MISCELLANEOUS NOTES

ON THE ORDER OF APPROXIMATION INVOLVED IN LAPLACE'S CENTRAL LIMIT THEOREM IN PROBABILITY

By MANINDRA NATH GHOSH

Statistical Laboratory, Calcutta

The central limiting theorem in probability is often put forward as the final justification for the frequent use of the normal distribution. As far as the present author is aware no thorough investigations have, however, been made about the exact order of approximation involved. Professor Cramer (1937) has shown that when absolute moments upto third order exist for the summands the approximation is of the order of $1/\sqrt{n}$. This is the generalisation of a previous result of Liapounoff. However, although the central limiting theorem was proved to hold when moments of the order of $2+\delta(\delta>0)$ exist by Liapounoff (1900) and also under the more general conditions of Lindburgh no approximation formula was established which might make the theorem suitable for statistical and numerical purposes. The object of this paper is to secure an approximation formula for the case where absolute moments of the order of $(2+\delta)$ $(\delta>0)$ are assumed to exist.

In order to arrive at the desired approximation we shall follow the lines of proof by Kolmogoroff of the central limiting theorem as given by Khintchine (1933). Let $F_1(x)$ be a distribution function with mean zero and variance unity and let $x'_1, x'_2, \ldots x'_n$ be n independant variates satisfying the same distribution law $F_1(x)$. For simplicity of analysis we shall introduce the variables $x_1 = x'_1/\sqrt{n}, \ldots, x_n = x'_n/\sqrt{n}$ satisfying the same distribution law $F_1(\sqrt{n}x) = F(x)$ with variance $1/n$ and mean zero. Denote by $U_k(x)$ the distribution function of the random variable $x = x_1 + \ldots + x_n$.

Then evidently we have

$$U_{k+1}(x) = \int_{-\infty}^{+\infty} U_k(x-\xi)\, dF(\xi) \qquad \ldots \quad (1)$$

We shall now show that $U_n(x)$ tends to the cumulative normal distribution function

$$\phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$$

as n becomes large. The first part of the proof is a modification of the proof of Khintchine (1933) to suit our present limited purpose.

Let
$$\phi(x/\sqrt{z}) = 1/\sqrt{2\pi z} \int_{-\infty}^{x} \exp(-t^2/2z) dt$$

It is easy to see that $\phi(x/\sqrt{z})$ satisfies the differential equation $\partial\phi/\partial z = 1/z.\ \partial^2\phi/\partial x^2$. Let us introduce the function $V(x,z)$ given by $V(x, z) = \phi(x/\sqrt{z}) + \epsilon z$ so that

$$\frac{\partial V}{\partial z} = \frac{1}{2} \frac{\partial^2 V}{\partial x^2} + \epsilon \qquad \ldots \quad (2)$$

We shall now prove that for a suitably chosen $\epsilon(\delta)$ in the whole halfplane $z>\delta$.

$$V(x, z+1/n) > \int_{-\infty}^{+\infty} V(x-\xi, z)\, dF(\xi)$$

Proof :
$$V(x-\xi, z) = V(x, z) - \xi \frac{\partial V}{\partial x} + \tfrac{1}{2} \xi^2 \frac{\partial^2 V}{\partial x^2} + \tfrac{1}{2} \xi^2 \rho(x,\xi,z)$$

where
$$\rho(x, \xi, z) = \left\{ \left[\frac{\partial^2 V}{\partial x^2}\right]_{x-\theta\xi, z} - \left[\frac{\partial^2 V}{\partial x^2}\right]_{x, z} \right\}_{0<\theta<1} \qquad \ldots \quad (3)$$

$$\int_{-\infty}^{+\infty} V(x-\xi, z)\, dF(\xi) = V(x,z) + \frac{1}{2n} \frac{\partial^2 V}{\partial x^2} + \tfrac{1}{2} \int_{-\infty}^{\infty} \xi^2 \rho(x,\xi,z)\, dF(\xi)$$

As
$$\int_{-\infty}^{+\infty} dF(\xi) = 1, \quad \int_{-\infty}^{+\infty} \xi\, dF(\xi) = 0 \text{ and } \int_{-\infty}^{+\infty} \xi^2\, dF(\xi) = 1/n.$$

From (2) it follows that if $\left| \int_{-\infty}^{\infty} \xi^3 \rho(x,\xi,z)\, dF(\xi) \right| < \frac{\varepsilon}{n}$ .. (4)

then $$\int_{-\infty}^{\infty} V(x-\xi,z)\, dF(\xi) < V(x,z) + \frac{1}{n}\frac{\partial V}{\partial z} - \frac{\varepsilon}{2n} \quad .. (5)$$

Again $$V\left(x, z+\frac{1}{n}\right) = V(x,z) + \frac{1}{n}\frac{\partial V}{\partial z} + \frac{1}{2n^2}\left[\frac{\partial^2 V}{\partial z^2}\right]_{x, z+\theta'/n, 0<\theta'<1} \quad .. (6)$$

Denoting by $M_1(\delta)$ the maximum value of $\frac{\partial^2 V}{\partial z^2}$ in the region $z>\delta$ when $M_1(\delta)/n<\varepsilon$ we have

$$V\left(x, z+\frac{1}{n}\right) > V(x,z) + \frac{1}{n}\frac{\partial V}{\partial z} - \frac{\varepsilon}{2n} \quad .. (7)$$

Hence from (5) and (7) we have

$$V\left(x, z+\frac{1}{n}\right) > \int_{-\infty}^{\infty} V(x-\xi, z)\, dF(\xi) \quad .. (8)$$

provided ε is chosen to satisfy (4) and $M_1(\delta)/n<\varepsilon$

From (8) we have $\phi(x/\sqrt{(k+1)/n}) + \varepsilon(k+1)/n > \int_{-\infty}^{+\infty} \phi((x-\xi)/\sqrt{k/n})\, dF(\xi) + k\varepsilon/n$ .. (9)

Now from (1) we have $U_{k+1}(x-2\alpha) = \int_{-\infty}^{+\infty} U_k(x-\xi-2\alpha)\, dF(\xi)$ .. (10)

Combining these two results we have

$$U_{k+1}(x-2\alpha) - \phi(x/\sqrt{(k+1)/n}) < \frac{\varepsilon}{n} + \int_{-\infty}^{+\infty} [U_k(x-\xi-2\alpha) - \phi(x/\sqrt{k/n})]\, dF(\xi) \quad .. (11)$$

Denote by P_k the least upper bound of $U_k(x) - \phi\{(x+2\alpha)/\sqrt{k/n}\}$ for values of x. Then we have $P_{k+1} \leqslant P_k + \varepsilon/n$. From this recurrence relation we find that

$$U_n(x) - \phi(x+2\alpha) \leqslant P_n \leqslant P_k + \varepsilon \text{ i.e. } U_n(x) - \phi(x) \leqslant 2\alpha/\sqrt{2\pi} + P_k + \varepsilon \quad .. (12)$$

We shall now see that P_k can be made small by a suitable choice of k and n and establish a lemma similar to the one given by Khintchine taking into consideration the existence of p-th order moments of the distribution function $F_1(x)$.

Lemma Given two random variables with mean zero and absolute pth moment $\leqslant \mu_p$ possessing distribution functions $F(x)$ and $G(x)$, then for $\alpha>0$, $F(x) - G(x+2\alpha) \leqslant \mu_p/\alpha^p$ holds for all values of x

Proof. When $x<-\alpha$, $F(x) \leqslant F(-\alpha) \leqslant \mu_p/\alpha^p$ and the lemma is true. When $x>-\alpha$ we have

$$F(x) - G(x+2\alpha) \leqslant 1 - G(x+2\alpha) \leqslant 1 - G(\alpha) \quad .. (13)$$

Now $G(\alpha) \geqslant 1 - \mu_p/\alpha^p$, and hence $F(x) \leqslant F(-\alpha) \leqslant \mu_p/\alpha^p$ (12) holds for all values of x. Putting $k/n=\delta$ we get from (12) by the above lemma

$$U_n(x) - \phi(x) \leqslant \frac{2\alpha}{\sqrt{2\pi}} + \varepsilon + \frac{\delta}{\alpha^2} \quad .. (14)$$

Since the variance of $U_k(x)$ and $\phi(x/\sqrt{\delta})$ are both δ therefore chosing $\alpha=\delta^{1/3}$ we have

$$U_n(x) - \phi(x) \leqslant \left(\frac{2}{\sqrt{2\pi}} + 1\right) \delta^{\frac{1}{3}} + \varepsilon \quad .. (15)$$

Now we can prove the opposite inequality by taking $V(x,z) = \phi(x/\sqrt{z}) - \varepsilon z$ and the corresponding values of δ and ε will be the same as above . Thus we may state

$$|U_n(x) - \phi(x)| \leqslant \left(\frac{2}{\sqrt{2\pi}} + 1\right) \delta^{\frac{1}{3}} + \varepsilon \quad .. (16)$$

which holds for all x. Now if we choose $\epsilon \geqslant \text{Max}\left[n\int_{-\infty}^{\infty}\xi^2|\rho(x,\xi,z)|\,dF(\xi), M_1(\delta)\right]$, the inequality (4) is satisfied. Now

$$\frac{\partial^2 V}{\partial z^2}=\int_{-\infty}^{x}\frac{t^4}{z^4}\cdot\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\,dt-\int_{-\infty}^{x}\frac{t^2}{z^3}\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\,dt+\int_{-\infty}^{x}\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\frac{dt}{z^2}$$

$$<\int_{-\infty}^{\infty}\frac{t^4}{z^4}\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\,dt+\int_{-\infty}^{\infty}\frac{t^2}{z^3}\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\,dt+\int_{-\infty}^{\infty}\frac{e^{-t^2/2z}}{\sqrt{2\pi z}}\frac{dt}{z^2} \quad \ldots (17)$$

$$<\frac{3}{4}+\frac{1}{z}+\frac{3}{4z^2}<\frac{1}{z^3} \text{ when } z \text{ is small}$$

Thus we have $M_1(\delta)=\underset{z>\delta}{\text{Max}}\,\partial^2V/\partial z^2<1/\delta^3$ when δ is small. Let us now consider the integral,

$$n\int_{-\infty}^{\infty}\xi^2|\rho(x,\xi,z)|\,dF(\xi)=n\int_{|\xi|<\tau}\xi^2|\rho(x,\xi,z)|\,dF(\xi)+n\int_{|\xi|>\tau}\xi^2|\rho(x,\xi,z)|\,dF(\xi) \quad \ldots (18)$$

$$\leqslant n|\rho(x,\theta\tau,z)|\int_{|\xi|<\tau}\xi^2\,dF(\xi)+n\,\underset{z>\delta}{\text{Max}}|\rho(x,\xi,z)|\int_{|\xi|>\tau}\xi^2\,dF(\xi)$$

where $0<\theta<1$ because of the continuity of $|\rho(x,\xi,z)|$ Now

$$\rho(x,\theta\tau,z)=\left(\frac{\partial^2V}{\partial x^2}\right)_{x+\theta\tau,z}-\left(\frac{\partial^2V}{\partial x^2}\right)_{x,z}=\theta\tau\left(\frac{\partial^3V}{\partial x^3}\right)_{x-\theta'\tau,z}$$

$$0<\theta'<1$$

$$\frac{\partial^3V}{\partial x^3}=\frac{(x^2/z)-1}{\sqrt{2\pi}}\cdot\frac{e^{-x^2/2z}}{z^{3/2}} \quad \ldots (19)$$

Thus the maximum value of $|\partial^3v/\partial x^3|$ when $z>\delta$ is found to be $1/\sqrt{2\pi}\,\delta^{3/2}$. On the other hand $\partial^2V/\partial x^2=-x/\sqrt{2\pi z^3}\,.\,\exp(-x^2/2z)$. Therefore $\underset{z>\delta}{\text{Max}}\frac{\partial^2V}{\partial z^2}<\frac{1}{\sqrt{2\pi\delta}}e^{-1/2}$ which gives

$$\underset{z>\delta}{\text{Max}}\left|\rho(x,\xi,z)\right|<\frac{xe^{-\frac{1}{2}}}{\sqrt{2\pi\delta}} \quad \ldots (20)$$

Consider now the integral, $$n\int_{|\xi|>\tau}\xi^2\,dF(\xi)=\int_{|\xi|>n\tau}\xi^2\,dF_1(\xi) \quad \ldots (21)$$

If $\sqrt{n}\tau\to\infty$ then from the existence of p-th order absolute moment of $F_1(\xi)$

$$\int_{|\xi|>\tau\sqrt{n}}\xi^p\,dF_1(\xi)<\mu_p=\int_{-\infty}^{\infty}\xi^p\,dF_1(\xi) \quad \ldots (22)$$

from which we have $$\int_{|\xi|>\sqrt{n}\tau}\xi^2\,dF_1(\xi)<\frac{\mu_p}{(\tau\sqrt{n})^{p-2}} \quad \ldots (23)$$

Thus finally we have $$\epsilon>\frac{\tau}{\sqrt{2\pi}\,\delta^{3/2}}+\frac{2e^{-1/2}}{\sqrt{2\pi\delta}}\frac{\mu_p}{(\tau\sqrt{n})^{p-2}} \quad \ldots (24)$$

and $\epsilon\geqslant 1/n\delta^3$ Thus chosing ϵ as above we have

$$\left|U_n(x)-\phi(x)\right|<\left(\frac{2}{\sqrt{2\pi}}+1\right)\delta^{1/2}+\frac{\tau}{\sqrt{2\pi}\,\delta^{1\cdot 2}}+\sqrt{\frac{2}{\pi e}}\cdot\frac{\mu_p}{(\tau/\sqrt{n})^{p-2}}$$

or $$\left|U_n(x)-\phi(x)\right|<\frac{1}{n\delta^3} \quad \text{whichever is greater} \quad \ldots (25)$$

We shall now chose δ and τ as functions of n i.e., $\delta = 1/n^{\alpha}$, $\tau = n^{\beta}$ in such a manner that the expression on the right has a minimum order in n.

Let $i_1 = -\alpha/3$, $i_2 = 2\alpha - 1$, $i_3 = 3\alpha/2 - \beta$, $i_4 = \beta(p-2) + \alpha - (p-2)/2$ denote the indices of n in the four quantities on the right. Then the problem is to minimise $m = Max\ (i_1, i_2, i_3, i_4)$ for values of α and β. It is clear that for an extremum, three of four indices i_1, i_2, i_3, i_4 should be equal or they should be equal in pairs. Among seven such sets of extremas we have to pick one that gives an absolute minimum for m. The different sets of solutions are given below :

(1) $i_1 = i_2 = i_3$

$$\alpha = \frac{3}{7},\ \beta = \frac{11}{14} \left\{ -\frac{1}{7}, -\frac{1}{7}, -\frac{1}{7}, \frac{2p-1}{7} \right\}$$

$m_1(p) = (2p-1)/7$ which is inadmissible as $m_1(p) > 0$

(2) $i_1 = i_2 = i_4$

$$\alpha = \frac{2}{7},\ \beta = \frac{1}{2} - \frac{4}{7(p-2)} \left\{ -\frac{1}{7}, -\frac{1}{7}, \frac{p+2}{7(p-2)}, -\frac{1}{7} \right\}$$

$m_2(p) = (p+2)/7(p-2)$ which is also inadmissible

(3) $i_1 = i_2$, $i_3 = i_4$

$$\alpha = \frac{3}{7},\ \beta = \frac{7p-11}{14(p-1)} \left\{ -\frac{1}{7}, -\frac{1}{7}, \frac{p+1}{7(p-1)}, \frac{p+1}{7(p-1)} \right\}$$

$m_3(p) = (p+1)/7(p-1)$ which is inadmissible

(4) $i_1 = i_3$, $i_2 = i_4$

$$\alpha = \frac{3(p-4)}{11p-21},\ \beta = \frac{1}{2} + \frac{\alpha-1}{p-2} \left\{ -\frac{4-p}{28-11p}, \frac{5p-4}{28-11p}, -\frac{4-p}{28-11p}, \frac{5p-4}{28-11p} \right\}$$

This is inadmissible for $2 < p < 4$

(5) $i_2 = i_3 = i_4$

$$\alpha = \frac{3(p-2)}{11p-14},\ \beta = \frac{11(p-2)}{22p-28} \left\{ -\frac{p-2}{11p-14}, -\frac{5p-2}{11p-14}, -\frac{p-2}{11p-14}, -\frac{p-2}{11p-14} \right\}$$

$m_5(p) = -(p-2)\ /\ (11p-14)$ which is admissible for all p

(6) $i_1 = i_4$, $i_2 = i_3$

$$\alpha = \frac{3(p-2)}{3p-14},\ \beta = 1 - \frac{\alpha}{2},\ \left\{ -\frac{p-2}{3p-14}, \frac{3p+2}{3p-14}, \frac{3p+2}{3p-14}, -\frac{p-2}{3p-14} \right\}$$

This is inadmissible

(7) $i_2 = i_3 = i_4$

$$\alpha = 1,\ \beta = \frac{1}{2} \left\{ -\frac{1}{3}, 1, 1, 1 \right\}$$

$m_7(p) = 1$ which is inadmissible

Thus we can write

$$|\phi(x) - U_n(x)| < (3/\sqrt{2\pi} + 1)\, n^{-(p-2)/(11p-14)} + \sqrt{2\pi e}\ \ \mu_p\, n^{-(5p-2)/(11p-14)} \qquad \ldots \quad (26)$$

We see, therefore, that when absolute moments upto the pth order exist the order of approximation depends upon p. For a given n therefore we can get the best numerical approximation by suitably chosing $\leqslant$ the highest order moment that exists. However since the index of $1/n$ is large for the second term for large values of n the first factor dominates. It is evident however that the above approximation is useless for $p \geqslant 3$ for in that case a much better order of approximation has been shown to exist by Liapounof and Cramer. The above approximation is interesting only when $2 < p < 3$.

Thanks are due to S. N. Roy of the Statistical Laboratory, Calcutta, for help and advice in the preparation of this note.

REFERENCES

1. CRAMER (1937) Random variables and probability distribution, Cambridge.
2. KHINTCHINE (1933) Asymptotische gesetze der wahrscheinlichkeitsrechnung Berlin.
3. LIAPOUNOFF (1900) Sur une proposition de la theorie des probabilities : *Bull. acad. Sci. St. Petersbourg* (5) 13 p. 359-386.

Paper received 25 August 1944.