Lecture 10: Central Limit Theorems

Emílio C. Silva

Department of Civil and Environmental Engineering, MIT

March 13, 2001

In a broad sense, a "Central Limit Theorem" states that the probability distribution $P_n(x)$ for the sum of random variables $\sum_{i=1}^{n} \Delta x_i$ converges to a gaussian as $n \to \infty$, under certain conditions. Some versions of this theorem are covered in this lecture.

1 Chebyshev (1890)

Let $X_n = \sum_{i=1}^{n} \Delta x_i$ be the sum of iid random variables, where all moments of the p.d.f. $p(\Delta x_i)$ are finite and $\langle \Delta x_i \rangle = 0$. Let also $P_n(X)$ be the p.d.f. of X_n. Defining $\sigma^2 = \langle \Delta x_i^2 \rangle$ and $\zeta = \frac{x}{\sigma\sqrt{n}}$, the pdf in the transformed coordinate becomes $\phi_n(\zeta) = \sigma P_n(\zeta\sigma)$ and its cumulative distribution function can be defined as

$$\Phi_n(\zeta) = \int_{-\infty}^{\zeta} \phi(\zeta) d\zeta \tag{1}$$

Defining also $\phi_0 = \frac{e^{-\zeta^2/2}}{\sqrt{2\pi}}$ as the gaussian, the Chebyshev statement may be written as

$$\Phi_n(\zeta) - \Phi_0(\zeta) \sim \frac{e^{-\zeta^2/2}}{\sqrt{2\pi}} \left(\frac{Q_1(\zeta)}{\sqrt{n}} + \frac{Q_2(\zeta)}{n} + \cdots \right) \tag{2}$$

where the first polynomials $Q_i(\zeta)$ are

$$Q_1(\zeta) = \frac{\lambda_3}{6}(\zeta^2 - 1) \tag{3}$$

$$Q_2(\zeta) = \frac{\lambda_3^2}{72}\zeta^5 - \frac{1}{8}\left(\frac{\lambda_4}{3} - \frac{10\lambda_3^2}{9}\right)\zeta^3 + \left(\frac{5}{24}\lambda_3^2 - \frac{\lambda_4}{8}\right)\zeta \tag{4}$$

and $\lambda_n = c_n/\sigma^n$ are the normalized cumulants. Deriving these, the polynomials previously shown in class should be obtained:

$$\phi_n(\zeta) \sim \phi_0(\zeta) \left(1 + \frac{f_1(\zeta)}{\sqrt{n}} + \frac{f_2(\zeta)}{n} + \cdots \right) \tag{5}$$

which is called "Gram-Charlier" expansion.

2 Berry-Eséen Theorem

Using the same definitions as in the previous section, but relaxing the condition that all moments must be finite, while still assuming $\langle \Delta x_i \rangle = 0$, $\sigma^2 = \langle \Delta x_i^2 \rangle$ and additionally $\langle |\Delta x_i|^3 \rangle = r_3 < \infty$, the Berry-Eséen theorem states that $\Phi(\zeta)$ tends to a normal satisfying

$$|\Phi_n(\zeta) - \Phi_0(\zeta)| \leq \frac{3r_3}{\sqrt{n}\sigma^{3/2}} \tag{6}$$

which is a very strong bound, since it is valid for all n.

3 Diverging third moment

The case $\langle |\Delta x_i|^3 \rangle \to \infty$ was shown in the previous lecture. The correction will be derived for this case. The p.d.f. is

$$p(x) = \frac{2/\sqrt{\pi}}{1 + x^4} \tag{7}$$

and the corresponding structure function is

$$\hat{p}(k) = e^{-|k|/2}\left(\cos\frac{k}{\sqrt{2}} + \sin\frac{|k|}{\sqrt{2}}\right) \sim 1 - \frac{k^2}{2} + \frac{|k|^3}{3\sqrt{2}} - \frac{5}{48}k^4 + O(|k|^5) \tag{8}$$

which is an asymptotic expansion around $k \to 0$. The logarithm of the structure function is

$$\log \hat{p}(k) \sim -\frac{k^2}{2} + \frac{|k|^3}{3\sqrt{2}} - \frac{11}{48}k^4 + \cdots \tag{9}$$

from where

$$\log \hat{P}_n(k) = \log \hat{p}(k)^n \sim -\frac{k^2 n}{2} + \frac{|k|^3 n}{3\sqrt{2}} - \frac{11}{48}k^4 n + \cdots \tag{10}$$

The term $-nk^2/2$ correspond to the gaussian, but in the general case, the following terms may have different coefficients, e.g. in the homework problem $p(x) = \frac{A}{(1+x^2)^2}$. Writing in a generalized form

$$\log \hat{P}_n(k) \sim -\frac{k^2 n}{2} + a|k|^3 n + \cdots \tag{11}$$

then the characteristic function of the position after n steps is

$$P_n(x) = \int_{\infty}^{\infty} e^{ikx}\hat{P}(k)\frac{dk}{2\pi} = \int_{-\infty}^{\infty} e^{ikx}e^{-k^2n/2}e^{a|k|^3n+\cdots}\frac{dk}{2\pi} \tag{12}$$

Changing the variables to $\zeta = x/\sigma\sqrt{n}$ and $\omega = k\sigma\sqrt{n}$, then

$$\phi_n(\zeta) = \sqrt{n}P_n(\zeta\sqrt{n}) \tag{13}$$
$$= \int_{\infty}^{\infty} e^{i\zeta\omega}e^{-\omega^2/2}e^{a|\omega|^3/\sigma^3\sqrt{n}+O(\omega^4/n)}\frac{d\omega}{2\pi} \tag{14}$$

which can be expanded for large n as

$$\phi_n(\zeta) = \int_{\infty}^{\infty} e^{i\zeta\omega}e^{-\omega^2/2}\left(1+\frac{a|\omega|^3}{\sigma^3\sqrt{n}}+\cdots\right)\frac{d\omega}{2\pi} \tag{15}$$

where the first term is simply the gaussian. Writing $|\omega|^3$ as $\omega^3 \operatorname{sgn}\omega$, the ω factors can be extracted as derivatives, yielding

$$\phi_n \sim \phi_0(\zeta) + \frac{a}{\sigma^3\sqrt{n}}\left(-i\frac{\partial^3}{\partial\zeta^3}\right)^3\int_{-\infty}^{\infty} e^{i\zeta\omega}e^{-\omega^2/2}\operatorname{sgn}\omega\frac{d\omega}{2\pi}+\cdots \tag{16}$$

where the familiar gaussian derivatives are obtained, albeit in a slightly different form. This can be written more generally as

$$\phi_n(\zeta) \sim \phi_0(\zeta) + \frac{a\phi_1(\zeta)}{\sigma^3\sqrt{n}} + \frac{b\phi_2(\zeta)}{\sigma^4 n}+\cdots \tag{17}$$

where

$$\phi_m(\zeta) = (const)\int_{-\infty}^{\infty} |\omega|^{m+2}e^{-\omega^2/2}e^{-\omega^2}\frac{d\omega}{2\pi} \tag{18}$$

.

Since $\operatorname{sgn}\omega$ is odd and, for real ζ, $e^{i\zeta\omega} = \cos\zeta\omega + i\sin\zeta\omega$ has even and odd parts, the integral in (16) may be rewritten using symmetry as

$$\int_{-\infty}^{\infty} e^{i\zeta\omega}e^{-\omega^2/2}\operatorname{sgn}\omega\frac{d\omega}{2\pi} = 2\int_0^{\infty} i\sin\zeta\omega e^{-\omega^2/2}\frac{d\omega}{2\pi} \tag{19}$$

or as the more convenient

$$\int_{-\infty}^{\infty} e^{i\zeta\omega}e^{-\omega^2/2}\operatorname{sgn}\omega\frac{d\omega}{2\pi} = -\frac{i}{\pi}\Im\int_0^{\infty} e^{-i\zeta\omega-\omega^2/2}d\omega \tag{20}$$

where the minus sign cancels the one at the exponent $-i\zeta\omega$. Calling the integral in the RHS I and completing the square in the exponent,

$$I = \int_0^{\infty} e^{-(\omega+i\zeta)^2/2-\zeta^2/2}d\omega \tag{21}$$
$$= e^{-\zeta^2/2}\int_0^{\infty} e^{-(\omega+i\zeta)^2/2}d\omega \tag{22}$$

what reduces the problem to the evaluation of $\int e^{z^2/2}dz$ in a contour segment.

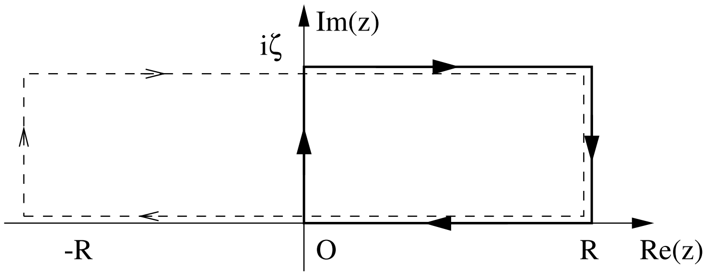

Figure 1: Integration contour for $e^{-z^2/2}$ (solid), when $\operatorname{sgn}\omega$ is present, and the usual (dashed).

Using the contour of Figure 1, it can be shown first that the right segment goes to zero as $R \to \infty$:

$$\left| \int e^{-z^2/2} dz \right| \leq \int |e^{-z^2/2}| dz = \zeta e^{-R^2/2} \tag{23}$$

what reduces the problem to the determination of the lower segment (a real integral), the upper segment ($Ie^{\zeta^2/2}$) and the left part, which is the special feature of this case, introduced by the $\operatorname{sgn}\omega$ function. Since the function $e^{-z^2/2}$ is entire, Cauchy theorem leads to

$$\oint e^{z^2/2} dz = Ie^{\zeta^2/2} - \int_0^\infty e^{-x^2/2} dx + i \int_0^\zeta e^{y^2/2} dy = 0 \tag{24}$$

therefore

$$I = e^{-\zeta^2/2} \left(\int_0^\infty e^{-x^2/2} dx - i \int_0^\zeta e^{y^2/2} dy \right) \tag{25}$$

$$= e^{-\zeta^2/2} \left(\frac{\sqrt{\pi}}{2} - i \int_0^\zeta e^{y^2/2} dy \right) \tag{26}$$

Substituting this back in (20) and noting that only the imaginary part is necessary,

$$\int_{-\infty}^\infty e^{i\zeta\omega} e^{-\omega^2/2} \operatorname{sgn}\omega \frac{d\omega}{2\pi} = -\frac{i}{\pi} \Im \left(-ie^{-\zeta^2/2} \int_0^\zeta e^{y^2/2} dy \right) \tag{27}$$

$$= \frac{i}{\pi} e^{-\zeta^2/2} \int_0^\zeta e^{y^2/2} dy \tag{28}$$

which can be written in terms of Dawson's integral, defined as

$$\mathrm{D}(x) = e^{-x^2} \int_0^x e^{y^2} dy \sim \frac{1}{2x} \left(1 + \frac{1}{2x^2} + \frac{3}{4x^4} + \sum_{m=3}^\infty \frac{(2m-1)!!}{2^m x^{2m}} \right) \tag{29}$$

resulting in

$$\int_{-\infty}^\infty e^{i\zeta\omega} e^{-\omega^2/2} \operatorname{sgn}\omega \frac{d\omega}{2\pi} = \frac{i\sqrt{2}}{\pi} \mathrm{D}\left(\frac{\zeta}{\sqrt{2}} \right) \tag{30}$$

and substitution of (30) in (16) gives

$$\phi_n(\zeta) \sim \phi_0(\zeta) + \frac{a\sqrt{2}}{\pi\sqrt{n}} \frac{\partial^3}{\partial\zeta^3} \mathrm{D}\left(\frac{\zeta}{\sqrt{2}}\right) + \cdots \tag{31}$$

where

$$\mathrm{D}'(x) = -2x\,\mathrm{D}(x) + 1 \tag{32}$$

$$\mathrm{D}''(x) = -2(x^2 + (1-2x^2)\,\mathrm{D}(x)) \tag{33}$$

$$\mathrm{D}'''(x) = -4(1 - x^2 + (2x^3 - 3x)\,\mathrm{D}(x)) \tag{34}$$

therefore, the first correction $\phi_1(\zeta)$ in (17) is

$$\phi_1(\zeta) = \frac{2}{\pi}\left[1 - \left(\frac{\zeta}{\sqrt{2}}\right)^2 + \left(2\left(\frac{\zeta}{\sqrt{2}}\right)^3 - 3\left(\frac{\zeta}{\sqrt{2}}\right)\right)\mathrm{D}\left(\frac{\zeta}{\sqrt{2}}\right)\right] \tag{35}$$

For large ζ, the leading order behavior is

$$\phi_1(\zeta) = \frac{6}{\pi\zeta^4} \tag{36}$$

which indicates that the $\frac{1}{x^4}$ behavior of the original p.d.f is preserved. The width of the central region may be estimated equating this to the gaussian:

$$\phi_0(\zeta_{\text{edge}}) \approx \frac{a\phi_1(\zeta_{\text{edge}})}{\sigma^3\sqrt{n}} \tag{37}$$

$$\frac{e^{-\zeta_{\text{edge}}^2/2}}{\sqrt{2\pi}} \approx \frac{(const)}{\sqrt{n}}\left(\frac{\zeta_{\text{edge}}^2}{2}\right)^{-2} \tag{38}$$

and, taking the log of both sides, ignoring the constants,

$$-\frac{\zeta_{\text{edge}}^2}{2} \approx -\frac{1}{2}\log n - 2\log\zeta_{\text{edge}}^2 + \cdots \tag{39}$$

$$\zeta_{\text{edge}}^2 \approx \log n + 4\log\zeta_{\text{edge}}^2 + \cdots \tag{40}$$

from which, recursively substituting ζ_{edge} in the RHS,

$$\zeta_{\text{edge}}^2 \approx \log n + 4\log\log n + \cdots \tag{41}$$

and finally the width of the central region has the order

$$\zeta_{\text{edge}} \approx \sqrt{\log n}\left(1 + 2\frac{\log\log n}{\log n} + \cdots\right) \tag{42}$$

The interesting point is that for "large" n, such as $n = 1000$, $\sqrt{\log n} = 2.63$, showing that the gaussian approximation is good only for a few standard devations. Even $n = 10^6$ is not "large enough" for convergence. This shows that $\frac{A}{a+x^4}$ converges very slowly to a gaussian, and care must be taken when fat-tailed p.d.f's are used.