risk. Since difficulties of this kind are not central to the theory or practice of decision making, it will typically be assumed in the following discussion that for any distribution P which might be considered, the Bayes risk $\rho^*(P)$ is attained for some decision $d \in D$.

We shall now discuss three examples.

EXAMPLE 1 Suppose that the parameter space Ω contains just the two numbers 0 and 1 and that the decision space D contains all the numbers d in the interval $0 \le d \le 1$. Suppose also that the loss function L is defined, for any values $w \in \Omega$ and $d \in D$, by the equation

$$L(w, d) = |w - d|^\alpha. \tag{2}$$

Here, α is a given positive integer. Finally, it is assumed that the probability distribution P of W is such that $\Pr(W = 0) = \frac{3}{4}$ and $\Pr(W = 1) = \frac{1}{4}$.

Consider now the problem in which $\alpha = 1$ in the loss function defined by Eq. (2). Then, for any decision $d \in D$, the risk $\rho(P, d)$ is given by the equation

$$\begin{aligned}\rho(P, d) &= L(0, d)\Pr(W = 0) + L(1, d)\Pr(W = 1) \\ &= \tfrac{3}{4}d + \tfrac{1}{4}(1 - d) = \tfrac{1}{2}d + \tfrac{1}{4}. \end{aligned} \tag{3}$$

It follows that $\rho(P, d)$ is minimized when $d = 0$. Hence, the value $d = 0$ is the unique Bayes decision, and the Bayes risk $\rho^*(P)$ has the value $\frac{1}{4}$.

The following point should be noted. If the decision space D is defined to be the interval $0 < d \leq 1$, where the end point 0 has now been deleted from D, then the Bayes risk $\rho^*(P)$ will still have the value $\frac{1}{4}$ but no decision in D will be a Bayes decision.

EXAMPLE 2 Consider again the problem in Example 1, but now suppose that $\alpha > 1$ in the loss function defined by Eq. (2). Then, for any decision $d \in D$,

$$\rho(P, d) = \tfrac{3}{4}d^\alpha + \tfrac{1}{4}(1 - d)^\alpha. \tag{4}$$

The value of d which minimizes the risk in Eq. (4) may be obtained by elementary differentiation. In this way, the unique Bayes decision d^* is found to have the value

$$d^* = (1 + 3^{1/(\alpha-1)})^{-1}. \tag{5}$$

EXAMPLE 3 Consider a problem in which the parameter is a vector $\mathbf{W} = (W_1, W_2)'$ which takes values in the plane R^2. Suppose that the decision space D contains just two decisions d_1 and d_2, and suppose that for any point $(w_1, w_2)' \in R^2$, the loss function L is defined by the equations

$$L(w_1, w_2, d_1) = w_1^2, \qquad L(w_1, w_2, d_2) = w_2^2. \tag{6}$$

Then, for any joint distribution P of W_1 and W_2, the Bayes decision against P is d_1 if $E(W_1^2) < E(W_2^2)$ and is d_2 if $E(W_1^2) > E(W_2^2)$. If $E(W_1^2) = E(W_2^2)$, then both d_1 and d_2 are Bayes decisions against P.

8.3 NONNEGATIVE LOSS FUNCTIONS

Suppose that the distribution of the parameter W in some decision problem is P. Let a be a given constant ($a > 0$), and let λ be a real-valued function on the parameter space Ω such that the integral $\int_\Omega \lambda(w)\,dP(w)$ is finite. Consider a new loss function L_0, which is defined in terms of

the original loss function L by the relation

$$L_0(w, d) = aL(w, d) + \lambda(w) \qquad w \in \Omega,\ d \in D. \tag{1}$$

For any decision $d \in D$, let $\rho(P, d)$ denote the risk which results from the original loss function L, as defined by Eq. (4) of Sec. 8.1, and let $\rho_0(P, d)$ denote the risk which results from the new loss function L_0. Then, for any two decisions $d_1 \in D$ and $d_2 \in D$, it follows that $\rho_0(P, d_1) \leq \rho_0(P, d_2)$ if, and only if, $\rho(P, d_1) \leq \rho(P, d_2)$. In particular, a decision d^* is a Bayes decision against P in the original problem with loss function L if, and only if, d^* is a Bayes decision against P in the new problem with loss function L_0.

Now consider the function λ_0 which is defined at each point $w \in \Omega$ by the equation

$$\lambda_0(w) = \inf_{d \in D} L(w, d). \tag{2}$$

If the integral of λ_0 satisfies the condition given at the beginning of this section, then we can replace L by a new loss function L_0, which is defined for any values $w \in \Omega$ and $d \in D$ by the equation

$$L_0(w, d) = L(w, d) - \lambda_0(w). \tag{3}$$

The loss function L_0 has the following properties for any values $w \in \Omega$ and $d \in D$:

$$L_0(w, d) \geq 0 \qquad \text{and} \qquad \inf_{d \in D} L_0(w, d) = 0. \tag{4}$$

It has been found convenient in many problems to work with nonnegative loss functions of this type, although the use of such functions makes it appear that the statistician must continually choose decisions from which he can never realize a positive gain.

It should be noted that the loss functions in Exercises 1 to 3 at the end of this chapter all satisfy the conditions given in relation (4).

8.4 CONCAVITY OF THE BAYES RISK

We shall now show that in any decision problem, the Bayes risk $\rho^*(P)$ must be a concave function of the distribution P of the parameter W. Here, as in Chap. 7, for any two distributions P_1 and P_2 of W and for any number α such that $0 \leq \alpha \leq 1$, we shall let $\alpha P_1 + (1 - \alpha)P_2$ denote the distribution which assigns the probability $\alpha P_1(A) + (1 - \alpha)P_2(A)$ to each event $A \in \mathcal{A}$. Hence, we must prove the following theorem.

Theorem 1 *For any distributions P_1 and P_2 of W and for any number α*