The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 256–257

NOTES

ESTIMATING THE TOTAL PROBABILITY OF THE UNOBSERVED OUTCOMES OF AN EXPERIMENT[1]

BY HERBERT E. ROBBINS

Columbia University

An experiment has the possible outcomes $E_1, E_2, \cdots$ with unknown probabilities $p_1, p_2, \cdots; p_i \geqq 0$, $\sum_i p_i = 1$. In n independent trials suppose that E_i occurs x_i times, $i = 1, 2, \cdots$, with $\sum_i x_i = n$. Let $\varphi_i = 1$ or 0 according as $x_i = 0$ or $x_i \neq 0$. Then the random variable $u = \sum_i p_i \varphi_i$ is the sum of the probabilities of the unobserved outcomes. How can we "estimate" u? (The quotation marks appear because u is not a parameter in the usual statistical sense.)

Suppose we make one more independent trial of the same experiment, and that in the total of $n + 1$ trials E_i occurs y_i times, $i = 1, 2, \cdots$, with $\sum_i y_i = n + 1$. (Each $y_i = x_i$, except for one value of the subscript.) Let $\psi_i = 1$ or 0 according as $y_i = 1$ or $y_i \neq 1$. Consider the statistic $v = (n + 1)^{-1} \sum_i \psi_i$, which is the number of "singleton" outcomes of the $n + 1$ trials, divided by $n + 1$. In contrast to u, v is observable. The idea of using something like v to estimate u goes back to A. M. Turing according to [1], where the problem is discussed from a somewhat different point of view. We shall show that v is a good "estimator" of u in the sense that setting $w = u - v$ we have always

$$Ew = 0, \qquad Ew^2 < (n + 1)^{-1}.$$

In fact,

$$Ew = \sum_i (p_i q_i^n - (n + 1)(n + 1)^{-1} p_i q_i^n) = \sum_i 0 = 0,$$

and a little algebra shows that $(n + 1)\ Ew^2$ is equal to the expression

$$\sum_i p_i q_i^n (1 + (n - 1)p_i) - \sum_{i \neq j} p_i p_j (1 - p_i - p_j)^n. \tag{1}$$

Hence since $1 - x \leqq e^{-x}$,

$$(n + 1)Ew^2 \leqq \sum_i p_i e^{-np_i} e^{(n-1)p_i} = \sum_i p_i e^{-p_i} < \sum_i p_i = 1.$$

In the special case in which some k of the p_i are equal to $1/k$ and all the others are 0, the expression (1) reduces to

$$(1 + (n - 1)k^{-1})(1 - k^{-1})^n - (1 - k^{-1})(1 - 2k^{-1})^n. \tag{2}$$

Set

$$a_n = \sup (1) \text{ for all probability vectors } (p_1, p_2, \cdots),$$
$$b_n = \sup (2) \text{ for all } k = 1, 2, \cdots;$$

Received 21 August 1967.

[1] This research was sponsored by the Office of Naval Research under Contract Nonr-4259(08) at Columbia University.

then $b_n \leqq a_n \leqq 1$. Putting $k = n/\lambda$, keeping λ fixed, and letting $n \to \infty$, we have

$$(2) \to (1 + \lambda)e^{-\lambda} - e^{-2\lambda} \leqq (1 + \lambda^*)e^{-\lambda^*} - e^{-2\lambda^*} = b \cong .61,$$

where $\lambda^* \cong .85$ is the root of $\lambda \doteq 2e^{-\lambda}$. Hence $b_n \to = b$ as $n \to \infty$. We do not know if $a_n = b_n$, $a_n \leq b$, or $a_n \to b$. In any case, the universal inequality $Ew^2 < (n + 1)^{-1}$ can certainly be improved, and can be used together with the Chebyshev inequality to obtain "confidence intervals" for u. (Similar inequalities for higher moments of w may yield shorter intervals.) There are many other statistics v based on the $n + 1$ trials such that always $E(u - v) = 0$. We do not know which is best in the sense of minimizing $E(u - v)^2$.

REFERENCE

[1] Good, I. J. (1953). On the population frequencies of species and the estimation of population parameters. *Biometrika* 40 237–264.