AN ASYMPTOTIC MINIMAX THEOREM FOR THE TWO ARMED BANDIT PROBLEM

By Walter Vogel[1]

University of Chicago and Universität Tübingen[2]

1. Introduction. Let Ex I and Ex II be two experiments, the outcomes of which are described by the two random variables X and Y. Let $P(X = 1) = p = 1 - P(X = 0)$, $P(Y = 1) = q = 1 - P(Y = 0)$ and $0 < p, q < 1$. An experimenter has to do n experiments, one after another, and at every step he may choose between Ex I or Ex II. He does not know the values of p and q and he wants to maximize the sum of all outcomes. Therefore he will choose a strategy, i.e. a procedure which tells him which experiment to use at the kth step as a function of his previous choices and the previous outcomes of the experiments. The question how to find a suitable strategy is known as the problem of the two armed bandit. For approaches other than the one used in this paper see [1], [2], [4], [5] and [6].

We will measure the value of a strategy by a loss function. Let Π_k be the unconditional probability of choosing Ex I at the kth step. The expected value of the performed experiment at the kth step will be

$$\Pi_k p + (1 - \Pi_k)q = p - (p - q)(1 - \Pi_k) = q - (q - p)\Pi_k .$$

We define as loss at the kth step:

$$\max (p, q) - (\Pi_k p + (1 - \Pi_k)q) = (p - q)(1 - \Pi_k)$$

if $p \geqq q$ or $(q - p)\Pi_k$ if $p \leqq q$. The loss $L(p, q)$ for the whole game is then

$$(p - q) \sum_{k=1}^{n} (1 - \Pi_k) \quad \text{or} \quad (q - p) \sum_{k=1}^{n} \Pi_k .$$

In $L(p, q)$ and $\Pi_k(p, q)$ the first argument is always related to Ex I. Let $\sigma = \max (p, q)$ and $\tau = \min (p, q)$; then

$$L(\sigma, \tau) = (\sigma - \tau) \sum_{k=1}^{n} (1 - \Pi_k(\sigma, \tau))$$

and

$$L(\tau, \sigma) = (\sigma - \tau) \sum_{k=1}^{n} \Pi_k(\tau, \sigma).$$

As we do not suppose any previous knowledge about p and q, it seems natural

Received April 2, 1959; revised January 18, 1960.

[1] Research carried out in part at the Statistical Research Center, University of Chicago, under the sponsorship of the Statistics Branch, Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Present address: Mathematisches Institut der Universität Tübingen, Germany.