A LOGIC OF INFERENCE IN SAMPLE SURVEY PRACTICE

B. V. Shah*, Research Triangle Institute

Abstract

This paper is an attempt to present the logic underlying some practices in sample surveys. The primary focus is on the nature of information in labels for estimating a distribution function F(X). Lack of sufficiency is an inadequate justification in practice for an argument that the marginal distribution of a variable should be influenced by its joint distribution with another variable such as the label.

The application of the maximum likelihood estimation (MLE) for a nonparametric distribution function leads to the Horvitz-Thompson estimator, in general. This result, coupled with the large sample properties of MLE, provides a justification for use of the Wald statistic, based on Horvitz-Thompson estimators, for tests of hypotheses and confidence intervals.

KEY WORDS: Foundations of survey sampling; Maximum likelihood; Sampling distribution function; Horvitz-Thompson estimator; Inference in finite populations.

1. INTRODUCTION

In the last twenty-five years, many "new" theories, foundations, and techniques have been proposed for estimation and inference in sample surveys. Good summaries appear in Smith (1976) and Cassel, Särndal and Wretman (1977).

Godambe (1955) confirmed the conjecture by Horvitz and Thompson (1952) that there is no linear estimator that is uniformly best ("minimum mean square error") for estimating population total over the set of all possible populations. Later formalizations of theory by Godambe (1966) and Basu (1969, 1971) show that Fisherian ideas of likelihood and sufficiency fail to provide any useful inference. Hartley and Rao (1969) and Royall (1968) have suggested alternative approaches based on the contention that labels are uninformative. Ericson (1969), Scott and Smith (1969), and Royall (1970) have suggested a super-population and/or Bayesian framework; these works suggest that the inference is independent of the probabilities of selection used in survey design. Basu (1977) has questioned even the need for randomization.

The major problems for inference in survey sampling are related to: (a) likelihood principle, and (b) labels. Hartley and Rao (1969) have suggested the "scale load approach" to derive maximum likelihood estimates. The objective in this paper is to show that maximum likelihood estimators are applicable in general for estimation of the distribution function.

The statistical parameter considered in this paper is the population distribution function, F(X), of a random vector X. For a simple random sample, the empirical distribution, $S_n(X)$, is the maximum likelihood estimate (MLE) of the true distribution function. A few formal results are presented in Section 2 as theorems for estimating F(X) under different conditions. These theorems are valid as per assumptions stated therein and are independent of sample survey practice. An application of the theorems to survey sampling is discussed in Section 3.

Section 4 addresses the problem of sufficient and maximum likelihood estimators of the marginal distribution of one of the variables (say $y = X_i$) in the vector X. The results in Section 4 indicate that the problem of inference in presence of labels is the same as the problem of inference about marginal distributions of a multivariate distribution, and hence, it is not unique to survey sampling.

In Section 5, estimation and tests of hypothesis for functionals of F are considered. A class of parametric functions is defined to which one can apply classical inference based on asymptotic properties of Wald statistics. The results obtained are consistent with those suggested intuitively and verified empirically by many survey researchers. Some suggestions for further research are presented in Section 6. The concluding section contains personal views.

2. ESTIMATION OF A DISTRIBUTION FUNCTION

This section presents a few theorems on estimation of a distribution function. These results are very general and not dependent on any specific aspect of survey sampling.

Theorem 2.1:
If $x_1, x_2, \ldots, x_n$ are independent random (vector) observations with identical distribution function F(X), then the empirical (sampling) distribution function is the maximum likelihood estimate of F(X) and is also a sufficient statistic.

The empirical distribution function $S_n(X)$ is defined as

$$S_n(X) = \sum_{i=1}^{n} e(X, x_i)/n \tag{2.1}$$

where

$$e(X, Z) = 1, \text{ if } Z \leq X; \quad = 0, \text{ otherwise.}$$

and the relation $\leq$ is any arbitrary partial ordering on the set {X}.

Proof:
For any X_0, the number r of the n observations satisfying $x_i \leq X_0$ has binomial distribution given by

$$P(r) = \binom{n}{r} F(X_0)^r \{1 - F(X_0)\}^{n-r}$$

and hence r/n is the MLE and a sufficient statistic for $F(X_0)$.

Theorem 2.2:
If $x_1, x_2, \ldots, x_n$ is a random sample of observations whose distribution function is G{F(X)}, and G is a one to one function, then

$$\hat{F}(X) = G^{-1}\{S_n(X)\} \tag{2.2}$$

is the maximum likelihood estimate and the jointly sufficient statistic for F(X).

Proof:
By Theorem 2.1, $S_n(X)$ is the maximum likelihood of G{F(X)}. Since G is one to one, G^{-1} exists. Hence, invariance of MLE implies $G^{-1}\{S_n(X)\}$ is the MLE of F(X). The same argument applies to sufficiency of the statistic.

*B. V. Shah is Chief Scientist, Statistical Sciences Group, Research Triangle Institute, P. O. Box 12194, Research Triangle Park, North Carolina 27709.

It should be noted that the sufficiency of $S_n(X)$ is obtained for each individual value of $X = X_0$; however, sufficiency for $G^{-1}\{S_n(X)\}$ is simultaneous for the entire range of values for X.

Theorem 2.3:

Let m independent observations from an r dimensional binomial random variable $(X_1, X_2, \ldots, X_r)$ be denoted by $(x_{1\ell}, x_{2\ell}, \ldots, x_{r\ell})$; $\ell = 1, 2, \ldots, m$. If the marginal distribution of each X_h $(h = 1, 2, \ldots, r)$ is binomial with the same probability P, and if every permutation of $(X_1, X_2, \ldots, X_r)$ has the same joint probability distribution, then the MLE of P is given by

$$\hat{P} = \sum_{h=1}^{r} \sum_{\ell=1}^{m} x_{h\ell}/mr. \tag{2.3}$$

Proof:

Let $n_{i_1 i_2, \ldots, i_r}$ $(i_j = 0, 1; j = 1, 2, \ldots, r)$ be the frequency of observations such that $X_j = i_j$, and let $\pi_{i_1 i_2 \ldots i_r}$ be the corresponding hypothetical probabilities. Then due to symmetry under permutations, all π's that have an equal number of one's in their subscripts are equal.

$$\pi_{i_1 i_2 \ldots i_r} = \lambda_k; \text{ if } \Sigma i_j = k.$$

Since the sum of all π's is one,

$$\sum_{k=0}^{r} \binom{r}{k} \lambda_k = 1 . \tag{2.4}$$

Maximizing likelihood with respect to the restrictions on λ_k yields

$$\hat{\lambda}_k = \Sigma_k^* n_{i_1 i_2 \ldots i_r} / \binom{r}{k} m , \tag{2.5}$$

where Σ_k^* is summation over all $i_1 i_2 \ldots i_r$ such that $\Sigma i_j = k$.

Now

$$P = \sum_{i_j=0}^{1} \pi_{1\, i_2 i_3 \ldots i_r}$$

$$= \sum_{k=0}^{r-1} \binom{r-1}{k} \lambda_{k+1}$$

Using (2.5), one obtains

$$\hat{P} = \sum_{k=1}^{r} k\, \Sigma_k^* n_{i_1 i_2 \ldots i_r}/mr .$$

Since, each $n_{i_1 i_2 \ldots i_k}$ in Σ_k^* has k of ij's equal to 1, for each k:

$$k\, \Sigma_k^* n_{i_1 i_2 \ldots i_r} = \sum_{\ell=1}^{m} x_{k\ell}$$

This leads to:

$$\hat{P} = \sum_{h=1}^{r} \sum_{\ell=1}^{m} X_{h\ell}/mr . \tag{2.6}$$

Corollary 2.1:

If $x_1, x_2, \ldots, x_n$ are binomial observations (not necessarily independent) and if it is assumed that all permutations of $x_1, \ldots, x_n$ have the same distribution, then the MLE of P (probability that any x is one) is given by $\bar{x} = \Sigma x_i/n$; and $\bar{x}$ is sufficient for P.

Proof:

The corollary immediately follows from Theorem 2.3 with $m = 1$ and $r = n$. Given $\bar{x}$, the conditional distribution of ones and zeroes over $x_1, \ldots, x_n$ is a function of $n\bar{x}$ only. Hence, $\bar{x}$ is sufficient for P.

Theorem 2.4:

If $x_1, x_2, \ldots, x_n$ are random observations with the same random distribution function F(X), and if every permutation $y_1, y_2, \ldots, y_n$ of $x_1, x_2, \ldots, x_n$ has the same joint distribution function; the estimator

$$\hat{F}(X) = \sum_{i=1}^{n} e(X, x_i)/n = S_n(X) \tag{2.7}$$

is the maximum likelihood estimate and a sufficient statistic.

Proof:

For a given value of X, $e(X, x_i)$ are binomial variables, with probability F(X), satisfying conditions of Corollary 2.1. Simple substitution of F(X) for P and $e(X, x_i)$ for x_i yields the result.

3. APPLICATION TO SURVEYS

Let a finite population consist of the vectors $X_1, X_2, \ldots, X_N$, then its distribution function F(X) with a given partial order relation $\leq$ on the set $\{X\}$, is given by

$$F(X) = \sum_{i=1}^{N} e(X, X_i)/N . \tag{3.1}$$

If a random sample of n observations with replacement is drawn, then each x_i has random distribution F(X) and Theorem 2.1 applies.

However, for the case of with replacement sampling, if one of the variables in the vector X represents the label or the identification, then with that additional information $S_n(x)$ may not be the sufficient statistic. Consider the hypothetical population of the responses X_{ir} that could have been realized for the ith individual on rth trial $(r = 1, 2, \ldots, n)$.

In a hypothetical case, all X_{ir} with the same i are identical; in practice they may slightly differ. In any case, the assumption is that the variance of $X_{i1}, X_{i2}, \ldots, X_{in}$ is small (equal to zero in hypothetical case). This assumption is equivalent to treating $X_{i1}, X_{i2}, \ldots, X_{in}$ as a cluster and assuming that within cluster variance is very small or equal to zero.

The utilization of information about clusters is well known to survey statisticians (Neyman 1934). If the within cluster variance is assumed to be zero, then optimal use of resources will require only one unit per cluster, or sampling without replacement. Of course, there may be additional information in labels beyond clustering (multiple observations with identical labels). The discussion of the other potential information is deferred to Section 4. The remainder of this section addresses sampling without replacement.

Theorem 3.1:

Let $x_1, x_2, \ldots, x_n$ be a random sample from a finite population without replacement and $P(x_i)$ be the probability that the observation x_i is included in the sample. Let us assume that the inclusion probability is greater than zero for each unit in the population, and that labels and any other variables of interest are included in the vector X. If the order of selection of $x_1, x_2, \ldots, x_n$ is ignored, then the Horvitz-Thompson estimator:

$$\hat{F}(X) = \left\{ \sum_{i=1}^{n} e(X, x_i)/P(x_i) \right\} \Big/ \left\{ \sum_{i=1}^{n} 1/P(x_i) \right\} \tag{3.2}$$

is the MLE of the population distribution function F(X), and $\hat{F}(X)$ is jointly sufficient for F(X).

Proof:

Let $y_1, y_2, \ldots, y_n$ be a random permutation of $x_1, x_2, \ldots, x_n$. Then ignoring the selection order of $x_1, x_2, \ldots, x_n$ is equivalent to assuming that inference under $x_1, x_2, \ldots, x_n$ is identical to inference under $y_1, y_2, \ldots, y_n$. The vectors $y_1, y_2, \ldots, y_n$ have identical joint distribution under permutation and, hence, satisfy the conditions of Theorem 2.4. It follows that $S_n(y)$ is the MLE and the sufficient statistic for estimation of the distribution function of each y_i.

The distribution function of each y_i is given by

$$G(Y) = \left\{\sum_{i=1}^{N} e(Y, X_i)/P(X_i)\right\} \Big/ \left\{\sum_{i=1}^{N} 1/P(x_i)\right\}. \quad (3.3)$$

The distribution function of interest is F(X) given by Equation (3.1). It can be shown that if P(X) is greater than zero, then the relationship between F and G is one to one. Hence, by Theorem 2.2, $G^{-1}\{S_n(Y)\}$ is the MLE and jointly sufficient statistic of F(Y),

$$\hat{F}(Y) = \left\{\sum_{i=1}^{n} e(Y, y_i)/P(y_i)\right\} \Big/ \left\{\sum_{i=1}^{n} 1/P(y_i)\right\}. \quad (3.4)$$

Due to invariance under permutation, Y's can be replaced by X's to obtain the result (3.2).

This completes the proof.

The following observations need to be emphasized for practical implications of Theorem 3.1.

a. The theorem is valid for any without replacement sample design provided inclusion probabilities are known and the inference space is repeated application of the same design to the same population. As a further illustration, an alternate approach to multistage designs is presented in the Appendix.

b. The statistic $\hat{F}$ is jointly sufficient for F. It is not true in general that a marginal or a set of marginals of $\hat{F}$ is sufficient for the corresponding marginals of F.

c. It should be noted that the conditions of Theorem 3.1 appear to be similar to the exchangeable prior of Ericson (1969). This concept has been used in finite populations by Madow and Madow (1944) and Kempthorne (1969), who have considered invariance under permutations of the population values. Theorem 3.1 requires invariance under permutations of the observed sample values only.

4. INFORMATION AND LABELS

Consider the bivariate distribution for the finite population Y_i, Z_i; i = 1,2,...W, one of which may represent labels. For a specific value of Y_0 and Z_0, define sets A, B, as follows:

$$A = \{Y_i \le Y_0\}$$

$$B = \{Z_i \le Z_0\}$$

the sets $\tilde{A}$ and $\tilde{B}$ are complements of A and B, respectively. Let the number of units M_{ij} in a finite population be as in the following bivariate table by A and B.

	B	$\tilde{B}$	
A	M_{11}	M_{12}	$M_{1.}$
$\tilde{A}$	M_{21}	M_{22}	$M_{2.}$
	$M_{.1}$	$M_{.2}$	$M_{..}$

If the results of a simple random sample are denoted by small m_{ij} in a similar table, the hypergeometric distribution yields the probability

$$P(m_{11}, m_{12}, m_{21}, m_{22}) = \pi\binom{M_{ij}}{m_{ij}} \Big/ \binom{M}{m}.$$

Assume that the parameter of interest is the marginal probability P for Y, given by

$$P_{1.} = M_{1.}/M_{..}$$

The maximum likelihood estimate of $P_{1.}$ is given by

$$\hat{P}_{1.} = m_{1.}/m_{..},$$

but $\hat{P}_{1.}$ is not a sufficient statistic. The conditional distribution of m_{ij}'s given $\hat{P}_{1.}$ is

$$P(m_{11}, m_{12}, m_{21}, m_{22}/P_{1.}) = \pi\binom{M_{ij}}{m_{ij}} \Big/ \pi\binom{M_{i.}}{m_{i.}},$$

which is not independent of $M_{1.}$ or $P_{1.}$. Of course, it is well known that $m_{11}, m_{12}, m_{21}, m_{22}$ are jointly sufficient for all parameters.

The nonacceptance of the estimator $\hat{P}_{1.}$ because it is not sufficient may seem to imply that there exists a "better" estimator which is a function of m_{ij}. If this were true for any variable in place of Z, the conclusion would be $\hat{P}_{1.}$ should be dependent on all variables that are observed in a survey.

Consider a hypothetical case where a hand-held calculator is available and produces a psuedorandom number between (0, 1) when a special key is pressed. It is assumed that the generated random numbers are all distinct. In a survey the following two variables are measured: height (Y) of the respondent and the number (Z) shown on the calculator when the respondent presses the random number key. A logical paradox, in attempting to use "all" the information about Y and Z while estimating the marginal distribution Y, is self evident in this case. Commonsense logic dictates that Z is irrelevant and hence $\hat{P}_{1.}$ has all the information; the sufficiency principle applied to likelihood implies that $\hat{P}_{1.}$ is not sufficient for $P_{1.}$, but we cannot logically conclude that Z has "some" information on $P_{1.}$. The unanswered question is: "Is there a criterion (besides common sense about the variables) that may determine relevance of the information on one variable while estimating the marginal distribution of the other?"

Of course, there may be some variables that are very relevant; for example, the height (Y) of a person and the sex (z) of a person. The practical operating principle is: "The information about all other variables is irrelevant, unless hypothesized or assumed otherwise, while estimating the marginal distribution of a variable."

The following summary points are worth noting:

a. Labels can be treated as values of a variable and the question of their relevance should be addressed in the same spirit as other variables like sex or a random number.

b. The sufficiency principle alone cannot resolve whether a given variable is "relevant" or not for estimating the marginal of another variable.

c. If the only assumption is that Z is relevant for Y, and no other assumptions or information are available, then the maximum likelihood estimate for the marginal distribution of Y is the same as the one produced by ignoring Z.

d. This problem does not arise in infinite populations because the $\hat{P}_{1.}$ is always the sufficient statistic. If N is very large, then the distribution of m_{ij} is approximated by the multinomial distribution and $\hat{P}_{1.}$ is sufficient for P in the multinomial distribution.

5. ESTIMATION AND TESTS OF HYPOTHESIS

Since MLE's are invariant under all functional transforms, and $\hat{F}(X)$ is the MLE of F(X), the MLE of any function of F(X) is the corresponding function of $\hat{F}(X)$. The mean of X (if X is numerical) is defined by

$$\mu = \int X dF(X), \tag{5.1}$$

and is estimated by

$$\hat{\mu} = \int X d\hat{F}(X). \tag{5.2}$$

The variance-covariance matrix of X is defined by

$$V = \int (X - \mu)^2 dF(X), \tag{5.3}$$

with the MLE given by

$$\hat{V} = \int (X - \hat{\mu})^2 d\hat{F}(X). \tag{5.4}$$

In general, any parameter that is expressed as a functional of the population distribution F(X) is estimated by the corresponding functional of the Horvitz-Thompson estimator $\hat{F}(X)$.

Furthermore, this formulation presents an opportunity to construct a large sample Wald statistic for a vector of parameters $\theta = (\theta_1 \ldots \theta_m)$ which are functions of F and are estimated by corresponding functions of $\hat{F}$. Such tests can be treated as a first approximation for testing hypotheses about finite populations from survey data.

The most commonly used test statistic in practice is

$$t = (\bar{x}_w - \mu)/(\hat{S}(\bar{x}_w)) \tag{5.5}$$

where $\bar{x}_w$ is the weighted sample mean or Horvitz-Thompson estimator of the population mean μ and $\hat{S}(\bar{x}_w)$ is a consistent estimator of the standard error of $\bar{x}_w$. If the sample size is large, it is generally concluded that the distribution t is approximated by Student's t-distribution. This result follows from asymptotic normality of MLE.

Kish and Frankel (1974) present simulation results for such tests for univariate hypothesis. For multivariate hypothesis, such a test was suggested by Koch, et al. (1975) for linear models of proportions in contingency tables.

Tests of hypotheses for statistics for linear models from survey samples have been considered by Shah, et al. (1977) and Fuller (1976). Rao and Scott (1979) and Holt et al. (1980) contain references to many other empirical investigations concerning tests for population proportions in contingency tables. These investigations confirm that for large samples, the Wald statistic using traditional sample survey techniques (based on Horvitz-Thompson estimators), is distributed approximately as a χ^2 variate with appropriate degrees of freedom.

The major difficulty in tests of hypothesis appears to be the need for defining parameters in terms of the distribution function F. The population total (T) is not expressible as a functional of the distribution function F(X), and hence, it is not directly estimable under the approach presented so far. If N is known, total can be estimated as $N\hat{\mu}$. If N is not known, it can be estimated from inclusion probabilities because

$$E(\Sigma 1/p_i) = N$$

$$\hat{T} = \Sigma(1/p_i)\hat{\mu}. \tag{5.6}$$

The distribution of inclusion probabilities created by a sampling statistician is known prior to observing the values of X. If the sample size n is variable, then the conditional expectation of the weighted mean given n is $\hat{\mu}$, but the equivalent statement is not true for $\hat{T}$ given in (5.6).

6. FURTHER RESEARCH

Theoretical support for the pivotal quantities based on functionals of the distribution function is provided in this paper, as well as empirical results cited in the previous section for the Wald statistic. Many results for the distribution function and its functionals have been derived for continuous distribution functions (see Boos and Serfling 1980).

If S_n is an empirical distribution function based on a simple random sample, and F is the true distribution function, then Kendall and Stuart (1967, Vol. 2, p. 451) state:

> "For each value of x, from the Strong Law of Large Numbers,
>
> $$\lim_{n \to \infty} P\{S_n(x) = F(x)\} = 1, \tag{30.98}$$
>
> and in fact stronger results are available concerning the convergence of sample distribution function to the true distribution function. In a sense, (30.98) is the fundamental relationship on which all statistical theory is based. If something like it did not hold, there would be no point in random sampling."

This long-run property is not considered quite compelling for inference on each specific occasion by many statistical philosophers. For example, Hacking (1965, p. 41) states: "Not only does the long-run rule not formally imply the unique case rule, but also there is no valid way of inferring the one from the other."

A distance function defined by Kolmogoroff (1941) is well known. Exact distribution of Kolmogoroff's distance has been obtained by Birnbaum and Tinger (1951). Walsh (1962, p. 308) has presented a method for obtaining bounds for confidence intervals in the discontinuous case. Kempthorne (1969) has suggested use of $\int (F - \hat{F})^2 dFx$ as a pivotal quantity for inference about F. Exact results or better approximation for the discontinuous case are needed. The next question concerns the possibility of translating confidence intervals on F into confidence intervals for functionals of F, and determining the impact of selection probabilities on these limits. Functional representation of statistical parameters and their asymptotic properties were first studied by Von Mises (1947); more recent references can be found in Boos and Serfling (1980).

The common approach is to consider the asymptotic results regarding MLE (Wald 1943). Sprott (1975) has considered applicability of asymptotic normality in finite samples. Empirical results suggest that these results are valid for a wide variety of cases. There is a need to determine the set or sets of conditions under which the Wald statistic is applicable to samples from finite populations. The problem of estimation of variance-covariance matrices of the functions of MLE needs further investigation.

REFERENCES

Basu, D. (1969), "Role of the Sufficiency and Likelihood Principles in Sample Survey Theory," *Sankhyā* A, 31, 441-454.

__________, (1971), "An Essay on the Logical Foundations of Survey Sampling, Part One," in V. P. Godambe and D. A. Sprott, Eds., *Foundations of Statistical Inference*. Toronto: Holt, Rinehart and Winston, 203-242.

__________, (1977), "Relevance of Randomization in Data Analysis," in N. Krishnan Namboodiri, Ed., *Survey Sampling and Measurement*, Academic Press, New York, 267-292.

Birnbaum, Z. W. and Tingey, F. H. (1951), "One-sided Confidence Contours for Probability Distribution Functions," *Annals of Mathematical Statistics*, 22, 592-596.

Bishop, Y. M. M. (1969), "Full Contingency Tables, Logits, and Split Contingency Tables," *Biometrics*, 25, 383-399.

Boos, D. D. and Serfling, R. J. (1980), "A Note on Differentials and the CLT and LIL for Statistical Functions with Applications to M-estimates," *Annals of Statistics*, 8, 618-624.

Cassel, Särndal, C. E. and Wretman, J. H (1977), *Foundations of Inference in Survey Sampling*, John Wiley and Sons, New York.

Des Raj and Khamis, S. H. (1958), "Some Remarks on Sampling with Replacement," *Annals of Mathematical Statistics*, 29, 550-557.

Ericson, W. A. (1969), "Subjective Bayesian Models in Sampling Finite Populations," *Journal of the Royal Statistical Society* B, 31, 195-224.

Fuller, W. A. (1974), "Regression Analysis for Sample Surveys," report prepared for the U. S. Bureau of the Census on work conducted under the Joint Statistical Agreement, Iowa State University, Ames, Iowa.

Godambe, V. P. (1955), "A Unified Theory of Sampling from Finite Populations," *Journal of the Royal Statistical Society* B, 17, 269-278.

__________, (1966), "A New Approach to Sampling from Finite Populations I, II," *Journal of The Royal Statistical Society* B, 28, 310-328.

Hacking, I. (1965), *Logic of Statistical Inference*. Cambridge University Press, Cambridge.

Hansen, M. H., Hurwitz, W. N. and Pritzker, L. (1963), "The Estimation and Interpretation of Gross Differences and the Simple Response Variance," in C. R. Rao, Ed., *Contributions to Statistics*, London: Pergamon Press and Calcutta: Statistical Publishing Society.

Hartley, H. O. and Rao, J. N. K. (1969), "A New Estimation Theory for Sample Surveys, II," in N. L. Johnson and H. Smith, Eds., *New Developments in Survey Sampling*, New York: Wiley-Interscience, 147-169.

Holt, D., Scott, A. J. and Ewings, P. D., (1980), "Chi-squared Tests with Survey Data," *Journal of the Royal Statistical Society* A, 143, 303-320.

Horvitz, D. G. and Thompson, D. J. (1952), "A Generalisation of Sampling Without Replacement from a Finite Universe," *Journal of the American Statistical Association*, 47, 663-685.

Kempthorne, O. (1952), *The Design and Analysis of Experiments*, John Wiley and Sons, New York.

__________, (1955), "The Randomization Theory of Experimental Inference," *Journal of the American Statistical Association*, 50, 946-967.

__________, (1969), "Some Remarks on Statistical Inference in Finite Sampling," in N. L. Johnson and H. Smith, Eds., *New Developments in Survey Sampling*, New York: Wiley-Interscience, 671-695.

Kendall, M. G., and Stuart, A. (1961), *The Advanced Theory of Statistics*, II, Charles Griffin and Company, London.

Kish, L. and Frankel, M. R. (1974), "Inference from Complex Samples," *Journal of the Royal Statistical Society*, B, 36, 1-37.

Koch, G. G., Freeman, D. H., Jr., and Freeman, J. L. (1975), "Strategies in the Multivariate Analysis of Data from Complex Surveys," *International Statistical Review*, 43, 59-78.

Kolmogoroff, A. N. (1941), "Confidence Limits for an Unknown Distribution," *Annals of Mathematical Statistics*, 12, 461-463.

Lessler, J. T. (1979), "An Expanded Error Model," in *Symposium on Incomplete Data: Preliminary Proceedings*, U. S. Department of Health, Education, and Welfare, Social Security Administration, Washington, D. C.

Madow, W. G. and Madow, L. H. (1944), "On the Theory of Systematic Sampling," *Annals of Mathematical Statistics*, 15, 1-24.

Namboodiri, K. N., Ed. (1978), *Survey Sampling and Measurement*, Academic Press, New York.

Neyman, J. (1934). "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and The Method of Purposive Selection," *Journal of the Royal Statistical Society*, 97, 558-625.

Parzen, E. (1962), "On Estimation of a Probability Density and Mode," *Annals of Mathematical Statistics*, 33, 1065-1076.

Rao, J. N. K. and Scott, A. J. (1979), "Chi Square Tests for Analysis of Categorical Data from Complex Surveys," *Proceedings of the American Statistical Association*.

Royall, R. M. (1968), "An Old Approach to Finite Population Sampling Theory," *Journal of the American Statistical Association*, 63, 1269-1279.

__________, (1970), "Finite Population Sampling—on Labels in Estimation," *Annals of Mathematical Statistics*, 41, 1774-1779.

Scott, A. J. and Smith, T. M. F. (1969), "Estimation in Multistage Surveys," *Journal of the American Statistical Association*, 64, 830-840.

Seeger, Paul (1970), "A Method of Estimating Variance Components in Unbalanced Designs," *Technometrics*, 12, 207-218.

Shah, B. V., Holt, M. M. and Folsom, R. E. (1977), "Inference About Regression Models from Sample Survey Data," *Bulletin of the International Statistical Institute*, 47(3), 43-57.

Sirken, Monroe G. (1970), "Household Surveys with Multiplicity," *Journal of the American Statistical Association*, 257-266.

Smith, T. M. F. (1976), "Foundations of Survey Sampling: A Review," *Journal of the Royal Statistical Society* A-139, 183-204.

Sprott, D. A. (1975), "Application of Maximum Likelihood Methods to Finite Samples," *Sankhyā* B-37, 259-270.

Von Mises, R. (1947), "On Asymptotic Distributions of Differentiable Statistical Functions," *Annals of Mathematical Statistics*, 18, 309-348.

Wald, A. (1943), "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54, 426.

Walsh, J. E. (1962), *Handbook of Nonparametric Statistics*, Von Nostrand, Princeton, New Jersey.

ALTERNATE APPROACH FOR MULTISTAGE DESIGNS

Theorem 3.1 is valid for any design; however, the alternate approach presented in this appendix clearly illustrates a rationale for stratification and clustering in survey practice.

In the case of stratified random samples, let there be H strata and F_h $(h = 1, 2, \ldots, H)$ be the distribution function of the hth strata. The population distribution function F is then a weighted average of the F_h given by

$$F = \Sigma N_h F_h / N,$$

where N_h is size of the hth strata, and N is the population size.

Since samples are drawn independently in each stratum, the sample distribution function f_h is the MLE of F_h. Hence, we have

$$\hat{F} = \Sigma N_h \hat{f}_h / N \ .$$

To extend our model to a two-stage sample with random selection, let there be H primary sampling units (PSU's), m of which are selected at random. From hth PSU, m_h individuals are selected and a value of X observed for each individual. Sample selection within each PSU is independent of the selection within any other PSU.

Now assume that the distribution function of X within the hth PSU is $F_h(X)$. Further assume that for each X, the $Y_h = F_h(X)$ $(h = 1, 2, \ldots, H)$ have a random distribution G(Y).

The population distribution function can now be written as

$$F(X) = \int Y dG(Y) \ .$$

The following conclusions follow: If f_h is a MLE of F_h, and *if* g(y) is the sample distribution function formed by values of f_h, $h = 1, 2, \ldots, m$, then g(y) is a maximum likelihood estimator of G(Y). Hence, the MLE of F is given by

$$\hat{F}(X) = \int y dg(y).$$

The reader can satisfy oneself by proving that estimates in this appendix are identical to the Horvitz-Thompson estimator in Theorem 3.1.