The Annals of Mathematical Statistics
1969, Vol. 40, No. 6, 2127–2142

BAYESIAN MODEL OF DECISION-MAKING AS A RESULT OF LEARNING FROM EXPERIENCE[1, 2]

BY BRUNO O. SHUBERT

Stanford University

1. Introduction. A statistical model of decision-making is formally described as follows: given is a set A of possible decisions or actions and a family $\mathcal{P} = \{\nu_\theta : \theta \varepsilon \Theta\}$ of probability distributions on a measurable space $(\Xi, \mathfrak{X})$. The decision-maker (statistician) observes a random variable X with values in the space $(\Xi, \mathfrak{X})$ and distributed according to some $\nu_\theta \varepsilon \mathcal{P}$, and on the basis of this observation decides for some action $a \varepsilon A$. The appropriateness of his decision then depends on the action chosen, and also on an unknown parameter θ (state of Nature), which specifies the distribution ν_θ of the random variable observed, and is measured by a numerical loss function L defined on the product space $\Theta \times A$. A rational decision-maker is then assumed to use a decision function δ, which assigns to every observed value (sample) $x \varepsilon \Xi$ an action $a \varepsilon A$ in such a manner that the resulting loss is as small as possible. It is clear, however, that no decision function can minimize the loss itself for all values of θ, since a decision function must not depend on this unknown parameter.

In the Bayesian approach, this problem is resolved by assuming that the parameter θ is also a random variable with distribution τ (prior distribution) known to the statistician, and the optimum decision function δ_τ^* (called Bayes decision function again τ) is then defined as that, for which the expected loss $E_\tau\{L(\theta, \delta(X))\}$ attains its minimum—the so-called Bayes risk $\rho(\tau) = \min_\delta E_\tau\{L(\theta, \delta(X))\}$.

To justify this Bayesian model as appropriate for studying decision-making we face a problem concerning both the adequacy of the assumption of randomness of θ and knowledge of the prior distribution as well as the question of interpretation of the minimum expected loss as optimum.

Two essentially different approaches have been taken in this respect. In the first (subjectivistic) approach the loss function is looked upon as a negative utility associated with all pairs (θ, a), $\theta \varepsilon \Theta$, $a \varepsilon A$, and satisfying Von Neumann-Morgenstern's (or other analogous) axioms. It is then shown that this is tantamount to the existence of a prior distribution, and the optimality of Bayes risk follows from the expected utility hypothesis.

The second (statistical) approach assumes, on the other hand, that the decision problem in question is a typical member of a large population of identical decision situations with parameters θ varying arbitrarily along the population.

Received 17 May 1968.

[1] Work on this paper has been done while the author was visiting professor at Morehouse College, Atlanta, Georgia.

[2] The work was partially supported under AFOSR contract AF 49 (638) 1517.