A DIFFERENT VIEW OF FINITE POPULATION ESTIMATION

Cathy Campbell, Arthur D. Little, Inc.

A. Introduction

General results concerning the behavior of nonlinear statistics calculated from complex survey are difficult to obtain although their need is well-documented (Kish and Frankel 1974). Even the presently available asymptotic results on variance estimation and asymptotic normality are limited by restrictive assumptions concerning independence of primary selections or equal selection probabilities.

In this work we outline how the theory of von Mises (1974) statistical functionals can be applied to finite population sampling to obtain some very general results concerning the asymptotic behavior of nonlinear estimators. Particular attention will be given to the Taylor and jackknife variance estimators. This effort stems from the work of Hinkley (1978) and Jaeckel (1972) who have shown that the von Mises expansion may be fruitfully employed to study the Taylor and jackknife variance estimators for iid data. First we present a brief review of the theory and then discuss estimating a distribution function for a finite population. The applicability of this theory to finite population sampling is then demonstrated. Finally we give a general definition of a jackknife pseudovalue that is applicable for unequal probability sampling, and generalize the jackknife variance estimate so that it applies to any sample design for which the variance of an estimated mean can be estimated.

B. Theory of Statistical Functionals

In this section we draw primarily on the work of Reeds (1976) while remembering that earlier, less general survey articles on von Mises functionals are the origninal work by von Mises (1947) and a later paper by Filippova (1962). Initially we consider the situation for which $X_1,\ldots, X_n$ (possibly vector-valued) are iid random variables with cumulative distribution function F. Let F_n be the empirical distribution function of $X_1,\ldots, X_n$. We consider only statistics T_n which can be expressed as $T_n=T(F_n)$. Some simple examples are

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \int x dF(x)$$

$$\frac{n-1}{n} s^2 = 1/2 \iint (x-y)^2 dF(x)dF(y).$$

Sample quantiles, maximum likelihood estimates, and many rank statistics can also be expressed in this format. The k^{th} order von Mises approximation of T_n can be expressed as

$$T(\hat{F}) = T(F) + \int t_1(x)dF_n(x) + \ldots$$
$$+ 1/2 \iint t_2(x,y)dF_n(x)dF_n(y)$$
$$+\cdots+ \frac{1}{k!}\int\cdots\int t_k(x_1,\ldots,x_k)dF_n(x_1)\ldots dF_n(x_k),$$

where $t_k(x_1\ldots,x_k)$ is the k^{th} von Mises derivative of the functional $T(\cdot)$ evaluated at F. We have adopted the convention that

$$E_{x_i}\left[t_k(X_1,\ldots,X_k) \mid X_j=x_j \quad j \neq i\right] = 0$$

In this case, the first derivative $t_1(\cdot)$ corresponds to the influence function of $T(\cdot)$ and can be defined as

$$\frac{d}{dt} T\left[(1-t)F + tG\right]_{t=0} = \int t_1(x)dG(x)$$

subject to the restriction $\int t_1(x)dF(x)= 0$.

Results of von Mises, Filippova, and Reeds have been directed primarily toward finding conditions on F and $T(\hat{F})$ so that the asymptotic distribution of $\sqrt{n}\,(T(\hat{F})-T(F))$ is the same as the asymptotic distribution $\sqrt{n}\int t_1(x)dF_n(x)$, the first-order term of the expansion. The conditions as given by Reed are

(i) an analytical conditon: $T(\cdot)$ is compactly differentiable at F.

(ii) stochastic conditions: $F_n = F + 0_p(1\sqrt{n})$, and the remainder term is a measurable random variable.

The differentiability condition will always be assumed in this work.

When $X_1,\ldots,X_n$ are iid random variables,

$$\int t_1(x)dF_n(x) = \frac{1}{n}\sum_{i=1}^{n} t_1(X_i).$$

The usefulness of the first-order approximation is immediately apparent: $\frac{1}{n}\sum_{i=1}^{n} t_1(X_i)$ is the sample mean of n iid random variables, hence

$$\sqrt{n}\int t_1(x)dF_n(x) \underset{L}{\to} N(0, \sigma_t^2).$$

A natural "estimator" of σ_t^2 would be

$$\tilde{\sigma}_t^2 = \frac{1}{n}\sum_{i=1}^{n} t_1^2(X_i)$$

but the realization of $t_1(X_i)$ is not known because it depends on the unknown distribution function F. To make this dependence explicit, we use the notation $t_1(X_i;F)$.

Different general methods of estimating the asymptotic variance of $T(F_n)$ have been proposed. Two of these, the Taylor or δ-method and the jackknife method can be better understood and compared via the von Mises expansion.

Each method relies on the use of a different predictor of $t_1(X_i;F)$. Hinkley (1978) has studied the jackknife variance estimator via the von Mises expansion and has shown that

$$P_i \sim T(F) + t_1(X_i;F)$$

as $n \to \infty$, where P_i is the i^{th} pseudovalue. The implied predictor of $t_1(X_i; F)$ is given by

either $\quad \tilde{t}_1(X_i;F) = P_i - T(F_n)$

or $\quad \tilde{\tilde{t}}_1(X_i;F) = P_i - \frac{1}{n} \sum_{i=1}^{n} P_i$.

Jaeckel (1972), and Gray, Schucany, and Watkins (1975), have shown, in different ways, that the δ-method variance estimator can be obtained as a limiting special case of the jackknife method when the weight on the "omitted" observation is $\frac{1}{n} - \varepsilon$ instead of 0. The limit is taken as $\varepsilon \to 0$. Jaeckel called this estimation procedure the infinitesimal jackknife and showed that this procedure leads to

$$\hat{\sigma}_t^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{t}_1^2(X_i;F)$$

where

$$\hat{t}_1(X_i;F) = t_1(X_i;F_n)$$

is the influence function evaluated at the empirical cdf F_n.

Most of the results on jackknife variance estimation and use of von Mises functionals have been developed for independent, indentically distributed random variables. A few exceptions do exist to the assumption of identical distribution: von Mises (1947), Arvesen (1969), Miller (1974), and Hinkley (1977). Nevertheless, the widespread use of Taylor and jackknife-type variance estimators in survey sampling led us to wonder if the theory of von Mises functionals has any general application in finite population sampling. Our initial attempt (Campbell, 1979) to apply this theory relied upon von Mises (1947) theorem which gave asymptotic results for $X_1,\dots,X_n$ independent, but not identically distributed. The results there applied only to a specific stratified sample design with independent primary selections within strata.

By using Reeds' more general formulation we have now developed a very general theorem giving the asymptotic behavior of nonlinear estimators for finite population sampling with unequal probabilities and without replacement. We have also developed a jackknife variance estimator for use in these same situations.

In the following section we examine implicit estimates of a distribution function in the finite population context. Then we discuss asymptotic behavior of parameter estimates, and finally develop a finite population jackknife.

C. Distribution Functions for Finite Populations

1. Population Distribution Function

To apply the von Mises theory it was first necessary to ascertain that finite population parameters could be written as functionals T(F) of some "distribution function" F and their corresponding estimators as $T(\hat{F})$, where $\hat{F}$ is an estimator of F.

Let $Y_1,\dots,Y_N$ be the values of a variable that are attached to the N elements of a finite population. The finite population version of F is found by noting that:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \int y dF(y)$$

where

$$F(y) = \frac{\# Y_i \le y}{N} \quad .$$

For notational simplicity we consider only scalar-valued Y_i since the multivariate generalization of F(y) is readily apparent. Simple inspection allows one to verify that other parameters such as ratios, variances, regression and correlation coefficients, can also be expressed as functionals of the appropriate distribution functions.

2. Implied Estimates of F

We examined estimators of $\bar{Y}$ for a variety of sample designs in order to determine what estimator of F is implied in each case. This was done by expressing each estimator of $\bar{Y}$ as $\hat{\bar{Y}} = \int y d\hat{F}$ to determine an expression for $\hat{F}$. As these examples are taken from Cochran (1977), we have followed his notation. In most cases, the implied estimator of F was intuitively appealing, but a few more interesting cases were also discovered.

a. Simple Random Sample

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$\hat{F}(y) = \frac{\# y_i \le y}{n}, \text{ the empirical cdf.}$$

b. Single Stage Cluster Sample

(i) Simple random sample of equal-sized clusters.

$$\hat{\bar{Y}} = \frac{1}{nM} \sum_{i=1}^{n} \sum_{j=1}^{M} y_{ij}$$

$$\hat{F}(y) = \frac{1}{n} \sum_{i=1}^{n} F_i(y)$$

where F_i is the distribution function for the M elements in cluster i.

(ii) Simple random sample of unequal-sized clusters.

The simple inflation estimator of $\bar{Y}$ is

$$\hat{\bar{Y}} = \frac{1}{M_o} \frac{N}{n} \sum_{i=1}^{n} M_i \bar{Y}_i$$

$$\hat{F}(y) = \frac{1}{M_o} \frac{N}{n} \sum_{i=1}^{n} M_i F_i(y)$$

Note that $\hat{F}(\infty) = \int d\hat{F}(y) = \dfrac{\sum_{i=1}^{n} M_i/n}{M_o/N}$ which is not, in general, equal to 1. If we standardize $\hat{F}(y)$ by dividing it by its integral, we obtain

$$\hat{F}_s(y) = \frac{\sum_{i=1}^{n} M_i F_i(y)}{\sum_{i=1}^{n} M_i} .$$

The corresponding standardized estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_s = \frac{\sum_{i=1}^{n} M_i \bar{Y}_i}{\sum_{i=1}^{n} M_i} ,$$

the usual ratio estimator of $\bar{Y}$.

(iii) PPS sample of clusters

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_i$$

$$\hat{F}(y) = \frac{1}{n} \sum_{i=1}^{n} F_i(y)$$

(iv) General unequal probability sample of clusters

Let $\pi_i = nz_i$ be the probability that cluster i enters the sample and let $W_i = M_i/M_o$ be the relative size of cluster i.

$$\hat{\bar{Y}} = \frac{1}{n} \sum_{i=1}^{n} \frac{W_i}{Z_i} \bar{Y}_i$$

$$\hat{F}(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{W_i}{Z_i} F_i(y).$$

Again, $\int d\hat{F} \neq 1$ and a standardized $\hat{F}$ can be found, as before.

c. Stratified Sample

$$\hat{\bar{Y}} = \sum_{i=1}^{L} W_i \bar{Y}_i$$

$$\hat{F}(y) = \sum_{i=1}^{n} W_i \hat{F}_i(y)$$

where $\hat{F}_i(y)$ is an estimator of F_i, the cdf for the elements in stratum i. The actual form of $\hat{F}_i(y)$ depends on the within-stratum sample design.

d. Multi-stage Sample

Implied estimators are obtained by replacing $F_i(y)$ with $\hat{F}_i(y)$ in the corresponding single-stage estimator. The form of $\hat{F}_i(y)$ depends on the within-cluster sample design.

e. General Unequal Probability Sample of Elements

$$\hat{\bar{Y}} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{\pi_i},$$ the Horvitz-Thompson estimator

$$\hat{F}(y) = \frac{1}{N} \sum_{\{i: y_i \leq y\}} \frac{1}{\pi_i}$$

This is another estimator of F with $\int d\hat{F} \neq 1$. Standardizing $\hat{F}$ by dividing by its integral gives

$$\hat{F}_s(y) = \frac{\sum_{\{i: y_i \leq y\}} \frac{1}{\pi_i}}{\sum_{i=1}^{n} \frac{1}{\pi_i}}$$

The corresponding estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_s = \frac{\sum_{i=1}^{n} \frac{y_i}{\pi_i}}{\sum_{i=1}^{n} \frac{1}{\pi_i}}$$

which is ratio-type estimator that we have not seen before.

3. Direct Estimation of F

For each element in the population, define a collection of variables

$$Z_i(y) = \begin{cases} 1 & Y_i \leq y \\ 0 & Y_i > y \end{cases} \qquad y \in (-\infty, \infty) .$$

Then for each y, $F(y) = \bar{Z}(y)$, the population mean of the variable $Z(y)$. Standard methods for estimating population means can, therefore, be used to estimate F(y) for any y. In each case, the direct estimate for a particular design and estimation procedure will be the same as the implied estimate given earlier.

While it is seldom of interest to estimate F(y) directly, expressing $\hat{F}(y)$ as an estimator of a mean allows easier study of its properties. For instance, if the sample design and estimator are such that

$$E(\hat{\bar{Y}}) = \bar{Y},$$

then

$$E[\hat{F}(y)]=F(y).$$

More importantly for our purposes, if

$$\hat{\bar{Y}} \underset{p}{\to} \bar{Y} \quad \text{as} \quad n,N\to \infty,$$

then

$$\hat{F}(y) \underset{p}{\to} F(y) \quad \text{as} \quad n,N\to \infty.$$

In general, the estimators of $F(y)$ assume all the properties of their companion estimators of population means.

D. Asymptotic Distribution of Nonlinear Estimators

We are now in a position to investigate the conditions under which

$$\sqrt{n}\,[T(\hat{F})-T(F)] \underset{L}{\to} \sqrt{n} \int t_1(y)d\hat{F}(y) \text{ as } n,N\to \infty.$$

The conditions, as stated in the following theorem, are quite general.

Theorem: Let the sample design and an estimator of $\bar{Y}$ be given. Let $\hat{F}$ be defined implicitly via

$$\hat{\bar{Y}}= \int y d\hat{F}.$$

If

(i) $\hat{\bar{Y}} \underset{p}{\to} \bar{Y}$ as $n,N \to \infty$

(ii) $V(\hat{\bar{Y}})= O\left(\frac{1}{n}\right)$ as $n,N\to \infty$

(iii) $T(\cdot)$ is compactly differentiable at F, then

$$\sqrt{n}[T(\hat{F})-T(F)] \underset{L}{\to} \sqrt{n} \int t_1(y)d\hat{F}(y) \text{ as } n,N\to \infty.$$

The proof follows directly from Theorem 4.2.1 of Reeds (1976), which gives, in a very general setting, three conditions that must be satisfied for the remainder term of the Taylor expansion to converge to 0. These conditions, appropriately reworded for the application here, are

(i) T is compactly differentiable

(ii) $\hat{F}$ converges uniformly to F and $V(\hat{F}(y))= O\left(\frac{1}{n}\right)$

(iii) The remainder term is a measurable random variable.

Condition (i) is an assumption of our theorem and condition (iii) is trivial.

The uniform convergence of $\hat{F}$ follows from two observations. Pointwise convergence of $\hat{F}(y)$ to $F(y)$ is guaranteed by the convergence of $\hat{\bar{Y}}$ to $\bar{Y}$, as noted in the previous section. Lemma 8.2.3 of Chow and Teicher (1978), states that pointwise convergence of $\hat{F}(y)$ to $F(y)$ for all $y \in (-\infty, \infty)$ implies that $\hat{F}$ converges uniformly to F in $(-\infty,\infty)$.

The theorem was, in a sense, already proved, because the very general framework used by Reeds does not involve the usual restrictive assumption that the data $y_1,\ldots,y_n$ are independent, identically distributed observations from a distribution F. The key, in this setting, is that the sequence of functions $\hat{F}$ converges uniformly and at the proper rate to a limit function F. Whether or not these are true distribution functions is irrelevant.

The power of the theorem lies in the weakness of the assumptions. Sampling with and without replacement, equal and unequal probability sampling all fall within its scope.

As in the case of iid random variables, the asymptotic behavior of $T(\hat{F})$ is the same as the asymptotic behavior of the estimated mean of the variable $t_1(Y_i;F)$. This can be seen by observing that $\hat{\bar{Y}} = \int yd\hat{F}(y)$ and $T(\hat{F}) - T(F)\sim \int t_1(y;F)d\hat{F}(y)$. An immediate implication of this fact is that $T(\hat{F})$ will usually have an asymptotically normal distribution for same sample designs that $\hat{\bar{Y}}$ does.

Some specific results concerning asymptotic normality of $\hat{\bar{Y}}$ are available. See Hajek (1960, 1964), Rosen (1972 a,b), and Holst (1973). With these results we can verify the asymptotic normality of $\hat{\bar{Y}}$ for several situations and hence assert the asymptotic normality of $T(\hat{F})$ for these same situations. Use of a ratio estimator for $\bar{Y}$, and hence for F, poses no problem here. If both the numerator and denominator, which are estimates of means, are normally distributed then application of the asymptotic approximation yields asymptotic normality and consistency for their ratio.

In the following section, we examine variance estimation for non-linear estimators $T(\hat{F})$.

E. Estimation of Asymptotic Variance of $T(\hat{F})$

We shall now assume that

$$T(\hat{F})-T(F) \sim \int t_1(y)d\hat{F}(y) = \sum_{i=1}^{n} w_i t_1(y_i;F),$$

i.e., that (n,N) are large enough that the asymptotic approximation is valid. We shall also assume that

$\sum_{i=1}^{n} w_i \equiv 1$ although only the consistency of $\hat{F}$ is necessary under our conditions. With the above formulation, one can see that the asymptotic variance of $T(F)$ is simply the variance of the estimated mean of $t_1(y_i;F)$.

The important properties of $t_1(\cdot;F)$, for purposes are:

(i) its form depends only on the parameter $T(F)$ being estimated and is independent of the sample design;

(ii) the value $t_1(Y_i;F)$ is well-defined, though unknown, for every element in the population

(iii) the population average of $t_1(Y_i;F)$ is equal to 0.

Although the function $t_1(\cdot;F)$ is often called the influence function, that name is not entirely appropriate for finite population sampling since the influence of an observation depends on its selection probability as well as the value of the observation.

Following the logic presented in section B, suppose

$\hat{\bar{Y}} = \sum_{i=1}^{n} w_i y_i$, and that $v(\hat{\bar{Y}}) = h(y_1, \ldots, y_n)$

is the proposed estimator of $V(\hat{\bar{Y}})$.

Since

$T(\hat{F}) - T(F) \sim \sum_{i=1}^{n} w_i t_1(y_i;F)$, we would

propose using

$$v(T(\hat{F})) = h\,(t_1(y_1;F), \ldots, t_1(y_n;F))$$

to estimate $V(T(\hat{F}))$ if the arguments were known. The function $h(\cdot, \ldots, \cdot)$ does depend on the sample design and will usually be an unbiased and/or consistent variance estimator for the particular sampling plan.

As pointed out earlier, the Taylor and jackknife methods of variance estimation can be differentiated by the implied estimates of $t_1(y_i;F)$ that are substituted in $h(\cdot, \ldots, \cdot)$. For example, if the Yates-Grundy-Sen variance estimator is proposed for estimating $V(\hat{\bar{Y}})$, then $t_1(y_i;\hat{F})$ would be substituted for y_i in that variance estimator to give the Taylor estimate of $V(T(\hat{F}))$.

Although the jackknife variance estimator has a finite population analog, as we shall develop, its derivation from the iid case is not as straightforward.

When $(y_1, \ldots, y_n)$ are realizations of iid random variables, jackknife pseudovalues are calculated as

$$P_i = nT(\hat{F}) - (n-1)\,T(\hat{F}_{-i}),$$

where $\hat{F}_{-i}$ is the empirical cdf obtained when the i^{th} observation is omitted. This leads to using either

$P_i - T(\hat{F})$ or $P_i - \bar{P}$ as the implied "estimate" of $t_1(y_i;F)$.

For unequal probability sampling, however, the usual definition of P_i does not lead directly to an estimator of $t_i(y_i;F)$ for use in $h(\cdot, \ldots, \cdot)$. We have found a more general definition of a pseudovalue that applies for both equal and unequal probability samples. Define

$$dF_{-i}(y) = \begin{cases} \dfrac{d\hat{F}(y) - w_i}{1 - w_i} & y = y_i \\[2ex] \dfrac{d\hat{F}(y)}{1 - w_i} & y \neq y_i \end{cases}.$$

Then defining the i^{th} pseudovalue as

$$P_i = T(\hat{F}) - \frac{1 - w_i}{w_i}\,[T(\hat{F}) - T(\hat{F}_{-i})]$$

leads to

$$\tilde{t}_1(y_i;F) = \frac{1-w_i}{w_i}\,[T(\hat{F}) - T(\hat{F}_{-i})]$$

as the jackknife estimator of $t_i(y_i;F)$.

These values may then be used in the appropriate variance estimator. The algebraic form of the unequal probability pseudovalue is the same as that given by Quenouille (1956) and Hinkley (1977) for use with linear regression where the observations are differentially weighted.

For stratified sampling, this definition of a pseudovalue differs from that used by Frankel (1971) and others. Their definition incorporates knowledge of the stratification. If the i^{th} observation is omitted from stratum p, Frankel would define pseudovalues based on

$$\hat{F}_{-(pr)} = \sum_{i \neq p} \frac{N_i}{N}\hat{F}_i + \frac{N_p}{N}\hat{F}_{p(-r)}$$

whereas our definition is equivalent to

$$\hat{F}_{-(pr)} = \left(1 - \frac{N_p}{n_p N}\right)^{-1}\left[\sum_{i \neq p} \frac{N_i}{N}\hat{F}_i + \frac{N_p}{N}\frac{(n_p - 1)}{n_p}\hat{F}_{p(-r)}\right].$$

The two methods can be shown to be asymptotically equivalent, however.

F. Summary

We have shown that results derived from the von Mises expansion can be fruitfully applied in finite population sampling. After defining a finite population "cdf" and its estimation under various sample designs, we then demonstrated that the general first-order approximation obtained by Reeds (1976) also applies for finite population sampling with unequal probability sampling. Using this approximation we were able to define jackknife pseudovalues and a jackknife variance estimatior for any design for which an estimator of

$V(\hat{\bar{Y}})$ exists.

We have not yet attempted to show the consistency of the Taylor or jackknife variance estimators, but conjecture that this will hold, under weak conditions, whenever

$$v(\sqrt{n}\,\hat{\bar{Y}}) \xrightarrow{p} V(\sqrt{n}\,\hat{\bar{Y}}).$$

G. References

1. Campbell, C. (1979). A New Approach to Second Order Comparisons of Variance Estimators for Complex Surveys. Technical Report #343, University of Minnesota, School of Statistics.
2. Chow, Y.S. and H. Teicher (1978). Probability Theory. Springer-Verlag.
3. Cochran, W.G. (1977). Sampling Techniques. 3rd edition. John Wiley and Sons.
4. Filippova, A.A. (1962). Mises' theorem on the asymptotic behavior of functionals of empirical distribution functions and its statistical applications. Theroy of Probability and its Applications, 7, 24-57.
5. Frankel, M.R. (1971). Inference from Sample Surveys. The University of Michigan.
6. Gray, H., W. Schucany and T. Watkins (1975). On the generalized jackknife and its relation to statistical differentials, Biometrika, 62, 637-642.
7. Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. Acta Mathematics Academy of Science of Hungary, 5, 361-74.
8. Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Annals of Mathematical Statistics, 35, 1491-1523.
9. Hampel, F.R. (1974). The influence curve and its role in robust estimation. Journal of the American Statistical Association, 69, 383-393.
10. Hinkley, D.V. (1977). Jackknifing in unbalanced situations. Technometrics, 19, 285-292.
11. Hinkley, D.V. (1978). Improving the jackknife with special reference to correlation estimation. Biometrika, 65, 13-21.
12. Horst, L. (1973). Some limit theorems with applications in sampling theory. Annals of Statistics, 1, 644-658.
13. Jaeckel, L. (1972). The infinitesimal jackknife. Technical memorandum MM-72-1215-11, Bell Laboratories. Murray Hill, New Jersey.
14. Kish, L. and Frankel, M.R. (1974). Inference from complex samples. Journal of the Royal Statistical Society. B, 36, 1-37.
15. Miller, R.G. (1974). The jackknife: a review Biometrika, 61, 1-15.
16. Miller, R.G. (1974). An unbalanced jackknife. Annals of Statistics, 2, 880-891.
17. Quenouille, M.H. (1956). Notes on bias in estimation. Biometrika, 43, 353-60.
18. Reeds, J.A. III (1976). On the definition of Von Mises functionals. Dissertation submitted to Department of Statistics, Harvard University.
19. Rosen, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement, I. Annals of Mathematical Statistics, 43, 373-397.
20. Rosen, B. (1972). Asymptotic theory for successive sampling with varying probabilities without replacement, II. Annals of Mathematical Statistics, 43, 748-776.
21. Von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions. Annals of Mathematical Statistics, 18, 309-348.