COMPLETENESS AND UNBIASED ESTIMATION

FOR SUM-QUOTA SAMPLING

BU-837-M   by   April, 1984

Walter K. Kremers

Biometrics Unit, Cornell, Ithaca, New York

ABSTRACT

Random variables, X_i, are sampled sequentially from a finite population until the sum, of the associated nonnegative random variables, C_i, is greater than or equal to a predetermined quota, Q. The objective of this paper is to show that the minimal sufficient statistic described by Pathak (1976,Ann.Statist.4:1012-1017) is also complete and to thereby obtain UMVUE's for population parameters. These estimates are similar to their fixed sample size counterparts except they have an adjustment for the sampling bias of the terminal observation. We also consider an adaptation of double sampling to sum-quota sampling in the bivariate case.

1.INTRODUCTION

Most statistical methodology in use today is designed for sampling schemes or experimental designs where the sample size or number of data points is regarded as a fixed preset parameter. This is in part for mathematical considerations but is also suggested by the intuition that sampling is restricted by a total cost of data collection and that the cost of sampling remains the same from unit to unit. However the cost of sampling may vary from unit to unit and the costs may be unknown before the sample is taken. If so, the objective of taking a sample of preset cost, Q, forces us to sample sequentially until achieving cost Q, and then stop. A cost may be monetary or general in character. Areas where sum-quota sampling may arise naturally include 1) the laboratory when sampling microscopic fields until a predetermined number of particles are observed, 2) packaging where we sample units until achieving a prespecified mass, or 3) nature where a big fish eats little fish until well satiated.

Pathak (1976) described sum-qouta sampling but used the phrase fixed cost sequential sampling. We choose the term sum-quota as it applies more naturally to the cases where cost is not the determinant of our stopping rule. Pathak derived unbiased estimates of population parameters based on all but the final observation. The objective of this paper is to show that the minimal sufficient statistic described by Pathak is also complete. We then obtain UMVUE's for population parameters in the univariate case by the use of completeness. These estimates are similar to their fixed sample size counterparts with an adjustment for the sampling bias of the final

observation. We also consider an adaptation of double sampling to sum-quota sampling in the bivariate case where we derive unbiased estimates of population parameters. Here, though, the minimal sufficient statistic is no longer complete.

2. MODEL AND THEORY FOR THE UNIVARIATE CASE

Consider a finite population of size N. For the j'th population unit define $U_j = (j, C_j, X_j)$ where j is the unit index and C_j ($C_j \geq 0$) is the cost of observing X_j. We assume j is not observable and denote the i'th unit sampled in a sequential fashion by (C_i, X_i).

Define a Sum-quota(Q, n_o) sampling scheme as a design where we begin by sampling n_o units randomly without replacement. If $\sum^{n_o} C_i \geq Q$ we stop. If $\sum^{n_o} C_i < Q$ we sample sequentially until $\sum^{(\nu-1)} C_i < Q \leq \sum^{\nu} C_i$ and then stop. We then have a sample size, ν, which is random.

We call (C_ν, X_ν) the terminal observation and all other observations collectively the preterminal observations. Define $\underline{X} = ((C_1, X_1), \ldots, (C_\nu, X_\nu))$ as the ordered sample and $\underline{X}_\nu = ((C_1, X_1), \ldots, (C_{\nu-1}, X_{\nu-1}))$ as the ordered sample of preterminal observations.

The preterminal observations behave much as if they are taken as a random sample of a fixed sample size. However the sampling distribution of C_ν is different from the sampling distribution of C_1 and incurs a sampling bias according to "length" or "size". Roughly speaking we expect the probability of observing $C_\nu = c$ to be approximately $c\, P(c)/\mu(C)$ where $P(c)$ is the probability of observing c when taking a simple random sample of size one from the C_j and $\mu(C) = \sum^N C_j$ (Cox, 1969) . This phenomenon is known as

length-biased sampling. If the X's and C's are correlated then there will

exist a sampling bias in X_ν as well. Thus our analysis will become simpler

if we consider the terminal and preterminal observations differently. A

second reason for distinguishing between the terminal and preterminal

observations is in practice we may stop sampling the moment we reach our

cost Q thus failing to completely determine (C_ν, X_ν). We thus consider the

statistics $T_\nu = (\{(C_1, X_1), \ldots, (C_{\nu-1}, X_{\nu-1})\}$, (C_ν, X_ν)) and T =

$\{(C_1, X_1), \ldots, (C_\nu, X_\nu)\}$. T_ν is an ordered pair, the first element being the

set of preterminal observations and the second element being the terminal

observation. T is simply the set of observations. Since the cost of

observing $C_1, \ldots, C_{\nu-1}$ is not dependent on the order of the observations,

conditional on T_ν all permutations of the preterminal observations are

equally as likely to be $\underline{X}_\nu$, the ordered sample of preterminal observations

(Pathak, 1976). Hence conditional on T_ν the preterminal observations are

exchangeable. We now consider the conditional distribution of $\underline{X}$ given T.

First if $\nu = n_o$ then all permutaions are equally as likely as $\underline{X}$. Next

consider the case $\nu > n_o$. If $(C', X') \varepsilon$ T and if $\sum^\nu C_i - C' < Q$ then possibly

$(C', X') = (C_\nu, X_\nu)$. Let $\#S$ be the number of such C' in T. Again since the

cost of taking observations is not dependent on the order of the

observations and since sampling is random for each observation, conditional

on T all such (C', X') are equally as likely to be (C_ν, X_ν). (Each

observation in our sequential sampling procedure is taken randomly. It is

only conditional on ν that the distributions of the (C_i, X_i)'s are no longer

the same as the population from which we are sampling.) Thus conditional on

T, $\underline{X}$ is distributed with mass $1/(\nu-1)!\#S$ for each point X such that X is

a permutation of T with $\sum^{\nu} C_i - C_\nu < Q$. Since the distribution of $\underline{X}$ conditional on T (or T_ν) is dependent only on T (or T_ν) ,T (and T_ν) are sufficient.

We now consider conditions under which T is also complete. Let $\theta = ((C_1,X_1),\ldots,(C_N,X_N))$ denote a parameterization of the finite population (of size N), where the j'th element represents the vector value of the j'th population unit. If we assume the set V of all possible values for the pair (C_j,X_j) is the same for all the population units, and does not depend on the values taken on by other units, then θ, the parameter space, consists of all θ a member of the cartesian product V^N.

Theorem 1. If the parameter space is given by θ above, $T = \{(C_1,X_1),\ldots,(C_n,X_n)\}$ is complete for sum-quota sampling.

Proof. Consider parameter spaces θ' and θ'', such that θ' is a subset of θ'', with supports $D(\theta')$ and $D(\theta'')$. Assume that $E[f(T)]=0$ for every θ ε θ' implies $f(T)=0$ a.e. θ'. It follows that $E[f(T)]=0$ for every θ ε θ'' implies $f(T)=0$ a.e. θ'' if we can show $f(T)=0$ a.e. θ'' on the set $D(\theta'')$ $-D(\theta')$. We show completeness of the measures induced on T by $\theta=V^N$ and sum-quota sampling with a sequence of such steps.

Assume f is such that $Ef(T)=0$ for every θ ε θ We begin by considering any realization $t=\{(c_1,x_1),\ldots,(c_n,x_n)\}$ which has positive probability for some θ of θ. We want to show that $f(t)=0$. Let $c_{max}=\max_{i=1,\ldots,n}(c_i)$ and x_{max} the corresponding x_i. Note that it is not necessary that $x_{max}=\max_{i=1,\ldots,\nu}(x_i)$. First consider $\theta_o=((c_{max},x_{max}),\ldots,(c_{max},x_{max}))$. Then $T=t_o$ with probability 1 where

$t_o = \{(c_{max}, x_{max}), \ldots, (c_{max}, x_{max})\}$. Therefore $f(T) = f(t_o)$, a constant so $f(t_o) = 0$. Observe that the number of elements of the set t_o is dependent on c_{max}, Q, and n_o. Next consider $\theta_i = ((c_i, x_i), (c_{max}, x_{max}), \ldots, (c_{max}, x_{max}))$ for each i, i ranging from 1 to n. θ_i introduces at most one additional point, $t_i = \{(c_i, x_i), (c_{max}, x_{max}), \ldots, (c_{max}, x_{max})\}$, to the sample space. By the assumption that $E[f(T)] = 0$, and since $f(t_o) = 0$ we have $Ef(T) = f(t_i)P(t_i) + f(t_o)P(t_o) = 0$ and $f(t_i) = 0$. We next consider $\theta_{i,j} = ((c_i, x_i), (c_j, x_j), (c_{max}, x_{max}), \ldots, (c_{max}, x_{max}))$ for each i,j pair where i and j range from 1 to n. Again we introduce only one new point to the sample space, call this $t_{i,j}$. By the assumption $E[f(T)] = 0$ we have $Ef(T) = f(t_{i,j})P(t_{i,j}) + OP(t_i) + OP(t_j) + OP(t_o) = 0$, implying $f(t_{i,j}) = 0$. Continuing in this way we expand our sample space one point at a time by expanding θ from $\theta_o = [\theta_o]$ to $\theta_t = V_i^N$, where $V_i = \{(c_1, x_1), \ldots, (c_n, x_n)\}$, and find that $f(t) = 0$. Since t was arbitrary we conclude that $f(T) = 0$ a.e., completing the proof.

Define $S = \{(c,x): (c,x) \in T \text{ and } \sum^{\nu} c_i - c < Q\}$ if $\nu > n_o$,

$S = T$ if $\nu = n_o$,

and again let #S be the cardinality of S. S is a random set and has the interpretation of being the set of all observations which could have been the terminal observation (C_ν, X_ν). μ, σ^2, and $\sigma^2_{\hat{\mu}}$ may all be estimated unbiasedly with uniform minimum variance by their usual fixed sample size estimates involving sample moments derived from T, and an adjustment for the sampling bias of the terminal observation involving sample moments derived from S and T. In particular we have the following theorem.

Theorem 2. For sum-quota(Q,n_o) sampling with $n_o >= 2$,

$$\hat{\mu} = \bar{X} - (\bar{X}_{[\nu]} - \bar{X})/(\nu-1)$$

is UMVUE for $\mu = {}^N\sum X_i/N$ where $\bar{X} = {}^\nu\sum X_i/\nu$ and $\bar{X}_{[\nu]} = \sum_S X_i/\#S$ where $\sum_S$ denotes sumation for all $(C_i, X_i) \in S$.

$$\hat{\sigma}^2 = S^2 - ((MS_{[\nu]} \cdot \nu/(\nu-1)) - S^2)/(\nu-2)$$

is UMVUE for $\sigma^2 = {}^N\sum(X_i-\mu)^2/(N-1)$ where $S^2 = {}^\nu\sum(X_i-\bar{X})^2/(\nu-1)$ and $MS_{[\nu]} = \sum_S(X_i-\bar{X})^2/\#S$

$$\hat{\sigma}^2_\mu = (1/(\nu-1) - 1/N)\hat{\sigma}^2 - S^2_{[\nu]}/(\nu-1)^2$$

is UMVUE for $\sigma^2_\mu = \text{var}(\hat{\mu})$ where $S^2_{[\nu]} = \sum_S(X_i - \bar{X}_\nu)^2/\#S$

Proof. X_1 is unbiased for μ. Taking the conditional expectation of X_1 with respect to the complete sufficient statistic T we have

$$E[X_1|T] = E[E[X_1|T_\nu]|T] = E[{}^{\nu-1}\sum X_i/(\nu-1)|T] = E[\bar{X}-(X_\nu-\bar{X})/(\nu-1)|T] = \hat{\mu}$$

The second equality is by the exchangeability of the preterminal observations conditional on T_ν, and the third by the algebraic equivalence of the terms within $E[\ |T]$. The forth equality is found by considering the distribution of X_ν conditional on T.

Similarly $(X_1-X_2)^2/2$ is unbiased for σ^2 and

$$E[(X_1-X_2)^2/2|T] = E[E[(X_1-X_2)^2/2|T_\nu|T]$$
$$= E[{}^{\nu-1}\sum(X_i - {}^{\nu-1}\sum X_i/(\nu-1))^2/(\nu-2)|T]$$
$$= E[S^2 - ((X_\nu-\bar{X})^2\nu/(\nu-1) - S^2)/(\nu-2)|T]$$
$$= \hat{\sigma}^2.$$

Since $\hat{\mu} = E[{}^{\nu-1}\sum X_i/(\nu-1)|T]$,

$$\text{var}(\hat{\mu}) = \text{var}(E[{}^{\nu-1}\sum X_i/(\nu-1)|T])$$
$$= \text{var}({}^{\nu-1}\sum X_i/(\nu-1)) - E[\text{var}({}^{\nu-1}\sum X_i/(\nu-1)|T)]$$

Now

$$\text{var}(^{\nu-1}\sum X_i/\nu-1)=E[(^{\nu-1}\sum X_i/(\nu-1))^2] - \mu^2 \quad \text{while}$$

$$E[X_1^2/N - X_1X_2/N(N-1)]= \mu^2 \quad \text{so that}$$

$$E[X_1^2/N - X_1X_2/N(N-1)|T_\nu]$$

$$=(^{\nu-1}\sum X_i/(n-1))^2 -(1/(\nu-1)-1/N)(^{\nu-1}\sum(X_i-\bar{X}_{\nu-1})^2 = \mu^2. \quad \text{Hence}$$

$$E[X_1^2/N - X_1X_2/N(N-1)|T_\nu]$$

$$=(^{\nu-1}\sum X_i/(n-1))^2 -(1/(\nu-1)-1/N)(s^2-((X_\nu-\bar{X})^2\nu/(\nu-1)-s^2)/(\nu-2))$$

is unbiased for μ^2 implying $(1/(\nu-1) - 1/N)\hat{\sigma}^2$ is unbiased for $\text{var}(^{\nu-1}\sum X_i/(\nu-1))$. Similarly $\text{var}(^{\nu-1}\sum X_i/(\nu-1)|T) =\text{var}(X_\nu/(\nu-1)|T)$ $=\sum_s(X_i-\bar{X}_\nu)^2/\#S(\nu-1)^2 =s_\nu^2/(\nu-1)^2$ is unbiased for $E[\text{var}(^{\nu-1}\sum X_i/(\nu-1)|T)]$. Therefore

$$\hat{\sigma}^2_{\hat{\mu}} = (1/(\nu-1) - 1/N)\hat{\sigma}^2 - s_\nu^2/(\nu-1)^2 \text{ is an unbiased estimator of}$$

$\sigma^2_{\hat{\mu}}$ based on the complete sufficient statistic and hence UMVUE.

3. DISCUSSION OF THE UNIVARIATE CASE

If we consider the estimates $\bar{X}_{\nu-1}=^{\nu-1}\sum X_i/(\nu-1)$ and $s^2_{\nu-1} = (1/(\nu-1)-1/N)^{\nu-1}\sum(X_i-\bar{X}_{\nu-1})^2/(\nu-2)$ we have the corresponding estimates based on T_ν, the preterminal observations , rather than T, the full sample. If ν is stochastically large these estimates may involve relatively little loss of information. However if ν is stochastically small the slightly more complicated UMVUE's may be markedly more efficient.

If the C_i and X_i are uncorrelated or independent we may expect the behavior of $\hat{\mu}$ and $\hat{\sigma}^2_{\hat{\mu}}$, conditional on ν, to be quite similar to their fixed sample size counterparts. If C_i and X_i are dependent we may expect a

stronger dependence between $\hat{\mu}$ and $\hat{\sigma}^2_{\hat{\mu}}$ than in the fixed sample size case. This should be considered if constructing approximate confidence intervals of the form $\hat{\mu} \overset{+}{-} z_\alpha \hat{\sigma}_{\hat{\mu}}$. An approach to lessen this dependence and improve coverage probabilities of approximate confidence intervals is to use an n_o larger than the nominal value of $n_o=2$ required. A similar modification is suggested when using a sequential procedure to estimate the mean of a normal with a preset standard error (Govindarajulu,1975).

We now wish to argue that for Q large $\hat{\mu}$ has indeed an approximate normal distribution. This is most easily considered if sampling is from an infinite population. Since sample size is random in sum-quota sampling it is not proper to consider asypmtotic relations as "$\nu \to \infty$". Instead asymptotic relations are considered as $Q \to \infty$. To denote the dependence of the random variable ν on Q, write $\nu(Q)$. From Billingsley (1968) (with a change in notation to be consistent with the present problem) is the following theorem concerning random selection of random sums. If

$$U_{[Q/\mu]} = {}^{[Q/\mu]}\sum(X_i - \mu)/\sigma_x[Q/\mu]^{1/2} \to W \text{ in distribution as } Q \to \infty$$

and if $\nu(Q)/[Q/\mu] \to k$, a constant in probability then,

$$V_{\nu(Q)} = {}^{\nu(Q)}\sum(X_i - \mu)/\sigma_x\nu(Q)^{1/2} \to W \text{ in distribution as } Q \to \infty.$$

($[Q/\mu]$ denotes the greatest integer less than or equal to Q/μ.) Also from Billingsley $(\nu(Q) - [Q/\mu_c])/(Q\sigma_c^2/\mu_c^3)^{1/2} \to N(0,1)$, as $Q \to \infty$. From these two results we may derive the following theorem.

Theorem 3. For sum-quota(Q,n_o) sampling $(n_o \geqslant 1)$ from an infinite population with $EX = \mu_x$, $var(x) = \sigma_x^2 < \infty$, $EC = \mu_c$, and $var(C) = \sigma_c^2 < \infty$,

$$(\nu(Q))^{1/2} (\hat{\mu} - \mu)/\sigma_x \to N(0,1) \text{ in distribution as } Q \to \infty.$$

This suggests or supports the intuition of setting confidence intervals using normal approximations.

4. BIVARIATE OR DOUBLE SAMPLING CASE

We now consider a multivariate sequential sum-quota design analogous to double sampling in the fixed sample size setting. We sample X_i's until $^{\nu-1}\sum C_i < Q_c <= {}^{\nu}\sum C_i$ and then stop sampling X_i's. We sample Y_i's until $^{\eta-1}\sum D_i < Q_d <= {}^{\eta}\sum D_i$ and then stop sampling Y_i's. If $^{\nu}\sum C_i < Q_c$ and $^{\eta}\sum D_i < Q_d$ we record the multivariate observation (C_i, X_i, D_i, Y_i). In sum-quota sampling we do not determine whether we will take a larger sample of the X_i's or the Y_i's. If we stop sampling the X_i's before the Y_i's the Y_i for $\nu < i < \eta$ may give us additional information about $\mu_x = {}^{N}\sum X_i / N$. An unbiased estimate of μ_x incorporating this additional information is found by adopting the Hartley-Ross estimate to sum-quota sampling and is described in Theorem 4. Define $\bar{X}_n = {}^{n}\sum X_i / n$, that is the mean of the first n observations. For $\eta < \mu$ let $HR = \bar{r}_{\nu-1}\bar{Y}_{\eta-1} + (\bar{X}_{\nu-1} - \bar{r}_{\nu-1}\bar{Y}_{\nu-1})(\nu-1)(\eta-2)/(\nu-2)(\eta-1)$ the HR estimate of $\bar{X}_{\eta-1}$ based on X_i, Y_i, and $r_i = X_i / Y_i$. Let $\hat{\sigma}^2_{HR}$ be an unbiased estimate of the variance of the HR estimate in the fixed sample size setting (Robson, 1957).

Theorem 4. Define $\hat{\mu}_x = HR$ if $\nu <= \eta$, $^{\nu-1}\sum X_i / (\nu-1)$ if $\nu > \eta$. $\hat{\mu}$ is unbiased for μ_x and $\hat{\sigma}^2 = (1/(n-1) - 1/N)\hat{\sigma}^2_x + \hat{\sigma}^2_{HR}$ is unbiased for $var(\hat{\mu}_x)$ where $\hat{\sigma}^2_x = {}^{\nu-1}\sum(X_i - \bar{X}_{\nu-1})^2/(\nu-2)$ and $n = max(\nu, \eta)$.

Proof. To show $\hat{\mu}_x$ is unbiased for μ_x we first consider $\tilde{\mu}_x = HR$ if $\nu <= \eta$, $^{\eta-1}\sum X_i / (\eta-1)$ if $\nu > \eta$. If $\nu <= \eta$ we may show that $E[HR] = {}^{\eta-1}\sum X_i / (\eta-1)$ using

conditional expectations and an argument similar to that in the fixed sample size case. We then have

$$E\tilde{\mu}_x = E[E[\tilde{\mu}_x|[Y_1,\ldots,Y_{\eta-1}],Y_\eta]]$$

$$= E[^{\eta-1}\sum X_i/(\eta-1)]$$

$$= \mu_x.$$

Considering the expectation of $\hat{\mu}_x$ we have

$$E\hat{\mu}_x = E[\hat{\mu}_x|\nu<=\eta]P[\nu<=\eta] + E[\hat{\mu}_x|\nu>\eta]P[\nu>\eta]$$

$$= E[\tilde{\mu}_x|\nu<=\eta]P[\nu<=\eta] + E[^{\nu-1}\sum X_i/(\nu-1)|\nu>\eta]P[\nu>\eta]$$

Thus it will follow that μ_x is unbiased once we show that

$$E[^{\nu-1}\sum X_i/(\nu-1)|\nu>\eta] = E[^{\eta-1}\sum X_i/(\eta-1)|\nu>\eta].$$

To this end consider the stopping rule which stops when $^n\sum C_i >= Q_c$ and $^n\sum D_i >= Q_d$. Sample both X_i and Y_i until one stops. A sufficient statistic is $T_m = ([(C_1,X_1,D_1,Y_1),\ldots,(C_m,X_m,D_m,Y_m)],[(C_{m+1},X_{m+1},D_{m+1},Y_{m+1}),\ldots,(C_n,X_n,D_n,Y_n)]$ for any $m<=n$. Since X_1 is unbiased we have

$$\mu_x = E[X_1|T_{n-1}] = E[^{\nu-1}\sum X_i/(\nu-1)|\nu>\eta]P[\nu>\eta] + E[^{\eta-1}\sum X_i/(\eta-1)|\nu<=\eta]P[\nu<=\eta]$$

$$\mu_x = E[X_1|T_{\eta-1}] = E[^{\eta-1}\sum X_i/(\eta-1)|\nu>\eta]P[\nu>\eta] + E[^{\eta-1}\sum X_i/(\eta-1)|\nu<=\eta]P[\nu<=\eta]$$

implying $E[^{\nu-1}\sum X_i|\nu>\eta] = E[^{\eta-1}\sum X_i/(\eta-1)|\nu>\eta]$.

To find an unbiased estimate of $\text{var}(\hat{\mu}_x)$ observe that

$$\text{var}(\hat{\mu}_x) = \text{var}E[\hat{\mu}_x|\nu,\eta,Y_1,\ldots,Y_{\eta-1}] + E\text{var}(\hat{\mu}_x|\nu,\eta,Y_1,\ldots,Y_{\eta-1})$$

$$= \text{var}(^{n-1}\sum X_i/(n-1)) + E\text{var}[\mu_x|\nu,\eta,Y_1,\ldots,Y_{\eta-1}].$$

That $E(1/(n-1) - 1/N)\hat{\sigma}_x^2 = \text{var}(^{n-1}\sum X_i/(n-1))$ is found in Theorem 2. Conditional on $Y_1,\ldots,Y_{\eta-1}$ the usual estimate of $\text{var}(HR)$ is unbiased, hence unconditionally unbiased for $E[\text{var}(\hat{\mu}_x|\nu,\eta,Y_1,\ldots,Y_{\eta-1})]$.

5. GENERAL DISCUSSION AND POSSIBLE EXTENSIONS

Sometimes, when sample size is random, estimates may be given assuming sample size is actually fixed by the sampler. However when conditioned on sample size in this manner, estimates no longer apply to parameters of interest but to new parameters involving the original parameters and the sample size. In sum-quota sampling, however, the estimates avoid this difficulty. When based on the preterminal observations the sum-qouta estimates are of the usual fixed sample size form and could be motivated assuming sample size to be fixed. However conditional on sample size the sum-quota estimates have a large bias while unconditionally they are found to possess certain optimal properties. It is this unconditional optimality which we desire as it considers a larger class of estimates containing the conditional estimates.

In finding estimators for μ, σ^2 and $\sigma_{\hat{\mu}}^2$ we have taken conditional expectations of unbiased estimates given T and T_ν. The basic property of sum-quota sampling we have drawn on in the formation of our estimates is the conditional exchangeability of the preterminal observations given T_ν. Other stopping rules which assure this exchangeability include multivariate quotas. For example if we observe costs C_i and D_i we may stop sampling when either $\sum^\nu C_i \geq Q_c$ or $\sum^\nu D_i \geq Q_d$ or we may stop sampling only when both $\sum^\nu C_i \geq Q_c$ and $\sum^\nu D_i \geq Q_d$. If the D_i's are degenerate at 1 we may employ a closed sequential procedure where we sample at most n_1 units. Thus, if we continue to sample the same variates throughout, then our set of observed vectors may be (depending on the stopping rule) complete sufficient and with slight modification the classical fixed sample size estimates are

UMVUE. Similarly in fixed-cost double sampling we may consider more general multivariate stopping rules which may arise in practice and still derive unbiased estimates of μ and $\sigma_{\hat{\mu}}^2$.

Stopping rules which assure the conditional exchangeability of the preterminal observations may be utilized to construct unbiased estimates of parameters other than μ and $\sigma_{\hat{\mu}}^2$. If g is any U-estimable function of the population and has kernel of degree n_o then any U-statistic of the preterminal observations is unbiased and may be improved upon by conditioning on T.

Directions for future work in sum-quota sampling include the testing problem for cross-classified data as well as estimation of variance components where unequal means imply unequal expected sample sizes.

REFERENCES

Billingsley, Patrick (1968).Convergence of Probability Measures. Wiley,
New York. 253 p.

Cox, D.R. (1969). Some Sampling Problems in Technology. In:Johnson,N.L.,
and Smith, H.Jr. (Eds.) New Developments in Survey Sampling, 506-527.
Wiley-Interscience, New York.

Govindarajulu, Z. (1975). Sequential Statistical Procedures. Academic
Press, New York. 565 p.

Pathak, P.K. (1976). Unbiased Estimation in Fixed Cost Sequential Sampling
Schemes. Ann. Statist., 4 1012-1017.

Robson, D.S. (1957). Application of Multivariate Polykays to the Theory of
Unbiased Ratio-type Estimation. J. Amer. Statist. Assoc., 52 511-522.