The Annals of Mathematical Statistics
1971, Vol. 42, No. 3, 1113–1117

LARGE-SAMPLE POSTERIOR DISTRIBUTIONS FOR FINITE POPULATIONS

ALASTAIR SCOTT
University of North Carolina[1]

0. Summary. One of the strongest features of conventional sample survey theory is that very little needs to be assumed about the form of the population distribution. Hartley and Rao (1968) and Ericson (1969) have recently developed a Bayesian approach to sampling from finite populations that shares this feature. In this note, the resulting posterior distribution of the population elements is shown to approach normality for a broad class of prior distributions when the population size N and sample size n increase so that $n \to \infty$ and $N-n \to \infty$. In most cases this leads to the same large-sample interval estimates for population moments as the usual approach invoking the Central Limit Theorem for random sampling from a finite population (Madow (1948), Erdös and Rényi (1958), Hájek (1960)).

1. Introduction. Following Godambe (1955), Hájek (1959), and others let U denote the finite population of N identifiable individuals, $U = \{1, 2, \cdots, N\}$ say, and let S be the set of all subsets $s \subset U$. There is an unknown value Y_i associated with each individual $i \in U$. A subset s is selected from S with probability prescribed by a given sample design (i.e. a given probability measure on S) and the value of Y_i is observed for each $i \in s$. We want to say something about the vector of population values $\mathbf{Y} = (Y_1, Y_2, \cdots, Y_N)$ given the outcome $(s, Y_i; i \in s)$.

The basic assumption made by Hartley and Rao (1968) and Ericson (1969) is that Y_i takes values only in a finite set $\{y_1, y_2, \cdots, y_k\}$. Since k can be made arbitrarily large this simply corresponds to the realities of practice. Any symmetric function of $(Y_1, Y_2, \cdots, Y_N)$ such as a population moment is a function of $\mathbf{N} = (N_1, \cdots, N_k)(\sum_1^k N_j = N)$ where N_j is the number of individuals $i \in U$ with $Y_i = y_j$. Let n_j be the number of sampled individuals $i \in s$ with $Y_i = y_j$. If $Y_1, Y_2, \cdots, Y_N$ are exchangeable random variables, so that every permutation of $Y_1, Y_2, \cdots, Y_N$ has the same prior distribution, it can be shown that the posterior distribution of $\mathbf{N}$ depends on the outcome $\{s, Y_i; i \in s\}$ only through $\mathbf{n} = (n_1, n_2, \cdots, n_k)(\sum_1^k n_j = n)$ and that the conditional density of $\mathbf{n}$ given $\mathbf{N}$ is

$$p(\mathbf{n} \mid \mathbf{N}) = \prod_1^k \binom{N_i}{n_i} \Big/ \binom{N}{n} \tag{1}$$

for any sample design $p(s)$. If the prior distribution is not exchangeable $\mathbf{n}$ is not sufficient for $\mathbf{N}$ and there is additional information in knowing which individuals were sampled and the particular value associated with each. Moreover the conditional distribution of $\mathbf{n}$ given $\mathbf{N}$, and hence any posterior inference based only on $\mathbf{n}$, depends on the sample design $p(s)$. It is common survey practice to deal with

Received January 23, 1970.

[1] On leave from the London School of Economics during 1969–70.

major departures from exchangeability by stratification and, within each stratum, to record only $\mathbf{n}$ or some function of $\mathbf{n}$ such as the sample mean and variance, regarding the loss of information as negligible. The conditional density (1) and the resulting inferences remain valid provided the sample design assigns equal probability to each subset of n individuals (i.e. simple random sampling).

The remaining element is the specification of a prior distribution for $\mathbf{N}$. Ericson (1969) has considered the broad class of exchangeable distributions generated by supposing that $Y_1, \cdots, Y_N$ are independent and identically distributed conditional on some parameter $\boldsymbol{\theta}$ and averaging over the marginal distribution of $\boldsymbol{\theta}$. This leads to a compound multinomial distribution for $\mathbf{N}$ here. In particular, Ericson takes the mixing distribution to be Dirichlet, which is mathematically convenient since the posterior distribution of $\mathbf{N}-\mathbf{n}$ given $\mathbf{n}$ is also of the Dirichlet-multinomial form. Hartley and Rao arrive at the same posterior distribution in a slightly different way. In general, a mixed multinomial distribution with arbitrary weight function, $W(p)$ say, has density

$$f(\mathbf{N}) = \frac{N!}{\prod_1^k N_j!}\int_R \prod_1^k p_j^{N_j}\, dW(\mathbf{p}) \tag{2}$$

where the integral is over the simplex $R = \{\mathbf{p}: p_j \geqq 0, \sum_1^k p_j = 1\}$. When this prior density is combined with the conditional density (1), it follows that the posterior distribution of $\mathbf{N}-\mathbf{n}$ is also of mixed multinomial type with weight function given by

$$dW_n(\mathbf{p}) = \prod_1^k p_j^{n_j}\, dW(\mathbf{p}) / \int_R \prod_1^k p_j^{n_j}\, dW(\mathbf{p}). \tag{3}$$

In the next section we examine the limiting form of the posterior distribution of $\mathbf{N}$ as n and $N-n$ become large with n_j/n approximately constant.

2. Limiting form of the posterior distribution. Following the usual formulation of the central limit theorem for finite populations (Hájek (1960)), we consider a sequence of finite populations indexed by v in which the number of sampled elements, n_v, and the number of unsampled elements, $N_v - n_v$, both increase without bound as $v \to \infty$. If the prior distribution of $\mathbf{N}_v$ has a density of the form (2) the posterior distribution of $\mathbf{N}_v - \mathbf{n}_v$ given $\mathbf{n}_v$ has characteristic function

$$E(\exp(i\textstyle\sum_j (N_{vj}-n_{vj})t_j)) = \int_S [p_k + \sum_{j=1}^{k-1} p_j e^{it_j}]^{N_v - n_v}\, dW_v(\mathbf{p}) \tag{4}$$

where $dW_v(\mathbf{p})$ has the form given in (3). Suppose that $n_{vj} = n_v \pi_j + O(1)$ as $v \to \infty$ (so that π_j is the limiting sample proportion of y_j) and that $W(\mathbf{p})$ is absolutely continuous with density $w(\mathbf{p})$ continuous in a neighbourhood of $\mathbf{p} = \boldsymbol{\pi}$. Let

$$u_{vj} = \frac{n_v^{\frac{1}{2}}(N_{vj} - N_v n_{vj}/n_v)}{N_v^{\frac{1}{2}}(N_v - n_v)^{\frac{1}{2}}}, \qquad \sum_1^k u_{vj} = 0, \qquad i = 1, \cdots, k.$$

THEOREM. *As* $\nu \to \infty$, $\mathbf{u}_\nu = (u_{\nu 1}, u_{\nu 2}, \cdots, u_{\nu k-1})'$ *converges in distribution to a* $(k-1)$*-variate normal distribution with mean zero and covariance matrix* $\Sigma = (\sigma_{ij})$ *where*

$$\sigma_{ij} = \pi_i(1-\pi_i), \qquad i = j$$
$$= -\pi_i\pi_j, \qquad i \neq j.$$

PROOF. From (4), the characteristic function of $\mathbf{u}_\nu$ is

$$Q_\nu(\mathbf{t}) = \int_{S_\nu} A_\nu(\mathbf{y}) B_\nu(\mathbf{y})\, d\mathbf{y} / \int_{S_\nu} A_\nu(\mathbf{y})\, d\mathbf{y} \tag{5}$$

with

$$y_j = n_\nu^{\frac{1}{2}}(p_j - \pi_j), \quad S_\nu = \{\mathbf{y}: \textstyle\sum_1^{k-1} y_j \leqq n_\nu^{\frac{1}{2}}\pi_k, y_j \geqq -n_\nu^{\frac{1}{2}}\pi_j, j = 1, \cdots, k-1\},$$
$$A_\nu(\mathbf{y}) = \textstyle\prod_1^k (1 + n_\nu^{-\frac{1}{2}} y_j/\pi_j)^{n_{\nu j}} w(\boldsymbol{\pi} + n_\nu^{-\frac{1}{2}}\mathbf{y})$$

and

$$B_\nu(\mathbf{y}) = \exp[-in_\nu^{-\frac{1}{2}}(1-f_\nu)^{\frac{1}{2}} \textstyle\sum_1^{k-1} n_{\nu j} t_j][\sum_1^{k-1} (\pi_j + n_\nu^{-\frac{1}{2}} y_j)$$
$$\cdot \exp(in_\nu^{\frac{1}{2}} N_\nu^{-\frac{1}{2}}(N_\nu - n_\nu)^{-\frac{1}{2}} t_j) + \pi_k + n_\nu^{-\frac{1}{2}} y_k]^{N_\nu - n_\nu}$$

where $f_\nu = n_\nu/\mathbf{N}_\nu$.

It follows that

$$A_\nu(\mathbf{y}) = [\textstyle\prod_1^k (1 + n_\nu^{-\frac{1}{2}} y_j/\pi_j)^{\pi_j + O(n_\nu^{-1})}]^{n_\nu} w(\boldsymbol{\pi} + n_\nu^{-\frac{1}{2}}\mathbf{y}),$$
$$= \{1 + [\textstyle\sum_1^{k-1} y_j^2(1-\pi_j^{-1}) + 2\sum\sum_{i<j} y_i y_j - (\sum_1^{k-1} y_j)^2(1+\pi_k^{-1})]$$
$$\div 2n_\nu + O(n_\nu^{-\frac{3}{2}})\}^{n_\nu} w(\boldsymbol{\pi} + n_\nu^{-\frac{1}{2}}\mathbf{y}),$$
$$= w(\boldsymbol{\pi}) \exp[-\mathbf{y}'\Sigma^{-1}\mathbf{y}/2] + \delta_\nu(\mathbf{y}),$$

and

$$B_\nu(\mathbf{y}) = \left\{1 + \frac{i(1-f_\nu)^{\frac{1}{2}} \sum_1^{k-1} y_j t_j}{N_\nu - n_\nu} - \frac{f_\nu[\sum_1^{k-1} \pi_j t_j^2 - ((\sum_1^{k-1} \pi_j t_j)^2/2)]}{N_\nu - n_\nu}\right.$$
$$\left. + \frac{O(n_\nu^{-\frac{1}{2}})}{N_\nu - n_\nu} + \frac{O(N_\nu - n_\nu)^{-\frac{1}{2}}}{N_\nu - n_\nu}\right\}^{N_\nu - n_\nu}$$
$$= \exp[-f_\nu \mathbf{t}'\Sigma\mathbf{t}/2 + i(1-f_\nu)^{\frac{1}{2}} \textstyle\sum_1^{k-1} y_j t_j] + \gamma_\nu(\mathbf{y})$$

where both $\delta_\nu(\mathbf{y})$ and $\gamma_\nu(\mathbf{y})$ converge to zero uniformly in any hypercube $H = \{\mathbf{y}: |y_j| \leqq L, j = 1, \cdots, k-1\}$ as $\nu \to \infty$. In addition, the contribution to the two integrals in (5) from values of $\mathbf{y}$ in H', the complement of H, can be made arbitrarily small simultaneously for all ν by choosing L large enough. For, letting $H_\nu' = \{\mathbf{p}: |p_j - \pi_j| > Ln_\nu^{-\frac{1}{2}}, p_j \geqq 0, \sum_1^k p_j = 1\}$,

$$|\textstyle\int_{H'} A_\nu(\mathbf{y}) B_\nu(\mathbf{y})\, d\mathbf{y}| \leqq \int_{H'} A_\nu(\mathbf{y})\, d\mathbf{y}$$
$$= n_\nu^{(k-1)/2} \textstyle\int_{H_\nu'} \prod_1^k (p_j/\pi_j)^{n_{\nu j}} w(\mathbf{p})\, d\mathbf{p}$$
$$\leqq K_1 n_\nu^{(k-1)/2} \textstyle\prod_1^k \pi_j^{-n_{\nu j}} \int_{H_\nu'} \prod_1^k p_j^{n_{\nu j}-1}\, d\mathbf{p}$$

for some constant K_1 since $w(p)$ is integrable. Using Tchebychev's inequality for a Beta random variable and Stirling's approximation, it follows that this last term is less than or equal to $K_2 L^{-2}$ for some constant K_2.

Therefore

$$\lim_{\nu\to\infty} \int_{S_\nu} A_\nu(\mathbf{y})\,d\mathbf{y} = (2\pi)^{(k-1)/2}|\Sigma|^{\frac{1}{2}}$$

and

$$\begin{aligned}\lim_{\nu\to\infty} \int_{S_\nu} A_\nu(\mathbf{y})B_\nu(\mathbf{y})\,d\mathbf{y} &= \lim_{\nu\to\infty} \exp(-f_\nu \mathbf{t}'\Sigma\mathbf{t}/2) \\ &\quad \cdot\int \exp(i(1-f_\nu)^{\frac{1}{2}} \textstyle\sum_1^{k-1} y_j t_j - \mathbf{y}'\Sigma^{-1}\mathbf{y}/2)\,d\mathbf{y} \\ &= (2\pi)^{(k-1)/2}|\Sigma|^{\frac{1}{2}} \exp(-\mathbf{t}'\Sigma\mathbf{t}/2)\end{aligned}$$

and this completes the proof.

If $k = 2$ and $w(\mathbf{p})$ is a Beta density the posterior distribution of $\mathbf{N}_\nu - \mathbf{n}_\nu$ belongs to Polya's class of contagious distributions. If we add the condition $f_\nu \to f(0 < f < 1)$ the theorem follows from the limiting form of this class given by Polya (1931). An analogous result in the classical frequency sense has been given by Hartley and Rao (1968) for the sample values $n_{\nu j}$ with simple random sampling.

COROLLARY. *If* $\bar{Y}_\nu = \sum_1^k N_{\nu j} y_j / N_\nu$ *is the population mean and* $\bar{y}_\nu = \sum_1^k n_{\nu j} y_j / n_\nu$ *and* $S_\nu^2 = (\sum_1^k n_{\nu j} y_j^2 - n_\nu \bar{y}_\nu^2)/(n_\nu - 1)$ *are the sample mean and variance, then* $n_\nu^{\frac{1}{2}}(\bar{Y}_\nu - \bar{y}_\nu)/(1-f_\nu)^{\frac{1}{2}} S_\nu$ *converges in distribution to a standard normal random variable.*

This gives a direct analogue of the central limit theorem for random sampling from finite populations, though the restriction that $Y_{\nu j}$ takes values only in a finite set means that $Y_{\nu j}$ is bounded, which is much stronger than the Lindberg-type condition imposed by Hájek (1960).

NOTE. If $w(\mathbf{p})$ is discrete with $dW(\mathbf{p}_{(i)}) = w_i (i = 1, \cdots, m)$ let $\mathbf{p}_\pi$ be the value of $\mathbf{p}_{(i)}$ that minimizes

$$D_i^2 = (\mathbf{p}_{(i)} - \boldsymbol{\pi})' \textstyle\sum^{-1} (\mathbf{p}_{(i)} - \boldsymbol{\pi}) \qquad (i = 1, \cdots, m),$$

and let

$$\bar{y}_\pi = \textstyle\sum_1^k p_{\pi j} y_j \quad \text{and} \quad S_\pi^2 = \sum_1^k p_{\pi j} y_j^2 - \bar{y}_\pi^2.$$

Then it follows that

$$\frac{N_\nu^{\frac{1}{2}}(\bar{Y}_\nu - f_\nu \bar{y}_\nu - (1-f_\nu)\bar{y}_\pi)}{(1-f_\nu)^{\frac{1}{2}} S_\pi}$$

converges in distribution to a standard normal random variable. If $f_\nu \to f$ as $\nu \to \infty$ the posterior variance of $\bar{Y}_\nu$ is of order $N_\nu^{-\frac{1}{2}}$ rather than $n_\nu^{-\frac{1}{2}}$.

3. Concluding note. Not all exchangeable distributions for N_ν belong to the class of mixed multinomial distributions. However it is meaningless to consider the limiting form of the posterior distribution if the members of the associated sequence of prior distributions are quite unrelated to each other. It is natural to require at least that the sequence be consistent in the sense that the marginal distribution of a subpopulation formed by choosing $N_{\nu 1}$ of the $N_{\nu 2}$ elements in the ν_2th population at random should be identical to distribution for the ν_1th population. Hald (1960) has called a sequence of distributions with this property reproducible, and shows that for $k = 2$ the condition implies that the sequence is equivalent to a sequence of mixed binomial distributions with constant weight function. The argument extends immediately to $k > 2$, as a consequence of de Finetti's Theorem.

REFERENCES

[1] ERDÖS, P. and RÉNYI, A. (1959). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungarian Acad. Sci.* **4** 49–57.

[2] ERICSON, W. A. (1969). Subjective Bayesian models in sampling finite populations. *J. Roy. Statist. Soc. B* **31** 195–233.

[3] GODAMBE, V. P. (1955). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc. B* **17** 267–278.

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

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

[6] HALD, A. (1960). The compound hypergeometric distribution and a system of single sampling inspection plans based on prior distributions and costs. *Technometrics* **2** 275–340.

[7] HARTLEY, H. O. and RAO, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika* **55** 547–557.

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

[9] POLYA, G. (1931). Sur quelques points de la théorie des probabilités. *Ann. Inst. H. Poincaré* 117–162.