ON ESTIMATING THE TOTAL PROBABILITY OF

THE UNOBSERVED OUTCOMES OF AN EXPERIMENT*

P.J. Bickel and J.A. Yahav

University of California at Berkeley

Robbins (1968) considered the problem of estimating the total probability of the unobserved outcomes of an experiment. In this paper we suggest an estimator, based on n trials, and show that under some regularity conditions one can construct asymptotic confidence intervals for the random quantity we look for.

Consider an experiment with positive outcomes $E_1, E_2, \ldots$ with unknown probabilities $\pi_1, \pi_2, \ldots, \pi_i > 0$, $\Sigma_i \pi_i = 1$. In n independent trials suppose that E_i occurs N_i times $i=1,2,3,\ldots$ with $\Sigma_i N_i = n$. Let $\psi_i = 1$ or 0 accordingly as $N_i = 0$ or $N_i > 0$. Then the random variable $U = \Sigma_i \psi_i \pi_i$ is the sum of the probabilities of the unobserved outcomes. How to estimate U? Robbins (1968) asked this question and suggested the following answer:

Suppose we make one more independent trial of the same experiment and that in the total of n + 1 trials, E_i occurs N_i', $i=1,2,\ldots$ with $\Sigma_i N_i' = n + 1$. Let $V' = \frac{1}{n+1} \Sigma_i I_{\{N_i' = 1\}}$, where I_A is the indicator function of A. In contrast to U, V' is observable, with n + 1 trials, and can be used to predict U (we use the word predict instead of estimate since U is r.v. and not a parameter).

For $W' = U - V'$ Robbins showed:

* Research partially supported by Office of Naval Research Contract N00014-80-C-0163.

Subject classifications: 62C12, 62G99.

Key phrases: Estimation, Asymptotic normality, Unobserved outcomes, Prediction.