2.16. Central limit theorem - statement, heuristics and discussion

If X_i are i.i.d with zero mean and finite variance σ^2, then we know that $\mathbf{E}[S_n^2] = n\sigma^2$, which can roughly be interpreted as saying that $S_n \approx \sqrt{n}$ (That the sum of n random zero-mean quantities grows like $\sqrt{n}$ rather than n is sometimes called the *fundamental law of statistics*). The central limit theorem makes this precise, and shows that on the order of $\sqrt{n}$, the fluctuations (or randomness) of S_n are independent of the original distribution of X_1! We give the precise statement and some heuristics as to why such a result may be expected.

Theorem 2.51. *Let X_n be i.i.d with mean μ and finite variance σ^2.Then, $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ converges in distribution to $N(0,1)$.*

Informally, letting χ denote a standard Normal variable, we may write $S_n \approx n\mu + \sigma\sqrt{n}\chi$. This means, the distribution of S_n is hardly dependent on the distribution of X_1 that we started with, except for the two parameter of mean and variance. This is a statement about a remarkable symmetry!

Heuristics: Why should one expect such a statement to be true? Without losing generality, let us take $\mu = 0$ and $\sigma^2 = 1$. As $\mathbf{E}\left[\left(\frac{S_n}{\sqrt{n}}\right)^2\right] = 1$ is bounded, we see that $n^{-\frac{1}{2}} S_n$ is tight, and hence has weakly convergent subsequences. Let us make a leap of faith and suppose that $\frac{S_n}{\sqrt{n}}$ converges in distribution. To what? Let Y be a random variable with the limiting distribution. Then, $(2n)^{-\frac{1}{2}} S_{2n} \xrightarrow{d} Y$ and further,

$$\frac{X_1 + X_3 + \ldots + X_{2n-1}}{\sqrt{n}} \xrightarrow{d} Y, \qquad \frac{X_2 + X_4 + \ldots + X_{2n}}{\sqrt{n}} \xrightarrow{d} Y.$$

But $(X_1, X_3, \ldots)$ is independent of $(X_2, X_4, \ldots)$. Therefore, by an earlier exercise, we also get

$$\left(\frac{X_1 + X_3 + \ldots + X_{2n-1}}{\sqrt{n}}, \frac{X_2 + X_4 + \ldots + X_{2n}}{\sqrt{n}}\right) \xrightarrow{d} (Y_1, Y_2)$$

where Y_1, Y_2 are i.i.d copies of Y. But then, by yet another exercise, we get

$$\frac{S_{2n}}{\sqrt{2n}} = \frac{1}{\sqrt{2}}\left(\frac{X_1 + X_3 + \ldots + X_{2n-1}}{\sqrt{n}} + \frac{X_2 + X_4 + \ldots + X_{2n}}{\sqrt{n}}\right) \xrightarrow{d} \frac{Y_1 + Y_2}{\sqrt{2}}$$

Thus we must have $Y_1 + Y_2 \stackrel{d}{=} \sqrt{2}Y$. Therefore, if $\psi(t)$ denotes the characteristic function of Y, then

$$\psi(t) = \mathbf{E}\left[e^{itY}\right] = \mathbf{E}\left[e^{itY/\sqrt{2}}\right]^2 = \psi\left(\frac{t}{\sqrt{2}}\right)^2.$$

Similarly, for any $k \geq 1$, we can prove that $Y_1 + \ldots Y_k \stackrel{d}{=} \sqrt{k}Y$, where Y_i are i.i.d copies of Y and hence $\psi(t) = \psi(tk^{-1/2})^k$. From this, by standard methods, one can deduce that $\psi(t) = e^{-at^2}$ for some $a > 0$ (**exercise**). By uniqueness of characteristic functions, $Y \sim N(0, 2a)$. Since we expect $\mathbf{E}[Y^2] = 1$, we must get $N(0,1)$.

It is an instructive exercise to prove the CLT by hand for specific distributions. For example, suppose X_i are i.i.d exp(1) so that $\mathbf{E}[X_1] = 1$ and $\text{Var}(X_1) = 1$. Then

$S_n \sim \Gamma(n,1)$ and hence $\frac{S_n - n}{\sqrt{n}}$ has density

$$\begin{aligned} f_n(x) &= \frac{1}{\Gamma(n)} e^{-n-x\sqrt{n}} (n + x\sqrt{n})^{n-1}\sqrt{n} \\ &= \frac{e^{-n} n^{n-\frac{1}{2}}}{\Gamma(n)} e^{-x\sqrt{n}} \left(1 + \frac{x}{\sqrt{n}}\right)^{n-1} \\ &\to \frac{1}{\sqrt{2\pi}} e^{-x^2} \end{aligned}$$

by elementary calculations. By an earlier exercise convergence of densities implies convergence in distribution and thus we get CLT for sums of exponential random variables.

Exercise 2.52. Prove the CLT for $X_1 \sim \text{Ber}(p)$. Note that this also implies CLT for $X_1 \sim \text{Bin}(k,p)$.

2.17. Central limit theorem - Proof using characteristic functions

We shall use characteristic functions to prove the CLT. To make the main idea of the proof transparent, we first prove a restricted version assuming third moments. Once the idea is clear, we prove a much more general version later which will also give Theorem 2.51. We shall need the following fact.

Exercise 2.53. Let z_n be complex numbers such that $nz_n \to z$. Then, $(1+z_n)^n \to e^z$.

Theorem 2.54. *Let X_n be i.i.d with finite third moment, and having zero mean and unit variance. Then, $\frac{S_n}{\sqrt{n}}$ converges in distribution to $N(0,1)$.*

PROOF. By Lévy's continuity theorem, it suffices to show that the characteristic functions of $n^{-\frac{1}{2}} S_n$ converge to the of $N(0,1)$. Note that

$$\psi_n(t) := \mathbf{E}\left[e^{itS_n/\sqrt{n}}\right] = \psi\left(\frac{t}{\sqrt{n}}\right)^n$$

where ψ is the c.f of X_1. Use Taylor expansion

$$e^{itx} = 1 + itx - \frac{1}{2}t^2x^2 - \frac{i}{6}t^3 e^{itx^*} x^3 \qquad \text{for some } x^* \in [0,x] \text{ or } [x,0].$$

Apply this with X_1 in place of x, $tn^{-1/2}$ in place of t, take expectations and recall that $\mathbf{E}[X_1] = 0$ and $\mathbf{E}[X_1^2] = 1$ to get

$$\psi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + R_n(t), \qquad \text{where } R_n(t) = -\frac{i}{6}t^3 \mathbf{E}\left[e^{itX_1^*} X_1^3\right].$$

Clearly, $|R_n(t)| \le Cn^{-3/2}$ for a constant C (that depends on t but not n). Hence $nR_n(t) \to 0$ and by Exercise 2.53 we conclude that for each fixed $t \in \mathbb{R}$,

$$\psi_n(t) = \left(1 - \frac{t^2}{2n} + R_n(t)\right)^n \to e^{-\frac{t^2}{2}}$$

which is the c.f of $N(0,1)$. ■

2.18. CLT for triangular arrays

The CLT does not really require the third moment assumption, and we can modify the above proof to eliminate that requirement. Instead, we shall prove an even more general theorem, where we don't have one infinite sequence, but the random variables that we add to get S_n depend on n themselves.

Theorem 2.55 (Lindeberg Feller CLT). *Suppose $X_{n,k}$, $k \le n$, $n \ge 1$, are random variables. We assume that*

1. *For each n, the random variables $X_{n,1}, \dots, X_{n,n}$ are defined on the same probability space, are* independent *and have finite second moments.*
2. $\mathbf{E}[X_{n,k}] = 0$ *and* $\sum_{k=1}^n \mathbf{E}[X_{n,k}^2] \to \sigma^2$, *as* $n \to \infty$.
3. *For any $\delta > 0$, we have* $\sum_{k=1}^n \mathbf{E}[X_{n,k}^2 \mathbf{1}_{|X_{n,k}|>\delta}] \to 0$ *as* $n \to \infty$.

Corollary 2.56. *Let X_n be i.i.d, having zero mean and unit variance. Then, $\frac{S_n}{\sqrt{n}}$ converges in distribution to $N(0,1)$.*

PROOF. Let $X_{n,k} = n^{-\frac{1}{2}} X_k$ fo r$k = 1,2,\dots,n$. Then, $\mathbf{E}[X_{n,k}] = 0$ while $\sum_{k=1}^n \mathbf{E}[X_{n,k}^2] = \frac{1}{n}\sum_{k=1}^N \mathbf{E}[X_1^2] = \sigma^2$, for each n. Further, $\sum_{k=1}^n \mathbf{E}[X_{n,k}^2 \mathbf{1}_{|X_{n,k}|>\delta}] = \mathbf{E}[X_1^2 \mathbf{1}_{|X_1|>\delta\sqrt{n}}]$ which goes to zero as $n \to \infty$ by DCT, since $\mathbf{E}[X_1^2] < \infty$. Hence the conditions of Lindeberg Feller theorem are satisfied and we conclude that $\frac{S_n}{\sqrt{n}}$ converges in distribution to $N(0,1)$. ■

Now we prove the Lindeberg-Feller CLT. As in the previous section, we need a fact comparing a product to an exponential.

Exercise 2.57. If z_k, w_k are complex numbers with absolute value bounded by θ, then $\left|\prod_{k=1}^n z_k - \prod_{k=1}^n w_k\right| \le \theta^{n-1}\sum_{k=1}^n |z_k - w_k|$.

PROOF. (**Lindeberg-Feller CLT**). The characteristic function of $S_n = X_{n,1} + \dots + X_{n,n}$ is given by $\psi_n(t) = \prod_{k=1}^n \mathbf{E}\left[e^{itX_{n,k}}\right]$. Again, we shall use the Taylor expansion of e^{itx}, but we shall need both the second and first order expansions.

$$e^{itx} = \begin{cases} 1 + itx - \frac{1}{2}t^2x^2 - \frac{i}{6}t^3 e^{itx^*}x^3 & \text{for some } x^* \in [0,x] \text{ or } [x,0]. \\ 1 + itx - \frac{1}{2}t^2 e^{itx^+}x^2 & \text{for some } x^+ \in [0,x] \text{ or } [x,0]. \end{cases}$$

Fix $\delta > 0$ and use the first equation for $|x| \le \delta$ and the second one for $|x| > \delta$ to write

$$e^{itx} = 1 + itx - \frac{1}{2}t^2x^2 + \frac{\mathbf{1}_{|x|>\delta}}{2}t^2x^2(1 - e^{itx^+}) - \frac{i\mathbf{1}_{|x|\le\delta}}{6}t^3x^3e^{itx^*}.$$

Apply this with $x = X_{n,k}$, take expectations and write $\sigma_{n,k}^2 := \mathbf{E}[X_{n,k}^2]$ to get

$$\mathbf{E}[e^{itX_{n,k}}] = 1 - \frac{1}{2}\sigma_{n,k}^2 t^2 + R_{n,k}(t)$$

where, $R_{n,k}(t) := \frac{t^2}{2}\mathbf{E}\left[\mathbf{1}_{|X_{n,k}|>\delta} X_{n,k}^2 \left(1 - e^{itX_{n,k}^+}\right)\right] - \frac{it^3}{6}\mathbf{E}\left[\mathbf{1}_{|X_{n,k}|\le\delta} X_{n,k}^3 e^{itX_{n,k}^*}\right]$. We can bound $R_{n,k}(t)$ from above by using $|X_{n,k}|^3 \mathbf{1}_{|X_{n,k}|\le\delta} \le \delta X_{n,k}^2$ and $|1 - e^{itx}| \le 2$, to get

$$|R_{n,k}(t)| \le t^2 \mathbf{E}\left[\mathbf{1}_{|X_{n,k}|>\delta} X_{n,k}^2\right] + \frac{|t|^3\delta}{6}\mathbf{E}\left[X_{n,k}^2\right]. \tag{2.17}$$

We want to apply Exercise 2.57 to $z_k = \mathbf{E}\left[e^{itX_{n,k}}\right]$ and $w_k = 1 - \frac{1}{2}\sigma_{n,k}^2 t^2$. Clearly $|z_k| \le 1$ by properties of c.f. If we prove that $\max_{k\le n} \sigma_{n,k}^2 \to 0$, then it will follow that $|w_k| \le 1$ and hence with $\theta = 1$ in Exercise 2.57, we get

$$\begin{aligned}
\limsup_{n\to\infty} \Big| \prod_{k=1}^{n} \mathbf{E}\left[e^{itX_{n,k}}\right] - \prod_{k=1}^{n}\left(1 - \frac{1}{2}\sigma_{n,k}^2 t^2\right) \Big| &\le \limsup_{n\to\infty} \sum_{k=1}^{n} |R_{n,k}(t)| \\
&\le \frac{1}{6}|t|^3 \sigma^2 \delta \quad \text{(by 2.17)}
\end{aligned}$$

To see that $\max_{k\le n} \sigma_{n,k}^2 \to 0$, fix any $\delta > 0$ note that $\sigma_{n,k}^2 \le \delta^2 + \mathbf{E}\left[X_{n,k}^2 \mathbf{1}_{|X_{n,k}|>\delta}\right]$ from which we get

$$\max_{k\le n} \sigma_{n,k}^2 \le \delta^2 + \sum_{k=1}^{n} \mathbf{E}\left[X_{n,k}^2 \mathbf{1}_{|X_{n,k}|>\delta}\right] \to \delta^2.$$

As δ is arbitrary, it follows that $\max_{k\le n} \sigma_{n,k}^2 \to 0$ as $n \to \infty$. As $\delta > 0$ is arbitrary, we get

$$\lim_{n\to\infty} \prod_{k=1}^{n} \mathbf{E}\left[e^{itX_{n,k}}\right] = \lim_{n\to\infty} \prod_{k=1}^{n}\left(1 - \frac{1}{2}\sigma_{n,k}^2 t^2\right). \tag{2.18}$$

For n large enough, $\max_{k\le n} \sigma_{n,k}^2 \le \frac{1}{2}$ and then

$$e^{-\frac{1}{2}\sigma_{n,k}^2 t^2 - \frac{1}{4}\sigma_{n,k}^4 t^4} \le 1 - \frac{1}{2}\sigma_{n,k}^2 t^2 \le e^{-\frac{1}{2}\sigma_{n,k}^2 t^2}.$$

Take product over $k \le n$, and observe that $\sum_{k=1}^{n} \sigma_{n,k}^4 \to 0$ (why?). Hence,

$$\prod_{k=1}^{n}\left(1 - \frac{1}{2}\sigma_{n,k}^2 t^2\right) \to e^{-\frac{\sigma^2 t^2}{2}}.$$

From 2.18 and Lévy's continuity theorem, we get $\sum_{k=1}^{n} X_{n,k} \xrightarrow{d} N(0, \sigma^2)$. ■