INVERSE BINOMIAL SAMPLING PLANS WHEN AN EXPONENTIAL DISTRIBUTION IS SAMPLED WITH CENSORING

By Jack Nadler

Bell Telephone Laboratories, Inc.

1. Introduction. Inverse binomial sampling plans are discussed in a number of papers in the literature, e.g., [3], [6], [7], [8], [13]. In addition, inverse sampling procedures have been proposed for some problems of drawing inferences from counted data about a parameter other than a population proportion, e.g., in [1] and [12]. In this note it is pointed out that inverse binomial sampling plans can be applied in certain situations where continuous observations must be taken with censoring. In particular, when the observations follow an exponential distribution and large values are censored, the use of such a sampling plan gives rise to a complete, sufficient statistic [10] that has a simple, well known distribution. This result is closely related to the work of Malmquist [11].

2. The statistical model. Let X be a random variable with c.d.f.

$$F_\theta(x) = \begin{cases} 1 - \exp(-x/\theta), & \text{if } x > 0 \\ 0, & \text{if } x \leqq 0, \end{cases} \tag{1}$$

where $0 < \theta < \infty$ is unknown. Suppose that observations are to be taken from this distribution with the restriction that for a known constant x_o, the value of X can be observed if and only if $X \leqq x_o$; otherwise, just the information that $X > x_o$ is provided. From a sample of observations with this sort of censoring (sometimes called "single censoring on the right"), it is required that an inference be made about the value of θ.

To apply a binomial sampling plan to this situation, define the random variable

$$U = \begin{cases} 1, & \text{if } X \leqq x_o \\ 0, & \text{if } X > x_o. \end{cases} \tag{2}$$

Clearly, U is a Bernoulli chance variable,

$$\Pr\{U = 1\} = 1 - \exp(-x_o/\theta) = p, \qquad \Pr\{U = 0\} = 1 - p = q, \tag{3}$$

that can be sampled only by taking observations on X subject to censoring. Thus, a sampling plan for the distribution of U specifies a procedure for sampling from the distribution of X with censoring.

The proposed sampling plan is defined in terms of taking observations on U: For a given, positive integer r, a sequence of independent observations, say $U_1, U_2, \cdots$, is sampled until for the first time $\sum_{i=1}^{N} U_i = r$. Then sampling is terminated. This inverse binomial sampling plan implies that independent observations on X are taken sequentially until r values of $X \leqq x_o$, say $Y_1, \cdots, Y_r$,

Received November 28, 1959; revised May 6, 1960.

are obtained. In the experiment the total number of observations N taken with censoring (including $Y_1, \cdots, Y_r$) is a random variable.

3. A sufficient statistic. The method of maximum likelihood suggests the use of the statistic

$$S = \sum_{i=1}^{r} Y_i + (N - r)x_o$$

for the analysis of such an experiment. Some important properties of S are now derived.

THEOREM: *With the proposed sampling plan, S is a sufficient statistic and $2S/\theta$ has a χ^2 distribution with $2r$ degrees of freedom.*

PROOF: Regard the given experiment as r independent repetitions of an experiment in which observations are taken one at a time with censoring until a single value of $X \leqq x_o$ is obtained. In the ith repetition ($i = 1, \cdots, r$), let Y_i be the observed value of X and $N_i - 1$, the number of censored observations immediately preceding Y_i. (Each censored observation is known to be greater than x_o.) Define

$$S_i = Y_i + (N_i - 1)x_o. \tag{5}$$

Since, with probability one,

$$N_i = 1 + [S_i/x_o], \qquad Y_i = S_i - x_o[S_i/x_o], \tag{6}$$

where $[x]$ denotes the integral part of x, S_i is sufficient for the joint distribution of Y_i and N_i.

To determine the distribution of S_i, note that there is no loss of generality in assuming that X is the waiting time up to a change of state in a homogeneous Poisson process, where changes occur at the mean rate of $1/\theta$ per unit time. From this point of view, taking an observation with censoring has the interpretation that starting from an arbitrary moment in time, the process is observed for an interval of $\min(X, x_o)$ units of time. Drawing a sample of observations with censoring then amounts to observing a sequence of these time intervals that are disjoint, but not necessarily adjacent. Using the properties of a Poisson process, write

$$P_0(t) = \exp(-t/\theta)$$

for the probability that no change occurs in an interval of length t.

For arbitrary $x > 0$, let $m = [x/x_o]$ and consider

$$\Pr\{S_i \leqq x\} = \Pr\{S_i \leqq m\} + \Pr\{m < S_i \leqq x\}. \tag{8}$$

The event $S_i \leqq m$ in terms of the Poisson process is the complement of the event that no change is observed in each of m non-overlapping intervals of length x_o. Furthermore, $m < S_i \leqq x$ is equivalent to the intersection of the event that no change is observed in m disjoint intervals with the event that at least one change is observed in an additional interval of length $x - mx_o$ that does not overlap any of the previous m. Hence,

$$\Pr\{S_i \leqq x\} = 1 - \{P_0(x_o)\}^m + \{P_0(x_o)\}^m\{1 - P_0(x - mx_o)\} \tag{9}$$
$$= 1 - \exp(-x/\theta).$$

Therefore, $S_1, \cdots, S_r$ form a random sample from the exponential distribution (1). Since $S = \sum_{i=1}^{r} S_i$, the theorem follows from well-known properties of the Γ distribution (see *e.g.*, Example 17.8 of [9])[1].

COROLLARY. *The distribution of S belongs to a complete family of distributions.*

A proof of the corollary is supplied by Example 3.5 of [10].

It is not difficult to verify that the theorem does not hold for the analogue of S in the single sample case [2], i.e., when the total number of observations taken with censoring is fixed and the number of values of $X \leqq x_o$ is a random variable. Indeed, the principal advantage of the inverse binomial sampling plan is that appropriate procedures are often well-known or easily derived. The completeness property provides easy proofs of the optimal character of many of these procedures.

4. Remarks.

(i) It is not essential to the sampling plan that observations are taken one at a time. The results of the theorem are valid, for example, in the following multiple stage experiment: In the first stage, a sample of r independent observations on X is drawn. The experiment is terminated if r values of $X \leqq x_o$ are observed; otherwise, the experiment is continued. At each successive stage a sample of observations is taken whose size is the additional number of values of $X \leqq x_o$ required to make the cumulative number of such values (i.e., the number observed in the entire experiment) equal to at most r. Experimentation is terminated as soon as r values of $X \leqq x_o$ are obtained.

(ii) The "nice" properties of S are shared by a statistic proposed by Epstein and Sobel [4], [5]. From the point of view of the experimental design, however, the restriction imposed on the observable values of X in this paper differs importantly from the one considered by those authors.

(iii) An inverse binomial sampling plan can, of course, be applied when X follows some distribution other than (1). Typically, results like the theorem above do not hold for such cases. At least from the point of view of maximum likelihood estimation, however, it is frequently advantageous to use such a sampling plan.

5. Acknowledgments. The author is grateful for the helpful comments and criticism received in connection with this work from M. H. DeGroot, R. B. Murphy, and W. L. Roach, Jr.

---

[1] M. Sobel, John W. Tukey, and the referee independently suggested that an earlier derivation of the distribution of S_i might be replaced by one proceding along the lines presented here.

REFERENCES

[1] Douglas G. Chapman, "Inverse, multiple and sequential sample censuses," *Biometrics*, Vol. 8 (1952), pp. 286–306.

[2] A. Clifford Cohen, Jr., "Maximum likelihood estimation of the dispersion parameter of a chi-distributed radial error from truncated and censored samples with applications to target analysis," *J. Amer. Stat. Assn.*, Vol. 50 (1954), pp. 1122–1135.

[3] Morris H. DeGroot, "Unbiased sequential estimation for binomial populations," *Ann. Math. Stat.*, Vol. 30 (1959), pp. 80–101.

[4] B. Epstein and M. Sobel, "Life testing," *J. Amer. Stat.* Assn., Vol. 48 (1953), pp. pp. 486–502.

[5] B. Epstein and M. Sobel, "Some theorems relevant to life testing from an exponential distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 373–381.

[6] D. J. Finney, "On a method of estimating frequencies," *Biometrika*, Vol. 36 (1949), pp. 233–234.

[7] M. A. Girshick, F. Mosteller, and L. J. Savage, "Unbiased estimates for certain binomial sampling problems with applications," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 13–23.

[8] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, Vol. 33 (1945), pp. 222–225.

[9] Maurice G. Kendall, *The Advanced Theory of Statistics*, Vol. II, 2nd Ed., Charles Griffin & Company Ltd., London, 1948.

[10] E. L. Lehmann and H. Scheffé, "Completeness, similar regions, and unbiased estimation—Part I," *Sankhyā*, Vol. 10 (1950), pp. 305–340.

[11] Sten Malmquist, "A statistical problem connected with the counting of radioactive particles," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 255–264.

[12] Martin Sandelius, "An inverse sampling procedure for bacterial plate counts," *Biometrics*, Vol. 6 (1950), pp. 291–292.

[13] M. C. K. Tweedie, "Inverse statistical variates," *Nature*, Vol. 155 (1945), p. 453.