Inequalities in Statistics and Probability
IMS Lecture Notes–Monograph Series Vol. 5 (1984), 206–210

SELECTING THE t BEST CELLS OF A MULTINOMIAL USING INVERSE SAMPLING

BY PINYUEN CHEN and MILTON SOBEL[1]
Syracuse University and University of California at Santa Barbara

An inverse sampling procedure R is proposed for selecting the t "best" cells (i.e., cells with the largest cell probabilities) from a multinomial distribution with k cells ($1 \leq t < k$). Two different formulations of this selection problem are considered and the measure of distance in both formulations is the ratio of the largest and second largest cell probabilities. One formulation is of the usual type based on an empty indifference zone; in the other (new) formulation any collection of t cells from the union of the preference zone (for selection) and the indifference zone is called a correct selection. Type 2-Dirichlet integrals are used (i) to express the probability of correct selection as an integral with parameters only in the limits of integration, and (ii) to prove that the least favorable configuration for each of the formulations under R is the so-called slippage configurations with k–t equal cell probabilities and t cell probabilities slipped to the right by a common amount.

1. Introduction. One of the important applications of ranking and selection techniques is to select (without respect to order) the t best cells of a multinomial distribution with k cells. For the special case $t = 1$ the fixed sample size problem was first considered by Bechhofer, Elmaghraby and Morse (1959) and the inverse sampling procedure was first considered by Cacoullos and Sobel (1966). We are presently discussing fixed subset size problems and not considering the random subset size problem which was considered by Gupta and Nagel (1971) and more recently by Hu (1982). It is well known by people working in this area that the generalization of the fixed subset size problem to arbitrary t ($1 < t < k$) presents some serious difficulties (cf. the work of Lee (1975) and Hwang, Hsuan and Parned (1980) on this topic). In this paper we consider the corresponding problem for general $t \geq 1$ with an inverse sampling procedure.

Actually we consider two different formulations of the ranking and selection problem. The measure of distance in both formulations is the ratio of cell probabilities as in the previous references. Let

$$p_{[1]} \leq p_{[2]} \leq \ldots \leq p_{[k-t]} \leq p_{[k-t+1]} \leq \ldots \leq p_{[k]} \tag{1.1}$$

denote the ordered cell probabilities which sum to one. Let $\delta^* > 1$ and $P^*(\binom{k}{t}^{-1} < P^* < 1)$ denote specified constants. In the usual (or first) formulation we require a procedure R such that

$$P\{CS|R\} > P^* \text{ whenever } \delta \geq \delta^*, \tag{1.2}$$

where $\delta = p_{[k-t+1]}/P_{[k-t]}$.

Actually we need only consider configurations (1.1) with $p_{[k-t]} < p_{[k-t+1]}$ and in this case the definition of correct selection (CS) is clear, namely that we select the t cells with largest p-values.

We shall say the p-value is in the indifference zone (IZ) if it lies strictly between $p_{[k-t+1]}/\delta^*$ and $p_{[k-t+1]}$. The p-values $\geq p_{[k-t+1]}$ will be said to lie in the preference zone (PZ).

---

[1] This paper was supported by National Science Foundation Grant Number MCS82–02247.
AMS 1980 subject classifications. Primary, 62F07; secondary, 62E15.
Key words and phrases: Ranking and selection, inverse sampling, multinomial distribution, Type 2-Dirichlet integral.