Central Limit Theorem (CLT)

Simple version of CLT:

If \( \{ Y_i \} \) is a sequence of i.i.d. r.v with common mean 0 and variance 1, then

\[
Z_n = \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \Rightarrow N,
\]

where \( N \) is the standard normal r.v.

**Proof**: If \( \varphi \) denotes the c.f. of \( Y_i \), then the c.f. \( \varphi_n \) of \( Z_n \) is given by

\[
\varphi_n(s) = \varphi^n(s/\sqrt{n}).
\]

Recall that as \( E(X^2) < \infty \),

\[
\varphi(s) = 1 + isE(X) - \frac{s^2}{2}E(X^2) + h(s)s^2,
\]

where \( h(s) \to 0 \) as \( s \to 0 \).

Since \( E(X) = 0 \) and \( E(X^2) = 1 \),

\[
\varphi(s) - 1 + (s^2/2) = o(s^2) \text{ as } s \to 0,
\]

and hence

\[
|\varphi_n(t) - e^{-t^2/2}| \leq n \left| \varphi \left( \frac{t}{\sqrt{n}} \right) - 1 + \frac{t^2}{2n} \right| + n \left| 1 - \frac{t^2}{2n} - e^{-t^2/2n} \right| \to 0.
\]

Why Gaussian limit in the CLT?

1. Why the limit distribution in the CLT is Normal? What is special about Normal distribution?

2. Can any other distribution be the limit in CLT under different conditions?

Let \( \{ X_n \} \) be i.i.d. r.v.s with a common symmetric distribution.

Put \( S_n = X_1 + \cdots + X_n \). Suppose \( \frac{1}{\sqrt{n}}S_n \Rightarrow F \).

Let \( g \) and \( f \) denote the characteristic functions of \( X_1 \) and \( F \).

Clearly \( g^n(t/\sqrt{n}) \to f(t) \) for all \( t \). So the c.f. \( g^{2n}(t/\sqrt{2n}) \) at \( t \) of

\[
\frac{S_{2n}}{\sqrt{2n}} = \left( \frac{S_n}{\sqrt{n}} + \frac{S_{2n} - S_n}{\sqrt{n}} \right) \frac{1}{\sqrt{2}}
\]

converges to \( f(t) \) as well as to \( f^2(t/\sqrt{2}) \). Thus \( f(t) = f^2(t/\sqrt{2}) \).

Similarly \( f(t) = f^m \left( \frac{t}{\sqrt{m}} \right) \) for all \( t \), as \( \frac{S_{mn}}{\sqrt{mn}} \Rightarrow F \) for any integer \( m \geq 1 \).
As $X_1$ has symmetric distribution, $g$ and $f$ are real valued functions and since $f(t) = f^2(t/\sqrt{2})$ for all $t$, $f(t) \geq 0$.

As $f$ is continuous at zero and $f(0) = 1$, $f(t) > 0$ for all $t$. Putting
\[ h(y^2) = \log f(y), \]
\[ f(t) = f^m \left( \frac{t}{\sqrt{m}} \right) \quad \text{implies} \quad h(y^2) = mh \left( \frac{y^2}{m} \right). \]

Hence for any integer $r \geq 1$,
\[ h(r) = mh(r/m), \quad h(1) = mh(1/m), \quad h(r) = rh(1). \]
Thus $h(r/m) = (r/m)h(1)$, and by continuity of $h$,
\[ h(x) = xh(1) \text{ for all } x > 0. \]
So $f(t) = e^{ct^2}$, where $c = \log f(1)$.
Clearly $c < 0$, and so $F$ is Gaussian.

By replacing the normalizing $\sqrt{n}$ with $1/\alpha$, $0 < \alpha < 2$, we get stable laws as the limiting distributions.

In the case of $\alpha = 1$, the sample mean of independent Cauchy random variables is Cauchy. (Use its c.f. $\phi(t) = e^{-|t|}$.)

### Lindeberg Theorem

**Theorem:** If $X_{n1}, \ldots, X_{nr_n}$ are independent, $S_n = X_{n1} + \cdots + X_{nr_n}$, $E(X_{nk}) = 0$, $\sigma_{nk}^2 = E(X_{nk}^2)$, $s_n^2 = \sum_{k=1}^{r_n} \sigma_{nk}^2$, and Lindeberg condition
\[ L_n(\epsilon) = \sum_{k=1}^{r_n} \frac{1}{s_n^2} \int_{\{|X_{nk}| \geq \epsilon s_n\}} X_{nk}^2 dP \to 0 \quad \text{holds, then} \quad S_n/s_n \Rightarrow N. \]

**Proof:** Assume $s_n = 1$. Lindeberg condition implies $\max_{k \leq r_n} \sigma_{nk}^2 \to 0$. Now use
\[ |z_1 \cdots z_n - w_1 \cdots w_n| \leq \sum_{k=1}^{n} |z_k - w_k| \text{ for } \max(|z_k|, |w_k|) \leq 1, \]
and
\[ |E(e^{itX_{nk}}) - (1 - \frac{1}{2} t^2 \sigma_{nk}^2)| \leq E(\min\{|tX_{nk}|^2, |tX_{nk}|^3\}) \]
\[ \leq \int_{\{|X_{nk}| < \epsilon\}} |tX_{nk}|^3 dP \]
\[ + \int_{\{|X_{nk}| \geq \epsilon\}} |tX_{nk}|^2 dP \]
\[ \leq \epsilon |t|^3 \sigma_{nk}^2 + t^2 \int_{\{|X_{nk}| \geq \epsilon\}} X_{nk}^2 dP. \]
**Theorem**: Lyapounov’s condition

\[
\sum_{k=1}^{r_n} s_n^{-2-\delta} E(|X_{nk}|^{2+\delta}) \to 0 \quad \text{for some } \delta > 0,
\]

then Lindeberg condition holds. In this case \( S_n/s_n \Rightarrow N \).

**Theorem**: If \( X_{n1}, \ldots, X_{nr_n} \) are independent with zero mean and \( S_n/s_n \Rightarrow N \), then Lindeberg condition holds provided

\[
\max_{k \leq r_n} \sigma_{nk}/s_n \to 0.
\]

(*)

The Lindeberg condition implies (*). The result is false without (*).

**Example**: Let \( \sigma_1 = 1 \) and \( \sigma_n^2 = ns_{n-1}^2 \). Take \( X_{nk} = X_k \) normal with mean 0 and variance \( \sigma_{nk}^2 = \sigma_k^2 \). \( S_n/s_n \) has the standard normal distribution for all \( n \). However, as \( s_n^2 = \sigma_n^2(n+n)/n \leq 2\sigma_n^2 \),

\[
L_n(\epsilon) = \frac{1}{s_n^2} \int \frac{1}{s_n^2} \int \{ |X_{nk}| \geq \epsilon s_n \} X_{nk}^2 dP \\
\geq \frac{1}{s_n^2} \int \{ |X_n| \geq \epsilon s_n \} X_n^2 dP \geq \frac{1}{2} \int \lim_{\epsilon \to \infty} \{ |X_n| \geq \epsilon \} X_n^2 dP.
\]

**δ-method**: If CLT holds for \( \{X_i\} \) \( (\sqrt{n}(X_n - c)/\sigma \Rightarrow N) \), and \( f \) has a non-zero derivative \( f'(c) \) at \( c \), then \( \sqrt{n}(f(\bar{X}_n) - f(c))/\sigma |f'(c)| \Rightarrow N \). Use Skorohod’s theorem to get \( Y_n, Y \) on a single probability space, with distributions of \( \bar{X}_n, N \) respectively, so that \( \sqrt{n}(Y_n(\omega) - c)/\sigma \to Y(\omega) \) for all \( \omega \) and hence \( Y_n(\omega) \to c \) for all \( \omega \). As

\[
\frac{f(Y_n(\omega)) - f(c)}{Y_n(\omega) - c} \to f'(c), \quad \frac{\sqrt{n}}{\sigma}(f(Y_n(\omega)) - f(c)) \to f'(c) Y(\omega).
\]

If \( f(x) = \frac{1}{x} \), and \( X_i \) i.i.d. exponential with mean \( \frac{1}{\alpha} \), then

\[
\frac{\sqrt{n}}{\alpha} \left( \frac{1}{\bar{X}_n} - \alpha \right) \Rightarrow N.
\]

**CLT for a random number of summands (Problem 27.14)**: \( \nu_t \) sequence of positive integer valued r.v., \( X_j \) i.i.d. r.v. \( S=S_n = X_1 + \cdots + X_n \). Suppose \( a_t \) positive constants tending to infinity, and \( \nu_t/a_t \to_p \theta \). Then

\[
\frac{S_{\nu_t}}{\sigma \sqrt{\nu_t}} \Rightarrow N, \quad \frac{S_{\nu_t}}{\sigma \sqrt{\theta a_t}} \Rightarrow N.
\]

Assume \( \theta = 1 \) and use

\[
P(|S_{\nu_t} - S_{[a_t]}| \geq \epsilon \sqrt{a_t}) \leq P(|\nu_t - a_t| \geq \epsilon a_t)
\]

\[
+ P \left[ \max_{k-a_t \leq a_t} |S_k - S_{[a_t]}| \geq \epsilon \sqrt{a_t} \right].
\]
Problem 29.8: Let $V_n = (V_{n1}, \ldots, V_{nk})$ be i.i.d. vectors with only one coordinate is 1 and the rest 0s. $Y_n = \sum_{r=1}^{n} V_r$ has $\text{Mult}(n; p_1, \ldots, p_k)$ distribution. $\sigma_{ij} = \left(\delta_{ij}p_j - p_i p_j\right)/\sqrt{p_i p_j} = \delta_{ij} - \sqrt{p_i} \sqrt{p_j}$ is the $ij$-th entry of the dispersion $\Sigma$ of $X_{ni} = (V_{ni} - p_i)/\sqrt{p_i}$.

$\sqrt{n} \overline{X}_n$ converges weakly to $X$, a centered normal distribution with dispersion $\Sigma = I - uu'$, where $u' = (\sqrt{p_1}, \ldots, \sqrt{p_k})$.

$\Sigma u = 0$, so 0 is an eigenvalue, and $\Sigma y = y$ if $y'u = 0$. So 1 is an eigenvalue of multiplicity $k - 1$. Let $y_1, \ldots, y_{k-1}$ be orthonormal ($\|y_i\| = 1$) vectors that are orthogonal to $u$.

Sum of squares of $m$ independent standard normal random variables has Chi-squared distribution with $m$-d.f.

Put $U = (y_1, \ldots, y_{k-1}, u)$, $Z = U'X$. Then $UU' = I$, and $E(ZZ') = U' \Sigma U = U' (I - uu') U$. So the first $k - 1$ coordinates of $Z$ are i.i.d. standard normal variables and $\sum_{i=1}^{k-1} Z_i^2 = \sum_{i=1}^{k} X_i^2$, as $Z_k = 0$. The ‘Chi-squared statistic’ $\sum_{i=1}^{k} (Y_{ni} - np_i)^2/np_i$ is thus asymptotically distributed as chi-squared distribution with $k - 1$ d.f.