Central limit theorems

Sayan Mukherjee

Given a sequence of random variables (assume iid for now) \( \{X_i, i \geq 1\} \) we would like to state that or when does

\[
\frac{S_n - n\mu}{\sigma \sqrt{n}} \Rightarrow \text{No}(0, 1), \quad S_n = \sum X_i, \quad \mu = \mathbb{E}X_i, \quad \sigma^2 = \text{Var}(X_i).
\]

**Sums of random variables** We start by studying the sum of iid random variables \( S_n = X_1 + \cdots + X_n \). Assume that the \( X_i \) are normal with mean \( \mu \) and variance \( \sigma^2 \), so \( S_n \) should have mean \( \sum_{i \leq n} \mu \) and variance \( \sum_{i \leq n} \sigma^2 \). Since the \( X_i \) have cdf \( F(x) \) then \( X_1 + X_2 \) will have cdf \( F(s) \) given by

\[
P(S_2 \leq s) = F(s) = \int_{x_1+x_2 \leq s} f_1(x_1) f_2(x_2) dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{s-x_2} f_1(x_1) f_2(x_2) dx_1 dx_2
\]

\[
f(s) = F'(s) = \int_{-\infty}^{\infty} f_1(s-x_2) f_2(x_2) dx_2 = \int_{-\infty}^{\infty} f_1(x_1) f_2(s-x_1) dx_1,
\]

the above is a convolution of \( f_1(x_1) \) and \( f_1(x_1) \) and

\[
f(s) = f_1(x_1) \ast f_2(x_2) \cdots \ast f_n(x_n), \quad F(s) = F_1(x_1) \ast F_2(x_2) \cdots \ast F_n(x_n).
\]

So the cdf of the sum seems not so simple and in general the above computation looks painful.

**Moment generating functions**

A simpler object to compute is based on the expectation of the sum, specifically the expectation of products of \( X_i \)

\[
\mathbb{E}e^{S_n} = \mathbb{E} \prod e^{X_i}.
\]

More generally for any complex number \( z \) we define the moment generating function (mgf)

\[
M_X(z) = \mathbb{E}e^{zS_n} = \prod_{i=1}^{n} \mathbb{E}e^{zX_i}.
\]

The function \( e^{zX} \) can explode so the expectation can be infinite even for well behaved functions, for example the Gamma distribution.

Examples of mgf for a few distributions follow

- **Binomial**: \( \text{Bi}(n, p) \) \( [1 + p(e^z - 1)]^N \) \( z \in \mathbb{C} \)
- **Neg Bin**: \( \text{NB}(\alpha, p) \) \( [1 + (p/q)(e^z - 1)]^{-\alpha} \) \( z \in \mathbb{C} \)
- **Poisson**: \( \text{Po}(\lambda) \) \( e^{\lambda(e^z-1)} \) \( z \in \mathbb{C} \)
- **Normal**: \( \text{No}(\mu, \sigma^2) \) \( e^{\mu+z\sigma^2/2} \) \( z \in \mathbb{C} \)
- **Gamma**: \( \text{Ga}(\alpha, \beta) \) \( (1 - z/\beta)^{-\alpha} \) \( \mathbb{R}(z) < \beta \)
- **Cauchy**: \( \text{Cauchy}(a, b) \) \( \frac{a}{\pi(a^2 + (x-b)^2)} e^{izb-a|z|} \) \( \mathbb{R}(z) = 0 \)

If the mgf exists it uniquely determines the cdf. So if it were to exist the following argument would allow us to prove a CLT for and iid \( \{X_n, n \geq 1\} \) with mean 0 and variance 1.

\[
\mathbb{E}e^{zS_n/\sqrt{n}} = \mathbb{E}e^{z \sum \frac{X_i}{\sqrt{n}}} = \mathbb{E} \prod_{i=1}^{n} e^{zX_i/\sqrt{n}} = (\mathbb{E}e^{zX_i/\sqrt{n}})^n,
\]
Taylor expand the above around zero

\[
\left(1 + \frac{z\mathbb{E}(X_1)}{\sqrt{n}} + \frac{z^2\mathbb{E}(X_1^2)}{2n} + \varepsilon\right)^n,
\]

and as \( n \to \infty \) the above goes to \( z^2/2 \) so the mgf converges to \( e^{z^2/2} \) which is the mgf for the standard normal.

The mgf gets its name because of the following property

\[
M'(z) = \mathbb{E}[X e^{zX}], \quad M''(z) = \mathbb{E}[X^2 e^{zX}],
\]

so \( M'(0) = \mathbb{E}(X) = \mu \) and \( M''(0) = \mathbb{E}(X^2) = \mu^2 + \sigma^2 \), so it generates moments, \( EX^k = M^{(k)}(0) \).

**Characteristic functions**

Note that the problem with the mgf exploding is in the real part of the complex number \( z = x + iy \) since

\[
e^{x+iy} = e^x \cos(y) + ie^x \sin(y),
\]

and \( \sin \) and \( \cos \) are bounded by one for any \( y \). So a solution is to drop the real part and look at the complex random variable \( e^{i\omega X} \) and define the characteristic function (ch. f)

\[
\phi_X(\omega) = E e^{i\omega X} = \int_{\mathbb{R}} e^{i\omega X} \mu_X(dx).
\]

The above is the Fourier transform of the density.

Examples of the ch. f for a few distributions follow

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PDF (( f ))</th>
<th>CH. F (( \phi ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial</td>
<td>( Bi(n, p) ) ( [1 + p(e^{i\omega} - 1)]^n )</td>
<td>( \phi_Bi(n, p)(\omega) = [1 + p(e^{i\omega} - 1)]^n )</td>
</tr>
<tr>
<td>Neg Bin</td>
<td>( NB(\alpha, p) ) ( [1 + (p/q)(e^{i\omega} - 1)]^{-\alpha} )</td>
<td>( \phi_NB(\alpha, p)(\omega) = [1 + (p/q)(e^{i\omega} - 1)]^{-\alpha} )</td>
</tr>
<tr>
<td>Poisson</td>
<td>( Po(\lambda) ) ( e^{\lambda(e^{i\omega} - 1)} )</td>
<td>( \phi_Po(\lambda)(\omega) = e^{\lambda(e^{i\omega} - 1)} )</td>
</tr>
<tr>
<td>Normal</td>
<td>( No(\mu, \sigma^2) ) ( e^{i\omega(\mu - \sigma^2 \omega^2/2)} )</td>
<td>( \phi_No(\mu, \sigma^2)(\omega) = e^{i\omega(\mu - \sigma^2 \omega^2/2)} )</td>
</tr>
<tr>
<td>Gamma</td>
<td>( Ga(\alpha, \beta) ) ( (1 - i\omega/\beta)^{-\alpha} )</td>
<td>( \phi_Ga(\alpha, \beta)(\omega) = (1 - i\omega/\beta)^{-\alpha} )</td>
</tr>
<tr>
<td>Cauchy</td>
<td>( Cauchy ) ( a\pi/(a^2+(x-b)^2) )</td>
<td>( \phi_Cauchy(x\omega) = a\pi/(a^2+(x-b)^2) )</td>
</tr>
</tbody>
</table>

One can use the ch. f to generate moments as well

\[
\phi(0) = 1, \quad \phi'(0) = i\mathbb{E}X, \quad \phi''(0) = -\mathbb{E}(X^2), \quad \phi^{(k)}(0) = i^k\mathbb{E}(X^k).
\]

It can also be used to get back centered quantities such as variance

\[
\log''(\phi)(0) = \frac{\phi''(0)\phi(0) - (\phi(0))^2}{\phi(0)^2} = -\mathbb{E}(X^2) + \mathbb{E}(X^2) = -\sigma^2.
\]

We will use the following Taylor expansion of the ch. f

\[
\log \phi(\omega) = 0 + i\mu\omega - \sigma^2/2 + O(\omega^3)
\]

\[
\phi(\omega) \approx e^{i\mu\omega - \sigma^2/2 + O(\omega^3)}
\]

The following fact about ch. f will be useful. For independent random variables \( X \) and \( Y \) and \( \alpha, \beta, \gamma \in \mathbb{R} \) and \( Z = \alpha + \beta X + \gamma Y \)

\[
\phi_Z(\omega) = E e^{i\omega(\alpha + \beta X + \gamma Y)} = e^{i\omega\alpha}\phi_X(\beta\omega)\phi_Y(\gamma\omega).
\]

**Existence and uniqueness of characteristic functions**

We need to show that two distributions \( \mu_1(x) \) and \( \mu_2(x) \) have the same Fourier transform \( \hat{\mu}(\omega) = \mathbb{E}[e^{i\omega X}] \) then \( \mu_1 = \mu_2 \). By this we mean uniqueness, existence is assumed by construction. If we show that the inverse transform

\[
\frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} \hat{\mu}(\omega)d\omega
\]
uniquely determines $\mu(x)$.

We show this for discrete and continuous random variables.

**Discrete distributions**

For integer valued discrete distributions observe $\phi(\omega + 2\pi) = \phi(\omega)$. We can recover $p_k = P(X = k)$ by inverting the Fourier series

$$
\phi(\omega) = E[e^{i\omega X}] = \sum p_k e^{ik\omega}
$$

$$
p_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ik\omega} \phi(\omega) d\omega.
$$

**Continuous distributions**

In the case of a continuous random variable we can write $\mu(dx) = f(x)dx$ and the ch. f

$$
\phi(\omega) = \hat{\mu}(\omega) = \int_{\mathbb{R}} e^{i\omega x} f(x) dx.
$$

Note that $|\phi(\omega)| \leq 1$ so for any $\varepsilon > 0$ the function $e^{-iy\omega - \varepsilon^2/2} \phi(\omega)$ is integrable. Define the function

$$
\gamma_\varepsilon(x) = \frac{1}{\sqrt{2\pi \varepsilon}} e^{-x^2 / 2\varepsilon}
$$

which has ch. f

$$
e^{-iy\omega - \varepsilon^2/2}.
$$

We now compute

$$
\frac{1}{2\pi} \int_{\mathbb{R}} e^{-iy\omega - \varepsilon^2/2} \phi(\omega) d\omega = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iy\omega - \varepsilon^2/2} \left[ \int_{\mathbb{R}} e^{i\omega x} f(x) dx \right] d\omega
$$

$$
= \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-i(x-y)\omega - \varepsilon^2/2} f(x) dx d\omega
$$

$$
= \frac{1}{2\pi} \int_{\mathbb{R}} \left[ \int_{\mathbb{R}} e^{-i(x-y)\omega - \varepsilon^2/2} d\omega \right] f(x) dx d\omega
$$

$$
= \frac{1}{2\pi} \int_{\mathbb{R}} f(x) dx d\omega = \frac{1}{\sqrt{2\pi \varepsilon}} \int_{\mathbb{R}} e^{-(x-y)^2/2\varepsilon} f(x) dx
$$

$$
= \gamma_\varepsilon * f(y).
$$

The question now is what does the following converge to

$$
\lim_{\varepsilon \to 0} \gamma_\varepsilon * f(y).
$$

The answer is

- if $f(x)$ is bounded and continuous the above converges to $f(y)$ uniformly,
- if $f(x)$ has a jump discontinuity at $x = y$ bounded the above converges pointwise to $\frac{f(y-) + f(y+)}{2}$
- to $\infty$ if $\mu(\{y\}) > 0$

This is the Fourier inversion formula and stats if the above integral exists

$$
f(x) = \lim_{\varepsilon \to 0} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iy\omega - \varepsilon^2/2} \phi(\omega) d\omega.
$$

The following two theorems about the Fourier transform are useful.

**Theorem 0.0.1** If $\int_{\mathbb{R}} |\mu(\omega)| d\omega < \infty$ then $\mu_\varepsilon \equiv \mu * \gamma_\varepsilon$ converges a.s. to an $L_1$ function $f(x)$, $\hat{\mu}_\varepsilon(\omega)$ converges to $\hat{f}(\omega)$, $\mu(A) = \int_A f(x) dx$, and $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\omega x} \hat{\mu}(\omega) d\omega$. 


Theorem 0.0.2  If \( \int_{\mathbb{R}} |x|^k \mu(dx) < \infty \) for an integer \( k > 0 \) then \( \hat{\mu} \) has continuous derivatives of order \( k \)

\[
\hat{\mu}^{(k)}(\omega) = \int_{\mathbb{R}} (ix)^k e^{i \omega x} \mu(dx).
\]

Conversely, if \( \hat{\mu}(\omega) \) has a derivative of finite even order \( k \) at \( \omega = 0 \) then \( \int_{\mathbb{R}} |x|^k \mu(dx) < \infty \) and \( E X^k = \int_{\mathbb{R}} x^k \mu(dx) = (-1)^{k/2} \hat{\mu}^{(k)}(0) \).

Central limit theorem

We now state and prove central limit theorems.

Theorem 0.0.3 Let \( \{X_n, n \geq 1\} \) be iid random variables with \( E(X_n) = \mu \) and \( V(X_n) = \sigma^2 < \infty \), then

\[
\lim_{n \to \infty} \frac{S_n - n\mu}{\sigma \sqrt{n}} \Rightarrow \mathcal{N}(0, 1).
\]

Proof.

\[
\phi_S(\omega) = \prod_{j=1}^{n} \phi \left( \frac{\omega}{\sqrt{n} \sigma^2} \right) e^{-i \mu \omega / \sqrt{n} \sigma^2}
\]

\[
= \prod_{j=1}^{n} \phi(s) e^{-i \mu s}, \quad s = \frac{\omega}{\sqrt{n} \sigma^2}
\]

\[
= e^{n (\log(s) - i s \mu)}
\]

\[
\log \phi_S(\omega) = n [ \log \phi(s) - i s \omega]
\]

\[
= n [ 0 + i \mu s - \frac{\sigma^2 s^2}{2} + O(S^3) ] - i n s \mu
\]

\[
= -\frac{n^2 \sigma^2}{2} \frac{\omega^2}{n \sigma^2} + O(n^{-1/2})
\]

\[
= -\frac{\omega^2}{2} + O(n^{-1/2}).
\]

This implies that

\[
\phi(\omega) \to e^{-\omega^2/2}, \quad \forall \omega \in \mathbb{R}
\]

\[
f(x) \to e^{-x^2/2} \square
\]

We now state a CLT for non iid random variables \( \{X_n, n \geq 1\} \) and \( E X_k = 0, E(X_n^2) = \sigma_k^2, s_n^2 = \sum \sigma_k^2 \).

A sequence \( \{X_k\} \) satisfies the Lindeberg condition if for all \( t > 0 \)

\[
\lim_{n \to \infty} \frac{1}{s_n^2} \sum_{k=1}^{n} E(X_k I_{|x_k| > t s_n}) = \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x| > t s_n} x^2 F_k(dx) = 0.
\]

The above condition implies the more intuitive condition that

\[
\max_{k \leq n} \frac{\sigma_k^2}{s_n^2} \to 0,
\]

no single random variable in the sequence dominates the total variance.

Theorem 0.0.4 The Lindeberg condition is necessary and sufficient for

\[
\lim_{n \to \infty} \frac{S_n}{s_n} \Rightarrow \mathcal{N}(0, 1).
\]