Statistics & Probability Letters 11 (1991) 463–468
North-Holland

June 1991

Large sample distribution of the sample total in a generalized rejective sampling scheme

S. Rao Jammalamadaka, G. Michaletzky * and P. Todorovic

Statistics and Applied Probability Program, University of California, Santa Barbara, CA 93106, USA

Received March 1990

Abstract: The weak convergence of a sample sum, in a generalized rejective sampling from a finite population, to a Poisson and Normal distribution is discussed. The generalization consists in assuming that the elements of the population are random variables, rather than fixed values.

1. Introduction

Consider a finite population of N units, $\{Y_j\}_1^N$, where the value of the jth unit, Y_j, is a non-negative integer. From this, a sample of size n is drawn according to a rejective sampling plan with parameters $p_1, \ldots, p_N$, with $\Sigma_1^N p_j = n$ (see Hajek, 1981, Chapter 7). Here p_j denotes the probability that the jth population unit is included in the sample — this event being represented by the indicator variable, I_j. Let S_{N_n} be the sample sum, then clearly $\mathscr{L}\{S_{N_n}\} = \mathscr{L}\{\Sigma_1^N Y_j I_j \mid \Sigma_1^N I_j = n\}$, where $\{I_j\}_1^N$ is a sequence of independent Bernoulli r.v.'s with $E\{I_j\} = p_j$.

In a recent paper (Praskova, 1985) the weak convergence of S_{N_n} to a Poisson r.v. as N, $n \to \infty$, was discussed in some detail. In this note, we extend and generalize the results of Praskova (1985). First, we show that the Poisson convergence still holds when $\{Y_j\}_1^N$ are non-negative independent integer valued random variables (r.v.'s), independent of $\{I_j\}_1^N$, such that $E\{Y_j\}^k < \infty$, $k = 1, 2$. The randomness of Y_j covers cases such as multistage sampling where Y_j is the value corresponding to the jth primary stage unit. Second, under mild regularity assumptions on $\{Y_j\}_1^N$, we also investigate the weak convergence of the standardized sample sum to a normal distribution.

2. Poisson convergence of the sample sum

Set $P_{N_n} = P\{\Sigma_1^N I_j = n\}$, $f_j(t) = E\{e^{itY_j}\}$ and

$$\varphi_{N_n}(t) = E\left(\exp\left\{it\sum_1^N Y_j I_j\right\} \middle| \sum_1^N I_j = n\right).$$

From a simple argument (see e.g. Holst, 1979, Theorem 1) we have:

$$\varphi_{N_n}(t) = \left(2\pi P_{N_n}\right)^{-1} \int_{-\pi}^{\pi} e^{-isn} \left[\prod_{j=1}^N E\left(\exp\left\{i\left(tY_j + s\right)I_j\right\}\right)\right] ds. \tag{2.1}$$

* Permanent Address: Department of Mathematics, Eötvös University, Budapest, Hungary.

Let Z_N be a Poisson r.v. with $E\{Z_N\} = \Sigma_1^N E(Y_j)p_j$, then its characteristic function $g_N(t)$ is

$$g_N(t) = \exp\left\{(\mathrm{e}^{\mathrm{i}t} - 1)\sum_1^N p_j E(Y_j)\right\}. \tag{2.2}$$

Our aim is to evaluate $|\varphi_{N_m}(t) - g_N(t)|$, which we do in Proposition 2.1. Throughout this paper, we assume that $p_j < 0.5$. Set $q_j = 1 - p_j$, then

$$E\left(\exp\{\mathrm{i}(tY_j + s)I_j\}\right) = \left(q_j + p_j\,\mathrm{e}^{\mathrm{i}s} f_j(t)\right). \tag{2.3}$$

This and (2.1) yield

$$\varphi_{N_n}(t) = \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}sn}\left[\prod_1^N \left(q_j + p_j\,\mathrm{e}^{\mathrm{i}s}\right)\left(q_j(s) + p_j(s)f_j(t)\right)\right]\mathrm{d}s \tag{2.4}$$

where $q_j(s) = q_j/(q_j + p_j\,\mathrm{e}^{\mathrm{i}s})$ and $p_j(s) = 1 - q_j(s)$. Since

$$1 = \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}sn} E\left(\exp\left\{\mathrm{i}s\sum_1^N I_j\right\}\right)\mathrm{d}s$$

it follows, using a well known identity for products, that

$$\begin{aligned}
\varphi_{N_n}(t) - g_N(t) &= \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}sn}\prod_1^N \left(q_j + p_j\,\mathrm{e}^{\mathrm{i}s}\right) \\
&\qquad\times\left\{\prod_1^N \left(q_j(s) + p_j(s)f_j(t)\right) - \prod_1^N \mathrm{e}^{p_j(\mathrm{e}^{\mathrm{i}t}-1)E(Y_j)}\right\}\mathrm{d}s \\
&= \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}sn}\prod_1^N \left(q_j + p_j\,\mathrm{e}^{\mathrm{i}s}\right) \\
&\qquad\times\left\{\sum_{j=1}^N \left[\prod_{l=1}^{j-1} \left(q_l(s) + p_l(s)f_l(t)\right)\right]\left[\prod_{k=j+1}^N \mathrm{e}^{p_k(\mathrm{e}^{\mathrm{i}t}-1)E(Y_k)}\right]\right. \\
&\qquad\left.\times\left[q_j(s) + p_j(s)f_j(t) - \mathrm{e}^{p_j(\mathrm{e}^{\mathrm{i}t}-1)E(Y_j)}\right]\right\}\mathrm{d}s.
\end{aligned} \tag{2.5}$$

To evaluate the absolute value of the difference in (2.5), the following lemma is needed.

Lemma 2.1.

$$\begin{aligned}
&\left|q_j(s) + p_j(s)f_j(t) - \mathrm{e}^{p_j(\mathrm{e}^{\mathrm{i}t}-1)E(Y_j)}\right| \\
&\quad\leqslant |\mathrm{e}^{\mathrm{i}t} - 1|^2\left[\frac{p_j EY_j(Y_j - 1)}{2(q_j - p_j)} + \frac{2p_j q_j|\sin\frac{1}{2}s|\,E(Y_j)}{(q_j - p_j)\,|\mathrm{e}^{\mathrm{i}t} - 1|} + \tfrac{1}{2}\left(p_j E(Y_j)\right)^2\right].
\end{aligned} \tag{2.6}$$

Proof. First, write the left hand side of (2.6) as follows:

$$\Big| p_j(s)\big[f_j(t) - 1 - E(Y_j)(e^{it} - 1)\big] + \big(p_j(s) - p_j\big)(e^{it} - 1)E(Y_j) - \Big(e^{p_j(e^{it}-1)E(Y_j)} - 1 - p_j(e^{it} - 1)E(Y_j)\Big)\Big|. \tag{2.7}$$

Clearly,

$$|p_j(s)| \leqslant \frac{p_j}{q_j - p_j}, \qquad |p_j(s) - p_j| \leqslant \frac{2p_j q_j}{q_j - p_j}|\sin \tfrac{1}{2}s|,$$

$$\begin{aligned}\big|f_j(t) - 1 - E(Y_j)(e^{it} - 1)\big| &= \left|\sum_{k=0}^{\infty} e^{ikt}P\{Y_j = k\} - 1 - (e^{it} - 1)\sum_{k=0}^{\infty} kP\{Y_j = k\}\right| \\ &\leqslant |e^{it} - 1| \sum_{k=0}^{\infty}\left|\left[\frac{e^{ikt} - 1}{e^{it} - 1} - k\right]\right| P\{Y_j = k\} \\ &\leqslant \tfrac{1}{2}|e^{it} - 1|^2 EY_j(Y_j - 1).\end{aligned}$$

Finally,

$$\Big|e^{p_j(e^{it}-1)E(Y_j)} - 1 - p_j(e^{it} - 1)E(Y_j)\Big| \leqslant \tfrac{1}{2}|e^{it} - 1|^2\big(p_j E(Y_j)\big)^2.$$

This proves the lemma. □

Set

$$A = \sum_1^N \frac{p_j E(Y_j)}{q_j - p_j}, \qquad B = \sum_1^N\left[\frac{p_j EY_j(Y_j - 1)}{q_j - p_j} + \big(p_j E(Y_j)\big)^2\right],$$

$$C = \sum_1^N \frac{p_j q_j}{q_j - p_j} E\{Y_j\}, \qquad d = \sum_1^N p_j q_j.$$

Then we have:

Proposition 2.1.

$$\big|\varphi_{N_n}(t) - g_N(t)\big| \leqslant \alpha\, e^{|e^{it}-1|A} \cdot \{\beta|e^{it} - 1|B + \gamma|e^{it} - 1|d^{-1/2}C\} \tag{2.8}$$

where α, β *and* γ *are positive constants.*

Proof. First, we have the following three inequalities:

(i) $$\begin{aligned}|q_l(s) + p_l(s)f_l(t)| &= |1 + (f_l(t) - 1)p_l(s)| \\ &\leqslant 1 + |e^{it} - 1|\, p_l(s)E(Y_l) \leqslant \exp\{p_l E(Y_l)\,|e^{it} - 1|/(q_l - p_l)\}.\end{aligned}$$

(ii) $$\big|e^{p_k(e^{it}-1)E(Y_k)}\big| \leqslant e^{p_k|e^{it}-1|E(Y_k)/(q_k - p_k)}.$$

(iii) $$\big|q_j + p_j e^{is}\big| \leqslant e^{-2p_j q_j \sin^2(s/2)}.$$

This and (2.5) yield:

$$\begin{aligned}&\big|\varphi_{N_n}(t) - g_N(t)\big| \\ &\quad\leqslant (2\pi P_{N_n})^{-1} e^{|e^{it}-1|A}\int_{-\pi}^{\pi} e^{-2d\sin^2(s/2)}\big[\tfrac{1}{2}B|e^{it} - 1|^2 + C|e^{it} - 1|\,|\sin \tfrac{1}{2}s|\big]\, ds.\end{aligned}$$

Since

$$\int_0^{\pi} e^{-2d\sin^2(s/2)}\, ds \le \left[\tfrac{1}{2}\pi\right]^{3/2} d^{-1/2}, \qquad \int_0^{\pi} e^{-2d\sin^2(s/2)} \sin\tfrac{1}{2}s\, ds \leqslant (1-e^{-2d})/d,$$

and $P_{N_n} \geqslant k \cdot d^{1/2}$, where $k>0$ is a constant, the proposition follows. □

An important consequence of the proposition is the Poisson convergence of the sample sum, stated formally in Corollary 2.1. Set ($\alpha = \alpha(N)$ is assumed to be a bounded sequence)

$$\alpha^{-1} = \min_{1 \leqslant j \leqslant N} (q_j - p_j).$$

Then clearly

$$A \leqslant \alpha \sum_1^N p_j E\{Y_j\},$$

$$B \leqslant \alpha \sum_1^N p_j E Y_j (Y_j - 1) + \max_{1 \leqslant k \leqslant N} (p_k E(Y_k)) \sum_1^N p_j E(Y_j),$$

$$C \leqslant \alpha \sum_1^N p_j E\{Y_j\}.$$

In addition, since $\Sigma_1^N p_j = n$ and $p_j < 0.5$ it follows that $d \geqslant 2n$.

Corollary 2.1. *If we set $I_j = I_{N_j}$, $p_j = p_{N_j}$, and assume that as $N \to \infty$, $n \to \infty$,*

$$\max_{1 \leqslant j \leqslant N} p_{N_j} E(Y_j) \to 0, \qquad \sum_1^N p_{N_j} E Y_j (Y_j - 1) \to 0$$

and

$$\sum_1^N p_{N_j} E(Y_j) \to \lambda,$$

then $\varphi_{N_n}(t) \to g(t)$, where $g(t) = e^{(e^{it}-1)\lambda}$. This implies that

$$\mathscr{L}\left\{ \sum_1^N Y_j I_{Ni} \,\middle|\, \sum_1^N I_{N_i} = n \right\} \to \mathscr{L}(Z)$$

where Z is a Poisson r.v. with $E\{Z\} = \lambda$. □

3. Convergence to a normal distribution

The result of this section can be formulated as:

Proposition 3.1. *Let $\{Y_j\}_1^\infty$ be independent real-valued r.v.'s such that $E\{Y_j\} = 0$ and* $\operatorname{Var}\{Y_j\} = 1$, $j = 1, 2, \ldots$, *independent of $\{I_j\}_1^\infty$. Then*

$$S_{N_n}/\sqrt{n} \xrightarrow{\mathscr{L}} Z$$

where $Z \sim N\ (0.1)$, *or equivalently* $\Psi_{N_n}(t) \to \phi(t)$, *uniformly where*

$$\Psi_{N_n}(t) = E\left(\exp\left\{\frac{it}{\sqrt{n}}\sum_1^N Y_j I_j \middle| \sum_1^N I_j = n\right\}\right) \quad \textit{and} \quad \phi(t) = e^{-t2/2}.$$

Proof. The method of proof is the one used in the previous proposition. Write

$$\Psi_{N_n}(t) = \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} e^{-isn}\prod_1^N\left(q_j + p_j\, e^{is} f_j\left(\frac{t}{\sqrt{n}}\right)\right) ds.$$

Then, we have

$$\begin{aligned}\Psi_{N_n}(t) - \phi(t) &= \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} e^{-isn}\left[\prod_1^N\left(q_j + p_j\, e^{is}\right)\right] \\ &\qquad\times\left[\prod_1^N\left(q_j(s) + p_j(s) f_j\left(\frac{t}{\sqrt{n}}\right)\right) - \prod_1^N e^{-(t^2/2n)p_j}\right] ds \\ &= \left(2\pi P_{N,n}\right)^{-1}\int_{-\pi}^{\pi} e^{-isn}\left[\prod_1^N\left(q_j + p_j\, e^{is}\right)\right] \\ &\qquad\times\left\{\sum_{j=1}^N\left[\prod_{k=1}^{j-1}\left(q_k(s) + p_k(s) f_k\left(\frac{t}{\sqrt{n}}\right)\right)\right]\left[\prod_{k=j+1}^N e^{-(t^2/2n)p_k}\right]\right. \\ &\qquad\qquad\left.\times\left(q_j(s) + p_j(s) f_j\left(\frac{t}{\sqrt{n}}\right) - e^{-(t^2/2n)p_j}\right)\right\} ds.\end{aligned}$$

But

$$\begin{aligned}&\left|q_j(s) + p_j(s) f_j\left(\frac{t}{\sqrt{n}}\right) - e^{-(t^2/2n)p_j}\right| \\ &\quad= \left|p_j(s)\left(f_j\left(\frac{t}{\sqrt{n}}\right) - 1\right) + \left(1 - e^{-(t^2/2n)p_j}\right)\right| \\ &\quad= \left|p_j(s)\left(f_j\left(\frac{t}{\sqrt{n}}\right)\right) + \left(p_j - p_j(s)\right)\frac{t^2}{2n} + 1 - \frac{t^2}{2n}p_j - e^{-(t^2/2n)p_j}\right| \\ &\quad\leqslant \frac{p_j}{q_j - p_j}o\left(\frac{1}{n}\right) + \frac{2p_j q_j\,|\sin\frac{1}{2}s|}{q_j - p_j}\frac{t^2}{2n} + o\left(\frac{1}{n}\right)\frac{t^2}{2}p_j,\end{aligned}$$

$$\begin{aligned}\left|q_k(s) + p_k(s) f_k\left(\frac{t}{\sqrt{n}}\right)\right| &= \left|1 + p_k(s)\left(f_k\left(\frac{t}{\sqrt{n}}\right) - 1\right)\right| \\ &\leqslant \exp\left\{|p_k(s)|\cdot\left|f_k\left(\frac{t}{\sqrt{n}}\right) - 1\right|\right\} \leqslant \exp\left\{\frac{p_k}{q_k - p_k}o\left(\frac{t^2}{2n}\right)\right\},\end{aligned}$$

and

$$\left|q_j + p_j\, e^{is}\right| \leqslant \exp\left\{2p_j q_j \sin^2\tfrac{1}{2}s\right\}.$$

From this, we obtain that

$$|\Psi_{N,n}(t)-\phi(t)| \leqslant \left(2\pi P_{N_n}\right)^{-1}\int_{-\pi}^{\pi} \mathrm{e}^{-(\Sigma_1^N q_j p_j)\sin^2(s/2)-(t^2/2n)\Sigma_1^N(p_k(q_k-p_k))+\mathrm{o}(1)}$$

$$\times\left[\mathrm{o}(1)+\frac{t^2}{2n}\left(\sum_1^N q_j p_j\right)\sin\tfrac{1}{2}s+\mathrm{o}(1)\right]\mathrm{d}s$$

$$\leqslant d^{-1/2}\left[\frac{\mathrm{o}(1)}{\sqrt{d}}+\frac{t^2}{2n}\right]=\mathrm{o}(1)+\frac{c}{\sqrt{d}}\to 0$$

which proves the assertion. □

References

Hajek, J. (1981), *Sampling from a Finite Population* (Dekker, New York).

Holst, L. (1979), Two conditional limit theorems with applications, *Ann. Statist.* 7, 551–557.

Praskova, Z. (1985), The convergence to the Poisson distribution in rejective sampling from a finite population, in: W. Grosman, J. Mogyorodi, I. Vincze and W. Wertz, eds., *Proc. 5th Panonian Symp. on Math. Statist., Visegrad Hungary, 1985.*