On the asymptotic error of a bivariate normal approximation with an application to simple random sampling

Paul Knottnerus

Discussion paper (09020)

Statistics Netherlands

The Hague/Heerlen, 2009

Explanation of symbols

.	= data not available
*	= provisional figure
x	= publication prohibited (confidential figure)
–	= nil or less than half of unit concerned
–	= (between two figures) inclusive
0 (0,0)	= less than half of unit concerned
blank	= not applicable
2007-2008	= 2007 to 2008 inclusive
2007/2008	= average of 2007 up to and including 2008
2007/'08	= crop year, financial year, school year etc. beginning in 2007 and ending in 2008
2005/'06–2007/'08	= crop year, financial year, etc. 2005/'06 to 2007/'08 inclusive

Due to rounding, some totals may not correspond with the sum of the separate figures.

Publisher
Statistics Netherlands
Henri Faasdreef 312
2492 JP The Hague

Prepress
Statistics Netherlands - Facility Services

Cover
TelDesign, Rotterdam

Information
Telephone .. +31 88 570 70 70
Telefax .. +31 70 337 59 94
Via contact form: www.cbs.nl/information

Where to order
E-mail: verkoop@cbs.nl
Telefax .. +31 45 570 62 68

Internet
www.cbs.nl

ISSN: 1572-0314

6008309020 X-10

On the asymptotic error of a bivariate normal approximation with an application to simple random sampling

Paul Knottnerus

Summary

This paper gives a formula for the limiting error of the central limit theorem for the bivariate case. Insight in this type of error simplifies the proofs of central limit theorems in probability sampling from finite populations.

Keywords: Central limit theorem; Characteristic function; Error bivariate normal approximation; Fourier transform; Simple random sampling.

1. Introduction

Among others, Feller (1971, page 538) gives a formula for the approximation error of the central limit theorem (CLT) for the normalized sum of N independent random variables x_k. The main aim of this paper is to derive a similar formula for the approximation error of the CLT when the x_k are mutually independent two-dimensional random vectors. To the author's best knowledge this kind of error formulas for multivariate cases are not readily found in the literature. As expected, the error of a bivariate normal approximation appears to be of the same small order as that of the univariate normal approximation. These error formulas are helpful for simplifying proofs of the central limit theorems in random sampling from finite populations.

The outline of the paper is as follows. Section 2 gives a formula for the CLT approximation error for the bivariate case when the two elements in x_k are *independent.* Section 3 derives the error formula for the more complicated case that the two elements in x_k are *dependent.* Based on the results of section 3, section 4 gives a relatively short proof of one of the central limit theorems in random sampling from finite populations. For more examples the reader is referred to Knottnerus (2008). For other CLT proofs in probability sampling, see Madow (1948), Erdös and Rényi (1959), and Hájek (1960 and 1964).

2. Bivariate approximation error for two independent variables

2.1 General remarks and notation

Let the x_k be independent and identically distributed two-dimensional random vectors ($k = 1, ..., N$). Denote their distribution function by $F(x) = F(x_1, x_2)$ and the corresponding marginal distribution functions by $F_1(x_1)$ and $F_2(x_2)$. To restrict the notational burden we assume that the elements in x_k, say x_{1k} and x_{2k}, have a zero expectation and a unit variance. In addition, x_{1k} and x_{2k} might be dependent. However, for our purposes it suffices to assume that they have a zero correlation. Furthermore, it is assumed that all fourth moments exist. Let $F_N(x)$ denote the joint distribution function of $\sqrt{N}\overline{x}_{1N}$ and $\sqrt{N}\overline{x}_{2N}$; $F_{1N}(.)$ and $F_{2N}(.)$ refer to the corresponding marginal distributions. Let $\Phi(u)$ denote the standard normal distribution function and $\varphi(u)$ its derivative. In the remainder the notation "$A \sim B$" is used to indicate that A/B tends to unity as $N \to \infty$.

2.2 Bivariate approximation error for two independent variables

In this subsection we look at the error formula for the particular case that x_{1k} and x_{2k} are independent so that $F_N(x)$ can be written as

$$F_N(x) = F_{N,ind}(x) = F_{1N}(x_1)F_{2N}(x_2) \sim \Phi(x_1)\Phi(x_2)$$

Using the error formula mentioned by Feller (1971, page 538) for the univariate case

$$\begin{aligned} F_{lN}(x_l) &= \Phi(x_l) + \omega_{lN}(x_l) + o(1/\sqrt{N}) \qquad (l = 1, 2) \\ \omega_{lN}(x_l) &= \frac{\mu_l^{(3)}}{6\sqrt{N}}(1 - x_l^2)\varphi(x_l) \\ \mu_l^{(3)} &= E(x_{lk}^3) \end{aligned} \tag{1}$$

it is seen that the approximation error of the bivariate normal distribution now becomes

$$\begin{aligned} & F_{1N}(x_1)F_{2N}(x_2) - \Phi(x_1)\Phi(x_2) \\ &= \omega_{N,ind}(x) + o(1/\sqrt{N}) \\ \omega_{N,ind}(x) &= \omega_{1N}(x_1)\Phi(x_2) + \omega_{2N}(x_2)\Phi(x_1) \end{aligned} \tag{2}$$

When F_{lN} is a lattice distribution ($l = 1, 2$), (1) is still true provided x_l is a midpoint of the lattice for F_{lN}. Similar results can be derived when the variances depend on k; see Feller (1971, pages 538-48).

3. Bivariate approximation error for two dependent variables

When x_{1k} and x_{2k} are dependent but uncorrelated, it is convenient to decompose the error of the bivariate normal approximation according to

$$\{F_N(x) - F_{N,ind}(x)\} + \{F_{N,ind}(x) - \Phi(x_1)\Phi(x_2)\} = \omega_N(x) + o(1/\sqrt{N}) \tag{3}$$

where $F_{N,ind}(x) = F_{1N}(x_1)F_{2N}(x_2)$ as before and

$$\omega_N(x) = \omega_{N,dep}(x) + \omega_{N,ind}(x) \tag{4}$$

$$\omega_{N,dep}(x) = \frac{\mu_{12}^{21}x_1 + \mu_{12}^{12}x_2}{2\sqrt{N}} \frac{\exp(-x'x/2)}{2\pi} \tag{5}$$

$\omega_{N,ind}(x)$ is given by (2) and $\mu_{12}^{ij} = E(x_{1k}^i x_{2k}^j)$.

The proof of (4) consists of three steps. The first step starts with the introduction of a function $G(x) = G(x_1, x_2)$. Let $G(x)$ be such that $[F(x) - G(x)] \to 0$ as $x \to \pm\infty$, where F is an arbitrary two-dimensional distribution function. Furthermore, we assume that there exists an m so that

$$|G(x) - G(z)| \leq m \|x - z\| + o(1/\sqrt{N}) \tag{6}$$

for every pair of vectors $x, z \in R^2$. If G depends on N, a possible rest term should be $o(1/\sqrt{N})$. This is somewhat different from the (one-dimensional) case described by Feller (1971) where G has a bounded derivative and is independent of N. Now we prove that for the difference between $F(x)$ and $G(x)$ it holds that for all $T > 0$

$$\eta \leq 3\eta_T + \frac{96m}{\pi T} + o(1/\sqrt{N}) \tag{7}$$

where

$$\begin{aligned}
\eta &= \sup |\Delta(x)| \\
\Delta(x) &= F(x) - G(x) \\
\eta_T &= \sup \left|{}^T\Delta(z)\right| \\
{}^T\Delta(z) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Delta(z - x) v_T(x) dx \qquad (8) \\
v_T(x) &= \frac{[1 - \cos(Tx_1)]\,[1 - \cos(Tx_2)]}{\pi^2 T^2 x_1^2 x_2^2}
\end{aligned}$$

The characteristic function of $v_T(x)$, say $\xi_T(t) = \xi_{1T}(t_1)\xi_{2T}(t_2)$, is zero for $\max(|t_1|, |t_2|) \geq T$; see Feller (1971, pages 503 and 536). Assume that at $x = x_0$ it holds that $|\Delta(x_0)| = \eta$. We may assume that $\Delta(x_0) = \eta$. Since F does not decrease, it follows from (6) that

$$\Delta(x_0 + s) \geq \eta - (s_1 + s_2)m + o(1/\sqrt{N}) \qquad (s > 0)$$

Define $h = \eta/3m$ and $z_0 = x_0 + d$ where $d = (h, h)'$. Since

$$\begin{aligned}
\Delta(x_0 + d) &\geq \eta - 2hm + o(1/\sqrt{N}) \\
&= \eta/3 + o(1/\sqrt{N})
\end{aligned}$$

we have for $-d \leq x \leq d$

$$\Delta(z_0 - x) \geq \frac{\eta}{3} + mx_1 + mx_2 + o(1/\sqrt{N}) \tag{9}$$

Furthermore, for $\max(|x_1|, |x_2|) \geq h$ it holds that $\Delta(z_0 - x) \geq -\eta$ while the corresponding mass contributed by $v_T(x)$ to the domain with $\max(|x_1|, |x_2|) \geq h$ is smaller than $2P(|x_1| \geq h; v_T) \leq 8/\pi Th$. Hence, we obtain for the convolution integral in (8) at $z = z_0$

$$\begin{aligned}\eta_T &\geq {}^T\Delta(z_0) \geq \frac{\eta}{3}(1 - \frac{8}{\pi Th}) - \eta\frac{8}{\pi Th} + o(1/\sqrt{N}) \\ &= \frac{\eta}{3} - \frac{32m}{\pi T} + o(1/\sqrt{N})\end{aligned}$$

from which (7) follows. Note that the contribution of the linear part in (9) to (8) is zero for reasons of symmetry.

In the second step it is shown that

$$\eta_T \leq \frac{1}{(2\pi)^2} \iint_{-T}^{T} \left| \frac{\chi(t) - \vartheta(t)}{t_1 t_2} \right| dt \tag{10}$$

where $\chi(t)$ and $\vartheta(t)$ are the Fourier-Stieltjes transforms of F and G, respectively. In analogy with the convolution ${}^T\Delta = V_T \star \Delta$ define

$$\begin{aligned}{}^T F &= V_T \star F \\ {}^T G &= V_T \star G\end{aligned}$$

The Fourier-Stieltjes transforms of these convolutions are $\chi(t)\xi_T(t)$ and $\vartheta(t)\xi_T(t)$, respectively. Hence, by the Fourier inversion theorem, the difference between the density ${}^T f$ and its counterpart from G becomes

$${}^T f - {}^T g = \frac{1}{(2\pi)^2} \iint_{-T}^{T} \exp(-it'x)\{\chi(t) - \vartheta(t)\}\xi_T(t) dt$$

Integrating with respect to x_1 and x_2, we obtain

$${}^T\Delta(x) = \frac{1}{(2\pi)^2} \iint_{-T}^{T} \exp(-it'x) \frac{\chi(t) - \vartheta(t)}{-t_1 t_2} \xi_T(t) dt \tag{11}$$

from which (10) follows. It is assumed here that the integrand in (11) is a continuous function in the domain of integration. Later this assumption should be verified; see the comments below (15). Moreover, similar to the left-hand side the right-hand side of (11) tends to zero under this assumption as $x \to \pm\infty$ because of the Riemann-Lebesgue Lemma; see Feller (1971, page 513). Combining (7) and (10) gives

$$|F(x) - G(x)| \leq \iint_{-T}^{T} \left| \frac{\chi(t) - \vartheta(t)}{t_1 t_2} \right| dt + \frac{96m}{\pi T} + o(\frac{1}{\sqrt{N}}) \tag{12}$$

In the third step we prove (5) or, equivalently,

$$F_N(x) - F_{N,ind}(x) - \omega_{N,dep}(x) = o(1/\sqrt{N}) \tag{13}$$

Since

$$F_{N,ind}(x) - F_{N,ind}(z) = \Phi(x_1)\Phi(x_2) - \Phi(z_1)\Phi(z_2)$$

$$+\omega_{N,ind}(x) - \omega_{N,ind}(z) + o(1/\sqrt{N})$$

where all first-order derivatives on the right-hand side are bounded, there exists an m so that

$$|F_{N,ind}(x) - F_{N,ind}(z)| \leq \frac{m}{2}\|x - z\| + o(1/\sqrt{N})$$

for every pair of vectors $x, z \in R^2$. Hence, $G(x) = F_{N,ind}(x) + \omega_{N,dep}(x)$ satisfies condition (6). Furthermore, let $\gamma_N(t)$ denote the Fourier transform of the second-order derivative $\partial^2\omega_{N,dep}(x)/\partial x_1\partial x_2$. That is,

$$\begin{aligned}
\gamma_N(t) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(it'x)\frac{\partial^2\omega_{N,dep}(x)}{\partial x_1\partial x_2}dx \\
&= \frac{t_1 t_2 \beta(t)}{\sqrt{N}}\exp(-t't/2) \qquad (14)\\
\beta(t) &= \frac{i^3(\mu_{12}^{21}t_1 + \mu_{12}^{12}t_2)}{2}
\end{aligned}$$

For a proof of (14), see the end of this section.

In order to prove (13), choose $T = a\sqrt{N}$ where a is so large that $96m < \varepsilon\pi a$; ε is an arbitrary positive number. According to (12) we now have to show that

$$\int\int_{|t_1|,|t_2|<a\sqrt{N}} \left|\frac{\chi^N(t/\sqrt{N}) - \chi_{ind}^N(t/\sqrt{N}) - \gamma_N(t)}{t_1 t_2}\right| dt + \frac{\varepsilon}{\sqrt{N}} = o(\frac{1}{\sqrt{N}}) \qquad (15)$$

where $\chi_{ind}(t)$ is the characteristic function of $F_{ind}(x) = F_1(x_1)F_2(x_2)$. Note that the power series of the difference $(\chi^N - \chi_{ind}^N)$ consists of terms $t_1^k t_2^l$ $(k, l \geq 1$ and $k + l \geq 3)$. Hence, the integrand is a continuous function at $t_1 = 0$ and at $t_2 = 0$ so that no problem of convergence arises. Using the triangle inequality, we obtain an upper bound for the integral in (15)

$$\int\int_{|t_1|,|t_2|<a\sqrt{N}} \left|\chi_{ind}^N(\frac{t}{\sqrt{N}})\right| \left|\frac{\exp\{N\theta(t/\sqrt{N})\} - 1 - t_1 t_2\beta(t)/\sqrt{N}}{t_1 t_2}\right| dt \qquad (16)$$

$$+ \int\int_{|t_1|,|t_2|<a\sqrt{N}} \left|\chi_{ind}^N(\frac{t}{\sqrt{N}}) - \exp(-t't/2)\right| \left|\frac{\beta(t)}{\sqrt{N}}\right| dt \qquad (17)$$

where

$$\theta(t) = \log\chi(t) - \log\chi_{ind}(t)$$

Now partitioning the domain of integration in (16) and (17) into subdomain 1 with $\max(|t_1|, |t_2|) > \delta\sqrt{N}$ $(0 < \delta < a)$ and subdomain 2 with $\max(|t_1|, |t_2|) \leq \delta\sqrt{N}$, it can be shown in analogy with the univariate case in Feller (1971, page 539-41) that the contributions from subdomain 1 to integrals (16) and (17) tend to zero faster than any power of $1/N$ for any $\delta \in (0, a)$. For instance, under the assumption that F_2 is not a lattice distribution $\max\{|\chi_2(t_2)|\}$ is strictly less than 1 for $\delta < t_2 < a$. Hence, χ_{ind}^N $(= \chi_1^N\chi_2^N)$ tends to zero faster than any power of $1/N$. When F_2 is a lattice distribution, the whole analysis should be based on $F_N^{\#}$ and $F_{2N}^{\#}$ defined by

$$F_N^{\#}(x) = \frac{1}{d_{2N}} \int_{-d_{2N}/2}^{d_{2N}/2} F_N(x_1, x_2 - y)dy$$
$$F_{2N}^{\#}(x_2) = \frac{1}{d_{2N}} \int_{-d_{2N}/2}^{d_{2N}/2} F_{2N}(x_2 - y)dy,$$

respectively, where d_{2N} is the span of the lattice for F_{2N}. By construction, it holds for a midpoint x_2 of the lattice that $F_{2N}^{\#}(x_2) = F_{2N}(x_2)$ and $F_N^{\#}(x_1, x_2) = F_N(x_{1,}x_2)$. For a proof that for $\delta \leq t_2 \leq a$ the corresponding characteristic function $\chi_{2N}^{\#}(t_2)$ tends to zero faster than any power of $1/N$, see Feller (1971, pages 540-1).

Now we examine the contribution to integral (16) from subdomain 2. Let $D_{12}^{jk}(.)$ denote $\partial^{j+k}(.)/\partial t_1^j \partial t_2^k$. It holds that $\theta(0) = 0$, $D_{12}^{k0}\theta(0) = D_{12}^{0k}\theta(0) = 0$ for all $k \geq 1$, $D_{12}^{11}\theta(0) = 0$ while $D_{12}^{21}\theta(0) = i^3\mu_{12}^{21}$ and $D_{12}^{12}\theta(0) = i^3\mu_{12}^{12}$. Using a third-order Taylor expansion, it is seen that

$$\left|\theta(t) - \frac{i^3(\mu_{12}^{21}t_1^2t_2 + \mu_{12}^{12}t_1t_2^2)}{2}\right| \leq \varepsilon(\left|t_1^2t_2\right| + \left|t_1t_2^2\right|) \tag{18}$$

for $|t_1|, |t_2| < \delta$ when δ is sufficiently small. In fact, we used a first-order Taylor expansion of the related power series of $h(t) = \{\exp(\theta) - 1\}/t_1t_2$ with $h(0) = 0$, and the fact that $\theta \approx t_1t_2h$. This explains the absence of the terms t_1^3 and t_2^3 on the right-hand side of (18). In addition, δ is chosen so small that for $t't < \delta$

$$|\theta(t)| \leq \frac{t't}{4} \tag{19}$$
$$|t_1t_2\beta(t)| \leq \frac{t't}{4} \tag{20}$$

Multiplying (18)-(20) throughout by N and replacing t by $t/\sqrt{N}$, it follows from the inequality mentioned by Feller (1971, page 534)

$$|\exp(\alpha) - 1 - \beta| \leq (|\alpha - \beta| + \left|\beta^2\right|/2)\exp\{\max(|\alpha|, |\beta|)\} \tag{21}$$

that the integrand in (16) in subdomain 2 is smaller than

$$\left|\chi_{ind}^N(t/\sqrt{N})\right| \left(\frac{\varepsilon(|t_1| + |t_2|)}{\sqrt{N}} + \frac{|t_1t_2|\{\beta(t)\}^2}{2N}\right)\exp(t't/4) \tag{22}$$
$$\leq \exp(-t't/8)\left(\frac{\varepsilon(|t_1| + |t_2|)}{\sqrt{N}} + \frac{|t_1t_2|\{\beta(t)\}^2}{2N}\right)$$

provided δ is sufficiently small. Since ε is arbitrary, the contribution to (16) from subdomain 2 is $o(1/\sqrt{N})$. Note that in the last inequality we used that for $N \to \infty$

$$\left|\chi_{ind}^N(\frac{t}{\sqrt{N}})\right| \sim \exp(-t't/2) \leq \exp(-3t't/8). \tag{23}$$

Likewise, the contribution to integral (17) from subdomain 2 is $o(1/\sqrt{N})$ because the integrand in (17) can be written as

$$\left|\frac{\chi_{ind}^N(t/\sqrt{N})}{\exp(-t't/2)} - 1\right| \frac{|\beta(t)|}{\sqrt{N}\exp(t't/2)} \leq \exp(-t't/4)\frac{\varepsilon t't}{2}\frac{|\beta(t)|}{\sqrt{N}}$$

provided that δ is small enough. For the last inequality we used that there exists a δ so that for $t't < \delta$

$$\left|\frac{\chi_{ind}^N(t/\sqrt{N})}{\exp(-t't/2)} - 1\right| \leq \frac{\varepsilon t't}{2}\exp(t't/4) \tag{24}$$

In order to prove this, define $\varepsilon_0 = \min(\varepsilon, 1/4)$ and

$$\psi(t) = \log \chi_{ind}(t) + t't/2$$

Since $\psi(0) = 0$, $D_{12}^{k0}\psi(0) = D_{12}^{0k}\psi(0) = 0$ $(k = 1, 2)$ and $D_{12}^{11}\psi(t) = 0$ for any t, we obtain for $t't < \delta$, using a second-order Taylor expansion, $|\psi(t)| \leq \varepsilon_0 t't/2$ provided that δ is small enough. Now applying (21) with $\alpha = N\psi(t/\sqrt{N})$ and $\beta = 0$ we obtain (24). Since (16) and (17) are $o(1/\sqrt{N})$, we have proved (5), (13), and (15).

We conclude this section with a proof of (14). Let I_2 denote the two-dimensional identity matrix. Recall from statistics theory that $\exp(-t't/2)$ is the characteristic function of the bivariate normal distribution $N(0, I_2)$. That is,

$$\begin{aligned} \iint_{-\infty}^{\infty} h(t,x)dx &= \exp(-t't/2) \\ h(t,x) &= \exp(it'x)\frac{\exp(-x'x/2)}{2\pi} \end{aligned} \tag{25}$$

Premultiplying both sides of (25) by $-i^3$ and taking the derivative of both sides with respect to t_2, we obtain

$$-\iint_{-\infty}^{\infty} x_2 h(t,x)dx = i^3 t_2 \exp(-t't/2) \tag{26}$$

Now taking on both sides the derivative $\partial^2(.)/\partial^2 t_1$ yields

$$\iint_{-\infty}^{\infty} x_1^2 x_2 h(t,x)dx = -i^3(t_2 - t_1^2 t_2)\exp(-t't/2) \tag{27}$$

Adding (26) and (27) gives

$$\iint_{-\infty}^{\infty} (x_1^2 x_2 - x_2)h(t,x)dx = i^3 t_1^2 t_2 \exp(-t't/2) \tag{28}$$

Likewise,

$$\iint_{-\infty}^{\infty} (x_2^2 x_1 - x_1)h(t,x)dx = i^3 t_2^2 t_1 \exp(-t't/2) \tag{29}$$

Multiplying (28) by μ_{12}^{21} and (29) by μ_{12}^{12} and adding the results, we obtain (14) with on the left-hand side the Fourier transform of $2\sqrt{N}D_{12}^{11}\omega_{N,dep}(x)$.

4. An application to simple random sampling

Consider a population U of N numbers $U = \{y_1, \ldots, y_N\}$. Suppose that as $N \to \infty$ the population mean $\overline{Y}_N$ and the population variance σ^2_{Ny} converge to $\overline{Y}$ and $\sigma^2_y > 0$, respectively. Let $\overline{y}_s$ denote the sample mean from a simple random sample of nonrandom size n $(= f_N N)$ without replacement (SRS). The sampling fraction f_N is such that $|f_N - f| < 1/N$ where f is a fixed number $(0 < f < 1)$.

Next, define for the Poisson sampling design with equal probabilities the following unbiased estimators for the population mean $\overline{Y}_N$ and n

$$\begin{aligned} \widehat{\overline{Y}}_{PO} &= \frac{1}{N}\sum_{k=1}^{N} a_k \frac{y_k}{f_N} \\ n_s &= \sum_{k=1}^{N} a_k, \end{aligned}$$

where the a_k are independent and identically distributed random variables. That is,

$$a_k = \begin{cases} 1 & \text{if the sample includes element } k \\ 0 & \text{otherwise} \end{cases}$$

with $P(a_k = 1) = f_N$ $(k = 1, \ldots, N)$.

In addition, define the two-dimensional random vector x_k and the matrix Σ_k by

$$\begin{aligned} x_k &= \begin{pmatrix} x_{1k} \\ x_{2k} \end{pmatrix} = a_k \begin{pmatrix} \frac{y_k}{f_N} \\ 1 \end{pmatrix} \\ \Sigma_k &= f_N(1-f_N)\begin{pmatrix} \frac{y_k^2}{f_N^2} & \frac{y_k}{f_N} \\ \frac{y_k}{f_N} & 1 \end{pmatrix}, \end{aligned}$$

respectively. Σ_k is the covariance matrix of x_k. Other useful definitions and formulas in this context are

$$\begin{aligned} \overline{x}_N &= \frac{1}{N}\sum_{k=1}^{N} x_k = \begin{pmatrix} \widehat{\overline{Y}}_{PO} \\ n_s/N \end{pmatrix} \\ E(\overline{x}_N) &= \begin{pmatrix} \overline{Y}_N \\ f_N \end{pmatrix} \\ \Sigma &= \lim_{N\to\infty} \overline{\Sigma}_N = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} \Sigma_k \\ &= (1-f)\begin{pmatrix} \frac{(\sigma_y^2 + \overline{Y}^2)}{f} & \overline{Y} \\ \overline{Y} & f \end{pmatrix} \end{aligned}$$

Denote the typical element ij of Σ by σ_{ij}, that of Σ^{-1} by σ^{ij}, and that of $\overline{\Sigma}_N$ of by $\overline{\sigma}_{Nij}$ $(1 \le i, j \le 2)$. Applying the central limit theorem for vectors to the mutually independent x_k yields the following theorem.

Theorem 1. As $N \to \infty$ the distribution of $\sqrt{N}\{\overline{x}_N - E(\overline{x}_N)\}$ tends to the bivariate normal distribution with zero expectation and covariance matrix Σ, provided that the data satisfy the Lindeberg condition.

For the Lindeberg condition, see Hájek (1964, page 1500) and Feller (1971, pages 262-3). Also note that the Lindeberg conditions for one dimension carry over to two-dimensional vectors due to the Cramér-Wold device; see Basu (2004, page 149).

In order to apply Theorem 1 to the problem of the limiting distribution of $\overline{y}_s$, define for an arbitrary constant u_0

$$\begin{aligned} P_{N0} &= P(\overline{y}_s \leq \overline{Y}_N + u_0 h_{N1}) \\ h^2_{N1} &= \frac{|\Sigma_N|}{N\overline{\sigma}_{N22}} \quad (= \frac{1}{N\overline{\sigma}^{11}_N} = (1-f_N)\frac{\sigma^2_{Ny}}{n}) \\ h^2_{N2} &= \frac{\sigma_{N22}}{N} \quad (= \frac{f_N(1-f_N)}{N}) \\ \psi(x;\Sigma) &= \frac{\exp(-x'\Sigma^{-1}x/2)}{2\pi\,|\Sigma|^{1/2}} \end{aligned}$$

where in the last line $x = (x_1, x_2)'$ and Σ =cov(x). Likewise, h^2_1 and h^2_2 are based on the elements in Σ. Recall from statistics theory that $1/\sigma^{11}$ is the conditional variance of x_1 given x_2. Furthermore, define λ_N by $\lambda_N = \mu^{(3)}_N/\sigma^3_N$ and λ by $\lambda = \lambda_\infty$ where, given N, σ^2_N and $\mu^{(3)}_N$ stand for the variance and the third central moment of x_{2k}, respectively. Note that $x_{2k} = a_k, \lambda_N = \lambda_N(f_N)$ and $\lambda = \lambda(f)$. Let σ^2 denote $f(1-f)$.

The following theorem specifies the error of the approximation of the lattice distribution of $\overline{x}_{2N}$ by the normal distribution.

Theorem 2. For a midpoint x of the lattice distribution F_{2N} of $\sqrt{N}\{\overline{x}_{2N} - E(\overline{x}_{2N})\}$ it holds that as $N \to \infty$

$$F_{2N}(x;\sigma_N,\lambda_N) - \Phi(\frac{x}{\sigma}) = \omega_{2N}(\frac{x}{\sigma};\lambda) + o(1/\sqrt{N}), \tag{30}$$

where

$$\omega_{2N}(x;\lambda) = \frac{\lambda}{6\sqrt{N}}(1-x^2)\varphi(x). \tag{31}$$

Proof. For the lattice distribution F_{2N} it holds that for a midpoint x of the lattice

$$F_{2N}(x;\sigma_N,\lambda_N) - \Phi(\frac{x}{\sigma_N}) = \omega_{2N}(\frac{x}{\sigma_N};\lambda_N) + o(1/\sqrt{N}). \tag{32}$$

For a proof, see Feller (1971, page 540). Note that Feller assumes that $\lambda_N = \lambda$ and $\sigma_N = \sigma$. However, his arguments are still valid when λ and σ^2 depend on N. Furthermore, since the derivatives of λ_N, σ_N and ω_{2N} as functions of f_N are continuous at the point f and $|f_N - f| < 1/N$, it holds that

$$|\lambda_N - \lambda| = O(1/N)$$

$$
\begin{aligned}
|\sigma_N - \sigma| &= O(1/N) \\
\left|\omega_{2N}(\frac{x}{\sigma_N};\lambda_N) - \omega_{2N}(\frac{x}{\sigma};\lambda)\right| &= O(1/N),
\end{aligned}
$$

from which (30) follows. This concludes the proof.

In addition, if F_{2N} were not a lattice distribution, (32) would be true for all x. Now the following theorem on the asymptotic behaviour of $\overline{y}_s$ can be proved.

Theorem 3. Under the Lindeberg condition it holds that as $N \to \infty$

$$P_{N0} \sim \Phi(u_0) \tag{33}$$

Proof. Denote $\overline{Y}_N + u_0 h_{N1}$ by m_{N0}. It follows from the equivalence of SRS sampling and Poisson sampling conditional on a given n or f_N that

$$
\begin{aligned}
P_{N0} &= P(\overline{x}_{1N} \leq m_{N0}\,|\overline{x}_{2N} = f_N) \\
&= \frac{P(\overline{x}_{1N} \leq m_{N0} \wedge \overline{x}_{2N} = f_N)}{P\{\overline{x}_{2N} = f_N\}}
\end{aligned} \tag{34}
$$

First we look at the relatively simple denominator of (34). Applying Theorem 1 yields

$$
\begin{aligned}
&P\{\overline{x}_{2N} = f_N\} = P\{\overline{x}_{2N} \leq (n+1/2)/N\} - P\{\overline{x}_{2N} \leq (n-1/2)/N\} \\
&= \Phi(\frac{1/2}{Nh_2}) - \Phi(\frac{-1/2}{Nh_2}) + \omega_{2N}(\frac{1/2}{Nh_2};\lambda) - \omega_{2N}(\frac{-1/2}{Nh_2};\lambda) + o(\frac{1}{\sqrt{N}})
\end{aligned} \tag{35}
$$

where the error $\omega_{2N}(x;\lambda)$ of the standard normal approximation at a midpoint x is given by (31). Since

$$(1-x^2)\{\varphi(x) - \varphi(-x)\} = 0$$

the total error in (35) is $o(1/\sqrt{N})$ and hence, using the first-order Taylor expansion $\Phi(x) = \Phi(0) + x\varphi(0) + O(x^2)$, (35) can be written as

$$P\{\overline{x}_{2N} = f_N\} = \frac{1}{Nh_2}\varphi(0) + o(\frac{1}{\sqrt{N}}) \tag{36}$$

In order to apply the results of section 3 to the numerator of (34), define the vector $\overline{z}_N$ by $\overline{z}_N = T_N\{\overline{x}_N - E(\overline{x}_N)\}$ with

$$T_N = \begin{pmatrix} 1 & -\beta_N \\ 0 & 1 \end{pmatrix},$$

where $\beta_N = \sigma_{N12}/\sigma_{N22} = \overline{Y}_N/f_N$. The numerator of (34) can now be written in terms of $\overline{z}_N$ as

$$P(\overline{z}_{1N} \leq u_0 h_{N1} - \beta_N \overline{z}_{2N} \wedge |\overline{z}_{2N}| \leq 1/2N)$$

Assuming that $\beta_N > 0$, an upper bound for the numerator is given by

$$P(\overline{z}_{1N} \leq u_0 h_{N1} + \beta_N/2N \wedge |\overline{z}_{2N}| \leq 1/2N)$$

Using $E(\overline{z}_N)=0$,

$$\overline{\Gamma}_N \equiv cov(\overline{z}_N) = T_N\overline{\Sigma}_N T_N'/N = diag(h_{N1}^2, h_{N2}^2)$$

and applying Theorem 1 with $\overline{\Gamma}_N$ instead of Σ, it is seen that the upper bound is equal to

$$\begin{aligned}
& \int_{-\infty}^{u_0 h_{N1}+\beta_N/2N} \int_{-1/2N}^{1/2N} \frac{\varphi(\frac{z_1}{h_{N1}})\varphi(\frac{z_2}{h_{N2}})}{h_{N1}h_{N2}} dz_1 dz_2 + \delta_N + o(\frac{1}{\sqrt{N}}) \\
&= \Phi(u_0 + \frac{\beta_N/2}{Nh_{N1}})\frac{\varphi(0)}{Nh_{N2}} + o(\frac{1}{\sqrt{N}}) \\
&= \Phi(u_0)\frac{\varphi(0)}{Nh_{N2}} + o(\frac{1}{\sqrt{N}}) \qquad (37) \\
\delta_N &= \omega_N^*(u_0 + \frac{\beta_N/2}{Nh_{N1}}, \frac{1/2}{Nh_{N2}}) - \omega_N^*(u_0 + \frac{\beta_N/2}{Nh_{N1}}, \frac{-1/2}{Nh_{N2}})
\end{aligned}$$

where $\omega_N^*(x)$ is based on (4) with a minor adjustment for the mutually different variances of the z_{1k}. Note that $\sqrt{N}\delta_N = o(1)$ as $N \to \infty$ and hence, $\delta_N = o(1/\sqrt{N})$. A sufficient condition for this is that the fourth moments exist; see Feller (1971, page 547-8). Since result (37) is also true for the corresponding lower bound, it is true for the numerator of (34). Combining (34), (36) and (37) gives (33). This concludes the proof.

References

Basu, A.K. (2004). *Measure Theory and Probability.* Prentice Hall of India, Delhi.

Erdös, P. and Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Magyar Tudoanyos Akademia Budapest Matematikai Kutato Intezet Koezlemenyei,* Trudy Publications, 4, 49−57.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. II. Wiley and Sons, New York.

Hájek, J. (1960). Limiting distributions in simple random sampling from a finite population. *Magyar Tudoanyos Akademia Budapest Matematikai Kutato Intezet Koezlemenyei,* Trudy Publications, 5, 361−374.

Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Annals of Mathematical Statistic*s, 35, 1491−1523.

Knottnerus, P. (2008). On asymptotic distributions in random sampling from finite populations. Statistics Netherlands, Discussion Paper, Voorburg.

Madow, W.G. (1948). On the limiting distributions of estimates based on samples from finite universes. *Annals of Mathematical Statistic*s, 19, 535−545.