UNBIASED ESTIMATES FOR CERTAIN BINOMIAL SAMPLING PROBLEMS WITH APPLICATIONS[1]

By M. A. Girshick, Frederick Mosteller, and L. J. Savage

U. S. Department of Agriculture; Statistical Research Group, Princeton University; and *Statistical Research Group, Columbia University*

1. Introduction. The purpose of this paper is to present some theorems with applications concerning unbiased estimation of the parameter p (fraction defective) for samples drawn from a binomial distribution. The estimate constructed is applicable to samples whose items are drawn and classified one at a time until the number of defectives i, and the number of nondefectives j, simultaneously agree with one of a set of preassigned number pairs. When this agreement takes place, the sampling operation ceases and an unbiased estimate of the proportion p of defectives in the population may be made. Some examples of this kind of sampling are ordinary single sampling in which n items are observed and classified as defective or nondefective; curtailed single sampling where it is desired to cease sampling as soon as the decision regarding the lot being inspected can be made, that is as soon as the number of defectives or nondefectives attain one of a fixed pair of preassigned values; double, multiple, and sequential sampling. In the cases of double and multiple sampling the subsamples may be curtailed when a decision is reached, while for sequential sampling the process may be truncated, i.e. an upper bound may be set on the amount of sampling to be done. In section 3 expressions are given for the unique unbiased estimates of p for single, curtailed single, curtailed double, and sequential sampling.

One or two of the illustrative examples of section 3 may be of interest because their rather bizarre results suggest that some estimate other than an unbiased estimate may be preferable; but the discussion of estimates other than unbiased ones is outside the scope of this paper.

2. The estimate $\hat{p}$. For the purposes of the present paper the word point will refer only to points in the xy-plane with nonnegative integral coordinates.

We shall need the following nomenclature. A region R is a set of points containing (0, 0). The point (x_2, y_2) is *immediately beyond* (x_1, y_1) if either $x_2 = x_1 + 1$, $y_2 = y_1$ or $x_2 = x_1$, $y_2 = y_1 + 1$. A *path in* R from the point α_0 to the point α_n is a finite sequence of points $\alpha_0, \alpha_1, \cdots, \alpha_n$ such that α_i $(i > 0)$ is immediately beyond α_{i-1}, and $\alpha_j \in R$ with the possible exception of α_n. A *boundary* point, that is, an element of the boundary B of R, is a point not in R which is the last point α_n of a path from the origin. *Accessible* points are the points in R which can be reached by paths from the origin, while *inaccessible* points are the points which cannot be reached by any path from the origin.

[1] This paper was originally written by Mosteller and Savage. A communication from M. A. Girshick revealed that he had independently discovered for the sequential probability ratio test the estimate $\hat{p}(\alpha)$ given here and demonstrated its uniqueness. For purposes of publication it seemed appropriate to present the results in a single paper.