ASYMPTOTICS IN FINITE POPULATION SAMPLING

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina at Chapel Hill

Institute of Statistics Mimeo Series No. 1483

March 1985

ASYMPTOTICS IN FINITE POPULATION SAMPLING

Pranab Kumar Sen

1. Introduction

The theory of (objective or probabilistic) sampling from a finite population plays a fundamental role in statistical inference in sample surveys. Indeed, in practice, mostly, one encounters a set or collection of a finite number (say, N) of objects or units comprising a population, and, on the basis of a subset of these units, called a sample (drawn in an objective manner), the task is to draw (valid) statistical conclusions on (various characteristics of) the population. The population size, N, though finite, needs not be small, and the sample size, say n, though presumably less than N, needs not be very saml1 compared to N (i.e., the sampling fraction n/N needs not be very saml1). In survey sampling, the sampling frame defines the units and the size of the population unambiguously. It also reconstructs a population having an uncountable (or infinite) number of natural units interms of a finite population by redefining suitable sampling units. Thus, given the sampling frame and units, one may like to draw inference on the population through an objective sampling scheme. In some other cases, though the units are clearly defined, the size of the population may not be known in advance, and one may therefore like to estimate the population size (along with its other characteristics), through objective sampling schemes. In either case, usually sampling is made without replacement (generally, leading to relatively samller margins of sampling fluctations), though, for sampling with repalcement, the theory is of relatively simpler in form. Again, in either with or without replacement schemes, the different units in the population may all have the common probability for inclusion in the sample (leading to equal probability or simple random sampling), or, they may be stratified into some subsets, for each one of which simple random sampling may be adopted. In the extreme case, the units in the

population, depending on their sizes or some other characteristics, may have possibly different probabilities for inclusion in the sample (i.e., varying probability sampling). There may be other variations in a sampling scheme (such as double sampling, interpenetrating sampling, successive sampling, etc.) However, all of these cases are characterized by an objective sampling procedure defined by a probability law governing the sampling distribution of suitable statistics based on the sample units.

When N is small, such a sampling distribution (of a statistic) may be studied, mostly, by direct enumeration of all possible cases. However, as N increases, this enumerational process, generally, becomes prohibitively laborious. On the other hand, in survey sampling and in other practical situations, usually, N is large and n/N may not be very small. In such a case, there may be a profound need to examine the generally anticipated and applicable large sample approximations for the sampling distributions (and related probability inequalities) with a view to prescribing them in actual applications. Our main interest is centered in these asymptotics in finite population sampling. Naturally, the asymptotic theory depends on the sampling design, and, for diverse sampling schemes, diverse techniques have been employed to achieve the general goals. It is intended to provide here a general account of these technqiues along with the related asymptotic theory. In line with the general objectives of the Handbook, mostly, the derivations will be replaced by motivations, and emphasis will be laid on the applications oriented theory only.

In Section 2, we start with the asymptotics in simple random sampling (with and without replacement) schemes. For general U-statistics, containing the sample mean and variances as special cases, asymptotic normality and related results are presented there. Some asymptotics on probability inequalities in simple random sampling (SRS) are then considered in Section 3. Asymptotics on jackknifing in finite population sampling (SRS) are presented in Section 4.

Capture-mark-recapture (CMR) techniques and asymptotic results on the estimation of the size of a finite population are considered in Section 5. Asymptotic results on sampling with varying probabilities (along with the allied coupon collector problem) are presented in Section 6. In this context, some limit theorems arising in the occupancy problem are also treated briefly. The concluding section deals with successive sub-sampling with varying probabilities, and the relevant asymptotic theory is discussed there.

2. Asymptotics in SRS

Let the N units with values $a_1,\ldots,a_N$ constitute the finite population. In a sample of size n ($\leq N$), drawn without replacement, the observation vector $\underset{\sim}{X}_n=(X_1,\ldots,X_n)$ is a (random) subset of $\underset{\sim}{A}_N=(a_1,\ldots,a_N)$, governed by the basic probability law:

$$P\{X_1=a_{i_1},\ldots,X_n=a_{i_n}\} = N^{-[n]}, \tag{2.1}$$

for every $1\leq i_1\neq\ldots\neq i_n\leq N$, where $N^{-[n]}=(N^{[n]})^{-1}$ and $N^{[n]}=N\ldots(N-n+1)$, for $n\leq N(N^{[0]}=1)$. Based on $\underset{\sim}{X}_n$, we may be interested in the estimation of the population mean

$$\bar{A}_N = N^{-1}\Sigma_{i=1}^{N} a_i \tag{2.2}$$

and the population variance

$$\sigma_N^2 = (N-1)^{-1}\Sigma_{i=1}^{N}[a_i-\bar{a}_n]^2, \tag{2.3}$$

among other characteristics of the population. The optimal sample estimators [viz., Nandi and Sen (1963)] are given by

$$\bar{X}_n = n^{-1}\Sigma_{i=1}^{n}X_i \quad \text{and} \quad s_n^2 = (n-1)^{-1}\Sigma_{i=1}^{n}(X_i-\bar{X}_n)^2 \tag{2.4}$$

respectively. Both these estimators are special cases of U-statistics, which may be introduced as follows.

For a symmetric kernel $g(x_1,\ldots,x_m)$ of degree $m(\geq 1)$, we define a (population) parameter

$$\theta_N = \theta(\underset{\sim}{A}_N) = N^{-[m]}\Sigma_{1\leq i_1<\ldots<i_m\leq N} g(a_{i_1},\ldots,a_{i_m}), \tag{2.5}$$

and the corresponding sample function, viz.,

$$U_n = n^{-[m]} \Sigma_{1\leq i_1 <\ldots< i_m \leq n} g(X_{i_1},\ldots,X_{i_m}) \tag{2.6}$$

termed a U-statistic, is an optimal unbiased estimator of Θ_N. We may note that for m=1, $g(x)=X$, $U_n=\bar{X}_n$, $\Theta_N=\bar{a}_n$, and, for m=2, $g(x,y)=\frac{1}{2}(x-y)^2$, $U_n=s_n^2$, $\Theta_N=\sigma_n^2$. In fact, as $g(\cdot)$ is assumed to be symmetric in its $m(\geq 1)$ arguments, in (2.5) and (2.6), we may take $i\leq i,\ldots<i_m\leq N$ (and $1\leq i,\ldots<i_m\leq n$) and replace $N^{-[m]}$ (and $n^{-[m]}$) by $\binom{N}{m}^{-1}$ (and $\binom{n}{m}^{-1}$), where $\binom{p}{q}=p!/(q!(p-q)!)$.

Note that the X_i are not independent random variables (r.v.). Nevertheless, they are symmetric dependent r.v.'s. For any given A_N, the (exact) sampling distribution of U_n may be obtained by direct enumeration of all possible $\binom{N}{n}$ samples of size n from A_N. This process, obviously, becomes prohibitively laborious for large N (and n). As such, there is a genuine need to provide suitable approximations to the large sample distribution, when N and n are both large, though $\alpha=n/N$ (the sampling fraction) needs not be very small. In this context, the permutational central limit theorems (PCLT) play a vital role. For the particular case of $\bar{X}_n$, Madow (1948) initiated the use of PCLT in finite population sampling, and, since then, this has been an active area of fruitful research. For general U_n, the asymptotic normality result in SRS has been studied by Nandi and Sen (1963), with further generalizations due to Sen (1972), Krewski (1978), Majumdar and Sen (1978), and others.

In an asymptotic setup, we conceive of a sequence $\{A_n\}$ of populations and a sequence $\{n\}$ of sample sizes, such that as N increases,

$$\alpha_N=n/N\to\alpha:0\leq\alpha<1 \tag{2.7}$$

Though, theoretically, the asymptotic theory is justified for N indefinitely large, in practice, the asymptotic approximations workout quite well, for N even moderately large.

For every N and $h:0\leq h\leq m$, let

$$g_h^{(N)}(a_{i_1},\ldots,a_{i_h}) = (N-h)^{-[m-h]} \Sigma^*_{Nh} g(a_{i_1},\ldots,a_{i_m}) \tag{2.8}$$

where the summation Σ^*_{Nh} extends over all distinct $i_{h+1},\ldots,i_m$ over the set $\{1,\ldots,N\}\setminus\{i_1,\ldots,i_h\}$. Note that $g_0^{(N)}\equiv\theta_N$ and $g_m^{(N)}\equiv g$. Let then

$$\bar{\zeta}_{h,N} = N^{-[h]} \Sigma_{1\le i_1<\ldots<i_h\le N} \{g_h^{(N)}(a_{i_1},\ldots,a_{i_h})\}^2 - \theta_N^2 , \tag{2.9}$$

for $h= 0,1,\ldots,m$ (where $\bar{\zeta}_{0,N}=0$). For the asymptotic theory, we assume that

$$\liminf_{N\to\infty} \bar{\zeta}_{1,N} > 0 \quad \text{and} \quad \limsup_{N\to\infty} \bar{\zeta}_{m,N} < \infty . \tag{2.10}$$

Then, as in Nandi and Sen (1963), we have, for every $n\ge m$,

$$\begin{aligned} \text{Var}(U_n) &= E[U_n-\theta_N]^2 \\ &= m^2(n^{-1}-N^{-1})\bar{\zeta}_{1,N} + O((n^{-1}-N^{-1})^2) \end{aligned} \tag{2.11}$$

Let us now assume that as N increases,

$$\{\max_{1\le i\le N} \{g(a_i) - N^{-1}\Sigma_{i=1}^n g(a_i)\}^2\}/\{N\bar{\zeta}_{1,N}\} \to 0 . \tag{2.12}$$

Also, let [k] denote the largest integer $\le k$, and let

$$Y_N(t) = Y_N(N^{-1}[Nt]) = [Nt](U_{[Nt]} - \theta_N)/\{m(N\bar{\zeta}_{1,N})^{\frac{1}{2}}\} , \; t\in[0,1], \tag{2.13}$$

where for $t<m/N$, we let $Y_N(t)=0$. Then $Y_N=\{Y_N(t),\ 0\le t\le 1\}$ is well defined for every N ($\ge m$). Finally, let $W^o=\{W^o(t),\ 0\le t\le 1\}$ be a Brownian bridge, i.e., $W^o(t)$ is Gaussian with $EW^o(t)=0$, $\forall 0\le t\le 1$ and $EW^o(s)W^o(t)=s\wedge t-st$, $\forall\ s,t\in[0,1]$, where $a\wedge b=\min(a,b)$. Then, we have the following general result, discussed in detail in Sen (1981, Section 3.5):

Under (2.10) and (2.12), as N increases, the stochastic process Y_N converges in distribution to W^o.

An immediate corollary to this basic result is the following:

Under (2.10 and (2.12), as n increases, satisfying (2.7),

$$n^{\frac{1}{2}}(U_n-\theta_N)/\{m\,\bar{\zeta}_{1,N}^{\frac{1}{2}}\} \sim N(0,1-\alpha). \tag{2.14}$$

It is also interesting to note that this weak convergence (of $\{Y_N\}$ to W^o) provides a very simple proof for the asymptotic normality result, when n is itself a positive integer-valued r.v. Suppose that $\{\nu_N\}$ is a sequence of non-negative integer valued r.v.'s, such that as N increases,

$$N^{-1}\nu_N \xrightarrow{P} \beta: \ 0<\beta<1. \tag{2.15}$$

Then, under (2.10) and (2.12),

$$N^{\frac{1}{2}}(U_{\nu_N} - \theta_N)/\{m\bar{\zeta}_{1,N}^{\frac{1}{2}}\} \quad \sim \quad \mathcal{N}(0, \beta^{-1} - 1). \tag{2.16}$$

This last result is useful in the case where the sample size n is determined by some other considerations, so that it may be stochastic in nature.

Note that for m=1 and g(x)=x, $\bar{\zeta}_{1,N}=(1-N^{-1})\sigma_N^2$. In general, $\bar{\zeta}_{1,N}$ is an (estimable) parameter. Knowledge of $\bar{\zeta}_{1,N}$ is useful in providing a confidence interval for θ_N, based on (2.14) or (2.15). The following estimator, due to Nandi and Sen (1963), is a variant form of the usual jackknifed estimator. For each i(=1,...n), let $U_{n-1}^{(i)}$ be the U-statistic based on $(X_1,\ldots,X_{i-1},X_{i+1},\ldots,X_n)$, and let

$$U_{n,i}=nU_n-(n-1)U_{n-1}^{(i)}, \ i=1,\ldots,n. \tag{2.17}$$

Let then

$$\tilde{s}_n^2 = (n-1)^{-1}\Sigma_{i=1}^{n}(U_{n,i}-U_n)^2. \tag{2.18}$$

Note that for m=1 and g(x)=x, $U_{n,i}=X_i$, $1\leq i\leq n$, so that $\tilde{s}_n^2=s_n^2$, defined by (2.4), while, in general, we have (cf. Nandi and Sen (1963)) under (2.10) and (2.12),

$$|\tilde{s}_n^2-m^2\bar{\zeta}_{1,N}| \xrightarrow{P} 0, \text{ as n increases.} \tag{2.19}$$

The asymptotic results considered here extend easily to the case of more than one U-statistic. Futher, we have studied, so far, the case of sampling without replacement. In the case of sampling with replacement, we have $X_1,\ldots,X_n$ independent and identically distributed (i.i.d.) r.v., where

$$P\{X_1=a_i\}= N^{-1}, \text{ for } i=1,\ldots, N. \tag{2.20}$$

As such, the classical central limit theorems and weak convergence results for U-statistics [viz., Hoeffding (1948), Miller and Sen (1972)], remain applicable. In particular, in such a case, for (2.14), $1-\alpha$ has to be replaced by 1, and, in (2.16), $\beta(1-\beta)$ has to be replaced by β alone; (2.19) follows from Sen (1960a).

3. Some Probability and Moment Inequalities for SRS.

In SRS with replacement, the sample observations are independent and identically distributed random variables, so that the usual probability and moment inequalities holding for sampling from an infinite population also remain valid in this case. On the other hand, in SRS without replacement, the sample observations are no longer independent (but, exchangeable) random variables. In (2.11) and (2.14), we have observed that the dependence in SRS without replacement leads to a smaller variance for the sample mean or,in general, for U-statistics. This feature is generally shared by a general class of statistics and the related inequality is termed the Hoeffding inequality.

Let $X_1,\ldots,X_n$ be a sample in SRS without replacement from a finite population, and let $Y_1,\ldots,Y_n$ be a sample in SRS with replacement from the same population. Then, for any convex and continuous function $\phi(x)$, we have

$$E\phi(X_1+\ldots+X_n) \leq E\phi(Y_1+\ldots+Y_n) . \tag{3.1}$$

We may refer to Hoeffding (1963) for a simple proof of (3.1). Rosen (1967) has extended the inequality in (3.1) for certain functions other than convex, continuous $\phi(.)$ and also for more general symmetric sampling plans which include the SRS with replacement as a particular case. A more general result in this direction is due to Karlin (1974). In his setup, $\phi(.)$ needs not be a function of the sum of the X_i (or Y_i). Let $\phi(x_1,\ldots,x_n)$ be a function, symmetric in its n arguments, such that

$$\phi(a,a,x_3,\ldots,x_n) + \phi(b,b,x_3,\ldots,x_n) \geq 2\,\phi(a,b,x_3,\ldots,x_n) , \tag{3.2}$$

for all $a,b,x_3,\ldots,x_n$. For all such $\phi(.)$ and any symmetric sampling plan $\mathcal{S}$,

$$E\phi(X_1,\ldots,X_n) \leq E\phi(Y_1,\ldots,Y_n) , \tag{3.3}$$

where the X_i are in SRS without replacement and the Y_i in the sampling plan $\mathcal{S}$. There are more general inequalities of this nature in Karlin(1974), and they should be of considerable theoretical interest. In passing, we may remark that the condition in (3.2) may not , in general, hold for functions of U-statistics.

To illustrate this point, let us consider the special case of the sample variance when $n = 2$. Here, $U_2 = (X_1-X_2)^2/2$, so that $EU_2 = \sigma_N^2$. Then, for $\phi(x_1,x_2) = (\frac{1}{2}(x_1-x_2)^2 - \sigma_N^2)^2$, we have $\phi(a,a) = \phi(b,b) = \sigma_N^4$ and $\phi(a,b) = (\frac{1}{2}(a-b)^2-\sigma_N^2)^2$. Thus, whenever $\frac{1}{2}(a-b)^2 > 2\sigma_N^2$, $2\phi(a,b) > \phi(a,a) + \phi(b,b)$. Since $E[\frac{1}{2}(X_1-X_2)^2] = \sigma_N^2$, in general, $\frac{1}{2}(X_1-X_2)^2$ exceeds $2\sigma_N^2$ with a positive probability, and hence, (3.2) does not hold. The same picture holds for $n \geq 2$. Nevertheless, for convex functions of U-statistics, we have some simple moment inequalities [due to Hoeffding (1963)]. Consider (as in Section 2) a kernel $g(X_1,\ldots,X_m)$ of degree $m(\geq 1)$, and for every $n \geq m$, define $k_n = [n/m]$ and let

$$U_n^* = k_n^{-1} \Sigma_{i=1}^{k_n} g(X_{(i-1)m+1},\ldots,X_{im}) . \tag{3.4}$$

Let $\mathcal{F}_n$ be the sigma-field generated by the ordered collection of $X_1,\ldots,X_n$, so that we have $U_n = E[U_n^* | \mathcal{F}_n]$, and, as a result, by the Jensen inequality, for any convex function ϕ (for which the expectation in (3.5) exists),

$$E[\phi(U_n)] \leq E[\phi(U_n^*)] . \tag{3.5}$$

On the other hand, for U_n^*, the inequality in (3.1) is directly applicable, so that we have for every continuous and convex ϕ,

$$E[\phi(U_n)] \leq E[\phi(\Sigma_{i=1}^{k_n} g(Y_{(i-1)m+1},\ldots,Y_{im})/k_n)] , \forall\, n \geq m , \tag{3.6}$$

where the Y_i are defined as in (3.1). For the right hand side of (3.6), the usual moment inequalities are applicable, and these are therefore adaptable for U_n in SRS without replacement too.

In SRS without replacement, the reverse martingale property of U-statistics (and hence, sample means) has been established by Sen(1970), and this enables one to derive other probability inequalities, which will be briefly discussed here. In passing we may also note that by virtue of this reverse martingale property, for any convex ϕ, $\{\phi(U_n), m \leq n \leq N\}$ has the reverse sub-martingale property, so that suitable moment inequalities may also be based on this fact.

We define $V_n = \mathrm{Var}(U_n)$ as in (2.11) and let $V_n^* = V_n - V_{n+1}$, for $n \geq m$.

Also, let $\{c_k; k \geq m\}$ be a nondecreasing sequence of positive numbers. Then, we have the following [cf. Sen(1970)]:

Whenever for some $r \geq 1$, $E|U_n - \theta_N|^r$ exists (for $n \geq m$), for every $t > 0$ and $m \leq n \leq n' \leq N$,

$$P\{\max_{n\leq k\leq n'} c_k|U_k - \theta_N| \geq t\} \leq t^{-r}\{c_n^r E|U_n-\theta_N|^r + \Sigma_{i=n+1}^{n'}(c_i^r-c_{i-1}^r)E|U_i-\theta_N|^r\}, \quad (3.7)$$

so that, in particular, we have

$$P\{\max_{n\leq k\leq n'} c_k|U_k - \theta_N| \geq t\} \leq t^{-2}\{c_n^2 V_n + \Sigma_{i=n}^{n'-1} c_i^2 V_i^*\}, \quad (3.8)$$

and

$$P\{\max_{n \leq k \leq N} |U_k - \theta_N| \geq t\} \leq t^{-2}V_n, \ \forall\, n \geq m. \quad (3.9)$$

For the particular case of sample means (or sums) i.e., kernels of degree 1, some related inequalities have also been studied by Serfling (1974). In this context, the following inequality [due to Sen(1979b)] is worth mentioning.

Let $\{d_{Ni}, 1 \leq i \leq N; N \geq 1\}$ be a triangular array of real numbers satisfying the normalizing constraints:

$$\Sigma_{i=1}^N d_{Ni} = 0 \quad \text{and} \quad \Sigma_{i=1}^N d_{Ni}^2 = 1. \quad (3.10)$$

Also, let $q = \{q(t): 0 < t < 1\}$ be a continuous, nonnegative, U-shaped and square integrable function inside $I = [0,1]$. Finally, let $\underset{\sim}{Q} = (Q_1,\ldots,Q_N)$ take on each permutation of $(1,\ldots,N)$ with the common probability $(N!)^{-1}$. Then

$$P\{\max_{1 \leq k \leq N-1} q(k/N)|\Sigma_{i=1}^k d_{NQ_i}| \geq 1\} \leq \int_0^1 q^2(t)dt. \quad (3.11)$$

Clearly, in SRS without replacement, (3.11) may be used to provide a simultaneous (in $k: 1\leq k\leq N$) confidence band for θ_N by choosing q in an appropriate way. For a related inequality (exploiting the 4th moment but not the inherent martingale structure) we may refer to Hájek and Šidák (1967,p.185):

$$P\{\max_{1\leq k\leq n} |\Sigma_{i=1}^k d_{NQ_i}| \geq t\} \leq (n/N)[\max_{1\leq i\leq N} d_{Ni}^2 + 3n/N]t^{-4}(1-n/N)^{-3}(1+\varepsilon_N), \quad (3.12)$$

where $\varepsilon_N \to 0$ as $N \to \infty$.

Generally, (3.7) with $r = 4$ (and $m=1$) provides a better bound than (3.12). We

may obtain even better bounds by exploiting the weak convergence results in Section 2 , for large values of N. As in Sen (1972), we consider the case of general U-statistics [with the same notations as in Section 2], so that the case of sample means (or sums) can be obtained as a particular one. Note that by virtue of the weak convergence result, stated after (2.13), we have for every $t > 0$ and $n : n/N \leq \alpha (0 < \alpha \leq 1)$,

$$\lim_{N\to\infty} P\{ \max_{m \leq k \leq n} k|U_k - \theta_N| \geq t m[N \bar{\zeta}_{1,N}]^{1/2} \}$$

$$= P\{ \sup_{0 \leq u \leq \alpha} |W^0(u)| \geq t \} , \qquad (3.13)$$

where $W^0 = \{W^0(t), 0 \leq t \leq 1\}$ is a Brownian bridge . Noting that $W^0(s/(s+1)) = (s+1)^{-1} W(s)$, $s \geq 0$, where $W = \{W(t), t \geq 0\}$ is a standard Brownian motion process on $[0,\infty)$, we may rewrite the right hand side of (3.13) as

$$P\{ \sup_{0 \leq u \leq \alpha/(1-\alpha)} |(u+1)^{-1} W(u)| \geq t \} . \qquad (3.14)$$

An upper bound for (3.14) is given by

$$P\{ \sup_{0 \leq u < \infty} |(u+1)^{-1} W(u)| \geq t \} = 2 \sum_{k=1}^{\infty} (-1)^{k+1} \exp(-2k^2 t^2) . \qquad (3.15)$$

For small values of α (as is usually the case encountered in practice), we may get a better bound :

$$P\{ \sup_{0 \leq u \leq \alpha/(1-\alpha)} |(1+u)^{-1} W(u)| \geq t \} \leq P\{ \sup_{0 \leq u \leq \alpha/(1-\alpha)} |W(u)| \geq t \}$$

$$\leq 4 P\{ W(\alpha/(1-\alpha)) \geq t \} = 4[1 - \Phi(t(1/\alpha - 1)^{\frac{1}{2}})] , \qquad (3.16)$$

where $\Phi(.)$ is the standard normal d.f. In particular, for kernels of degree 1, (3.16) may be compared to (3.12), and , as $1 - \Phi(x)$ converges to 0 exponentially, as $x \to \infty$, usually (3.16) performs much better than (3.12). The same conclusion holds for the comparison between ((3.8) and (3.16) [or (3.9) and (3.15). We conclude this section with the remark that for moderate values of N, the equality sign in (3.13) may , generally, be replaced by a less than or equality sign, so that we may have a conservative property for small values of N.

4. Jackknifing in Finite Population Sampling.

In SRS or other sampling plans, for ratio, regression or other estimators,

'jackknifing' was mainly introduced to serve a dual purpose : To reduce the bias of estimators (which are typically of the non-linear form) and to provide an efficient (and asymptotically normally distributed) estimator of the sampling variance of the (jackknifed) estimator. In the same setup as in Section 2, for a general estimator $T_n = T(X_1,\ldots,X_n)$ (containing U_n as a special case), we may define the pseudo values $T_{n,i} = nT_n - (n-1)T_{n-1}^{(i)}$, $i=1,\ldots,n$, as in (2.17). Then the jackknifed estimator is defined by

$$T_n^* = n^{-1}(T_{n,1} + \ldots + T_{n,n}) , \tag{4.1}$$

and the (Tukey form of the) jackknifed variance estimator is given (as in (2.18)) by

$$S_n^2 = (n-1)^{-1} \Sigma_{i=1}^n (T_{n,i} - T_n^*)^2 . \tag{4.2}$$

To motivate the jackknifed estimator, we may start with a possibly biased estimator T_n for which we may have

$$ET_n = \theta_N + n^{-1}a_1(N) + n^{-2}a_2(N) + \ldots . , \tag{4.3}$$

where the $a_j(N)$ are real numbers depending possibly on the population size N and the set A_N . Using n-1 for n in (4.3) for each $T_{n-1}^{(i)}$ and (4.1), we obtain that under (4.3),

$$ET_n^* = \theta_N - a_2(N)/n(n-1) + \ldots . = \theta_N + O(n^{-2}) . \tag{4.4}$$

Thus, the bias of T_n is reduced from $O(n^{-1})$ to that of $O(n^{-2})$ for T_n^* . In addition to this important feature of 'bias reduction' , the variance estimator S_n^2 also plays a very important role in drawing statistical conclusions on θ_N . Since in this chapter, we are primarily concerned with the asymptotics in finite population sampling, we shall mainly restrict ourselves to the discussion of the large sample properties of T_n^* and S_n^2 ; hopefully, in some other chapter(s), there will be complementary discussions on other aspects of jackknifing .

Keeping in mind the ratio, regression and other estimators (which are all expressible as functions of some U-statistics), we conceive of a general estimator T_n of the form $T_n = h(\underset{\sim}{U}_n)$ where $h(.)$ is a smooth function and $\underset{\sim}{U}_n$ ia a vector

of U-statistics, defined as in Section 2. Also, we keep in mind the conditional (permutational) distribution generated by the n! equally likely permutations of $X_1,\ldots,X_n$ among themselves, and define $\mathcal{F}_n$ as in after (3.4). Then, it follows from the basic results in Majumdar and Sen (1978) that

$$T_n^* = T_n + (n-1)E[\,(T_n - T_{n-1}) \mid \mathcal{F}_n\,] \;,\; \forall\, n > m \;, \tag{4.5}$$

$$S_n^2 = n(n-1)\mathrm{Var}[\,(T_n - T_{n-1}) \mid \mathcal{F}_n\,] \;,\; \forall\, n > m \;. \tag{4.6}$$

Thus, for both the jackknifed estimator T_n^* and the variance estimator S_n^2, the inherent permutational distributional structure provides the access for the necessary modifications. This theoretical justification for jackknifing has been elaborately studied in Sen (1977).

To fix the notations, we let $\underset{\sim}{\mu}_N = E\underset{\sim}{U}_n$, and, in addition to (2.10) (in a matrix setup), we assume that

$$\sup_N E\,\|\, \underset{\sim}{g}(X_1,\ldots,X_m)\,\|^4 < \infty \quad ; \; m = \max(m_1,\ldots,m_p) \;, \tag{4.7}$$

where $\underset{\sim}{g}(.)$ stands for the vector of kernels of degrees $m_1,\ldots,m_p$, respectively. Further, we assume that $h(\underset{\sim}{u})$ has bounded second order (partial) derivatives (with respect to $\underset{\sim}{u}$) in some neighbourhood of $\underset{\sim}{\mu}_N$ and $h(\underset{\sim}{\mu}_N)$ is finite. Finally, let us define

$$\sigma_{Nn}^2 = E[(T_n^* - \theta_N)^2] \;,\; n \geq n_o \;, \text{ where } n_o (\geq m) \text{ is finite,} \tag{4.8}$$

and assume that there exists a sequence $\{\sigma_N^2\}$ of positive numbers, such that

$$n\,\sigma_{Nn}^2 - [(N-n)/(N-1)]\,\sigma_N^2 \to 0 \text{ as } n \text{ increases, where } \underline{\lim}\; \sigma_N^2 > 0 \;. \tag{4.9}$$

Now, parallel to that in (2.13), we consider a stochastic process $Y_N = \{Y_N(t);\ 0 \leq t \leq 1\}$ by letting

$$Y_N(t) = Y_N(N^{-1}[Nt]) = [Nt]\,(\,T^*_{[Nt]} - \theta_N\,)/\sigma_N \;,\; \frac{m}{N} \leq t \leq 1 \;, \tag{4.10}$$

where, for $t < m/N$, we complete the definition of $Y_N(t)$, by letting $Y_N(t) = 0$. Further, as in after (2.13), we define a Brownian bridge $W^o = \{W^o(t);\ 0 \leq t \leq 1\}$. Then, we have the following result [viz., Majumdar and Sen (1978)]:

For the jackknifed estimator, under the assumed regularity conditions, Y_N converges in distribution (or law) to W^o, as N increases. (4.11)

Further, under the same regularity conditions, $S_n^2 - \sigma_N^2$ strongly converges to 0 as n increases; this strong convergence is in the sense that for every ε (> 0) and δ (> 0), there exists a positive integer $n_o = n_o(\varepsilon,\delta)$, such that

$$P\{\max_{n_o \leq n \leq N} | S_n^2 - \sigma_N^2 | > \varepsilon \} < \delta \quad , \quad N \geq n_0 \; . \tag{4.12}$$

This strong convergence result enables us to replace in (4.10) σ_N by $S_{[Nt]}$, for every $t > 0$. Thus, if we denote such a studentized process by $Y_N^*(t)$, $t > 0$, then, we conclude that for every $\eta > 0$,

$\{Y_N^*(t); t \in [\eta,1]\}$ converges in law to $\{W^o(t); t \in [\eta,1]\}$, as N increases. (4.13)

In particular, it follows that for any (fixed) $\alpha : 0 < \alpha \leq 1$, if $n/N \to \alpha$, as N increases, then

$$n^{\frac{1}{2}} (T_n^* - \theta_N)/S_n \text{ is asymptotically normal } (0, 1- \alpha) \; . \tag{4.14}$$

Further, if ν_N be nonnegative integer valued random variable, such that $N^{-1}\nu_N$ converges in probability to α ($0 < \alpha \leq 1$), then

$$N^{\frac{1}{2}} (T_{\nu_N}^* - \theta_N)/S_{\nu_N} \text{ is asymptotically normal } (0,(1-\alpha)/\alpha) \; . \tag{4.15}$$

The last two results are very useful in setting up a confidence interval for the parameter θ_N or to test for a null hypothesis H_0: $\theta_N = \theta_0$ (specified). As a simple illustration, consider a typical <u>ratio-estimator</u> of the form:

$$T_n = U_n^{(1)}/U_n^{(2)} \; ; \; U_n^{(j)} = n^{-1}\Sigma_{i=1}^n g_j(X_i) \; , \; j=1,2 \; , \tag{4.16}$$

where the functions $g_1(.)$ and $g_2(.)$ may be of quite general form. In fact, we may even consider some U-statistics for $U_n^{(1)}$ and $U_n^{(2)}$ (of degrees ≥ 1). In such a setting, T_n is not generally an unbiased estimator of the population parameter $\theta_N = \mu_N^{(1)}/\mu_N^{(2)}$, though the $U_n^{(j)}$ may unbiasedly estimate the $\mu_N^{(j)}$, j=1,2. Typically, the bias of T_n is of the form in (4.3), and hence, jackknifing reduces the bias to the order n^{-2} . Further, here $h(a,b) = a/b$, so that $(\partial^2/\partial a^2)h(a,b) = 0$, $(\partial^2/\partial a \partial b)h(a,b) = -b^{-2}$ and $(\partial^2/\partial b^2)h(a,b)=2b^{-2}h(a,b)$. Consequently, whenever $\mu_N^{(2)}$ is strictly positive and finite , for finite θ_N , the regularity conditions are all satisfied, and hence, (4.11) through (4.15) hold. For some specific cases, we may refer to Majumdar and Sen (1978). The

basic advantage of using (4.11), (4.12) and (4.13) , instead of (4.14) or (4.15), is that these asymptotics are readily adoptable for sequential testing and estimation procedures. Further, the asymptotic inequalities discussed in the preceeding section also remain applicable for the jackknifed estimators. In particular, (3.13) through (3.16) also hold when we replace the U-statistics and their variances by the T_k^* and the Tukey estimator of their variances. Finally, the results are easily extendable to the case where the T_n are q-vectors, for some $q \geq 1$. In that case, instead of (4.11) or (4.13), we would have a tied-down Brownian sheet approximation (in law), and instead of (4.12), we would have the strong convergence of the matrix of jackknifed variance covariances. For (4.14) and (4.15), we would have an analogous result involving a multivariate normal distribution.

5. Estimation of Population Size : Asymptotics

The estimation of the total size of a population (of mobile individuals, such as the number of fish in a lake etc.) is of great importance in a variety of biological, environmental and ecological studies. Of the methods available for obtaining information about the size of such populations, the ones based on capture, marking, release and recapture (CMRR) of individuals, originated by Petersen (1896), have been extensively studied and adapted in practice. The Petersen method is a two-sample experiment and amounts to marking (or tagging) a sample of a given number of individuals from a closed population of unknown size (N) and then returning it into the population. The proportion of marked individuals appearing in the second sample estimates the proportion marked in the population, providing in turn, the estimate of the population size N. Schnabel (1938) considered a multi-sample extension of the Petersen method , where each sample captured commencing from the second is examined for marked members and then every member of the sample is given another mark before being

returned to the population. For this method, the computations are simple, successive estimates enable the field worker to see his method as the work progresses, and the method can be adapted for a wide range of capture conditions.

For the statistical formulation of the CMRR procedure, we use the following notations. Let N = total population size (finite and unknown), k = number of samples ($k \geq 2$), n_i = size of the ith sample, $i \geq 1$, m_i = number of marked individuals in n_i , $i=1,\ldots,k$, $u_i = n_i - m_i$, $i=1,\ldots,k$, and M_i = number of marked individuals in the population just before the ith sample is drawn (i.e., $M_i = \Sigma_{j=1}^{i-1} u_j$), $i=1,\ldots,k$. Conventionally, we let $M_1 = u_1 = 0$ and $M_{k+1} = M_k + n_k - m_k = \Sigma_{j=1}^{k}(n_j - m_j)$. Now, the conditional distribution of m_i , given M_i and n_i , is given by

$$L_N^{(i)}(m_i \mid M_i, n_i) = \binom{M_i}{m_i}\binom{N-M_i}{n_i-m_i}\Big/\binom{N}{n_i} \quad , \; i=2,\ldots,k \; , \tag{5.1}$$

so that the (partial) likelihood function is

$$L_N(n_1,\ldots,n_k) = \Pi_{i=2}^{k} L_N^{(i)} = \Pi_{i=2}^{k}\{\binom{M_i}{m_i}\binom{N-M_i}{n_i-m_i}\Big/\binom{N}{n_i}\} \quad . \tag{5.2}$$

Note that

$$\begin{aligned} L_N/L_{N-1} &= N^{-(k-1)}\{\Pi_{i=2}^{k}(N-n_i)\}(N-M_2)/(N-M_{k+1}) \\ &= N^{-(k-1)}\{\Pi_{i=1}^{k}(N-n_i)\}(N-M_{k+1})^{-1} \; , \end{aligned}$$

so that

$$L_N/L_{N-1} \text{ is } \gtreqless 1 \text{ according as } (1-N^{-1}M_{k+1}) \text{ is } \lesseqgtr \prod_{j=1}^{k}(1-N^{-1}n_j) \; . \tag{5.3}$$

Now, (5.3) provides the solution for the maximum likelihood estimator (MLE) of N. For the Petersen scheme (i.e., $k = 2$) , (5.3) reduces to

$$L_N/L_{N-1} \text{ is } \gtreqless 1 \text{ according as } N \text{ is } \lesseqgtr n_1n_2/m_2 \; , \tag{5.4}$$

so that $[n_1n_2/m_2] = \hat{N}_2$ is the MLE of N. For $k \geq 3$, in general, (5.3) needs an iterative solution for locating the MLE of N. Note that based on $L_N^{(i)}$, the MLE of N is given by $\hat{N}_i = [n_iM_i/m_i]$, for $i = 2,\ldots,k$. It is of natural interest to study the relationship between the MLE $\hat{N}$ (from (5.3)) and the $\hat{N}_j$, $j=2,\ldots,k$, when $k \geq 3$. Before doing so, we may note that by virtue of (5.1),

$P(m_i = 0 \mid M_i, n_i) = \binom{N-M_i}{n_i} / \binom{N}{n_i} > 0$, for every $i = 2,\ldots,k$, so that the MLE $\hat{N}_i$ do not have finite moments of any positive order. To eliminate this drawback, we may proceed as in Chapman (1951) and consider the modified MLE

$$\check{N}_i = (n_i+1)(M_i+1)/(m_i+1) - 1 \ , \text{ for } i = 2,\ldots,k \ . \tag{5.5}$$

Asymptotically (as $N \to \infty$), both $\hat{N}_i$ and $\check{N}_i$ behave identically, and hence, this modification is well recommended. Using the normal approximation to the hypergeometric distribution , one readily obtains from (5.5) that

$$N^{-\frac{1}{2}}(\hat{N}_2 - N) \text{ is asymptotically normal } (0, \gamma^2(\alpha_1,\alpha_2)) \ , \tag{5.6}$$

whenever for some $0 < \alpha_1, \alpha_2 \leq 1$, $n_1/N \to \alpha_1$ and $n_2/N \to \alpha_2$, as $N \to \infty$, where

$$\gamma^2(a,b) = (1-a)(1-b)/ab \geq [(2-a-b)/(a+b)]^2 \ , \quad 0 < a,b \leq 1, \tag{5.7}$$

and where the equality sign in (5.7) holds when $a = b$.

For the case of $k \geq 3$, a little more delicate treatment is needed for the study of the asymptotic properties of the MLE's as well as their interrelations. Using some martingale characterizations, such asymptotic studies have been made by Sen and Sen (1981) and Sen(1982a,b). First, it follows from Sen and Sen (1981) that a very close approximation N* to the actual MLE $\hat{N}$ [in (5.3)] is given by the solution :

$$N^* = [\Sigma_{s=2}^k \hat{N}_s m_s/(N^*-M_s)(N^*-n_s)]/[\Sigma_{s=2}^k m_s/(N^*-M_s)(N^*-n_s)] \ , \tag{5.8}$$

where the MLE N_i are defined as in after (5.4). Two other approximations, listed in Seber (1973), are given by

$$\tilde{N} = [\Sigma_{s=2}^k \hat{N}_s m_s/(\tilde{N} - M_s)]/[\Sigma_{s=2}^k m_s/(\tilde{N} - M_s)] \ , \tag{5.9}$$

and

$$\approx{N} = [\Sigma_{s=2}^k \hat{N}_s m_s]/[\Sigma_{s=2}^k m_s] \ . \tag{5.10}$$

$\tilde{N}$ works out well when the $n_j, 2 \leq j \leq k$ are all equal or $N^{-1}n_j$ are all small, while, (5.10) is quite suitable, when in addition, the $N^{-1}M_i$, $i=2,\ldots,k$ are all small. For both (5.8) and (5.9), an interative solution works out very well, and has been discussed in Sen and Sen(1981). Schumacher and Eschmeyer (1943) considered another estimator (pertaining to the same scheme):

$$\dot{N} = [\ \Sigma_{s=2}^{k} \hat{N}_s\, m_s M_s\ ]/[\ \Sigma_{s=2}^{k} m_s M_s\ ], \tag{5.11}$$

which is also an weighted average of the Petersen estimators. Sen and Sen (1981) considered an alternative estimator

$$\ddot{N} = [\ \Sigma_{s=2}^{k} \hat{N}_s (m_s/M_s)]/[\ \Sigma_{s=2}^{k} (m_s/M_s)]\ . \tag{5.12}$$

If we let

$$n_i = N\alpha_i \ \ (0 < \alpha_i \leq 1) \text{ and } \beta_i = \Pi_{j=1}^{i}(1-\alpha_j),\ i=1,\ldots,k\ , \tag{5.13}$$

then, for the MLE $\hat{N}$ as well as the approximate MLE N^* , we have [viz., Sen and Sen (1981)] :

$$N^{-\frac{1}{2}}(N^* - N) \text{ asymptotically normal } (0,\ \sigma^{*2})\ , \tag{5.13}$$

$$\sigma^{*2} = [\ \Sigma_{s=2}^{k} \alpha_s (1-\beta_{s-1})/\beta_s\ ]^{-1}\ . \tag{5.14}$$

Parallel results for the other estimators are :

$$N^{-\frac{1}{2}}(\tilde{N} - N) \text{ asymptotically normal}(0,\ \tilde{\sigma}^2)\ , \tag{5.15}$$

$$\tilde{\sigma}^2 = [\ \Sigma_{s=2}^{k}\alpha_s(1-\alpha_s)(1-\beta_{s-1})/\beta_{s-1}\ ][\ \Sigma_{s=2}^{k} \alpha_s(1-\beta_{s-1})/\beta_{s-1}\ ]^{-1}\ ;$$

$$N^{-\frac{1}{2}}(\tilde{\tilde{N}} - N) \text{ asymptotically normal } (0,\ \tilde{\tilde{\sigma}}^2)\ ,$$

$$\tilde{\tilde{\sigma}}^2 = [\ \Sigma_{s=2}^{k} \beta_s^*\alpha_s(1-\alpha_s)(1-\beta_{s-1})\beta_{s-1}]/[\ \Sigma_{s=2}^{k} \alpha_s(1-\beta_{s-1})]^2\ , \tag{5.16}$$

$$\beta_s^* = 1 + \Sigma_{j=s+1}^{k} \alpha_j\ ,\ s=2,\ldots,k-1\ ,\ \beta_k^* = 1;$$

$$N^{-\frac{1}{2}}(\dot{N} - N) \text{ asymptotically normal } (0,\ \dot{\sigma}^2)\ ,$$

$$\dot{\sigma}^2 = [\Sigma_{s=2}^{k}\{(\gamma_s+\Sigma_{j>s}\alpha_j\gamma_j)^2\alpha_s(1-\alpha_s)\gamma_s(1-\gamma_s)\}]/[\Sigma_{s=2}^{k}\alpha_s\gamma_s^2\ ]^2\ , \tag{5.17}$$

$$\gamma_s = 1-\beta_{s-1}\ ,\ \text{for } s = 2,\ldots,k\ ;$$

$$N^{-\frac{1}{2}}(\ddot{N} - N) \text{ asymptotically normal } (0,\ \ddot{\sigma}^2)\ ,$$

$$\ddot{\sigma}^2 = (\Sigma_{s=2}^{k}\ \alpha_s)^{-1}[\Sigma_{s=2}^{k}(\gamma_s^{-1} + \Sigma_{j>s}\ \alpha_j/\gamma_j)^2\alpha_s(1-\alpha_s)\gamma_s(1-\gamma_s)\ ]. \tag{5.18}$$

It follows from Sen and Sen (1981) that $\tilde{\sigma}$, $\tilde{\tilde{\sigma}}$, $\dot{\sigma}$, $\ddot{\sigma}$ are all greater than or equal to σ^* , where $\ddot{\sigma} = \sigma^*$ iff the α_s $(2 \leq s \leq k)$ are all equal, while, in the other three cases, an approximate equality sign holds when the α_s are all small. Numerical comparison of these asymptotic variances reveals that over the entire domain of variation of the α_s , none of the estimators $\tilde{\tilde{N}}$,$\dot{N}$ and $\ddot{N}$ is uniformly

better than the others. In (5.1) and (5.2), we have considered the so called sampling without replacement scheme. If we draw the 2nd,...,kth samples with replacement, we need to replace the hypergeometric distributions in (5.1)-(5.2) by the corresponding binomial distributions, and this will lead to some simplifications in the formulae for the asymptotic variances.

In many situations when the n_j are very small compared to N, the m_j are also very small (may even be equal to 0 with a positive probability). This may push up the variability of the estimators considered earlier. For this reason, often, an inverse sampling scheme is recommended. In this setup, at the sth stage , the sample units are drawn one by one , until a preassigned number m_s of the marked units appear, so that the sample size n_s is a random variable, while m_s is fixed in advance, for s = 2,...,k. For this inverse sampling scheme, parallel to (5.1), we have

$$L_N^{(i)}(n_i \mid M_i, m_i) = \binom{N}{n_i-1}^{-1} \binom{M_i}{m_i-1} \binom{N-M_i}{n_i-m_i} \{(M_i-m_i+1)/(N-n_i+1)\}$$
$$= \{m_i \binom{M_i}{m_i} \binom{N-M_i}{n_i-m_i}\} / \{n_i \binom{N}{n_i}\}, \quad i=2,\ldots,k, \qquad (5.19)$$

and (5.2) can be modified accordingly. Note that (5.3) and (5.4) are not affected , so that the MLE remains the same. It follows from Bailey (1951) that $\check{N}_i = (M_i+1)n_i/m_i - 1$, i=2,...,k are unbiased estimators of N. Note that the exact variance of $\check{N}_2$ is equal to $(n_1-m_2+1)(N+1)(N-n_1)/m_2(n_1+2)$, so that on letting $m_2 = \alpha_1^* n_1 = \alpha_1^* . \alpha_1 N$, we have parallel to (5.6)-(5.7) that

$$N^{-\frac{1}{2}}(\check{N}_2 - N) \text{ asymptotically normal } (0, (1-\alpha_1)(1-\alpha_1^*)/\alpha_1\alpha_1^*). \qquad (5.20)$$

Note that in (5.20), α_1^* plays the same role as α_2 in (5.6)-(5.7). With a similar modification for the other m_s , the results considered earlier for the direct sampling scheme all go through for the inverse sampling scheme too (when N is large); the main advantage of this inverse sampling scheme is that the estimates have finite moments of positive orders, although, the amount of sampling (i.e, $n_2+\ldots+n_k$) is not predetermined (but , is a random variable).

Inverse sampling schemes are the precursors of sequential sampling tagging considered by Chapman (1952), Goodman (1953), Darroch (1958) and others. Darling and Robbins (1967) and Samuel (1968) have studied some related problems on stopping times arising in sequential sampling tagging for the estimation of the population size N, and the asymptotic theory plays a vital role in this context. Lack of stochastic independence of the random variables at successive stages of drawing and nonstationarity of their marginal distributions call for a nonstandard approach for a rigorous study of the asymptotic properties of the MLE of N in a multi-stage or sequential sampling procedure. Using a suitable martingale characterization, this asymptotic theory has been developed in Sen (1982a,b), and is presented below.

Individuals are drawn randomly one by one, marked and released before the next drawing is made. Let M_k be the number of marked individuals in the population just before the kth drawal, for $k \geq 1$. Thus, $M_0 = M_1 = 0$, $M_2 = 1$, $M_{k+1} \geq M_k$, for every $k \geq 1$, and $M_{k+1} = M_k + 1 - X_k$, $k \geq 1$, where, for every $k (\geq 1)$, X_k is equal to 1 or 0 according as the kth drawal yields a marked individual or not. Now, the conditional probability function for X_k, given $X_1, \ldots, X_{k-1}$ is $f_k(X_k \mid X_1, \ldots, X_{k-1}) = N^{-1} M_k^{X_k} (N - M_k)^{1-X_k}$, $k > 1$, so that at the nth stage, the (partial) likelihood function is given by

$$L_n(N) = \prod_{k=2}^{n} f_k(X_k \mid X_1, \ldots, X_{k-1}) = N^{-(n-1)} \prod_{k=2}^{n} \{M_k^{X_k} (N - M_k)^{1-X_k}\}. \quad (5.21)$$

Note that

$$(\partial/\partial N) \log L_n(N) = \sum_{k=2}^{n} (1 - X_k)/(N - M_k) - (n-1)/N \ . \quad (5.22)$$

The summands in (5.22) are neither independent nor identically distributed random variables. Nevertheless, thery lead to a simple martingale-difference structure on which the asymptotic theory has been built in . The MLE $\hat{N}_{Sn}$ of N, based on $L_n(N)$, is a solution of (5.22) (equated to 0), and one is interested in the asymptotic behaviour of the partial sequence $\{\hat{N}_{Sn} ; n \leq n^*\}$, where n^* is large , and the sequence is suitably normalized. For this purpose, we define

$$Z_n^*(N) = \Sigma_{k=2}^{n} (N-M_k)^{-1} - N^{-1}(n-1),\ n \geq 2\ ,\quad Z_0^*(N) = Z_1^*(N) = 0\ , \tag{5.23}$$

and, for every N,n, such that n = [Nα], for some α >0, we define

$$n(t) = \max\{\ k : Z_k^*(N) \leq tZ_n^*(N)\ \}\ ,\ 0 \leq t \leq 1\ . \tag{5.24}$$

Then, for each (N,n) and every ε :0 < ε < 1, we may consider a stochastic process $W_{Nn}^{*\varepsilon} = \{W_{Nn}^*(t), \varepsilon \leq t \leq 1\}$, by letting

$$W_{Nn}^*(t) = N^{-\frac{1}{2}}(\ \hat{N}_{Sn(t)} - N\ )(e^{\alpha} - \alpha - 1)^{\frac{1}{2}}\ ,\varepsilon \leq t \leq 1. \tag{5.25}$$

Further, defining the standard Brownian motion process W = {W(t);t ε [0,1]} as in before (3.14), we let $W^{*\varepsilon} = \{W^*(t) = t^{-1}W(t),\ \varepsilon \leq t \leq 1\}$. Then, we have the following : For every ε : 0 < ε < 1 , as N increases,

$$W_{Nn}^{*\varepsilon} \text{ converges in law to } W^{*\varepsilon}\ \text{, whenever } n = [N\alpha] \text{ for some } \alpha > 0\ . \tag{5.26}$$

A direct consequence of (5.26) is that whenever n = [Nα], for some α > 0,

$$N^{-\frac{1}{2}}(\ \hat{N}_{Sn} - N\ ) \text{ is asymptotically normal } (0,\ (\ e^{\alpha} - \alpha - 1)^{-1}). \tag{5.27}$$

Further, if $\{\nu_n\}$ be any sequence of positive integer valued random variables, such that $n^{-1}\nu_n \to 1$, in probability, as n increases, then, we have

$$N^{-\frac{1}{2}}(\ \hat{N}_{S\nu_n} - N) \text{ asymptotically normal } (0,\ (e^{\alpha} - \alpha - 1)^{-1}\ ). \tag{5.28}$$

We are now in a position to compare (5.6) and (5.27), where we put $\alpha = \alpha_1 + \alpha_2$. By virtue of (5.7), the asymptotic variance in (5.6) is a minimum when $\alpha_1 = \alpha_2$. Comparing this minimum value with (5.27), we conclude that the asymptotic relative efficiency (A.R.E.) of the two-sample Petersen estimator (for $\alpha_1 = \alpha_2$) with respect to the sequential estimator is given by

$$E(P,S) = \alpha^2/\{\ (2-\alpha)^2(\ e^{\alpha} - \alpha - 1\ )\ \}. \tag{5.29}$$

As α goes to 0, (5.29) converges to 1/2, so that for small values of α , the Petersen estimator is about 50% efficient compared to $\hat{N}_{Sn}$. On the other hand, as α increases, E(P,S) also increases, and in fact, for α ≥ .7657, (5.29) exceeds 1 and it can be quite large when α is close to 2. However, in all practical situations, α is generally quite small, and hence, the sequential estimator can be recommended with full confidence. From the operational point of view, often, sequential schemes are not very practical, and hence, the Petersen estimator may be used .

In the context of sequential estimation of the total size of a finite population, the following urn model arises typically. Suppose that an urn contains an unknown number N of white balls and no others. We repeatedly draw a ball at random, observe its colour and replace it by a black ball, so that before each draw, there are N balls in the urn. Let W_n be the number of white balls observed in the first n drawals. Note that W_k is nondecreasing in k , $W_k \leq k$, for every $k \geq 1$ and $W_0 = 0$, $W_1 = 1$. For every $c > 0$, consider a stopping variable

$$t_c = \inf\{ n : n \geq (c+1)W_n \} , \tag{5.30}$$

Note that t_c can take on only the values $[(c+1)k]$, for $k = 1,2,\ldots$, and $W_{t_c} = m$ whenever $t_c = [m(c+1)]$. In this situation, one is not only interested in the study of the asymptotic properties of the MLE $\hat{N}_{St_c}$, but also of the standardized form of the stopping variable t_c . Samuel(1968) made some conjectures , and general results in this direction are due to Sen (1982b).

We may note that $W_n = W_{n-1} + w_n$, where w_n is equal to 1 or 0 according as the ball appearing at the nth draw is white or not, for $n \geq 1$; $W_0 = w_0 = 0$. For every K $(0 < K < \infty)$ and N , we consider a stichastic process $Z_N = \{Z_N(t), t \in [0,K]\}$ by letting

$$Z_N(t) = N^{-\frac{1}{2}} \left(W_{[Nt]} - N(1-(1-N^{-1})^{[Nt]}) \right), \; t \in [0,K] , \tag{5.31}$$

where [s] denotes the largest integer contained in s. Also, let $Z = \{Z(t), t \in [0,K]\}$ be a Gaussian process with 0 drift and covariance function

$$EZ(s)Z(t) = e^{-t}\{ 1 - (1+s)e^{-s} \} , \text{ for } 0 \leq s \leq t \leq K. \tag{5.32}$$

Then, as N increases, Z_N converges in law to Z. Since $(1-N^{-1})^n$ is close to $e^{-n/N}$, this suggests that a convenient estimator of N (based on n draws) is given by the solution (N_n^*) of the equation : $W_n = N(1-e^{-n/N})$, and the asymptotic properties of this estimator can then be studied by incorporating the convergence of Z_N to Z (in law). For the associated stopping time, we define now

$$I^* = [a,b] \text{ where } 0 < a < b < 1 . \tag{5.33}$$

Also, for every $m \in I^*$, we define t_m^* as the solution of the equation

$$m = (1 - e^{-t_m^*})/t_m^* , \; m \in I^*. \tag{5.34}$$

Note that $mt_m^* \leq 1$ for every $m \in [0,1]$, $t_1^* = 0$ and t_m^* monotonically goes to ∞ as m moves from 1 to 0. For every N , we consider a stochastic process $Y_N = \{Y_N(m) , m \in I^* \}$ by letting

$$\tau_{Nm} = \inf\{ n(\geq 1): mn \geq W_n\} \quad \text{and} \quad Y_N(m) = N^{-\frac{1}{2}}(\tau_{Nm} - Nt_m^*), \; m \in I^*. \quad (5.35)$$

Note that W_n $(= W_{Nn})$ depends on N as well and m plays the role of $(1+c)^{-1}$ in (5.30). Let then $Y = \{Y(m), m \in I^*\}$ be a Gaussian process on I^* with 0 drift and covariance function

$$EY(m)Y(m') = e^{-t_m^*}\{1 - (1+t_{m'}^*)e^{-t_{m'}^*}\}/\{(m-e^{-t_m^*})(m' - e^{-t_{m'}^*})\} , \; m \geq m'. \quad (5.36)$$

Then, as N increases, Y_N converges in law to Y. This convergence result, in turn, provides the asymptotic normality of $Y_N(m)$ for every fixed m ($\in I^*$) as well as for any sequence $\{m_n\}$ of positive random variables for which $m_n \to m$ ($\in I^*$) , in probability, as n increases. For m very close to 1 (i.e., c in (5.30) very close to 0), Poisson approximations for τ_{Nm} , suggested by Samuel(1968),works out well.

In the two or multi-sample capture-recapture model, and, more critically, in the sequential tagging scheme, there are certain basic assumptions which may not always match the practical applocations. For example, effects of migration need to be taken into account when sampling is conducted over a period of time and new individuals may enter into the scheme as well as some existing ones may exit. Also, the catchability of an individual in the tagging scheme may depend on some other characteristics . Moreover, once caught, an individual may develop some trap-shyness or trap-addictions, so that at the subsequent stage(s), the capture-probabilities are affected. Farm(1971) considered the asymptotic normality in a capture-recapture problem when catchability is affected by the tagging procedure. Seber(1973) studied the robustness of CMRR procedures against possible departures from these basic homogeneity assumptions. For some CMRR models allowing some relaxations of these basic assumptions, large sample theory has been neatly developed in Rosen(1979). The basic problem with this development is that there are so many unknown parameters involved in the final structure that from the statistical inferential point of view, there is little encouragement in their possible adoptions.

6. Sampling with Varying Probabilities : Asymptotics

Hansen and Hurwitz (1943) initiated the use of unequal selection probabilities leading to more efficient estimators of the population total. If N and n stand for the number of units in the population and sample, respectively, and if $Y_1,\ldots,Y_N$ and $y_1,\ldots,y_n$ denote the values of these units in the population and sample respectively, then, one may consider the following sampling with replacement scheme. Let $P = (P_1,\ldots,P_N)$ be positive numbers which are normalized in such a way that $P_1+\ldots+P_N = 1$. Typically, one may consider a measure S_i of the size of the ith unit in the population and set $P_i = S_i/(\Sigma_{i=1}^N S_i)$, for i=1,...,N. Now, corresponding to the sample entries $y_1,\ldots,y_n$, the associated P's are denoted by $p_1,\ldots,p_n$, respectively. Here, sampling is made with replacement and the jth unit in the population is drawn with the probability P_j , for j=1,...,N. Then, the Hansen-Horvitz estimator of the population total $Y = Y_1 + \ldots+Y_N$ is

$$\hat{Y}_{HH} = n^{-1}(y_1/p_1 + \ldots + y_n/p_n) . \tag{6.1}$$

This estimator is unbiased and its sampling variance is given by

$$\begin{aligned} \mathrm{Var}(\hat{Y}_{HH}) &= n^{-1}\{ \Sigma_{i=1}^N Y_i^2/P_i - Y^2 \} \\ &= (2n)^{-1} \Sigma_{1\le i\ne j\le N} P_iP_j\{ Y_i/P_i - Y_j/P_j\}^2 \end{aligned} \tag{6.2}$$

We may further note that

$$\begin{aligned} S_{nHH}^2 &= [n(n-1)]^{-1} \Sigma_{i=1}^n(y_i/p_i - \hat{Y}_{HH})^2 \\ &= [2n^2(n-1)]^{-1}\Sigma_{1\le i\ne j\le n} \{y_i/p_i - y_j/p_j \}^2 \end{aligned} \tag{6.3}$$

is an unbiased estimator of $\mathrm{Var}(\hat{Y}_{HH})$. Since sampling is made with replacement and the y_i/p_i are independent with mean Y and variance $\Sigma_{i=1}^N Y_i^2/P_i - Y^2$ (= σ_{NHH}^2, say), standard large sample theory is adoptable to verify that as n increases,

$$nS_{nHH}^2/\sigma_{NHH}^2 \text{ converges to one, in probability,} \tag{6.4}$$

$$n^{\frac{1}{2}}(\hat{Y}_{HH} - Y) \text{ is asymptotically normal } (0, \sigma_{NHH}^2), \tag{6.5}$$

so that, by (6.4) and (6.5),

$$n^{\frac{1}{2}}(\hat{Y}_{HH} - Y)/S_{nHH} \text{ is asymptotically normal } (0, 1) . \tag{6.6}$$

The situation becomes quite different when sampling is made without replacement. In one hand, one has generally more efficient estimators; on the other hand, the exact theory becomes so complicated that one is narurally inclined to rely mostly on the asymptotics. To encompass diverse sampling plans (without replacements), we identify the population with the set $N = \{1,\ldots,N\}$ of natural integers and denote the sample by s . A sampling design may then be defined by the probabilities $p(s)$, $s \in S$, associated with all possible samples. In particular, we let

$$\pi_i = P\{i \in s\} = \Sigma_{\{\text{all s containing } i\}}\, p(s) \ , \ i=1,\ldots,N. \tag{6.7}$$

These are termed the first order inclusion probabilities. Similarly, the second order inclusion probabilities are defined as

$$\pi_{ij} = P\{ i,j \in s\} \ = \Sigma_{\{\text{all s containing } (i,j)\}}\, p(s), \ i\neq j=1,\ldots,N \ . \tag{6.8}$$

The classical Horvitz-Thompson (1952) estimator of the population total Y is then expressible as

$$\hat{Y}_{HT} \ = \Sigma_{i\in s}(\ y_i/\pi_i) \ . \tag{6.9}$$

Various properties of this estimator are discussed in some other chapters, and, hence, we shall not repeat the discussion here. We shall mainly concentrate on the asymptotic theory. The sampling variance of this unbiased estimator of Y is

$$\text{var}(\ \hat{Y}_{HT}) = \Sigma_{i=1}^{N} \ (\pi_i^{-1}-1)Y_i^2 + \Sigma_{1\leq i\neq j\leq N} \ (\pi_{ij}/\pi_i\pi_j \ -1)Y_iY_j \ . \tag{6.10}$$

When the number of units (n) in the sample s is fixed, an alternative expression for the variance in (6.10), due to Sen(1953) and Yates and Grundy(1953), is

$$\Sigma_{1\leq i<j\leq N} \ (\pi_i\pi_j \ - \ \pi_{ij} \ )(\ Y_i/\pi_i \ - \ Y_j/\pi_j \ )^2 \ . \tag{6.11}$$

It is clear that if the Y_i are all (exactly or closely) proportional to the corresponding π_i , then (6.11) is (exactly or closely) equal to 0 ; this point advocates the choice of the π_i as proportional to the size of the units, and on that count, 'probability proportional to size' (pps) sampling is quite a reasonable option.

Now, in the context of sampling (without replacement) with varying probabilities, various sampling designs have been considered by various workers. Among these, rejective sampling may be defined as in Hájek(1964) as sampling with replacement

with drawing probabilities $\alpha_1,\ldots,\alpha_N$ at each draw, conditioned on the requirement that all drawn units are distinct. The α_i are positive numbers adding upto 1. As soon as one obtains a replication, one rejects the whole partially built up sample and starts completely new. In this scheme, the inclusion probabilities π_i can be computed, as in Hájek (1964), in terms of the α_i . A related sampling plan, known as the Samford-Durbin sampling, is defined in a similar manner, where the first unit in the sample is drawn from the population with the probabilities $\alpha_i^{(1)} = n^{-1}\pi_i$, i=1,...,N, and in the subsequent (n-1) draws, one considers the drawing probabilities as $\alpha_i^{(*)} = \alpha\pi_i(1-\pi_i)^{-1}$, i=1,...,N where α is so selected that $\Sigma_{i=1}^{N} \alpha_i^{(*)} = 1$. Here also, a sample is accepted only if all selected units are distinct. For both these schemes, a rejection of the accumulating sample is made when at any intermediate stage, a repetition occurs. In a successive sampling plan one draws units one by one with drawing probabilities $P_1,\ldots, P_N$, and, if a replication occurs at any draw, that particular one is rejected, and the drawing is continued in this manner until one has the prefixed number n of distinct units in the sample . Following Rosen(1972,1974), let I_1 ,.. .,I_n be the indices in the (random) order in which they appear in the sample of size n , and let

$$\Delta(r,n) = \text{Probability that item r is included in the sample of size n drawn accprding to the successive sampling plan,} \qquad (6.12)$$

for r= 1,...,N. For this scheme, the Horvitz-Thompson estimator in (6.9) reduces to the following:

$$\hat{Y}_{HT} = \Sigma_{i=1}^{n} Y_{I_i}/\Delta(I_i,n) \ . \qquad (6.13)$$

In passing, we may remark that if $P_1=\ldots=P_N = N^{-1}$, the $\Delta(r,n)$ all reduces to n/N, so that (6.13) is given by $N(n^{-1}\Sigma_{i=1}^{n} y_{I_i})$, and hence, relates to the usual equal probability sampling scheme (without replacemeamt). In the more general case where the P_i are not all equal, the $\Delta(r,n)$ can be obtained as in Rosen (1972) in terms of a set of inclusion probabilities. However, these expressions

(as given below) are generally quite complicated and call for asymptotic considerations. For every $n(\geq 1)$ and $r_1,\ldots,r_n : 1 \leq r_1 \neq \ldots \neq r_n \leq N$, let

$$P(r_1,\ldots,r_n) = P_{r_1}.\{\Pi_{k=2}^{n} P_{r_k}[1- \Sigma_{j=1}^{k-1} P_{r_j}]^{-1}\}. \tag{6.14}$$

Then,

$$\Delta(r,n) = \Sigma_{j=1}^{n} \{ \Sigma_{(j)} P(r_1,\ldots,r_n)\}, \tag{6.15}$$

where the summation $\Sigma_{(j)}$ extends over all permutations of $(r_1,\ldots,r_n)$ over $(1,\ldots,N)$, subject to the constraint that $r_j = r$; $j=1,\ldots,n$, for $r =1,\ldots,N$.

The varying probability structure and the complications underlying the $\Delta(r,n)$ in (6.14)-(6.15) introduce certain complications in the study of the asymptotic distribution theory of the Horvitz-Thompson estimator (or other estimators available in the literature). Rosen (1970,1972) considered an alternative approach (through the coupon collector's problem) and provided some deeper results in this context. To illustrate this approach, we first consider a coupon collector problem. Let

$$\Omega_N = \{(a_{N1},P_{N1}) ,\ldots , (a_{NN},P_{NN})\} , \quad N \geq 1 , \tag{6.16}$$

be a sequence of coupon collector's situations, where the a_{Nj} and P_{Nj} are real numbers, the P_{Nj} are positive and $\Sigma_{j=1}^{N} P_{Nj} = 1$, $\forall\ N \geq 1$. Consider also a (double) sequence $\{ J_{Nk}, k \geq 1\}$ of (row-wise) independent and identically distributed random variables, where, for each $N(\geq 1)$, $k \geq 1$,

$$P\{ J_{Nk} = s \} = P_{Ns} , \text{ for } s=1,\ldots,N . \tag{6.17}$$

Let then

$$X_{Nnk} = \begin{cases} a_{NJ_{Nk}}/\Delta(J_{Nk},n) & , \text{ if } J_{Nk} \notin \{J_{N1},\ldots,J_{Nk-1}\}, \\ 0, & \text{otherwise ;} \end{cases} \qquad k \geq 1. \tag{6.18}$$

$$\nu_{Nm} = \inf\{ n : \text{number of distinct } J_{N1},\ldots,J_{Nn} = m \}, \ m \geq 1. \tag{6.19}$$

Note that for each $N(\geq 1)$, the ν_{Nm} are positive integer-valued random variables. Then, Rosen(1970,1972) has shown that on identifying $\{Y_1,\ldots,Y_N; P_1,\ldots,P_N\}$ with Ω_N in (6.16) ,

$$Y_{HT} \overset{\mathcal{D}}{=} \Sigma_{k=1}^{\nu_{Nn}} X_{Nnk} = Z_{n\nu_{Nn}} , \text{ say,} \tag{6.20}$$

where $\overset{\mathcal{D}}{=}$ stands for the equality in distributions. Now, for a coupon collector situation Ω_N in (6.16), $B_{Nn} = \Sigma_{k=1}^{n} a_{NJ_{Nk}}$ is termed the Bonus sum after n coupons, for $n \geq 1$. Thus, corresponding to the situation Ω_N in (6.16), if we consider another situation $\Omega_{Nn}^{*} = \{(a_{N1}/\Delta(1,n), P_{N1}), \ldots, (a_{NN}/\Delta(N,n), P_{NN})\}$, and, as in before, identify (a_{Nj}, P_{Nj}) with (Y_j, P_j), $j=1,\ldots,N$, then, for a given n, $Z_{n\nu_{Nn}}^{*}$ in (6.20) is the bonus sum after ν_{Nn} coupons in the collector's situation Ω_{Nn}^{*}. Thus, the asymptotic normality of (randomly stopped) bonus sums (for the reduced coupon collector's situation) provides the same result for the Horvitz-Thompson estimator. A similar treatment holds for many other related estimators in successive sampling with varying probabilities (without replacement). Towards this goal, we may note that as in Rosen(1972), under some regularity conditions on the a_{Ni} as well as the P_{Ni}, as N increases,

$$\Delta(s,n) = 1 - \exp\{-P_{Ns}t(n)\} + o(N^{-\frac{1}{2}}), \quad s=1,\ldots,N, \tag{6.21}$$

where the function $t(.) = \{t(x), x \geq 0\}$, is defined implicitly by

$$N - x = \Sigma_{k=1}^{N} \exp\{-t(x)P_{Nk}\}, \quad x \geq 0, \tag{6.22}$$

(and therefore, depends on $P_{N1},\ldots,P_{NN}$). Given this asymptotic relation, we may write for every $s(=1,\ldots,N)$,

$$a_{Ns}^{*} = a_{Ns}/\Delta(s,n) = (1 - \exp\{-P_{Ns}t(n)\})^{-1} a_{Ns} + o(N^{-\frac{1}{2}}). \tag{6.23}$$

It also follows from Rosen(1970) that under the same regularity conditions,

$$\nu_{Nn}/t(n) \text{ converges in probability to one,} \tag{6.24}$$

whenever n/N is bounded away from 0 and 1. Consequently, if we define the bonus sum for the reduced coupon collector's situation by $B_{Nnk}^{*} = \Sigma_{i=1}^{k} a_{NJ_{Ni}}^{*}$, $k \geq 1$, then, we need to verify that (i) the normalized version of $B_{Nnt(n)}^{*}$ is asymptotically normal, and (ii) $n^{-\frac{1}{2}} \max\{|B_{Nnk}^{*} - B_{Nnt(n)}^{*}| : |k/t(n) - 1| \leq \delta\}$ converges in probability to 0 (the later condition is known in the literature as the Anscombe (1952) 'uniform continuity in probability' condition). A stronger result, which ensures both (i) and (ii), relates to the weak convergence of the partial sequence $\{(B_{Nnk}^{*} - EB_{Nnk}^{*})/\{var(B_{Nnt(n)}^{*})\}^{\frac{1}{2}}; k \leq t(n)\}$, and, this has

been established by Sen (1979a) through a martingale approach. For simplicity of presentation, we consider the case of the original coupon collector's situation (and the same result continues to hold for the reduced situation too). Let us denote by

$$\phi_{Nn} = \Sigma_{s=1}^{N} a_{Ns}[\,1 - \exp\{-nP_{Ns}\}], \quad n \geq 0, \tag{6.25}$$

$$d_{Nn}^2 = \Sigma_{s=1}^{N} a_{Ns}^2 \exp\{-nP_{Ns}\}[1 - \exp\{-nP_{Ns}\}] - n\Big(\Sigma_{s=1}^{N} a_{Ns}P_{Ns}\exp\{-nP_{Ns}\}\Big)^2, \quad n \geq 0. \tag{6.26}$$

Then, under the usual (Rosen-) regularuty conditions, it follows that

$$d_{Nn}^2 = O_e(n) \text{ whenever } n/N \text{ is bounded away from 0 and } \infty. \tag{6.27}$$

Further, under the same regularity conditions,

$$(B_{Nn} - \phi_{Nn})/d_{Nn} \text{ is asymptotically normal } (0,1). \tag{6.28}$$

In fact, if we consider any finite number, say, q, of the sample sizes, i.e., $n_1,\ldots,n_q$, where the n_j all satisfy the condition that $0 < n_j/N < \infty$, for $j=1,\ldots,q$, then, (6.28) readily extends to the multinormal case. Further, if we define $W_N = \{ W_N(t),\ t \in T \}$, where $T = [0,K]$ for some finite K, and

$$W_N(t) = N^{-\frac{1}{2}}\,(B_{N[Nt]} - \phi_{N[Nt]}),\ 0 \leq t \leq K, \tag{6.29}$$

then, it follows from Sen (1979a) that W_N converges in law to a Gaussian function on T, and this ensures the tightness of W_N as well (so that the Anscombe condition holds). In passing, we may remark that if the a_{Ns} are all nonnegative, the bonus sum B_{Nn} is then nondecreasing in n, so that, we may define

$$U_N(t) = \min\{\, k : B_{Nk} \geq t \,\},\ \text{for every } t \geq 0. \tag{6.30}$$

Then, $U_N(t)$ is termed the waiting time to obtain the bonus sum t in the coupon collector's situation Ω_N. Note that, by definition,

$$P\{\, U_N(t) > x \,\} = P\{\, B_{N[x]} < t \,\},\ \text{for all } x, t > 0. \tag{6.31}$$

Therefore, the asymptotic distribution of the normalized version of the waiting time can readily be obtained from (6.28), and, moreover, the weak convergence result on W_N also yields a parallel result for a similar stochastic process constructed from the $U_N(t)$.

Note that by (6.7), (6.8), (6.14) and (6.15), $\pi_r = \Delta(r,n)$ for every $r = 1,\ldots,N$, while for every $r \neq s$ $(=1,\ldots,N)$,

$$\pi_{rs} = \Sigma_{1\leq i\neq j\leq n} \{ \Sigma_{(ij)} \; P(r_1,\ldots,r_n)\}, \tag{6.32}$$

where the summation $\Sigma_{(ij)}$ extends over all permutations of $(r_1,\ldots,r_n)$ over $(1,\ldots,N)$ subject to the constraints that $r_i = r$ and $r_j = s$, for $i\neq j=1,\ldots,n$. Further, the expression for the variance in (6.10) can be rewritten as

$$\Sigma_{i=1}^{N} \Sigma_{j=1}^{N} \pi_{ij} Y_i Y_j/(\pi_i \pi_j) - Y^2 \quad \text{where } Y = \Sigma_{i=1}^{N} Y_i \;, \tag{6.33}$$

so that the expressions for the π_{ij} and π_i may be incorporated to evaluate (6.33). This, however, is quite complicated (in view of (6.14), (6.15) and (6.32)), and therefore, we proceed to obtain simpler expressions. We may note that for the reduced coupon collector's situation Ω_N^* and for n replaced by $t(n)$, we have parallel to (6.26),

$$\begin{aligned} d_{Nt(n)}^{*2} &= \Sigma_{s=1}^{N} a_{Ns}^{*2} \exp\{-t(n)P_{Ns}\}[1- \exp\{-t(n)P_{Ns}\}] \\ &\quad - t(n)[\Sigma_{s=1}^{N} a_{Ns}^{*} P_{Ns} \exp\{-t(n)P_{Ns}\}]^2 \\ &= \Sigma_{s=1}^{N} Y_s^2 \exp\{-t(n)P_{Ns}\}/[1- \exp\{-t(n)P_{Ns}\}] \\ &\quad - t(n)[\Sigma_{s=1}^{N} Y_s P_{Ns} \exp\{-t(n)P_{Ns}\}/(1-\exp\{-t(n)P_{Ns}\})]^2, \end{aligned} \tag{6.34}$$

where, we may note that the P_{Ns} are all specified numbers, so that by (6.22), $t(n)$ is a known quantity. As a result, we obtain that as N increases and n/N is bounded away from 0 (and is finite too),

$$(\hat{Y}_{HT} - Y)/d_{Nt(n)}^{*} \text{ is asymptotically normal } (0,1). \tag{6.35}$$

Further, if we take the sample observations as y_j $(= Y_{I_{Nj}})$, $j=1,\ldots,n$ and denote by $P_{NI_{Nj}} = p_{Nj}$, $j=1,\ldots,n$, we may set

$$U_{Nn}^{(1)} = n^{-1}\Sigma_{j=1}^{n} y_j^2 \exp\{-t(n)p_{Nj}\} /[1 - \exp\{-t(n)p_{Nj}\}]^2 \;, \tag{6.36}$$

$$U_{Nn}^{(2)} = n^{-1}\Sigma_{j=1}^{n} y_j p_{Nj} \exp\{-t(n)p_{Nj}\} /[1 - \exp\{-t(n)p_{Nj}\}]^2 \;, \tag{6.37}$$

$$V_{Nn} = U_{Nn}^{(1)} - t(n)[U_{Nn}^{(2)}]^2. \tag{6.38}$$

Then, it follows that as n increases, $V_{Nn}/d_{Nt(n)}^{*2}$ converges, in probability, to 1, so that in (6.35), $d_{Nt(n)}^{*}$ may also be replaced by $V_{Nn}^{\frac{1}{2}}$.

Besides the sampling strategies considered so far, there are some others considered by other workers. Among these, mention may be made of one approach proposed by Rao, Hartley and Cochran (1962). They suggested that the population of N units be first divided randomly into n sub-populations of predetermined sizes $N_1,\ldots,N_n$, respectively. Within each sub-population (or rather, group), using a convenient varying probability selection, a Horvitz-Thompson estimator is used to estimate the group total, and the sum of these estimates then taken as an estimator of the population total. When n, the number of groups , is large, and the groupings are made randomly, the asymptotic theory remains applicable under quite general conditions. However, there is some arbitrariness in this division of N units into subsets of $N_1,\ldots,N_n$ units which may make this procedure rather unappealing to practicians. For some further work in this direction, we may refer to Krewski (1978). Systematic procedures (random or ordered) for sampling with varying probabilities were also considered by Madow (1949), Hartley (1966) and Rao and Hartley (1962), among others. Some of these procedures are discussed in detail in some other chapters of this volume, and hence, we shall not elaborate on their related asymptotics. Besides the systematic procedures, there are other procedures due to Narain (1951), Midzuno (1952) , Yates and Grundy (1953), Sen (1953) , and others. Most of these procedures work out well for small values of n (viz. n = 2, 3 or 4), and as n increases, the procedures become prohibitively cumbrous. We may refer to Brewer and Hanif (1983) for some detailed discussions of these procedures when n need not be large. However, as regards the asymptotic theory is concerned, a lot of work remains to be accomplished.

7. Successive Sub-sampling with Varying Probabilities : Asymptotics.

Sub-sampling or multi-stage sampling is often adopted in practice and has a great variety of applications in survey sampling. These are elaborated in some other chapters of this volume. Typically, we may consider a finite population of

N units with variate values $a_{N1},\ldots,a_{NN}$, respectively. Consider a successive sampling scheme where items are sampled one after the other (without replacement) in such a way that at each draw the probability of drawing item s is proportional to a number P_{Ns} if item s has not already appeared in the earlier drawals, for s=1,...,N, where $P_{N1},\ldots,P_{NN}$ are a set of positive numbers, adding upto 1 . We like to consider a multi-stage extension of this sampling scheme . Here, each of the N items in the population (called the primary units) is composed of a number of smaller units (sub-units), and it may be more economic to select first a sample of n primary units, and then to use sub-samples of sub-units in each of these selected primary units. Suppose that the sth primary unit has M_s sub-units with variate values b_{sj} , j=1,...,M_s , so that $a_{Ns} = b_{s1} + \ldots + b_{sM_s}$, for s = 1,..., N. For each s, we conceive of a set $\{ P^{o}_{sj} , 1\leq j\leq M_s\}$ of positive numbers (such that $\Sigma_{j=1}^{M_s} P^{o}_{sj} = 1$) and consider a successive sampling sheme (without replacement), where m_s (out of M_s) sub-units are chosen. Then, as in (6.13), an estimator of a_{Ns} can be framed, for each of the n selected primary units. Finally, these estimates can be combided as in (6.13) to yield the estimator of the total $A_N = a_{N1} + \ldots + a_{NN}$. The procedure can be extended to the multi-stage case in a similar way. This scheme may be termed the successive sub-sapling with varying probabilities (without replacement) or SSSVPWR. To study the asymptotic theory, first, we may note that an Horvitz-Thompson estimator of a_{Ns} is

$$\hat{a}_{Ns} = \Sigma_{j=1}^{M_s} \omega^{*}_{sj}\, b_{sj}/\Delta^{*}_{s}(j,m_s) , \qquad (7.1)$$

where the b_{sj} are defined as before, ω^{*}_{sj} is equal to 1 or 0 according as the jth sub-unit in the sth primary unit belongs to the sub-sample of size m_s or not, j=1,...,M_s and $\Delta^{*}_{s}(j,m_s)$ is the probability that the jth sub-unit belongs to the sub-sample of m_s sub-units from the sth primary unit, $1 \leq j \leq M_s$, s=1,...,N. Combining (6.13) and (7.1), we may consider the natural estimator

$$\begin{aligned}\hat{A}_{N(HT)} &= \Sigma_{s=1}^{N} \omega_{Ns}\, \hat{a}_{Ns}/\Delta(s,n) \\ &= \Sigma_{s=1}^{N} \Sigma_{j=1}^{M_s} \omega_{Ns}\, \omega^{*}_{sj}\, b_{sj}/[\Delta(s,n)\Delta^{*}_{s}(j,m_s)] , \qquad (7.2)\end{aligned}$$

where ω_{Ns} is equal to 1 or 0 according as the sth primary unit is in the sample of n primary units from the population, s=1,...,N, and the inclusion probabilities $\Delta(s,n)$ are degined as in (6.12). Note that for each (selected) primary unit s, for the estimator $\hat{a}_{Ns}$ in (7.1), one may use the theory discussed in Section 6. This, however, leads to a multitude of stopping numbers and thereby introduces complications in a direct extension of the Rosen approach to SSVPWR. A more simple approach (based on some martingale constructions) has been worked out in Sen(1980), and we may present the basic asymptotic theory as follows.

Our primary interest is to present the asymptotic theory of the estimator $\hat{A}_{N(HT)}$ in (7.2). In this context, as in earlier sections, we allow N to increase. As $N \to \infty$, we assume that n, the primary sample size, also increases, in such a way that n/N is bounded away from 0 and ∞ , while the m_s (i.e., the sub-sample sizes) for the selected primary units may or may not be large. For this situation, the asymptotic theory rests heavily on the structure of the primary unit sampling, and , we may also allow the sampling scheme for the sub-units to be rather arbitrary (not necessarily a SSVPWR), while we assume that the primary units are sampled in accordance with a SSVPWR scheme. A second situation may arise where the number of primary units (i.e., N) is fixed or divided in to a fixed number of strata, and within each strata a sample of secondary units is drawn according to a SSVPWR scheme. This situation, however, is congruent to the stratified sampling scheme under SSVPWR , for which the theory in Section 6 extends readily. Hence, we shall not enter into the detailed discussions on this second scheme.

With the notations introduced before, we set now

$$a^{o}_{Ns} = E(\hat{a}_{Ns}) \quad \text{and} \quad \sigma^{2}_{Ns} = \mathrm{Var}(\hat{a}_{Ns}) \ , \ \text{for } s=1,\ldots,N \ ; \tag{7.3}$$

$$A^{o}_{N} = \Sigma_{s=1}^{N} a^{o}_{Ns} = E(\hat{A}_{N(HT)}) \ . \tag{7.4}$$

In order that $A^{o}_{N} = A_N$, it is therefore preferred to have unbiased estimators at the sub-unit stage, so that $a^{o}_{Ns} = a_{Ns}$, for every s. Otherwise, the bias may not be negligible. Also, for every N, we consider a nondecreasing function $t_N = \{t_N(x): 0 \leq x \leq N\}$ by letting

$$N - x = \sum_{s=1}^{N} \exp\{-P_{Ns} t_N(x)\} \quad , x \in (0, N) . \tag{7.5}$$

Let then

$$\delta_{Nn}^2 = \sum_{s=1}^{N} [a_{Ns}^o]^2 \exp\{-P_{NS}\, t_N(n)\}[1 - \exp\{-P_{Ns} t_N(n)\}]^{-1}$$
$$+ \sum_{s=1}^{N} \sigma_{Ns}^2 [1 - \exp\{-P_{Ns} t_N(n)\}]^{-1}$$
$$- t_N(n)[\sum_{s=1}^{N} a_{Ns}^o P_{Ns} \exp\{-P_{Ns}\, t_N(n)\}/(1 - \exp\{-P_{Ns} t_N(n)\})]^2 . \tag{7.6}$$

Finally, we assume that the sub-unit estimators $\hat{a}_{Ns}$ satisfy a Lindeberg-type condition, namely, that for every $\eta > 0$,

$$\max_{1 \le s \le N} E[(\hat{a}_{Ns} - a_{Ns}^o)^2 I(|\hat{a}_{Ns} - a_{Ns}^o| > \eta N^{\frac{1}{2}})] \to 0 , \text{ as } N \to \infty . \tag{7.7}$$

The other regularity conditions are, of course, the compatibility of the probabilities $P_{N1}, \ldots, P_{NN}$ and the sizes $a_{N1}, \ldots, a_{NN}$ (in the sense that for each sequence the ratio of the maximum to the minimum entry is asymptotically finite). Then, we have the following:

$$(\hat{A}_{N(HT)} - A_N^o)/\delta_{Nn} \text{ is asymptotically normal } (0, 1). \tag{7.8}$$

Actually, parallel to (6.29), we may consider a stochastic process $\xi_N = \{\xi_N(t); c \le t \le 1\}$ (where $c > 0$), by letting $\xi_N(t) = N^{-\frac{1}{2}} (\hat{A}_{N(HT)}^{(t)} - A_N^o)$, where $\hat{A}_{N(HT)}^{(t)}$ is the estimator in (7.2) based on the sample size $n = [Nt]$ (for the primary sample), $t \in [c,1]$. Then, the process ξ_N converges in law to a Gaussian function on $[c,1]$. The proofs of these results are based on some asymptotic theory for an extended coupon collector's problem, where in (6.16) through (6.19), the real (non-stochastic) elements a_{Ns} are replaced by suitable random variables X_{Ns} , $s=1,\ldots,N$. For details of these developments, we may refer to Sen (1980).

Note that in the above development, apart from the uniform integrability condition in (7.7), we have not imposed any restriction on the estimates $\hat{a}_{Ns}$. Thus, we are allowed to make the sub-sample sizes m_s arbitrary, subject to the condition that (7.7) holds. In this context, we may note that if these m_s are also chosen to be large, then the σ_{Ns}^2 , defined by (7.3), will be small, so that in (7.6), the second sum on the right hand side will be of smaller order of magnitude (compared to the first sum), and hence, in (7.8), δ_{Nn} may be replaced

by d^{*}_{Nn} , defined by (6.34), where the Y_s are to be replaced by the a^{o}_{Ns} . In this limiting case, we observe therefore that sub-sampling does not lead to any significant increase of the variance (compared to SSVPWR for the primary units); although in many practical problems, sub-sampling is more suitable, because it does not presuppose the knowledge of the values of the primary units $\{a_{Ns}\}$ and a complete census for these may be much more expensive than the estimates $\{\hat{a}_{Ns}\}$ based on a handful of sub-units.

So far, we have considered sampling without replacement. In SSVP sampling with replacement, the theory of successive VP sampling with replacement, discussed in the beginning of Section 6, readily extends. In (6.1), instead of the primary units y_j , we need to use their estimates $\hat{y}_j$, derived from the respective sub-samples. As in (7.6), this will result in an increaded variability due to the individual variances of the second-stage estimators. However, with replacement strategy yields simplifications in the treatment of the relevant asymptotic theory , and (6.5) and (6.6) both extend to this sub-sampling scheme without any difficulty.

In conclusion, we may remark that in finite population sampling, the usual treatment for the asymptotic theory (valid for independent random variables) may not be directly applicable. But, in most of these situations, by appeal to either some appropriate permutation structures(for equal probability sampling) or to some martingale theory (for VP sampling as well), the asymptotic theory has been established under quite general regularity conditions. These provide theoretical justifications of the asymptotic normality of different estimators (under diverse sampling schemes) when the sample size(s) may or may not be non-stochastic. In particular, for optimal allocation based on pilot data, often, we end up with sample sizes being random (positive integer valued) variables. In such a case, the asymptotic results on the stochastics processes referred to earlier are useful. These results are also useful for quasi-sequential or repeated significance testing problems in finite population sampling .

References

Arnab, R. (1979). On strategies of sampling finite populations on successive occasions with varying probabilities. Sankhya, Ser.C 41 , 140-154.

Asok, C. and Sukhatme, B.V. (1976). On Samford's procedure of unequal probability sampling without replacement. J. Amer. Statist. Asso. 71, 912-918.

Brewer, K.R.W. and Hanif, M. (1983). Sampling with Unequal Probabilities. Lecture Notes in Statistics 15 , Springer-Verlag, New York.

Chapman, D.G. (1951). Some properties of the hypergeometric distribution with applications to zoological census. Univ. Calif. Publ. Statist. 1, 131-160.

Chapman, D.G. (1952). Inverse, multiple and sequential sample censuses. Biometrics 8, 286-306.

Chaudhuri, A. (1975). Some properties of estimators based on sampling schemes with varying probabilities. Austral. J. Statist. 17, 22-28.

Chaudhuri, A. (1976). On the choice of sampling strategies. Calcutta Statist.Asso. Bull. 25, 119-128.

Cochran, W.G. (1977). Sampling Techniques (3rd Ed.). John Wiley, New York.

Cochran, W.G., Hartley, H.O. and Rao, J.N.K. (1962). On a simple procedure of unequal probability sampling without replacement. J. Roy. Statist. Soc. Ser. B 24 , 482-491.

Cormack, R.M. (1968). The statistics of capture-recapture methods. Oceanogr. Mar. Biol.Ann. Rev. 6, 455-506.

Darling, D.A. and Robbins, H. (1967). Finding the size of a finite population. Ann. Math. Statist. 38, 1392-1398.

Darroch, J.N. (1958). The multiple recapture census,I: estimation of closed population. Biometrika 45 , 343-359.

Das, A.C. (1951). On two phase sampling and sampling with varying probabilities. Bull. Internat. Statist. Inst. 33, 105-112.

Durbin, J. (1953). Some results in sampling when units are selected with unequal probabilities. J. Roy. Statist. Soc. Ser. B15, 262-269.

Farm, A. (1971). Asymptotic normality in a capture-recapture problem when catchability is affected by the tagging procedure. Tech.Rep. TRITA-MAT-1971-10, Roy. Inst. Tech., Stockholm, Sweden.

Goodman, L.A. (1953). Sequential sampling tagging for population size problems. Ann. Math. Statist. 24, 56-69.

Hájek, J.(1959). Optimum strategy and other problems in probability sampling. Casopis Pro Pěstování Matematiky 84, 387-423.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. Ann. Math. Statist. 35, 1491-1523.

Hájek, J. (1974). Asymptotic theories of sampling with varying probabilities without replacement. Proc. Prague Confer. Asymp. Statist.(Ed. J. Hájek),pp.127-138.

Hájek, J. (1981). Sampling From a Finite Population. Marcel Dekker, New York.

Hájek, J. and Šidák, Z. (1967). Theory of Rank Tests. Academic Press, New York.

Hanif, M. and Brewer, K.R.W. (1980). On unequal probability sampling without replacement: A review. Internat. Statist. Rev. 48, 317-335.

Hansen, M.H. and Hurwitz, W.N. (1943). On the theory of sampling from a finite population. Ann. Math. Statist. 14, 333-362.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample Survey Methods and Theory. John Wiley, New York.

Hartley, H.O. (1966). Systematic sampling with unequal probability and without replacement. J. Amer. Statist. Asso. 61, 739-748.

Hartley, H.O. and Rao, J.N.K. (1962). Sampling with unequal probabilities and without replacement. Ann. Math. Statist. 33, 350-374.

Hoeffding, W. (1948). On a class of statistics with asymptotically normal distribution. Ann. Math. Statist. 19, 293-325.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. J. Amer. Statist. Asso. 58, 13-30.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Asso. 47, 663-685.

Karlin, S. (1974). Inequalities for symmetric sampling plans, I. Ann. Statist. 2 , 1065-1094.

Krewski, D. (1978). Jackknifing U-statistics in finite populations. Commun.Statist. Theor. Meth. A 7 1-12.

Lahiri, D.B. (1951). A method for selection providing unbiased estimates. Bull. Internat. Statist. Inst. 33, 133-140.

Madow, W.G. (1949). On the theory of systematic sampling, II. Ann. Math. Statist. 20, 333-354.

Majumdar, H. and Sen, P.K. (1978). Invariance principles for jackknifing U-statistics for finite population sampling and some applications. Commun. Statist. Theor. Meth. A7 , 1007-1025.

Midzuno, H. (1952). On the sampling system with probability proportionate to sum of sizes. Ann. Inst. Statist. Math. 3, 99-107.

Miller, R.G., Jr. and Sen, P.K. (1972). Weak convergence of U-statistics and von Mises' differentiable statistical functions. Ann. Math. Statist. 43, 31-41.

Murthy, M.N. (1967). Sampling Theory and Methods. Statist. Pub. Soc., Calcutta.

Nandi, H.K. and Sen, P.K. (1963). On the properties of U-statistics when the observations are not independent. Part Two : Unbiased estimation of the parameters of a finite population. Calcutta Statist. Asso. Bull. 12, 125-148.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. J. Indian Soc. Agricul. Statist. 3, 169-175.

Pathak, P.K. (1967a). Asymptotic efficiency of Des Raj's strategy, I. Sankhya , Ser. A 29, 283-298.

Pathak, P.K. (1967b). Asymptotic efficiency of Des Raj's strategy, II. Sankhya, Ser. A 29 , 299-304.

Petersen, C.G.J. (1896). The yearly immigration of young plaice into the Limfjord from the German sea. Rep. Danish. Biol. Sta. 6, 1-48.

Raj, Des (1956). Some estimators in sampling with varying probabilities without replacement. J. Amer. Statist. Asso. 51, 269-284.

Rao, C.R. (1977). Some problems of sample surveys. Sankhya, Ser. C 39, 128-139.

Rao, J.N.K. (1961). On sampling with varying probabilities and with replacement in subsampling designs. J. Indian Soc. Agri. Soc. 13, 211-217.

Rao, J.N.K. (1966). On the comparison of sampling with and without replacement. Rev. Internat. Statist. Inst. 34, 125-138.

Rosen, B. (1967). On an inequality of Hoeffding. Ann. Math. Statist. 38, 382-392.

Rosen, B. (1970). On the coupon collector's waiting time. Ann. Math. Statist. 41, 1952-1969.

Rosen, B. (1972a). Asymptotic theory for successive sampling with varying probabilities without replacement, I . Ann. Math. Statist. 43, 373-397.

Rosen, B. (1972b). Asymptotic theory for successive sampling with varying probabilities without replacement, II. Ann. Math. Statist. 43, 748-776.

Rosen, B. (1974). Asymptotic theory for Des Raj estimators. Proc. Prague Confer. Asymptotic Statist. (ed. J. Hajek), pp.313-330.

Rosen, B. (1979). A tool for derivation of asymptotic results for sampling and occupancy problems, illustrated by its application to some general capture-recapture sampling problems. Proc. 2nd Prague Confer. Asymptotic Statist. (ed. P. Mandl and M. Hušková), pp.67-80.

Samuel, E. (1968). Sequential maximum likelihood estimation of the size of a population. Ann. Math. Statist. 39,1057-1068.

Schnabel, Z.E. (1938). The estimation of the total fish population of a lake. Amer. Math. Mon. 45, 348-352.

Seber, G.A.F. (1973). The estimation of animal abundance, Griffin: London.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. J. Indian Soc. Agri. Statist. 5, 119-127.

Sen, A.R. and Sen, P.K. (1981). Schnabel type estimators for closed populations with multiple markings. Sankhya, Ser. B 43, 68-80.

Sen, P.K. (1960a). On some convergence properties of U-statistics. Calcutta Statist. Asso. Bull. 10, 1-19.

Sen, P.K. (1960b). On the estimation of population size by the capture-recapture method. Calcutta Statist. Asso. Bull. 10 , 91-110.

Sen, P.K. (1970). The Hajek-Renyi inequality for sampling from a finite population. Sankhya, Ser. A 32 , 181-188.

Sen, P.K. (1972). Finite population sampling and weak convergence to a Brownian bridge. Sankhya, Ser. A 34, 85-90.

Sen, P.K. (1977). Some invariance principles relating to jackknifing and their role in sequential analysis. Ann. Statist. 5, 315-329.

Sen, P.K. (1979a). Invariance principles for the coupon collector's problem : A martingale approach. Ann. Statist. 7, 372-380.

Sen, P.K. (1979b). Weak convergence of some quantile processes arising in progressively censored tests. Ann. Statist. 7, 414-431.

Sen, P.K. (1980). Limit theorems for an extended coupon collector's problem and for successive sub-sampling with varying probabilities. Calcutta Statist. Asso. Bull. 29, 113-132.

Sen, P.K. (1981). Sequential Nonparametrics: Invariance Principles and Statistical Inference. John Wiley, New York.

Sen, P.K. (1982a). On the asymptotic normality in sequential sampling tagging. Sankhya, Ser.A 44, 352-363.

Sen, P.K. (1982b). A renewal theorem for an urn model. Ann. Probability 10, 838-843.

Sen, P.K. (1983). On permutational central limit theorems for general multivariate linear rank statistics. Sankhya, Ser. A 45, 141-149.

Serfling, R.J. (1974). Probability inequalities for the sum in sampling without replacement. Ann. Statist. 2, 39-48.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. J. Roy. Statist. Soc. Ser.B 15, 253-261.