ASYMPTOTIC THEORY OF REJECTIVE SAMPLING WITH VARYING PROBABILITIES FROM A FINITE POPULATION[1]

By Jaroslav Hájek

Czechoslovak Academy of Science and University of California, Berkeley

0. Summary. In [3] the author established necessary and sufficient conditions for asymptotic normality of estimates based on simple random sampling without replacement from a finite population, and thus solved a comparatively old problem initiated by W. G. Madow [8]. The solution was obtained by approximating simple random sampling by so called Poisson sampling, which may be decomposed into independent subexperiments, each associated with a single unit in the population. In the present paper the same method is used for deriving asymptotic normality conditions for a special kind of sampling with varying probabilities called here rejective sampling. Rejective sampling may be realized by n independent draws of one unit with fixed probabilities, generally varying from unit to unit, given the condition that samples in which all units are not distinct are rejected. If the drawing probabilities are constant, rejective sampling coincides, with simple random sampling without replacement, and so the present paper is a generalization of [3].

Basic facts about rejective sampling are exposed in Section 2. To obtain more refined results, Poisson sampling is introduced and analyzed (Section 3) and then related to rejective sampling (Section 4). Next three sections deal with probabilities of inclusion, variance formulas and asymptotic normality of estimators for rejective sampling. In Section 8 asymptotic formulas are tested numerically and applications to sample surveys are indicated. The paper is concluded by short-cuts in practical performance of rejective sampling.

The readers interested in applications only may concentrate upon Sections 1, 8 and 9. Those interested in the theory of mean values and variances only, may omit Lemma 4.3 and Section 7.

1. Introduction. Consider a population U consisting of N identifiable units of arbitrary nature, so that they may be represented by integers $1, 2, \cdots, N$, $U = \{1, 2, \cdots, N\}$. There are many different formal definitions of a sample from the population U, no one satisfactory from all points of view. Here we shall adhere to the simplest possible one: *a sample is a subset of* U. Thus there are 2^N possible samples from U. A typical sample of this kind will be denoted by s, $s \subset U$.

The above definition does not reflect some features of common sampling designs, such as ordering of units (due to carrying out sampling by successive draws of one unit), repetition of units (due to replacement or interpenetration),

Received June 23, 1962; revised March 30, 1964.

[1] Prepared with the partial support of the National Science Foundation, Grant G-14648.

and hierarchy of units (due to stages). Consequently, within the theory of s-samples some common terms such as "sampling with or without replacement" or "two-stage sampling" lose their meaning. However, this is justified, at least in theoretical considerations, by a theorem proving that in cases where the sample is defined so as to distinguish ordering, repetitions, stages, random strata, etc., the sample s together with values ascertained thereon represent a sufficient statistic (see [2] or [11], for example). In multistage sampling, s should be understood as a subset of the population of ultimate units.

Some sampling procedures define directly a probability distribution on the space of all s-samples. This is true about systematic sampling, rejective sampling and Poisson sampling, for example. In other cases the outcome of the sampling experiment, say ω, is described in more detail, and the sample s is merely an abstract function of it, $s = s(\omega)$. For example, if ω denotes a sequence of n units from U with possible repetitions (there are N^n such sequences), then $s = s(\omega)$ will denote the subset of U consisting of all *distinct* units appearing at least once in ω. Outcomes of this kind are assumed in successive sampling and in sampling with replacement, for example. In all cases, however, we can project the probability distribution $P_1(\cdot)$ defined on the outcomes ω into the space of probability distributions defined on the samples s. In doing that we put

$$P(s) = P_1(\omega \,\varepsilon\, \Omega_s) \tag{1.1}$$

where Ω_s consists of all ω's which yield the given subset of distinct units s. $P(s)$, $s \,\varepsilon\, U$, given by (1.1) will be called *a projection* of $P_1(\omega)$.

Thus any sampling experiment may be described by a probability distribution $P(s)$ of the samples s, and by a set of conditional distributions $P_1(\omega \mid \Omega_s)$, $s \subset U$. The conditional distributions $P_1(\omega \mid \Omega_s)$, however, are irrelevant for estimation purposes, and may be neglected in theoretical considerations. For this reason, any probability distribution $P(s)$ defined on the space of the samples s will be called a *probability sampling*, and probability sampling so defined will be considered as a mathematical model of what is usually called "sampling plan", or "sampling design", or "sampling procedure", or "sampling experiment", or "sampling method".

The probability of selecting a sample s which contains the unit i, say π_i, equals

$$\pi_i = \sum_{s \ni i} P(s), \qquad (i \,\varepsilon\, U) \tag{1.2}$$

where $\sum_{s \ni i}$ denotes summation extended over all samples containing the unit i. Similarly,

$$\pi_{ij} = \sum_{s \ni i,j} P(s), \qquad (i, j \,\varepsilon\, U) \tag{1.3}$$

is the probability of including both units i and j in the sample. The probabilities π_i and π_{ij} will be called *probabilities of inclusion*.

For current needs of practice it is not necessary to know the probabilities $P(s)$ for every s. All we need is to know the probabilities of inclusion π_i and π_{ij}, at least approximately, and be certain that the estimates are approximately normal.

A particular kind of probability sampling is given by a formula for $P(s)$, $s \subset U$, or for $P_1(\omega)$ to be projected into $P(s)$, containing some free parameter, which may be controlled by the statistician. Most frequently these parameters are some non-negative numbers $\alpha_1, \cdots, \alpha_N$ related in some way to the "size" of the units, and we may assume that

$$\sum_{i=1}^{N} \alpha_i = 1. \tag{1.4}$$

For example, *rejective sampling* of size n corresponding to $(\alpha_1, \cdots, \alpha_N)$, say $R(s \mid n, \alpha_1, \cdots, \alpha_N)$, is defined by

$$\begin{aligned} R(s \mid n, \alpha_1, \cdots, \alpha_N) &= c(n, \alpha_1, \cdots, \alpha_N) \prod_{i \varepsilon s} \alpha_i, \qquad \text{if } s \text{ contains } n \text{ units} \\ &= 0, \qquad \text{otherwise,} \end{aligned} \tag{1.5}$$

where the constant $c(n, \alpha_1, \cdots, \alpha_N)$ is chosen so that $\sum R(s \mid n, \alpha_1, \cdots \alpha_N) = 1$, s running through all subsets of size n.

In accordance with (1.5) rejective sampling may be regarded as projected sampling with replacement of size n with probabilities α_i given the condition that the number of distinct units in the sample equals n.

Successive sampling is originally defined on ordered sequences $\omega = (r_1, \cdots, r_n)$ of distinct units, and depends on n and numbers $\alpha_1, \cdots, \alpha_N$ as follows.

$$\begin{aligned} P_1(\omega \mid n, \alpha_1, \cdots, \alpha_N) &= \frac{\alpha_{r_1} \cdots \alpha_{r_n}}{(1 - \alpha_{r_1}) \cdots (1 - \alpha_{r_1} - \cdots - \alpha_{r_{n-1}})}, \\ \omega &= (r_1, \cdots, r_N). \end{aligned} \tag{1.6}$$

Sampling with replacement is originally defined on ordered sequences $\omega' = (r_1, \cdots, r_n)$ of not necessarily distinct units by the formula

$$\begin{aligned} P_1'(\omega' \mid n, \alpha_1, \cdots, \alpha_N) &= \alpha_{r_1} \cdots \alpha_{r_n}, \qquad \omega' = (r_1, \cdots r_n), \\ & \qquad 1 \leqq r_1, \cdots, r_n \leqq N. \end{aligned} \tag{1.7}$$

For all three kinds of probability sampling just introduced, the correspondence between the parameters $(n, \alpha_1, \cdots, \alpha_N)$ and probabilities of inclusion π_i, π_{ij}, $1 \leqq i, j \leqq N$, is rather complicated, and even approximations are not too simple, unless n is of lower order than N.

Next two kinds of probability sampling are distinguished by a simple relation between the parameters $(n, \alpha_1, \cdots, \alpha_N)$ and the probabilities of inclusion $\pi_1, \cdots, \pi_N$. First, *Poisson sampling*, is defined by

$$P(s \mid n, \alpha_1, \cdots, \alpha_N) = \prod_{i \varepsilon s} (n\alpha_i) \prod_{j \varepsilon U - s} (1 - n\alpha_j), \qquad s \subset U, \tag{1.8}$$

(where $\alpha_i < n^{-1}$, $1 \leqq i \leqq N$). Second, *randomized systematic sampling*, is defined by the following experiment: permute randomly the units in the population and denote the result by $U^+ = \{R_1, \cdots, R_N\}$; then take an observation ξ from the uniform distribution over [0, 1] and include in s each unit i such that for some $k = 0, 1, \cdots, n - 1$ and $m = 1, \cdots, N$ we have $R_m = i$ and

$$(1.9) \qquad n(\alpha_{R_1} + \cdots + \alpha_{R_{m-1}}) \leqq \xi + k < n(\alpha_{R_1} + \cdots + \alpha_{R_m}).$$

For both Poisson and randomized systematic sampling

$$(1.10) \qquad \pi_i = n\alpha_i, \qquad i = 1, \cdots, N,$$

provided that the right side does not exceed 1 for any $i = 1, \cdots, N$. On the other hand, Poisson sampling is not very suitable in practice, because it makes the sample size a random variable, and even an empty sample may occur. Also randomized systematic sampling has some disadvantages: First, the procedure of randomly permuting the population may be very tedious if N is very large. Second, the dependence of the π_{ij}'s on the π_i's is very complicated, and the recent asymptotic results by Hartley and Rao [5] are applicable only if n is of smaller order than N.

For illustration, let us consider a population consisting of 4 units, $U = \{1, 2, 3, 4\}$, and compute the probabilities $P(s)$ for all five kinds of sampling with $n = 2$ and $\alpha_1 = .1$, $\alpha_2 = .2$, $\alpha_3 = .3$ and $\alpha_4 = .4$. The results are given in Table 1. The successive sampling, randomized systematic sampling and sampling without replacement have been projected according to (1.1).

TABLE 1

The Sample s	The Kind of Sampling				
	Rejective	Poisson	Projected successive	Projected with replacement	Projected randomized systematic
0		.038			
1		.010		.010	
2		.026		.040	
3		.058		.090	
4		.154		.160	
12	.057	.006	.047	.040	.067
13	.086	.014	.076	.060	.067
14	.114	.038	.111	.080	.067
23	.171	.038	.161	.120	.067
24	.229	.102	.233	.160	.266
34	.343	.230	.371	.240	.466
123		.010			
124		.026			
134		.058			
234		.154			
1234		.038			
	1.000	1.000	.999	1.000	1.000

$n = 2, \quad \alpha_1 = .1, \quad \alpha_2 = .2, \quad \alpha_3 = .3, \quad \alpha_4 = .4$

The rest of this section will be devoted to the question of motivation and integration with the previous literature.

Sampling with unequal probabilities presents no especial difficulties, if it amounts, within individual strata, either to sampling with equal probabilities, as in Neyman [9], or to sampling just one unit, as in Hansen and Hurwitz [4]. The general case, however, raises very difficult mathematical problems, which has stimulated attempts to establish a unified and modernized theory of probability sampling (another name for sampling from finite populations). One aspect of this theory should be an appropriate definition of a sample, and of the class of admissible probability samplings and estimators of, say, the population total

$$Y = \sum_{i=1}^{N} y_i . \tag{1.11}$$

Horwitz and Thompson [6] showed that

$$\hat{Y} = \sum_{i \varepsilon s} (y_i/\pi_i) \tag{1.12}$$

is an unbiased estimator of Y, if $\pi_i > 0$, $i = 1, \cdots, N$, for any sampling design, and established the variance of $\hat{Y}$ in terms of the probabilities π_{ij}, $1 \leqq i, j \leqq N$. Yates and Grundy [12] noted that for sampling design where the sample is of fixed size, the variance of Y may be rewritten in following form:

$$\text{var}\ (\hat{Y}) = \sum_{i<j}\sum [(y_i/\pi_i) - (y_j/\pi_j)]^2(\pi_i\pi_j - \pi_{ij}). \tag{1.13}$$

However, in order to be able to utilize the general formulas, we have to study in more detail some particular kinds of probability sampling, and solve for them, at least approximately, the following three problems:

P 1. To establish relation between controlled parameters $(n, \alpha_1, \cdots, \alpha_N)$ and the probabilities of inclusion $\pi_1, \cdots, \pi_N$.

P 2. To establish relation between the probabilities $\pi_1, \cdots, \pi_N$ and the probabilities π_{ij}, $1 \leqq i < j \leqq N$.

P 3. To find conditions for asymptotic normality of estimators such as $\hat{Y}$.

Without solving these problems, we cannot control the π_i's in (1.12), simplify and estimate the variance (1.13), and justify the confidence intervals based on the assumption of normality.

All three above problems are solved below for rejective sampling, and the author hopes to do the same for successive sampling in some subsequent paper. As for the previous literature, there is just one paper in this line, namely by Hartley and Rao [5]. They presented a solution of the Problem P 2 for randomized systematic sampling by a quite different method than one used in the present paper (recall that P 1 is trivial for this kind of sampling in view of (1.10)). Moreover, their approach is asymptotic in a quite different sense, because, roughly speaking, they are interested in the case when n is fixed and $N \to \infty$, while the present paper assumes that $n \to \infty$ and $(N - n) \to \infty$. As a matter of fact under the conditions assumed by Hartley and Rao,

$$n(\pi_i\pi_j - \pi_{ij})/(\pi_i\pi_j) \to 1 \tag{1.14}$$

for randomized systematic sampling as well as for rejective and successive sampling. Formula (1.14), however, is false for all three kinds of sampling, if $n \rightarrow \infty$ and $(N - n) \rightarrow \infty$, or more precisely, if

$$\sum_{i=1}^{N} \pi_i(1 - \pi_i) \rightarrow \infty. \tag{1.15}$$

For rejective sampling under condition (1.15),

$$\left[\sum_{i=1}^{N} \pi_i(1 - \pi_i)\right] (\pi_i\pi_j - \pi_{ij})/[\pi_i(1 - \pi_i)\pi_j(1 - \pi_j)] \rightarrow 1 \tag{1.16}$$

uniformly in i, j, as is shown in Section 5. Hartley and Rao leave open the problem what happens under condition (1.15), and center their attention upon deriving expansions for π_{ij} in terms of π_i accurate to higher orders in N than that given by (1.14). Consequently, their formulas are applicable, if N is much larger than n only, and it is not reasonable to compare them with those obtained here. To make a fruitful comparison possible, one should solve the problem P 2 for rejective sampling under the Hartley-Rao conditions, and for randomized systematic sampling under condition (1.15).

In conclusion, let us mention the paper [10] by Rao, Hartley and Cochran, where they present a clever attempt to circumvent the above problems by introducing random strata and selecting just one unit from each stratum with unequal probabilities. However the simplicity of their approach is somewhat invalidated by the fact that the estimator they propose is not admissible, i.e. there exists another estimator with uniformly smaller variance. Actually, the outcome of the sampling experiment suggested in [10] may be described as $\omega = (S_1, \cdots, S_L, s)$ where S_h are random strata and s the selected subset of distinct units. Now s, together with the values y_i, $i \varepsilon s$, represents sufficient statistics, and the conditional mean values of the Rao-Hartley-Cochran estimator with respect to this sufficient statistics has the above-mentioned superior property. Unfortunately this conditional mean value is very difficult to compute and thus only indicates a possibility of improvement. If the Rao-Hartley-Cochran design were used with the estimator (1.12), one should first solve the problems P 1 and P 2. Thus also in this line further development is necessary.

Finally, as noted by the referee, the sampling referred to here as *Poisson sampling* is but a generalization of *binomial sampling*, introduced by L. A. Goodman [1], for the case of unequal probabilities.

2. Rejective sampling. We shall begin with a definition of rejective sampling equivalent to (1.4):

DEFINITION 2.1. Let $p_1, \cdots, p_N$ be some fixed numbers such that $0 < p_i < 1$, $i = 1, \cdots, N$. A probability sampling $R(s)$ will be called *rejective sampling of size n with probabilities* $p_1, \cdots, p_N$ if

$$\begin{aligned} R(s) &= C\prod_{i\varepsilon s} p_i \prod_{i\varepsilon U-s} (1 - p_i), \qquad \text{if } s \text{ contains } n \text{ units}, \\ &= 0, \qquad \text{otherwise}. \end{aligned} \tag{2.1}$$

The representation (2.1) will be called *canonical*, if

$$\sum_{i=1}^{N} p_i = n. \tag{2.2}$$

Of course, (2.1) is equivalent to (1.5), if

$$\alpha_i = p_i/(1 - p_i)\left[\sum_{j=1}^{N} p_j/(1 - p_j)\right]^{-1}, \qquad i = 1, \cdots, N. \tag{2.3}$$

There exist infinitely many other probabilities $p_1^*, \cdots, p_N^*$ and constants C^* such that (2.1) may be rewritten as

$$\begin{aligned} R^*(s) &= C^* \prod_{i\varepsilon s} p_i^* \prod_{i\varepsilon U-s} (1 - p_i^*), \qquad \text{if } s \text{ contains } n \text{ units,} \\ &= 0, \qquad \text{otherwise.} \end{aligned} \tag{2.4}$$

Obviously $R^*(s) = R(s)$ for all s, if, and only if,

$$p_i^*/(1 - p_i^*) = c[p_i/(1 - p_i)], \qquad 1 \leqq i \leqq N, \tag{2.5}$$

for some constant c.

LEMMA 2.1. *Let the numbers* $p_1, \cdots, p_N$ *and* $p_1^*, \cdots, p_N^*$ *satisfy* (2.5). *Put*

$$h = \sum_{i=1}^{N} p_i, \tag{2.6}$$

$$h^* = \sum_{i=1}^{N} p_i^*, \tag{2.7}$$

and

$$d = \sum_{i=1}^{N} p_i(1 - p_i). \tag{2.8}$$

Then

$$p_i^*(1 - p_i) = p_i(1 - p_i^*)\{1 + (h^* - h)/d + o[(h^* - h)/d]\} \tag{2.9}$$

where $o(x)$ *means that* $o(x)x^{-1} \to 0$ *for* $x \to 0$.

PROOF. From (2.5) it easily follows that

$$p_i - p_i^* = (1 - c)p_i(1 - p_i^*), \tag{2.10}$$

and

$$p_i - p_i^* = (c^{-1} - 1)p_i^*(1 - p_i). \tag{2.11}$$

Assume that $h^* > h$, so that $p_i^* > p_i$ and $1 - p_i^* < 1 - p_i$. Carrying out a summation in (2.10) and (2.11) with respect to i we get

$$h - h^* = (1 - c)\sum_{i=1}^{N} p_i(1 - p_i^*) < (1 - c)\, d, \tag{2.12}$$

$$h - h^* = (c^{-1} - 1)\sum_{i=1}^{N} p_i^*(1 - p_i) > (c^{-1} - 1)\, d, \tag{2.13}$$

from which we immediately obtain

$$\{1 - ([h^* - h)/d]\}^{-1} < c < 1 + [(h^* - h)/d]. \tag{2.14}$$

Now, (2.9) is an easy consequence of (2.14) and (2.5). The case when $h^* < h$ may be treated quite similarly. The proof is finished.

Now, let $p_1, \cdots, p_N$ be any numbers such that $0 < p_i < 1$, $1 \leqq i \leqq N$.

LEMMA 2.2. *If* $\sum_{i=1}^N p_i = n$ *and* s_n *is a sample of size* n, *then*

$$\sum_{i \varepsilon U - s_n} p_i = \sum_{i \varepsilon s_n} (1 - p_i) > \tfrac{1}{2}d. \tag{2.15}$$

If $\sum_{i=1}^N p_i = h$ *and* s_k *is a sample of size* k, *then*

$$h - k + \sum_{i \varepsilon s_k} (1 - p_i) = \sum_{i \varepsilon U - s_k} p_i. \tag{2.16}$$

PROOF. The proof is immediate. A property of rejective sampling is that conditional probabilities $P(s \mid s \ni i)$, $P(s \mid s \ni i, j)$ etc. and "complementary" probabilities $P^*(s) = P(U - s)$ again represent rejective sampling with properly modified parameters.

3. Poisson sampling. Also here we need to rewrite definition (1.8):

DEFINITION 3.1. A probability sampling will be called *Poisson sampling with probabilities* $p_1, \cdots, p_N$, if

$$P(s) = \prod_{i \varepsilon s} p_i \prod_{i \varepsilon U - s} (1 - p_i), \qquad (s \subset U). \tag{3.1}$$

Let $K(s)$ denote the number of units contained in s.

From formula (3.1) it follows that Poisson sampling may be carried out by N independent subexperiments, the ith of them deciding with probabilities p_i or $1 - p_i$ whether the ith unit will be included in s or not, respectively.

Obviously $K = K(s)$ is a random variable in Poisson sampling, and it ranges from 0 to N. Now, introduce zero-one random variables $\xi_i = \xi_i(s)$ characterizing the presence or the absence of the ith unit in s:

$$\begin{aligned} \xi_i &= 1, \quad \text{if } s \ni i, \\ &= 0, \quad \text{otherwise.} \end{aligned} \qquad (1 \leqq i \leqq N) \tag{3.2}$$

Then

$$K = \sum_{i=1}^N \xi_i, \tag{3.3}$$

and the distribution of K is one considered in Poisson's generalization of Bernoulli's scheme, since the random variables ξ_i are independent, as many be easily seen from (3.1). Hence our label "Poisson sampling".

From (3.3) it follows that

$$EK = h = \sum_{i=1}^N p_i, \tag{3.4}$$

and

$$\text{var } K = d = \sum_{i=1}^{N} p_i(1 - p_i). \tag{3.5}$$

The following Lemma follows from asymptotic normality of K for $d \to \infty$ (Cramér, pp. 216–217), and the unimodality of the distribution of K which can be demonstrated.

LEMMA 3.1. *For any constant C it holds that*

$$\max_{h - Cd^{\frac{1}{2}} < k < h + Cd^{\frac{1}{2}}} \left| \frac{(2\pi\, d)^{\frac{1}{2}} P(K = k)}{\exp\,[-\frac{1}{2}(k - h)^2 d^{-1}]} - 1 \right| \to 0 \tag{3.10}$$

if $d = \sum_{i=1}^{N} p_i(1 - p_i) \to \infty$.

Now, let $y_1, \cdots, y_N$ be some real values associated with the units $1, \cdots, N$. In Poisson sampling with probabilities $p_1, \cdots, p_N$ the variance of the estimator

$$T' = \sum_{i \varepsilon s} y_i/p_i \tag{3.11}$$

will be due to (a) the variability of the ratios y_i/p_i and (b) the variability of the sample size $K = K(s)$. The latter source of variability is not present in rejective sampling and, hence, undesirable in our consideration. To eliminate it, let us shift the values y_i/p_i by a constant c chosen so that we may obtain minimum variance. Using the random variables ξ_i defined by (3.3), we easily see that

$$\begin{aligned} \text{var } \Big\{\sum_{i \varepsilon s} [(y_i/p_i) - c]\Big\} &= \text{var}\left\{\sum_{i=1}^{N} [(y_i/p_i) - c]\xi_i\right\} \\ &= \sum_{i=1}^{N} [(y_i/p_i) - c]^2 \text{ var } \xi_i = \sum_{i=1}^{N} [(y_i/p_i) - c]^2 p_i(1 - p_i) \\ &\geqq \sum_{i=1}^{N} [(y_i/p_i) - R]^2 p_i(1 - p_i) = \sum_{i=1}^{N} (y_i - Rp_i)^2(p_i^{-1} - 1) \end{aligned}$$

where

$$R = \sum_{i=1}^{N} y_i(1 - p_i) \Big/ \sum_{i=1}^{N} p_i(1 - p_i) \tag{3.12}$$

and the minimum is obtained for $c = R$. This invites us to replace (3.11) by

$$T = \sum_{i \varepsilon s} [(y_i/p_i) - R] + R \sum_{i=1}^{N} p_i\,. \tag{3.13}$$

The quantity T will have an auxiliary but important role in our considerations. Note that T is not an estimator, since it depends on y_i, $i \varepsilon U - s$, unless the size of s is $\sum_{i=1}^{N} p_i$. However, for samples of size $\sum_{i=1}^{N} p_i$ it coincides with T'.

Throughout the present paper we formulate the asymptotic relations for the whole class of Poisson or rejective samplings corresponding to all possible $N = 1, 2, \cdots$ and $p_1, \cdots, p_N$, $0 < p_i < 1$, $i = 1, \cdots, N$. This approach has been applied in Lemmas 2.1 and 3.2. To be able to state the condition of asymptotic normality for T in the same way, we need a modification of the Lindeberg condition implicit in the following

LEMMA 3.2. *With the above notation, we have*

$$ET = \sum_{i=1}^{N} y_i \tag{3.14}$$

and

$$\operatorname{var} T = \sum_{i=1}^{N} (y_i - Rp_i)^2(p_i^{-1} - 1). \tag{3.15}$$

Moreover,

$$\sup_x \left| P(T < x) - \Phi\left(\frac{x - ET}{(\operatorname{var} T)^{\frac{1}{2}}}\right) \right| \to 0 \tag{3.16}$$

where Φ denotes the normed normal distribution function, if

$$e = e(y_i, \cdots, y_N, p_1, \cdots, p_N) \to 0, \tag{3.17}$$

where

$$e = \inf \{\epsilon: L_\epsilon^* \leqq \epsilon\}, \tag{3.18}$$

and

$$L_\epsilon^* = (\operatorname{var} T)^{-1} \sum_{i \varepsilon A_\epsilon} (y_i - Rp_i)^2(p_i^{-1} - 1), \tag{3.19}$$

where

$$A_\epsilon = \{i: (y_i - Rp_i)^2 > \epsilon^2 p_i^2 \operatorname{var} T\}. \tag{3.20}$$

PROOF. Putting $T = \sum_{i=1}^N [(y_i/p_i) - R]\xi_i + R\sum_{i=1}^N p_i$ we readily see that the Lindeberg condition for T amounts to $L_\epsilon \to 0$ for every $\epsilon > 0$, where

$$\begin{aligned} L_\epsilon = (\operatorname{var} T)^{-1} \sum_{i \varepsilon C_\epsilon} [(y_i/p_i) - R]^2(1 - p_i)^2 p_i \\ + \sum_{i \varepsilon B_\epsilon} [(y_i/p_i) - R]^2(1 - p_i)p_i^2, \end{aligned} \tag{3.21}$$

where

$$C_\epsilon = \{i: |(y_i/p_i) - R|(1 - p_i) > \epsilon(\operatorname{var} T)^{\frac{1}{2}}\}$$

and

$$B_\epsilon = \{i: |(y_i/p_i) - R|p_i > \epsilon(\operatorname{var} T)^{\frac{1}{2}}\}.$$

Now, it may be easily verified that $L_\epsilon \leqq L_\epsilon^* \leqq 2L_{\frac{1}{2}\epsilon}$. Consequently $L_\epsilon \to 0$ for every $\epsilon > 0$ is equivalent to $L_\epsilon^* \to 0$ for every $\epsilon > 0$. Finally, as $[e < \frac{1}{2}\epsilon] \Rightarrow [L_\epsilon^* \leqq L_{2e} \leqq 2e]$, condition (3.17) entails $L_\epsilon^* \to 0$ for every $\epsilon \to 0$ for any sequence of Poisson samplings. The proof is finished.

REMARK. Following the pattern of the proof of Theorem 3.1 in [3], we could show that (3.17) is also necessary for (3.16).

4. Lemmas connecting rejective sampling and Poisson sampling. A comparison of (2.1) and (3.1) shows that rejective sampling of size n may be regarded as Poisson sampling given the condition that the sample size equals n. If the probabilities $p_1, \cdots, p_N$ are chosen so that $n = \sum_{i=1}^{N} p_i$ the sample size of the rejective sampling will equal the average sample size of the Poisson sampling.

We have also seen that the rejective sampling may be regarded as projected sampling with replacement of size n given the condition that the number of distinct units equals n.

Poisson sampling and sampling with replacement may both be decomposed into independent subexperiments, and, for this reason, they both possess simple variance formulas and simple conditions for asymptotic normality. However, whereas all the relevant properties of Poisson sampling may be carried over to rejective sampling, at least asymptotically, sampling with replacement gives formulas diverging from those for rejective sampling even in the limit, unless the probability of rejection tends to zero. This somewhat surprising result is probably due to the fact that rejective sampling may represent an "average result" of Poisson sampling, whereas in sampling with replacement rejective sampling always represents an extreme case (all units must be distinct). Thus the parallelism of rejective sampling and sampling with replacement, though at first sight very inviting, is misleading.

First we show that convergence in probability in Poisson sampling implies the same in rejective sampling for an important class of "hereditary" events.

LEMMA 4.1. *Consider Poisson sampling $P(\cdot)$ and rejective sampling $R(\cdot)$ of size n, both with probabilities, $p_1, \cdots, p_N$ such that $\sum_{i=1}^{N} p_i = n$. Let A be an event such that either*

$$[s \in A,\ s^* \subset s] \Rightarrow [s^* \in A] \qquad \text{for all } s,\ s^* \in U, \tag{4.1}$$

or

$$[s \in A,\ s^* \supset s] \Rightarrow [s^* \in A] \qquad \text{for all } s,\ s^* \in U. \tag{4.2}$$

Then there is a d_0 such that for $d > d_0$

$$P(A) \geqq \tfrac{1}{20} R(A). \tag{4.3}$$

PROOF. Assume that (4.1) holds, for example. Then, in view of (2.16),

$$P(A, K = k) = \sum_{s_k \in A} P(s_k) \geqq \sum_{s_{k+1} \in A} P(s_{k+1}) \sum_{j \in s_{k+1}} \frac{1 - p_j}{n - k - 1 + p_j + \sum_{i \in s_{k+1}} (1 - p_i)}.$$

Furthermore, from (2.15) it follows that

$$\sum_{j \in s_{k+1}} \frac{1 - p_j}{n - k - 1 + p_j + \sum_{i \in s_{k+1}} (1 - p_i)} \geqq 1 - 4(n - k)/d, \quad (k < n). \tag{4.4}$$

On combining (4.3) and (4.4), we get

$$
\begin{aligned}
P(A, K = k) &\geqq (1 - 4(k-n)/d)P(A, K = k+1) \\
&\geqq (1 - 4(k-n)^2/d)P(A, K = n), \quad (k < n).
\end{aligned} \tag{4.5}
$$

After easy computations, (4.5) yields

$$
\begin{aligned}
P(A) &= \sum_{k=0}^{N} P(A, K = k) \geqq \sum_{n - \frac{1}{3}d^{\frac{1}{2}} < k < n} P(A, K = k) \\
&\geqq \tfrac{5}{9}(\tfrac{1}{3}d^{\frac{1}{2}} - 1)P(A, K = n).
\end{aligned} \tag{4.6}
$$

Now, noting that

$$
R(A) = P(A \mid K = n) = [P(A, K = n)]/[P(K = n)], \tag{4.7}
$$

and recalling (3.10), we see that (4.6) implies (4.3) immediately.

The case, where (4.2) holds, may be carried over to the previous one by considering the complementary rejective sampling $P^*(s) = P(U - s)$ of size $N - n$ with probabilities $1 - p_1, \cdots, 1 - p_N$. The proof is finished.

From Lemma 4.1 we obtain almost immediately

LEMMA 4.2. *Consider rejective sampling $R(\cdot)$ of size n with probabilities $p_1, \cdots, p_N$, $\sum_{i=1}^{N} p_i = n$, and some numbers $b_1, \cdots, b_N$ such that*

$$
0 \leqq b_i \leqq 1, \qquad (1 \leqq i \leqq N). \tag{4.8}
$$

Denote $d = \sum_{i=1}^{N} p_i(1 - p_i)$.

Then

$$
\left(d^{-1} \sum_{i \varepsilon s} b_i - d^{-1} \sum_{i=1}^{N} b_i p_i\right) \to 0 \tag{4.9}
$$

in R-probability if $d \to \infty$, uniformly in $b_1, \cdots, b_N$.

PROOF. The convergence in probability follows for the corresponding Poisson sampling $P(\cdot)$ by Tchebyshev inequality, since

$$
E^P(d^{-1} \sum_{i \varepsilon s} b_i) = d^{-1} \sum_{i=1}^{N} b_i p_i
$$

and, for $b_i^2 \leqq 1$,

$$
\operatorname{var}^P (d^{-1} \sum_{i \varepsilon s} b_i) = d^{-2} \sum_{i=1}^{N} b_i^2 p_i(1 - p_i) \leqq d^{-1}
$$

where E^P and var^P refer to $P(\cdot)$. Now the events

$$
d^{-1} \sum_{i \varepsilon s} b_i > d^{-1} \sum_{i=1}^{N} b_i p_i + \epsilon \tag{4.10}
$$

and

$$
d^{-1} \sum_{i \varepsilon s} b_i < d^{-1} \sum_{i=1}^{N} b_i p_i - \epsilon \tag{4.11}
$$

are of the hereditary type considered in Lemma 4.1, for $b_i \geqq 0$. Consequently the probability of (4.10) and (4.11) converges to 0 also for rejective sampling if $d \to \infty$, in view of (4.3).

EXAMPLES. Lemma 4.2 entails, for example, that

$$d^{-1} \sum_{i \varepsilon s} (1 - p_i) \to 1$$

and

$$\left[d^{-1} \sum_{i \varepsilon s} p_i(1 - p_i) - d^{-1} \sum_{i=1}^{N} p_i^2(1 - p_i) \right] \to 0$$

in probability for rejective sampling in canonical form, if $d \to \infty$.

Now, we shall establish an important result, which may be regarded as the corner-stone of the present paper. Our aim is to construct an experiment yielding jointly a rejective sample s_n and Poisson sample s_K in such a way that

$$\text{either } s_K \subset s_n \quad \text{or} \quad s_K \supset s_n \tag{4.12}$$

with probability 1. The experiment will consist of two successive steps:

EXPERIMENT. *First, we select a rejective sample s_n in accordance with its distribution, and, independently, ascertain the size $K = k$ of the Poisson sample. Second, if $k = n$, we put $s_K = s_n$; if $k > n$, we select a rejective sample s_{k-n} from $U - s_n$, considered as a population, with probabilities $(k - n)D_n^{-1}p_i$, $i \varepsilon U - s_n$, where*

$$D_n = D(s_n) = \sum_{i \varepsilon s_n} (1 - p_i) = \sum_{i \varepsilon U - s_n} p_i, \tag{4.13}$$

and put $s_K = s_n \cup s_{k-n}$; if $k < n$, we select a rejective sample s_{n-k} from s_n, considered as a population, with probabilities $(n - k)D_n^{-1}(1 - p_i)$, and put $s_K = s_n - s_{n-k}$.

In the above experiment a pair (s_k, s_n) such that either $s_k \subset s_n$ or $s_n \supset s_k$ will have probability

$$\begin{aligned} Q(s_k, s_n) &= R(s_n)P(K = k)R^*(s_k - s_n \mid U - s_n, k - n), && \text{if } k > n, \\ &= R(s_n)P(K = n), && \text{if } k = n, \\ &= R(s_n)P(K = k)R^+(s_n - s_k \mid s_n, n - k), && \text{if } k < n, \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} R^*(s_k - s_n \mid U - s_n, k - n) = C^*(U - s_n, k - n) \prod_{i \varepsilon s_k - s_n} (k - n) \\ \cdot D_n^{-1}p_i \prod_{i \varepsilon U - s_k} [1 - (k - n)D_n^{-1}p_i], \end{aligned} \tag{4.15}$$

and

$$\begin{aligned} R^+(s_n - s_k \mid s_n, n - k) = C^+(s_n, n - k) \prod_{i \varepsilon s_n - s_k} (n - k) \\ \cdot D_n^{-1}(1 - p_i) \prod_{i \varepsilon s_k} [1 - (n - k)D_n^{-1}(1 - p_i)]. \end{aligned} \tag{4.16}$$

Denoting the Poisson samplings corresponding to R^* and R^+ by P^* and P^+, respectively, and by K^* and K^+ the sizes of respective Poisson samples, we can write

$$C^*(U - s_n, k - n) = [P^*(K^* = k - n)]^{-1} \tag{4.17}$$

and

$$C^+(s_n, n - k) = [P^+(K^+ = n - k)]^{-1}.$$

Furthermore, $EK^* = \sum_{i\varepsilon U - s_n} (k - n)D_n^{-1}p_i = k - n$

$$\operatorname{var} K^* = \sum_{i\varepsilon U - s_n}{}' (k - n)D_n^{-1}p_i[1 - (n - k)D_n^{-1}p_i]$$

so that, in view of (2.15),

$$(k - n)[1 - 2(k - n)/d] < \operatorname{var} K^* < (k - n). \tag{4.18}$$

Combining (4.17) and (4.18) and applying Lemma 3.1, we get

$$[d \to \infty, |k - n|/d \to 0] \Rightarrow [C^*(2\pi(k - n))^{-\frac{1}{2}} \to 1]. \tag{4.19}$$

The same result holds for C^+.

Now, return to the joint distribution (4.14) and ask whether the marginal distribution of s_k, say $P_0(s_k)$, corresponds to Poisson sampling $P(\cdot)$ with probabilities $p_1, \cdots, p_N$, as we wished. Unfortunately, this is not so, unless $p_1 = p_2 = \cdots = p_N$, as in [3]. On the other hand, we shall be able to prove that the difference $P_0 - P$ tends to zero in variation as $d \to \infty$. Consequently any random variable T has the same limiting distribution, if any, for P_0 as for P. We shall denote the variation of $P_0 - P$ by $\|P_0 - P\|$ and define it as follows:

$$\|P_0 - P\| = 2 \sum_{P_0(s) < P(s)} [1 - (P_0(s)/P(s))]P(s) \tag{4.20}$$

where the sum extends over all samples s satisfying the condition under the sign $\sum$.

LEMMA 4.3. *Let $P(\cdot)$ denote Poisson sampling with probabilities $p_1, \cdots, p_N$, $\sum p_i = n$, and define $P_0(\cdot)$ by*

$$\begin{aligned} P_0(s_k) &= \sum_{s_n \subset s_k} Q(s_k, s_n) && \textit{if } k > n \\ &= P(s_k) && \textit{if } k = n \\ &= \sum_{s_n \supset s_k} Q(s_k, s_n) && \textit{if } k < n \end{aligned} \tag{4.21}$$

where Q is given by (4.14) and the sums extend over all s_n satisfying the indicated conditions.

Then $d = \sum_{i=1}^N p_i(1 - p_i) \to \infty$, implies

$$\|P_0 - P\| \to 0. \tag{4.22}$$

PROOF. Let $a \approx b$ and $a \sim b$ denote that $a/b \to 1$ and $(a - b) \to 0$ in P-prob-

ability, respectively, if $d \to \infty$. Furthermore, let $a \gtrapprox b$ denote that for every $\epsilon > 0$, $a > (1 - \epsilon) b$ holds with P-probability tending to 1 if $d \to \infty$.
If a and b depend also on s_n, the convergences are assumed uniform with respect to s_n. For, example, Lemma 3.1 implies that

$$[P(K = k)/P(K = n)] \approx \exp [-\tfrac{1}{2}d^{-1}(k - n)^2] \tag{4.23}$$

and

$$d^{-2}(k - n)^3 \sim 0. \tag{4.24}$$

In both above relations the left side depends on s_k only, namely on the size k of s_k. However, (4.24) and (2.15) imply that also

$$D_n^{-2}(k - n)^3 \sim 0, \tag{4.25}$$

uniformly in s_n (recall (4.13)).

In view of (4.20), our task is to prove that

$$P_0(s_k)/P(s_k) \gtrapprox 1. \tag{4.26}$$

From now on we shall assume that $k > n$, so that probability and the signs $\approx$, $\sim$, $\gtrapprox$ should be understood as conditional given $k > n$. Note that

$$\begin{aligned} R(s_n) &= P(s_n)[P(K = n)]^{-1} \\ &= P(s_k)[P(K = n)]^{-1} \prod_{i \varepsilon s_k - s_n} (1 - p_i)/p_i \end{aligned} \tag{4.27}$$

and put Q in the following form:

$$\begin{aligned} Q(s_k, s_n) = P(s_k)&[P(K = k)/P(K = n)]M(s_k, s_n) \\ &\cdot C^*(U - s_n, k - n) \prod_{i \varepsilon s_k - s_n} (k - n)D_\kappa^{-1}(1 - p_i) \\ &\cdot \prod_{i \varepsilon s_n} [1 - (k - n)D_k^{-1}(1 - p_i)], \end{aligned} \tag{4.28}$$

where

$$D_k = \sum_{i \varepsilon s_k} (1 - p_i), \tag{4.29}$$

and

$$\begin{aligned} M(s_k, s_n) = (D_k/D_n)^{k-n} &\prod_{i \varepsilon U - s_k} [1 - (k - n)D_n^{-1}p_i] \\ &\cdot \prod_{i \varepsilon s_n} [1 - (k - n)D_k^{-1}(1 - p_i)]^{-1}. \end{aligned} \tag{4.30}$$

Now using the usual development $1 - x = \exp [-x - \tfrac{1}{2}x^2 + O(x^3)]$, we can write

$$\begin{aligned} M(s_k, s_n) \approx \exp \big[&(k - n)[(D_k - D_n)/D_n] \\ &- (k - n)D_n^{-1} \sum_{i \varepsilon U - s_k} p_i + (k - n)D_k^{-1} \sum_{i \varepsilon s_n} (1 - p_i) \\ &- \tfrac{1}{2}(k - n)^2 D_n^{-2} \sum_{i \varepsilon U - s_k} p_i^2 + \tfrac{1}{2}(k - n)^2 D_k^{-2} \sum_{i \varepsilon s_n} (1 - p_i)^2\big] \end{aligned} \tag{4.31}$$

where we have left out terms dominated (uniformly with respect to s_n) by a multiple of $d^{-2}(k-n)^3$. Now, by (2.16),

$$(4.32)\quad \begin{aligned}(k-n)(D_k-D_n)/D_n - (k-n)D_n^{-1}\sum_{i\varepsilon U-s_k} p_i + (k-n)D_k^{-1}\sum_{i\varepsilon s_n}(1-p_i)\\ = D_n^{-1}(k-n)^2 - (k-n)D_k^{-1}\sum_{i\varepsilon s_k-s_n}(1-p_i)\end{aligned}$$

and, further,

$$(4.33)\quad |D_n^{-1}-D_k^{-1}| = (D_k-D_n)/(D_nD_k) \leqq 4d^{-2}(k-n).$$

Consequently, (4.31) may be continued as follows

$$(4.34)\quad \begin{aligned}M(s_k, s_n) \approx \exp [-(k-n)D_k^{-1}\sum_{i\varepsilon s_k-s_n}(1-p_i)\\ + (k-n)^2D_k^{-2}\sum_{i\varepsilon s_k}(1-p_i)^2 + D_k^{-1}(k-n)^2\\ - \tfrac{1}{2}(k-n)^2D_k^{-2}\cdot(\sum_{i\varepsilon U-s_k} p_i^2 + \sum_{i\varepsilon s_k}(1-p_i)^2)].\end{aligned}$$

Now, by Tchebyshev inequality, $D_k \approx d$, $D_k^{-1}(k-n)^2$ is bounded in probability and

$$d^{-1}[\sum_{i\varepsilon U-s_k} p_i^2 + \sum_{i\varepsilon s_k}(1-p_i)^2] \sim d^{-1}\sum_{i=1}^{N}[p_i^2(1-p_i)+(1-p_i)^2p_i] = 1.$$

Thus, finally,

$$(4.35)\quad \begin{aligned}M(s_k, s_n) \approx \exp [\tfrac{1}{2}d^{-1}(k-n)^2]\\ \cdot\exp [-(k-n)D_k^{-1}\sum_{i\varepsilon s_k-s_n}(1-p_i) + (k-n)^2D_k^{-2}\sum_{i\varepsilon s_k}(1-p_i)^2].\end{aligned}$$

As for C^*, it follows from (4.19) that

$$(4.36)\quad C^*(U-s_n, k-n) \approx [2\pi(k-n)]^{\frac{1}{2}}.$$

On combining (4.23), (4.28), (4.35) and (4.36), we get

$$(4.37)\quad \begin{aligned}Q(s_k, s_n) \approx P(s_k)[2\pi(k-n)]^{\frac{1}{2}}\\ \cdot\exp [(k-n)^2D_k^{-2}\sum_{i\varepsilon s_k}(1-p_i)]^2 \prod_{i\varepsilon s_k-s_n}(k-n)D_k^{-1}(1-p_i)\\ \cdot\prod_{i\varepsilon s_n}[1-(k-n)D_k^{-1}(1-p_i)] \exp [-(k-n)D_k^{-1}\prod_{i\varepsilon s_k-s_n}(1-p_i)].\end{aligned}$$

Let $R^+(\cdot \mid s_k, k-n)$ denote rejective sampling of size $k-n$ with probabilities $(k-n)D_k^{-1}(1-p_i)$ from the population s_k, i.e.

$$(4.38)\quad \begin{aligned}R^+(s_k-s_n \mid s_k, k-n) = C^+(s_k, k-n)\cdot\prod_{i\varepsilon s_k-s_n}(k-n)\\ \cdot D_k^{-1}(1-p_i)\prod_{i\varepsilon s_n}[1-(k-n)D_k^{-1}(1-p_i)],\ (s_n \subset s_k).\end{aligned}$$

By the same considerations as we used in the proof of (4.19), we could show that

$$C^{+}(s_k, k - n) \approx [2\pi(k - n)]^{\frac{1}{2}}. \tag{4.39}$$

On comparing (4.21), (4.37), (4.38) and (4.39), we easily see that

$$\begin{aligned} P_0(s_k) \approx P(s_k) \exp \Big[(k - n)^2 D_k^{-2} \sum_{i \varepsilon s_k} (1 - p_i)^2\Big] \\ \cdot E^{+} \Big\{\exp \Big[-(k - n) D_k^{-1} \sum_{i \varepsilon s_k - s_n} (1 - p_i)\Big] \mid s_k, k - n\Big\} \end{aligned} \tag{4.40}$$

where $E^{+}(\cdot \mid s_k, k - n)$ denotes the mean where via $R^{+}(\cdot \mid s_k, k - n)$.

To conclude the proof for $k > n$, it remains to show that

$$\begin{aligned} E^{+} \Big\{\exp \Big[-(k - n) D_k^{-1} \sum_{i \varepsilon s_k - s_n} (1 - p_i)\Big] \mid s_k, k - n\Big\} \\ \gtrapprox \exp \Big[-(k - n)^2 D_k^{-2} \sum_{i \varepsilon s_k} (1 - p_i)^2\Big]. \end{aligned} \tag{4.41}$$

This is the most intricate part of the whole proof. First, introduce d_+ corresponding to R^{+}

$$\begin{aligned} d_+ = d_+(s_k) &= \sum_{i \varepsilon s_k} (k - n) D_k^{-1} (1 - p_i)[1 - (k - n) D_k^{-1} (1 - p_i)] \\ &= (k - n) - (k - n)^2 D_k^{-2} \sum_{i \varepsilon s_k} (1 - p_i)^2. \end{aligned} \tag{4.42}$$

Obviously

$$d^{+} \approx (k - n), \tag{4.43}$$

and $d^{+} \to \infty$ in P-probability. In view of Lemma 4.2 with U and p_i replaced by s_k and $(k - n) D_k^{-1} (1 - p_i)$, respectively, for any $\epsilon > 0$ there exists an $a(\epsilon)$ such that for $d_+ > a(\epsilon)$

$$\Big|d_+^{-1} \sum_{i \varepsilon s_k - s_n} (1 - p_i) - d_+^{-1} \sum_{i \varepsilon s_k} (k - n) D_k^{-1} (1 - p_i)^2\Big| < \epsilon, \tag{4.44}$$

with R^{+}-probability greater than $1 - \epsilon$. Further, there exists a $b(\epsilon)$ such that for $d > b(\epsilon)$

$$d_+(s_k) > a(\epsilon) \tag{4.45}$$

and

$$d_+(k - n) D_k^{-1} < \epsilon^{-\frac{1}{2}} \tag{4.46}$$

hold with P-probability greater than $1 - \epsilon$. Now, the relations (4.44) through (4.46) imply that

$$\begin{aligned} (1 - \epsilon) \exp \Big[-\epsilon^{\frac{1}{2}} - (k - n)^2 D_k^{-2} \sum_{i \varepsilon s_k} (1 - p_i)^2\Big] \\ \leqq E^{+} \Big\{\exp \Big[-(k - n) D_k^{-1} \sum_{i \varepsilon s_k - s_n} (1 - p_i)\Big] \mid s_k, k - n\Big\} \end{aligned} \tag{4.47}$$

with P-probability greater than $1 - \epsilon$ if $d > b(\epsilon)$. However, (4.47) is equivalent to (4.41). Thus the proof is finished for $k > n$. The proof for $k < n$ would be the same in principle, and is left to the reader.

REMARK 4.1. By a similar method we could obtain a sample s_n with approximately rejective distribution by starting from a Poisson sample s_k, and then extracting (if $k > n$) or adding (if $k < n$) $|k - n|$ units by rejective sampling of size $|k - n|$ from s_k or $U - s_k$ with probabilities $(k - n)D_k^{-1}(1 - p_i)$ or $(k - n)D_k^{-1}p_i$, respectively. This possibility will be considered in Section 9.

5. Probabilities of inclusion π_i and π_{ij}. We begin with

THEOREM 5.1. *Consider rejective sampling of size n with probabilities p_i, $1 \leqq i \leqq N$, such that $\sum_{i=1}^N p_i = n$, and denote by π_i the probability of including the* ith *unit in the sample, $i = 1, \cdots, N$ (see (1.2)).*

Then

$$\pi_i(1 - p_i) = p_i(1 - \pi_i)\,[1 - [(\bar{\bar{\pi}} - \pi_i)/d_0] + o(d_0^{-1})], \quad (1 \leqq i \leqq N), \tag{5.1}$$

$$\pi_i(1 - p_i) = p_i(1 - \pi_i)[1 - [(\bar{\bar{p}} - p_i)/d] + o(d^{-1})], \quad (1 \leqq i \leqq N), \tag{5.2}$$

where

$$d_0 = \sum_{i=1}^N \pi_i(1 - \pi_i), \tag{5.3}$$

$$d = \sum_{i=1}^N p_i(1 - p_i), \tag{5.4}$$

$$\bar{\bar{\pi}} = d_0^{-1}\sum_{i=1}^N \pi_i^2(1 - \pi_i), \tag{5.5}$$

and

$$\bar{\bar{p}} = d^{-1}\sum_{i=1}^N p_i^2(1 - p_i) \tag{5.6}$$

and $d \cdot o(d^{-1}) \to 0$ and $d_0 \cdot o(d_0^{-1}) \to 0$ if $d \to \infty$ and $d_0 \to \infty$, respectively, uniformly in $i = 1, \cdots, N$.

Furthermore

$$\max_{1 \leqq i \leqq N} |(\pi_i/p_i) - 1| \to 0, \tag{5.7}$$

$$\max_{1 \leqq i \leqq N} |[(1 - \pi_i)/(1 - p_i)] - 1| \to 0, \tag{5.8}$$

$$\max_{1 \leqq i \leqq N} |[(\pi_i(1 - \pi_i))/(p_i(1 - p_i))] - 1| \to 0, \tag{5.9}$$

and

$$|(d_0/d) - 1| \to 0, \tag{5.10}$$

if either $d \to \infty$ or $d_0 \to \infty$.

PROOF. Consider $P(s)$ given by (3.1) and recall the fact that for samples of size n, $K(s) = n$,

$$\sum_{h \varepsilon U - s} [p_h / \sum_{j \varepsilon s} (1 - p_j)] = [\sum_{h \varepsilon U - s} p_h / \sum_{j \varepsilon s} (1 - p_j)] = 1, \quad (K(s) = n),$$

(see (2.15)). Thus we have

$$\begin{aligned} P(s) &= P(s) \sum_{h \varepsilon U - s} [p_h / \sum_{j \varepsilon s} (1 - p_j)] \\ &= p_i/(1 - p_i) \sum_{h \varepsilon U - s} P(s \cup \{h\} - \{i\})[(1 - p_h)/\sum_{j \varepsilon s}(1 - p_j)], \quad (K(s) = n), \end{aligned} \tag{5.11}$$

and, consequently,

$$\sum_{s \ni i} P(s) = p_i/(1 - p_i) \sum_{U - s \ni i} P(s) \sum_{h \varepsilon s} [(1 - p_h)/(p_h - p_i + \sum_{j \varepsilon s} (1 - p_j))] \tag{5.12}$$

where s on the right side is the subset denoted by $s \cup \{h\} - \{i\}$ in (5.11). From the definition (1.2) of π_i it follows that

$$(1 - \pi_i) \sum_{s \ni i} P(s) = \pi_i \sum_{U - s \ni i} P(s) \qquad [K(s) = n] \tag{5.13}$$

since in rejective sampling the probability of s, $R(s)$ is proportionate to $P(s)$ (see (2.1)). On combining (5.12) and (5.13) we get

$$\begin{aligned} \pi_i(1 - p_i) &= p_i(1 - \pi_i) [\sum_{U - s \ni i} P(s)]^{-1} \\ &\quad \cdot \sum_{U - s \ni i} P(s) \left\{ \sum_{h \varepsilon s} \frac{1 - p_h}{p_h - p_i + \sum_{j \varepsilon s} (1 - p_j)} \right\}, \end{aligned} \tag{5.14}$$

$$(1 \leqq i \leqq N).$$

Now denote by $R_i(\cdot)$ the rejective sampling of size n with probabilities p_j, $j \neq i$, from $U - \{i\}$. Then (5.14) may be rewritten as

$$\pi_i(1 - p_i) = p_i(1 - \pi_i) E_i \left\{ \sum_{h \varepsilon s} \frac{1 - p_h}{p_h - p_i + \sum_{j \varepsilon s} (1 - p_j)} \right\} \tag{5.15}$$

$$(1 \leqq i \leqq N),$$

where E_i denotes the mean value via $R_i(\cdot)$.

Next step is to analyse the rejective sampling $R_i(\cdot)$. To put it into canonical form let us introduce probabilities p_j^*, $j \neq i$, such that (2.5) holds and $\sum_{j \neq i} p_j^* = n$. According to (2.9) we have, denoting $d_i = \sum_{j \neq i} p_j(1 - p_j)$

$$p_j^*(1 - p_j) = p_j(1 - p_j^*)[1 + (p_i/d_i) + o(d_i^{-1})], \qquad (j \neq i), \tag{5.16}$$

since $\sum_{j \neq i} p_j^* - \sum_{j \neq i} p_j = n - (n - p_i) = p_i$. Obviously, d_i may be replaced by d in (5.16), since $d/d_i = 1 + o(1)$ if $d \to \infty$. Thus we can modify (5.16) as follows:

$$p_j^*(1 - p_j) = p_j(1 - p_j^*)[1 + (p_i/d) + o(d^{-1})], \qquad (j \neq i). \tag{5.17}$$

Since $p_j^* \geqq p_j$, $j \neq i$, (5.17) implies

$$1 \leqq p_j^*/p_j \leqq 1 + (p_i/d) + o(d^{-1}) \tag{5.18}$$

and

$$1 \leqq (1 - p_j)/(1 - p_j^*) \leqq 1 + (p_i/d) + o(d^{-1}). \tag{5.19}$$

Consequently,

$$\max_{j \neq i} |[p_j^*(1 - p_j^*)/p_j(1 - p_j)] - 1| \to 0, \qquad \text{if } d \to \infty. \tag{5.20}$$

This means that $d \to \infty$ implies

$$d_{0i}^* = \sum_{j \neq i} p_j^*(1 - p_j^*) \to \infty \tag{5.21}$$

uniformly in i (for brevity, the dependence on i is not indicated in the symbol p_j^*).

Now, on account of (2.15)

$$\sum_{h \varepsilon s} \frac{1 - p_h}{p_h - p_i + \sum_{j \varepsilon s} (1 - p_j)} = 1 - d^{-1}B_i + o(d^{-1}) \tag{5.22}$$

where

$$B_i = d \frac{\sum_{j \varepsilon s} (1 - p_j)(p_j - p_i)}{[\sum_{j \varepsilon s} (1 - p_j)]^2} \tag{5.23}$$

with

$$|B_i| \leqq 4, \qquad (1 \leqq i \leqq N). \tag{5.24}$$

Now, we infer from (5.18) through (5.21), (5.23) and Lemma 4.2, that

$$(B_i - \bar{\bar{p}} + p_i) \to 0 \tag{5.25}$$

in R_i-probability as $d \to \infty$, uniformly with respect to i. From (5.24) and (5.25) it follows that $E_i(B_i) = \bar{\bar{p}} - p_i + o(1)$, if $d \to \infty$, which inserted into (5.15) gives (5.2).

Now from (5.2) it follows that

$$p_i - \pi_i = [(\bar{\bar{p}} - p_i)/d + o(d^{-1})]p_i(1 - \pi_i)$$

and, consequently,

$$|(\pi_i/p_i) - 1| = (1 - \pi_i)|(\bar{\bar{p}} - p_i)/d + o(d^{-1})|, \quad (1 \leqq i \leqq N). \tag{5.26}$$

However, (5.26) is equivalent to (5.7). Furthermore, (5.8) may be proved in a similar way, and (5.9) is an easy consequence of (5.7) and (5.8). Finally (5.10) follows from (5.9), and (5.1) follows from (5.2) and (5.7), (5.9) and (5.10). The proof is finished.

The following theorem is basic for applications.

THEOREM 5.2. *Let π_{ij} be the probability of including both the* ith *and the* jth *unit in the sample for rejective sampling of size n with probabilities p_i, $1 \leqq i \leqq N$, such that $\sum_{i=1}^N p_i = n$.*

Then

$$\pi_i\pi_j - \pi_{ij} = d_0^{-1}\pi_i(1 - \pi_i)\pi_j(1 - \pi_j)[1 + o(1)], \tag{5.27}$$

where $d_0 = \sum_{i=1}^N \pi_i(1 - \pi_i)$ *and* $o(1) \to 0$ *uniformly for* $d_0 \to \infty$.

PROOF. First observe that π_{ij}/π_i may be regarded as the probability of inclusion of the jth unit in rejective sampling of size $n - 1$ with probabilities p_j, $j \neq i$, from the population $U - \{i\}$. Corresponding canonical probabilities p_j^+ satisfy, in accordance with (2.9),

$$p_j^+(1 - p_j) = p_j(1 - p_j^+)[1 + [(p_i - 1)/d] + o(d^{-1})], \quad (j \neq i), \tag{5.28}$$

since $\sum_{j\neq i} p_j^+ - \sum_{j\neq i} p_j = n - 1 - (n - p_i) = p_i - 1$, and $\sum_{j\neq i} p_j(1 - p_j)$ may be replaced by $d = \sum_{j=i}^N p_j(1 - p_j)$. On the other hand we may apply (5.2) to the present sampling (i.e. put π_{ij}/π_i for π_i and p_j^+ for p_i), which yields

$$(\pi_{ij}/\pi_i)(1 - p_j^+) = p_j^+[1 - (\pi_{ij}/\pi_i)][1 - [(\bar{\bar{p}}^+ - p_j^+)/d_+] + o(d_+^{-1})], \tag{5.29}$$

where

$$d_+ = \sum_{j\neq i} p_j^+(1 - p_j^+).$$

Utilizing (5.28) in the same way as we did (5.2) in proving (5.7) through (5.10), we may show that $\bar{\bar{p}}^+$, p_j^+ and d_+ in the bracket in (5.29) may be replaced by $\bar{\bar{p}}$, p_j and d, which, in turn, may be replaced by $\bar{\bar{\pi}}$, π_j and d_0, in view of (5.7) through (5.10). Thus, (5.28) and (5.29) may be rewritten as follows:

$$p_j^+(1 - p_j) = p_j(1 - p_j^+)[1 + [(\pi_i - 1)/d_0] + o(d_0^{-1})], \quad (j \neq i), \tag{5.30}$$

$$(\pi_{ij}/\pi_i)(1 - p_j^+) = p_j^+[1 - (\pi_{ij}/\pi_i)][1 - [(\bar{\bar{\pi}} - \pi_j)/d_0] + o(d_0^{-1})], \quad (j \neq i). \tag{5.31}$$

Eliminating p_j^+ and p_j from (5.1), (5.30) and (5.31), we get

$$(\pi_{ij}/\pi_i)(1 - \pi_j) = \pi_j[1 - (\pi_{ij}/\pi_i)][1 - [(1 - \pi_i)/d_0] + o(d_0^{-1})], \quad (j \neq i). \tag{5.32}$$

However, carrying out the multiplications, (5.32) becomes

$$\pi_i\pi_j - \pi_{ij} = d_0^{-1}\pi_i\pi_j[1 - (\pi_{ij}/\pi_i)][1 - \pi_i + o(1)], \quad (j \neq i). \tag{5.33}$$

Since (5.32) also implies that

$$|\{[1 - (\pi_{ij}/\pi_i)]/[1 - \pi_j]\} - 1| \to 0, \qquad \text{if } d_0 \to \infty,$$

uniformly in $j \neq i$, (5.33) is equivalent to (5.27) provided that $\pi_i \leqq \frac{1}{2}$. If $\pi_i \geqq \frac{1}{2}$, it suffices to consider the complementary rejective sampling with $P'(s) = P(U - s)$, which has the same d_0 and probabilities of inclusion $\pi_i' = 1 - \pi_i$ and $\pi_{ij}' = 1 + \pi_{ij} - \pi_i - \pi_j$, so that $\pi_i(1 - \pi_i) = \pi_i'(1 - \pi_i')$ and $\pi_i'\pi_j' - \pi_{ij}' = \pi_i\pi_j - \pi_{ij}$. The proof is finished.

6. The variance of the simple linear estimator. We shall first assume that we know the exact values of the π_i's.

THEOREM 6.1. *Consider a rejective sample s of size n selected with probabilities of inclusion* $\pi_1, \cdots, \pi_N$, *and a population of values* $y_i, \cdots, y_N$. *Put*

$$\hat{Y}_0 = \sum_{i \varepsilon s} y_i/\pi_i . \tag{6.1}$$

Then

$$E(\hat{Y}_0) = \sum_{i=1}^{N} y_i \tag{6.2}$$

and

$$\operatorname{var}(\hat{Y}_0) = [1 + o(1)]\sum_{i=1}^{N}(y_i - R\pi_i)^2[(1/\pi_i) - 1] \tag{6.3}$$

where

$$R = \sum_{i=1}^{N} y_i(1 - \pi_i)/\sum_{i=1}^{N} \pi_i(1 - \pi_i) \tag{6.4}$$

and $o(1) \to 0$ *if* $\sum_{i=1}^{N} \pi_i(1 - \pi_i) \to \infty$.

PROOF. The assertion (6.2) is obvious and (6.3) follows immediately from (5.27) and (1.13).

Now assume, as in practice, that we want the π_i's to be proportionate to some nonnegative numbers x_i such that

$$0 < nx_i < X = \sum_{i=1}^{N} x_i , \qquad (1 \leqq i \leqq N). \tag{6.5}$$

Since $\sum_{i=1}^{N} \pi_i = n$, we should have

$$\pi_i = nx_i/X, \qquad (1 \leqq i \leqq N). \tag{6.6}$$

However, we are only able to control the probabilities p_i, of which the π_i's are rather complicated functions. None the less we have established the relations (5.9) and (5.27) from which it follows that

$$\pi_i\pi_j - \pi_{ij} = d^{-1}p_i(1 - p_i)p_j(1 - p_j)[1 + o(1)] \tag{6.7}$$

So, replacing (6.6) by

$$p_i = nx_i/X \qquad (1 \leqq i \leqq N) \tag{6.8}$$

we still have the following modification of Theorem 6.1.

THEOREM 6.2. *Consider rejective sampling of size n from a population of values* $y_1, \cdots, y_N$ *with probabilities* $p_i, \cdots, p_N$ *given by* (6.8), *and the estimator*

$$\hat{Y} = X/n\sum_{i \varepsilon s} y_i/x_i . \tag{6.9}$$

Then

$$E(\hat{Y} - Y)^2 = [1 + o(1)]\sum_{i=1}^{N}(y_i - Rx_i)^2[(X/nx_i) - 1] \tag{6.10}$$

where $o(1) \to 0$ *if* $d \to \infty$,

$$d = \sum_{i=1}^{N} nx_i/X[1 - (nx_i/X)] \tag{6.11}$$

and

$$R = \sum_{i=1}^{N} y_i[1 - (nx_i/X)]/\sum_{i=1}^{N} x_i[1 - (nx_i/X)]. \tag{6.12}$$

Proof. Obviously,

$$E(\hat{Y} - Y)^2 = \text{var}(\hat{Y}) + (E\hat{Y} - Y)^2. \tag{6.13}$$

Now, in accordance with (6.7) and (6.8),

$$\begin{aligned} \text{var } \hat{Y} &= \tfrac{1}{2}\sum\sum_{i \neq j} [(y_i/x_i) - (y_j/x_j)](X^2/n^2)(\pi_i\pi_j - \pi_{ij}) \\ &= \tfrac{1}{2}\sum\sum_{i \neq j} [(y_i/x_i) - (y_j/x_j)]x_ix_j[1 - (nx_i/x_i)][1 - (nx_j/x)] \\ &\qquad \cdot d^{-1}[1 + o(1)] \\ &= [1 + o(1)]\sum_{i=1}^{N} (y_i - Rx_i)^2[(X/nx_i) - 1]. \end{aligned} \tag{6.14}$$

Furthermore, in view of (6.12) and (5.26)

$$\begin{aligned} (E\hat{Y} - Y)^2 &= \left[\sum_{i=1}^{N} y_i((\pi_i X/nx_i) - 1)\right]^2 \\ &= \left[\sum_{i=1}^{N} (y_i - Rx_i)\,\frac{\pi_i - nx_i/X}{nx_i/X}\right]^2 \\ &= \left[\sum_{i=1}^{N} (y_i - Rx_i)(1 - \pi_i)\left(\frac{\bar{\bar{p}} - p_i}{d} + o(d^{-1})\right)\right]^2 \\ &\leqq \sum_{i=1}^{N} (y_i - Rx_i)^2(\pi_i^{-1} - 1) \sum_{i=1}^{N} \pi_i(1 - \pi_i)\left[\frac{\bar{\bar{p}} - p_i}{d} + o(d^{-1})\right]^2 \\ &= \sum_{i=1}^{N} (y_i - Rx_i)^2[(X/nx_i) - 1][o(1)] \end{aligned}$$

which shows that

$$(E\hat{Y} - Y)^2/\text{var } \hat{Y} \to 0.$$

Consequently, the asymptotic expression for var $\hat{Y}$ (6.14) is correct for $E(\hat{Y} - Y)^2$ as well. The proof is finished.

REMARK 6.1. In Section 8 we shall suggest instead of (6.8) another relation between the p_i's and the x_i's yielding a better approximation of (6.6). Generally, it suffices that

$$(6.10)\qquad p_i[1 - (nx_i/X)] = (nx_i/X)(1 - p_i)[1 + o(1)], \qquad 1 \leqq i \leqq N$$

where $o(1) \to 0$, uniformly in i, if $d \to \infty$.

7. Necessary and sufficient conditions for asymptotic normality. Under the conditions of Theorem 6.2 consider the estimator $\hat{Y}$ given by (6.9) and simultaneously the quantity

$$(7.1)\qquad T = (X/n)\sum_{i \varepsilon s}[(y_i/x_i) - R] + XR$$

where R is given by (6.12) and s denotes a Poisson sample with the same probabilities (6.8). From (3.15) and (6.10) it easily follows that

$$(7.2)\qquad \text{var}^P(T) = E^R(\hat{Y} - Y)^2[1 + o(1)] \qquad \text{if } d \to \infty,$$

where var^P and E^R refer to Poisson and rejective sampling, respectively.

There is, however, a still deeper relation between T and $\hat{Y}$, which may be heuristically explained as follows: Let s_n and s_k denote a typical rejective and Poisson sample, respectively. If, for example, $s_n \subset s_k$, then

$$(7.3)\qquad T - \hat{Y} = (X/n)\sum_{i \varepsilon s_k - s_n}[(y_i/x_i) - R].$$

Consequently, if the difference $s_k - s_n$ contains a relatively small number of units, the difference $T - \hat{Y}$ will be small, too, and may be asymptotically negligible.

THEOREM 7.1. *Consider rejective sampling of size n with probabilities* $p_i = nx_i/X$ *from a population of values* $y_1, \cdots, y_N$. *Define* Y *by* (6.9) *and put*

$$(7.4)\qquad B^2 = \sum_{i=1}^{N}(y_i - Rx_i)^2[(X/nx_i) - 1]$$

where R *is given by* (6.12). *Moreover put*

$$(7.5)\qquad e = e(y_1, \cdots, y_N, x_i, \cdots x_N) = \inf\{\epsilon : L_\epsilon^* \leqq \epsilon\}$$

where

$$(7.6)\qquad L_\epsilon^* = B^{-2}\sum_{i \varepsilon A_\epsilon}(y_i - Rx_i)^2[(X/nx_i) - 1]$$

with

$$(7.7)\qquad A_\epsilon = \{i : |y_i - Rx_i| > \epsilon(nx_i/X)B\}$$

Then

$$(7.8)\qquad e \to 0$$

implies

$$\sup_x |P(\hat{Y} < x) - \Phi[(x - Y)/B]| \to 0, \tag{7.9}$$

where $P(\cdot)$ *denotes the probability referring to the rejective sampling and* Φ *denotes the normed normal distribution function.*

PROOF. We know from Lemma 3.3 that the theorem holds for corresponding Poisson sampling and the statistic (7.1). Now, according to Lemma 4.3 the asymptotic distribution of T will be the same also for the distribution $P_0(\cdot)$ considered there. We also may consider the distribution $Q(s_k, s_n)$ from Lemma 4.2, since P_0 is its marginal distribution for s_k. However, under Q we always have either $s_k \supset s_n$ or $s_k \subset s_n$, and $s_k - s_n$ (or $s_n - s_k$) represents a rejective sample of size $k - n$ (or $n - k$) from $U - s_n$ (or s_n). Thus it suffices to show that the difference (7.3) is asymptotically negligible in the following sense:

$$B^{-2}(T - \hat{Y})^2 \sim 0, \tag{7.11}$$

which means that left side converges to the right side in Q-probability. Furthermore, for $k > n$ (7.11) is equivalent to

$$B^{-2} \operatorname{var}^*(T - \hat{Y} \mid U - s_n, k - n) \sim 0 \tag{7.12}$$

and

$$B^{-1}E^*(T - \hat{Y} \mid U - s_n, k - n) \sim 0 \tag{7.13}$$

where var^* and E^* refer to the rejective sampling R^* of size $k - n$ from $U - s_n$ given by (4.15). Leaving the case $k < n$ to the reader, we shall conclude the proof by establishing (7.12) and (7.13).

According to (1.13) and (7.3)

$$\begin{aligned} &\operatorname{var}^*(T - \hat{Y} \mid U - s_n, k - n) \\ &\qquad = \tfrac{1}{2}(X/n)^2 \sum\sum_{i,j \varepsilon U - s_n} [(y_i/x_i) - (y_j/x_j)]^2(\pi_i^*\pi_j^* - \pi_{ij}^*) \end{aligned} \tag{7.14}$$

and, furthermore,

$$E^*(T - Y \mid U - s_n, k - n) = X/n \sum_{i \varepsilon U - s_n} [(y_i/x_i) - R]\pi_i^*. \tag{7.15}$$

Denote by $\approx$ that the ratio of both sides tends to 1 in Q-probability if $d \to \infty$ and note that (4.18) entails

$$d^* = \sum_{i \varepsilon U - s_n} \pi_i^*(1 - \pi_i^*) = \operatorname{var}^* K^* \approx (k - n).$$

Since $|K - n| \to \infty$ in Q-probability, and in view of (5.1) and (5.7), we have

$$\begin{aligned} &\pi_i^* \approx (k - n)D_n^{-1}p_i \approx (k - n)d^{-1}(nx_i/X), \\ &\qquad\qquad (\text{uniformly in } i = 1, \cdots, N), \end{aligned} \tag{7.16}$$

where $d = \sum_{i=1}^N (nx_i/X)(1 - nx_i/X)$. Furthermore, in view of (5.27)

$$\pi_i^*\pi_j^* - \pi_{ij}^* \approx (k - n)^{-1}\pi_i^*\pi_j^* \approx (k - n)d^{-2}(n/X)^2 x_i x_j. \tag{7.17}$$

Now, (7.14) in connection with (7.16) and (7.17) yields

$$\begin{aligned}
&\operatorname{var}^*(T - Y \mid U - s_n, k - n) \\
&\qquad \approx \tfrac{1}{2}(X/n)^2(k - n)^{-1}\sum\sum_{i,j\varepsilon U-s_n}[(y_i/x_i) - (y_j/x_j)]^2\pi_i^*\pi_j^* \\
&\qquad \leqq (X/n)^2 \sum_{i\varepsilon U-s_n}[(y_i/x_i) - R]^2\pi_i^* \\
&\qquad \approx (k - n)d^{-1}\sum_{i\varepsilon U-s_n}(y_i - Rx_i)^2(X/nx_i).
\end{aligned} \tag{7.18}$$

Consequently,

$$\begin{aligned}
&B^{-2}\operatorname{var}^*(T - \hat{Y} \mid U - s_n, k - n) \\
&\qquad \approx (k - n)d^{-1}B^{-2}\sum_{i\varepsilon U-s_n}(y_i - Rx_i)^2(X/nx_i)
\end{aligned} \tag{7.19}$$

where

$$(k - n)d^{-1} \sim 0 \tag{7.20}$$

and the remaining factor is bounded in Q-probability, since

$$\begin{aligned}
&E\left\{B^{-2}\sum_{i\varepsilon U-s_n}(y_i - Rx_i)^2(X/nx_i)\right\} \\
&\qquad = B^{-2}\sum_{i=1}^{N}(y_i - Rx_i)^2(X/nx_i)(1 - \pi_i) \approx 1.
\end{aligned} \tag{7.21}$$

Obviously (7.19) through (7.21) entails (7.12).

It remains to prove (7.13), which is somewhat more intricate. In view of (7.15) and (7.16), we have

$$\begin{aligned}
&E^*(T - \hat{Y} \mid U - s_n, k - n) \\
&\qquad = X/n\sum_{i\varepsilon U-s_n}[(y_i/x_i) - R](k - n)d^{-1}(nx_i/X)[1 + o_i(1)] \\
&\qquad = (k - n)d^{-1}\sum_{i\varepsilon U-s_n}(y_i - Rx_i)[1 + o_i(1)],
\end{aligned} \tag{7.22}$$

where $o_i(1) \to 0$ uniformly in i if $d \to \infty$. Thus, denoting by E_0 the mean value referring to s_n, we have

$$\begin{aligned}
&E_0[E^*(T - \hat{Y} \mid U - s_n, k - n)] \\
&\qquad = (k - n)d^{-1}\sum_{i=1}^{N}(y_i - Rx_i)[1 + o_i(1)](1 - \pi_i) \\
&\qquad = (k - n)d^{-1}\sum_{i=1}^{N}(y_i - Rx_i)[1 - (nx_i/X)][1 + o_i(1)] \\
&\qquad = (k - n)d^{-1}\sum_{i=1}^{N}(y_i - Rx_i)[1 - (nx_i/X)][o_i(1)].
\end{aligned} \tag{7.23}$$

Consequently, by the Cauchy inequality,

(7.24) $\{E_0[E^*(T - \hat{Y} \mid U - s_n, k - n)]\}^2 = (k - n)^2 d^{-1} B^2[o(1)]$

Since $(k - n)^2 d^{-1}$ is bounded in Q-probability, we have

$$B^{-2}\{E_0[E(T - \hat{Y} \mid U - s_n, k - n)]\}^2 \sim 0 \tag{7.25}$$

if $d \to \infty$.

Further, denoting the variance referring to s_n by var_0, we obtain from (7.22)

$$\begin{aligned} &\text{var}_0[E^*(T - \hat{Y} \mid U - s_n, k - n)] \\ &\approx \tfrac{1}{2}(k - n)^2 d^{-3} \sum_{i=1}^{N}\sum_{j=1}^{N}[(y_i - Rx_i)(1 + o_i(1)) \\ &- (y_j - Rx_j)(1 + o_j(1))]^2 \cdot [(nx_i nx_j)/XX][1 - (nx_i/X)][1 - (nx_i/X)] \\ &\lessapprox (k - n)^2 d^{-2} \sum_{i=1}^{N} (y_i - Rx_i)^2 (nx_i/X)[1 - (nx_i/X)] \\ &\leqq (k - n)^2 d^{-2} B^2. \end{aligned} \tag{7.26}$$

Consequently, also

$$B^{-2}\, \text{var}_0[E(T - \hat{Y} \mid U - s_n, k - n)] \sim 0. \tag{7.27}$$

However, the relations (7.27) and (7.25) are sufficient for (7.13) in the same way as (7.13) and (7.12) were sufficient for (7.11). The proof is finished.

REMARK. Condition (7.8) is also necessary.

8. Numerical investigations and application. Given some numbers x_i, $1 \leqq i \leqq N$, we generally do not know any method of performing rejective sampling such that $\pi_i = nx_i/X$ would hold exactly. However, there are satisfactory approximations. We have seen that rejective sampling with probabilities $p_i = nx_i/X$ will yield probabilities of inclusion such that

$$\pi_i/(nx_i/X) \to 1, \qquad 1 \leqq i \leqq N, \tag{8.1}$$

uniformly in i, if $d \to \infty$, $d = \sum_{i=1}^{N} (nx_i/X)[1 - (nx_i/X)]$. Convergence of $\pi_i/(nx_i/X)$ to 1 may be accelerated by replacing $p_i = nx_i/X$ by some more intricate relationship obtained from (5.1).

Before doing that, let us recall that definition (2.1) of rejective sampling is associated with the possibility of performing it as Poisson sampling with probabilities $p_1, \cdots . p_N$ given the condition that the sample size is n. The alternative definition (1.5) is associated with the possibility of performing rejective sampling as sampling with replacement in which the probability of selecting the ith unit in a single draw equals α_i, $1 \leqq i \leqq N$. The parameters p_i and α_i both may be controlled by the statistician and are related by (2.3). Thus, for example, $p_i = nx_i/X$ is equivalent to

$$\alpha_i = (\lambda_1 x_i)/[1 - (nx_i/X)], \qquad 1 \leqq i \leqq N, \tag{8.2}$$

where λ_1 is to be chosen so that $\sum_{i=1}^{N} \alpha_i = 1$.

Now, from (5.1) it follows that

$$\begin{aligned} \alpha_i &= \lambda_2 p_i/(1 - p_i) \doteq [(\lambda_2\pi_i)/(1 - \pi_i)]\{1 - [(\bar{\bar{\pi}} - \pi_i)/d_0]\}^{-1} \\ &\doteq [(\lambda_2\pi_i)/(1 - \pi_i)]\{1 + [(\bar{\bar{\pi}} - \pi_i)/d_0]\} \qquad (8.3) \\ &= [(\lambda_3\pi_i)/(1 - \pi_i)]\{1 - [\pi_i/(d_0 + \bar{\bar{\pi}})]\}. \end{aligned}$$

So, if we put $\pi_i = nx_i/X$ and recall (5.3) and (5.5), we get for the α_i's the following approximation

$$\alpha_i = \lambda x_i/[1 - (nx_i/X)][1 - (x_i/X^*)], \qquad (1 \leqq i \leqq N) \qquad (8.4)$$

where

$$X^* = \sum_{i=1}^N x_i[1 - (nx_i/X)] + \frac{\sum_{i=1}^N x_i^2[1 - (nx_i/X)]}{\sum_{i=1}^N x_i[1 - (nx_i/X)]} \qquad (8.5)$$

and λ is determined from $\sum_{i=1}^N \alpha_i = 1$. The α_i's obtained from (8.4) bring $\pi_i/(nx_i/X)$ close to 1 even for small n and N (see example below). The relation (8.4) may be simplified in various ways. For example, we may take

$$\alpha_i = \lambda x_i/\{1 - [(n - 1)x_i/X]\}, \qquad (1 \leqq i \leqq N) \qquad (8.6)$$

or

$$\alpha_i = \lambda x_i[1 - (x_i/X)]/[1 - (nx_i/X)], \qquad (1 \leqq i \leqq N). \qquad (8.7)$$

In Table 2 there are numerical values of ratios of the true α_i's and the approximate α_i's computed from above formulas for $N = 10$, $n = 4$ and $x_i = \pi_i$,

TABLE 2

A comparison of the α_i's with the probabilities of inclusion π_i and with approximation α_i^I, α_i^{II}, α_i^{III}, α_i^{IV} based on formulas (8.4), (8.7), (8.6), (8.2) with $x_i = \pi_i$, respectively, for rejective sampling ($N = 10$, $n = 4$, $\alpha_i = i/55$)

Formula for α_i^Q ...			(8.4)	(8.7)	(8.6)	(8.2)
i	$\alpha_i = \frac{i}{55}$	$\frac{\frac{1}{4}\pi_i}{\alpha_i}$	$\frac{\alpha_i^I}{\alpha_i}$	$\frac{\alpha_i^{II}}{\alpha_i}$	$\frac{\alpha_i^{III}}{\alpha_i}$	$\frac{\alpha_i^{IV}}{\alpha_i}$
1	.0182	1.47	.999	.94	1.03	.85
2	.0364	1.37	.999	.95	1.03	.88
3	.0545	1.28	1.000	.96	1.04	.91
4	.0727	1.19	1.000	.98	1.03	.94
5	.0909	1.12	1.001	.99	1.03	.96
6	.1091	1.05	1.001	1.00	1.03	.99
7	.1273	.98	1.001	1.00	1.01	1.01
8	.1455	.92	1.001	1.01	.99	1.03
9	.1636	.87	.999	1.02	.98	1.04
10	.1818	.82	.998	1.02	.96	1.06

$1 \leqq i \leqq 10$. Denoting the probabilities of inclusion referring to α_i, α_i^I, α_i^{II}, α_I^{III}, α_i^{IV} by π_i, π_i^I, π_i^{II}, π_i^{III}, π_i^{IV}, respectively, we may assume that π_i^g/π_i would be even less variable than α_i^g/α_i, $g = I, II, III, IV$. This assumption is justified by (8.4).

Hence we can conclude that formula (8.4) and its simplification solve the problem P1 formulated in Section 1 for any situation which may in practice occur.

Now, as for the problem P2, let us calculate the values of $\pi_i\pi_j - \pi_{ij}$ for the same example as in Table 2 and compare them with the approximate values given by (5.27):

$$\pi_i\pi_j - \pi_{ij} \doteq \left[\sum_{i=1}^{N} \pi_i(1 - \pi_i)\right]^{-1} \pi_i(1 - \pi_i)\pi_j(1 - \pi_j). \tag{8.8}$$

The results are presented in Tables 3 and 4. In Table 4 we can observe that the ratios tend to be smaller than 1, which is due to the fact that, denoting $d = \sum \pi_i(1 - \pi_i)$,

$$\sum_{i<j}\sum (\pi_i\pi_j - \pi_{ij}) = \tfrac{1}{2}d$$

whereas

$$\sum_{i<j}\sum \pi_i(1 - \pi_i)\pi_j(1 - \pi_j)d^{-1} = \tfrac{1}{2}d - \tfrac{1}{2}d^{-1}\sum \pi_i^2(1 - \pi_i)^2 < \tfrac{1}{2}d.$$

The relative difference is, however, of order $1/N$ and is not worth correcting for the N's encountered in practice.

Consequently, we may hold that the problem P2 also is satisfactorily solved by formula (8.8).

Now in practice we may proceed as follows. Wanting the π_i's to be proportionate to some numbers x_i, we carry out rejective sampling with the α_i's given (8.6), since this formula is both simple and satisfactory in usual cases. Then we use the estimate

$$\hat{Y} = (X/n)\sum_{i\varepsilon s} y_i/x_i \tag{8.9}$$

TABLE 3

$\pi_i\pi_j - \pi_{ij}$ for rejective sampling ($N = 10$, $n = 4$, $\alpha_i = i/55$)

i	j = 2	3	4	5	6	7	8	9	10
1	.0066	.0087	.0101	.0111	.0116	.0119	.0120	.0119	.0117
2		.0151	.0175	.0191	.0200	.0204	.0205	.0203	.0200
3			.0228	.0247	.0258	.0263	.0263	.0260	.0255
4				.0284	.0296	.0300	.0299	.0295	.0288
5					.0318	.0321	.0319	.0314	.0306
6						.0330	.0328	.0321	.0313
7							.0328	.0321	.0313
8								.0317	.0308
9									.0301

TABLE 4

$\dfrac{\pi_i(1-\pi_i)\pi_j(1-\pi_j)}{(\pi_i\pi_j-\pi_{ij})\sum_{i=1}^{10}\pi_i(1-\pi_i)}$ *for rejective sampling* ($N = 10$, $n = 4$, $\alpha_i = i/55$)

i	j								
	2	3	4	5	6	7	8	9	10
1	1.07	1.03	.99	.97	.95	.93	.92	.91	.91
2		.99	.96	.93	.92	.91	.90	.89	.89
3			.93	.91	.90	.89	.88	.88	.88
4				.89	.88	.88	.87	.87	.88
5					.87	.87	.87	.87	.88
6						.87	.87	.88	.88
7							.88	.88	.89
8								.89	.90
9									.91

whose mean square deviation equals, approximately,

$$E(\hat{Y} - Y)^2 = \sum_{i=1}^{N} (y_i - Rx_i)^2[(X/nx_i) - 1] \tag{8.10}$$

with

$$R = \frac{\sum_{i=1}^{N} y_i[1 - (nx_i/X)]}{\sum_{i=1}^{N} x_i[1 - (nx_i/X)]} \doteq Y/X. \tag{8.11}$$

Furthermore, (8.10) may be estimated by

$$e(\hat{Y} - Y)^2 = X^2/[n(n-1)]\sum_{i\varepsilon s} [(y_i/x_i) - r]^2[1 - (nx_i/X)] \tag{8.12}$$

where

$$r = \frac{\sum_{i\varepsilon s} (y_i/x_i)[1 - (nx_i/X)]}{\sum_{i\varepsilon s} [1 - (nx_i/X)]}\ ; \tag{8.13}$$

with a slight overestimation we may put

$$r = n^{-1}\sum_{i\varepsilon s} y_i/x_i\,. \tag{8.14}$$

Formula (8.12) is a simplified version of the Yates-Grundy estimator

$$e(\hat{Y} - Y)^2 = \tfrac{1}{2}\sum_{i,j\varepsilon s}\sum [(y_i/\pi_i) - (y_j/\pi_j)]^2(\pi_i\pi_j - \pi_{ij})/\pi_{ij}$$

$$\doteq \tfrac{1}{2}(X^2/n^2)\sum_{i,j\varepsilon s}\sum [(y_i/x_i) - (y_j/x_j)]^2[1 - (nx_i/X)]$$

$$\cdot [1 - (nx_j/X)][d_0 - (1 - nx_i/X)(1 - nx_j/X)]^{-1}$$

where we have used $\pi_i \doteq nx_i/X$ and the formula (8.8) for π_{ij}. Now, further approximations

$$[d_0 - (1 - nx_i/X)(1 - nx_j/X)]^{-1} \doteq d_0^{-1}[n/(n-1)]$$

and (see examples following Lemma 4.2) $d_0^{-1}\sum_{i\varepsilon s}[1 - (nx_i/X)] \doteq 1$ lead to formula (8.12).

If we decide to utilize the numbers x_i by ratio estimation, we still may use unequal probabilities in order to make use of the variability of expected values of $(y_i - Rx_i)^2$. Then the π_i's should be proportionate to some numbers z_i (usually $z_i = x_i^g$, $1 \leqq g \leqq 2$), which may be accomplished by using again (8.6) with the x_i's replaced by the z_i's. The estimator becomes

$$\check{Y} = X \frac{\sum_{i\varepsilon s} y_i/z_i}{\sum_{i\varepsilon s} x_i/z_i} \tag{8.15}$$

and its mean square deviation may be approximated by

$$\begin{aligned} E(\check{Y} - Y)^2 &= E\left\{\left(\frac{Xn}{Z\sum_{i\varepsilon s} x_i/z_i}\right)^2 \left(Z/n \sum_{i\varepsilon s} \frac{y_i - (Y/X)x_i}{z_i}\right)^2\right\} \\ &\doteq E\left(Z/n \sum_{i\varepsilon s} \frac{y_i - (Y/X)x_i}{z_i}\right)^2 \\ &= \sum_{i=1}^{N} [y_i - (Y/X)x_i - \varphi z_i]^2[(Z/nz_i) - 1], \end{aligned} \tag{8.16}$$

where

$$\varphi = \frac{\sum_{i=1}^{N} (y_i - (Y/X)x_i)[1 - (nz_i/Z_i)]}{\sum_{i=1} z_i[1 - (nz_i/Z)]} \doteq 0 \tag{8.17}$$

and $Z = \sum_{i=1}^{N} z_i$. Obviously, we have used (8.10) with y_i and x_i replaced by $y_i - (Y/X)x_i$ and z_i, respectively. Usually the approximation with $\varphi = 0$ will do as well.

For estimated mean square error we can take, admitting a slight upward bias,

$$\begin{aligned} e(\hat{Y} - Y)^2 = (X/\hat{X})^2[Z^2/n(n-1)]\sum_{i\varepsilon s} [(y_i/z_i) - (\hat{Y}/\hat{X})(x_i/z_i)]^2 \\ \cdot [1 - (nz_i/Z)] \end{aligned} \tag{8.18}$$

where

$$\hat{Y} = Zn^{-1}\sum_{i\varepsilon s} y_i/z_i, \qquad \hat{X} = Zn^{-1}\sum_{i\varepsilon s} x_i/z_i. \tag{8.19}$$

9. Short-cuts in performing rejective sampling. Most frequently rejective sampling would be performed by n successive independent (i.e. with replacement) draws of one unit with single-draw probabilities α_i given by, say, (8.6), rejecting the whole partly built up sample, if any two units in it coincide (in contra distinction of successive sampling, where only the last unit is rejected, if it coincides with some previously drawn).

In carrying out the single draws, the following short-cuts may prove useful:

METHOD 1. We choose a number Q such that

$$Q = \max_{1 \leqq i \leqq N} x_i/\{1 - [(n-1)x_i/X]\}$$

and select a random number i in the range from 1 to N and, independently, a random number q in the range from 1 to Q. If

$$q \leqq x_i/\{1 - [(n-1)x_i/X]\}, \tag{9.1}$$

we draw the unit i but reiterate the whole process otherwise.

When applying Method 1, we do not need to compute the numbers $x_i/(1 - (n-1)x_i/X)$ except for the units selected in the course of the performance, and even for a great portion of them we are able to check (9.1) by eye.

If the sample is rejected with a high probability, then another method based on Remark 4.1 is worth considering:

METHOD 2. We select N independent random numbers $q_1, \cdots, q_N$ in the range from 1 to $n^{-1}X$ and include in a provisory sample every unit i such that $q_i \leqq x_i$. Denote the provisory sample by s_k and suppose it contains k units. Now, if $k = n$, we accept s_k for the definitive sample; if $k < n$ we select from the population $U - s_k$ a rejective sample s^0_{n-k} of size $n - k$ with single-draw probabilities

$$\alpha_i^0 = \lambda x_i[1 - (n-k-1)x_i d^{-1}]^{-1}, \qquad (i \varepsilon U - s_k)$$

where $d = \sum_{i=1}^N (nx_i/X)[1 - (nx_i/X)]$, and put $s = s_k \cup s^0_{n-k}$; if $k > n$, we select from s_k a rejective sample s^0_{k-n} of size $k - n$ with single-draw probabilities

$$\alpha_i^0 = \lambda[1 - (nx_i/X)]\{1 - (k-n-1)d^{-1}[1 - (nx_i/X)]\}^{-1}$$

and put $s = s_k - s^0_{k-n}$.

If N is large, the first step in Method 2 may be facilitated as follows: We select a number Q such that $Q \geqq \max_{1 \leqq i \leqq N} x_i$ and then draw a simple random sample s_m of size m, where m is a realization of a binomial random variable with number of trials N and probability of success nQ/X (the binomial distribution may be replaced appropriately by normal or Poisson). Then we proceed with s_m in the same way as with U except that the numbers $q_1, \cdots, q_N$ are selected in the range from 1 to Q.

10. Acknowledgment. I wish to thank Professor Douglas G. Chapman for his valuable suggestions concerning the structure and intelligibility of the paper.

REFERENCES

[1] GOODMAN, LEO A. (1949). On the estimation of the number of classes in a population. *Ann. Math. Statist.* **20** 572–579.

[2] HÁJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Časopis Pěst. Mat.* **84** 387–423.

[3] HÁJEK, J. (1960). Limiting distributions in simple random sampling from a finite population. *Publ. Math. Inst. Hung. Acad. Sci.* **5** 361–374.

[4] HANSEN, M. H. and HURWITZ, W. N. (1943). On the theory of sampling from finite populations. *Ann. Math. Statist.* **20** 332–362.

[5] HARTLEY, H. O. and RAO, J. N. K. (1962). Sampling with unequal probabilities and without replacement. *Ann. Math. Statist.* **33** 350–374.

[6] HORWITZ, D. G., and THOMPSON, D. J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* **47** 663–685.

[7] LOÈVE, M. (1960). *Probability Theory.* Van Nostrand, Princeton.

[8] MADOW, W. G. (1948). On the limiting distributions of estimates based on samples from finite universes. *Ann. Math. Statist.* **19** 535–545.

[9] NEYMAN, J. (1934). On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.* **97** 558–606.

[10] RAO, J. N. K., HARTLEY, H. O. and COCHRAN, W. G. (1962). On a simple procedure of unequal probability sampling without replacement. *J. Roy. Statist. Soc. Ser. B.* **24** 482–491.

[11] ROY, J. and CHAKRABORTI, I. M. (1960). Estimating the mean of a finite population. *Ann. Math. Statist.* **31** 392–398.

[12] 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.