§ 8.2 - 8.4 Laws of Large Numbers, Central Limit Theorem

**Theorem 1** (Weak law of large numbers)

If \( X_1, X_2, \ldots, X_n, \ldots : \mathbb{S} \rightarrow \mathbb{R} \) are independent r.v.s, identically distributed and \( E[X_i] = \mu \) then

for all \( \varepsilon > 0 \)

\[
\lim_{n \to \infty} P \left( \left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \mu \right| \geq \varepsilon \right) = 0
\]

**Theorem 2** (Strong law of large numbers)

If \( X_1, X_2, \ldots, X_n, \ldots : \mathbb{S} \rightarrow \mathbb{R} \) are independent r.v.s, identically distributed and \( E[X_i] = \mu \) then

\[
P \left( \lim_{n \to \infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu \right) = 1
\]
Theorem 3 (Central Limit Theorem)

If $X_1, X_2, \ldots, X_n : S \rightarrow \mathbb{R}$ are independent r.v.,
identically distributed and $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$
then, for all $\alpha \in \mathbb{R}$

$$
\lim_{n \to \infty} \Pr \left( \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma \sqrt{n}} \leq \alpha \right) = \Phi(\alpha)
$$

$$
= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} \, dx
$$

Remarks (i) Central limit theorem states that
the distribution function for

$$
\frac{\sum_{i=1}^{n} (X_i - \mu) / \text{Var}(\sum_{i=1}^{n} X_i)}{\text{Var}(\sum_{i=1}^{n} X_i)^{1/2}}
$$

converges to the distribution function of the
standard normal random variable. We have
already used it to approximate probabilities of
sum of (independent) random variables.

(ii) Strong law of large numbers
says that the set of outcomes $\omega \in \Omega$ on
which the sample mean \( \frac{\sum_i x_i}{m} \) converges to the expected value forms a set of probability one, or, equivalently, the sample mean converges to the expected value almost surely!

In particular if \( A \) is an event and

\[ x_i = 1 \quad \text{if} \quad A \text{ occurs at } i \text{th trial} \]

then the sample mean = frequency = \( \frac{\sum x_i}{m} \) converges to \( \mathbb{E}[x_i] = P(A) \) almost surely!

(iii) Weak law of large numbers is a weaker version of Theorem 2 and states that the sample mean \( \frac{x_1 + x_2 + \ldots + x_n}{n} \) converges in probability to the expected value \( \mu \), i.e. the probability that \( \frac{x_1 + x_2 + \ldots + x_n}{n} \) is not in \((\mu - \varepsilon, \mu + \varepsilon)\) goes to zero as \( n \to \infty \), for any \( \varepsilon > 0 \).
In what follows I will sketch the proofs using the characteristic function:

\[ \phi_X(t) = \mathbb{E} \left[ e^{itX} \right] \text{, see Lecture 2.} \]

This method does not require extra hypotheses such as: finite variance, see the proof of Theorem 1 in the textbook; finite fourth moment, see the proof of Theorem 2 in the textbook; the existence of moment-generating function, in particular all moments must be finite, see the proof of Theorem 3 in the textbook.

All will require the following Levy Continuity Theorem:

Lemma 1. If the characteristic functions of the sequence \( \{X_n\} \) of random variables satisfies

\[ \lim_{n \to \infty} \phi_{X_n}(t) = \phi_X(t) \text{ for all } t \text{, and } \phi_X(t) \text{ is the characteristic function of a r.v. } Y \text{ then} \]

\[ \lim_{n \to \infty} F_{X_n}(x) = F_Y(x) \text{ for all } x \in \mathbb{R} \text{ at which } F_Y \text{ is continuous.} \]
Proof of Th 1.: Let
\[ \Phi_{X_1}(t) = \mathbb{E} \left[ e^{itX_1} \right] \]
then by independence:
\[ \Phi_{X_1 + X_2 + \ldots + X_n}(t) = \mathbb{E} \left[ e^{it\sum_{i=1}^{n} X_i} \right] \]
\[ = \Phi_{X_1}(\frac{t}{n}) \Phi_{X_2}(\frac{t}{n}) \cdots \Phi_{X_n}(\frac{t}{n}) \]
\[ = \left( \Phi_{X_1}(\frac{t}{n}) \right)^n \]
Let \( L_n(t) = \log \left( \Phi_{X_1}(\frac{t}{n}) \right)^n = n \log \Phi_{X_1}(\frac{t}{n}) \)
\[ \lim_{n \to \infty} L_n(t) = \lim_{n \to \infty} \frac{1}{n} \log \frac{\Phi_{X_1}^n(\frac{t}{n})}{\Phi_{X_1}(\frac{t}{n})} = t \frac{\Phi'_{X_1}(0)}{\Phi_{X_1}(0)} = it \]
L'Hopital
\[ \lim_{n \to \infty} \frac{\Phi_{X_1}(\frac{t}{n}) - \frac{t}{n}}{\Phi_{X_1}(\frac{t}{n})} = t \frac{\Phi'_{X_1}(0)}{\Phi_{X_1}(0)} \]
[where \( \text{used } \Phi'_{X_1}(0) \text{ exists} \) and is continuous due to \( \mathbb{E} \left[ X_1^2 \right] > 0 < \infty \), see lecture 21.]
So \[ Y_{\frac{x_1 + x_2 + \ldots + x_n}{n}} (t) \xrightarrow{n \to \infty} e^{\lambda t} = Y_{X \equiv \lambda} (t) \]

\[ \Rightarrow F_{\frac{x_1 + x_2 + \ldots + x_n}{n}} (a) \xrightarrow{n \to \infty} F (a) = \begin{cases} 0 & a < \mu \\ 1 & a \geq \mu \end{cases} \]

at any \( a \neq \mu \).

W.e.r., for any \( \varepsilon > 0 \),

\[
P \left\{ \left| \frac{X_1 + \ldots + X_n}{n} - \mu \right| > \varepsilon \right\} = F_{\frac{X_1 + \ldots + X_n}{n}} (\mu - \varepsilon) + 1 - F_{\frac{X_1 + \ldots + X_n}{n}} (\mu + \varepsilon - \delta)
\]

Passing to the limit as \( n \to \infty \) we get

\[
\lim_{n \to \infty} P \left\{ \left| \frac{X_1 + \ldots + X_n}{n} - \mu \right| > \varepsilon \right\} = F (\mu - \varepsilon) + 1 - F (\mu + \varepsilon - \delta)
\]

\[ = 0 + 1 - 1 = 0 \]
Remark. A known result in measure theory states that if \( \frac{\sum_{i=1}^{n} x_i}{n} \) converges in probability to \( \mu \), then there is a subsequence converging to \( \mu \) almost surely, i.e., Theorem 2 holds on a subsequence. To show that the entire sequence converges almost surely to \( \mu \) we need careful estimates which are consequences of Chebyshev's inequality, a corollary of Markov's inequality:

**Proposition 1 (Markov Inequality)** If \( X \geq 0 \) then for any \( a > 0 \)

\[
P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}
\]

**Proof** Let \( Y : S \rightarrow \mathbb{R} \) \( Y = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases} \)

\[
\Rightarrow \mathbb{E}[Y] = P(X \geq a) \quad \text{and} \quad Y \leq \frac{X}{a}
\]

\[
\Rightarrow P(X \geq a) = \mathbb{E}[Y] \leq \frac{1}{a} \mathbb{E}[X] \quad \blacksquare
\]
Proposition 2 (Chebyshev’s Inequality) \( \frac{1}{\sigma^2} \mathbb{E} [X] = \mu \) and \( \text{Var} (X) = \sigma^2 \) then for any \( a > 0 \)

\[ P \left( |X - \mu| \geq a \right) \leq \frac{\sigma^2}{a^2} \]

Proof. Let \( Z = (X - \mu)^2 \geq 0 \) \( \Rightarrow \) \( \mathbb{E} [Z] = \sigma^2 \)

By Markov’s Inequality, for any \( a^2 > 0 \) we have

\[ P \left( Z \geq a^2 y \right) \leq \frac{\mathbb{E} [Z]}{a^2} \quad \text{i.e.} \]

\[ P \left( (X - \mu)^2 \geq a^2 y \right) \leq \frac{\sigma^2}{a^2} \]

But \( \{ (X - \mu)^2 \geq a^2 y \} = \{ |X - \mu| \geq a \} \)

Sketch of Proof of Theorem 2:

We first prove the result under the assumption \( X_u \geq 0 \) for all \( u \). The general result is then obtained using

\[ X_u = X_u^+ - X_u^- \quad \text{where} \quad I(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \]
$$X^{+}_n = X_n I(x_n); X^{-}_n = -X_n I(-x_n), \text{ hence}$$

$$x^{+}_n, x^{-}_n \geq 0 \text{ and applying the result for non-negative}$$

$$n \times n \text{ case we get}$$

$$\frac{1}{n} \sum_{i=1}^{n} x^{+}_i - \frac{1}{n} \sum_{i=1}^{n} x^{-}_i \xrightarrow{P} \mu^+ - \mu^- = \mu \text{ with probability 1}$$

i.e. the statement of the theorem.

So, assume $X_n \geq 0 \text{ for all } n$. Then

$$S_n = x_1 + x_2 + \ldots + x_n \text{ is increasing}!!$$

To better understand the convergence properties of

$$S_n / n$$

We are going to use Chebyshev's Inequality:

$$P\left( \left| \frac{S_n}{n} - \mu \right| \geq \epsilon \right) \leq \frac{\text{Var} \left( \frac{S_n}{n} \right)}{\epsilon^2} = \frac{\text{Var} (X_i)}{\epsilon^2}$$

But for it to be useful we need to replace $X_i$ with something for sure less finite variation:
Let \( Y_n = X_n \cdot I_{(X_n)} \) where

\[ I_{(x)}(x) = \begin{cases} 1 & \text{if } -n \leq x \leq n \\ 0 & \text{otherwise} \end{cases} \]

Then \( |Y_n| \leq n \) and \( Y_n \leq X_n \) hence

\[ E[I_{(X_n)}] \leq E[I_{(Y_n)}] = n \]

\[ \lim_{n \to \infty} E[I_{(Y_n)}] = \lim_{n \to \infty} \mu_n = \mu \] (since, for example, for a continuous random variable)

\[ E[I_{(X_n)}] = \int_{-\infty}^{\infty} x f_{X_n}(x) \, dx \]

\[ = \int_{-n}^{n} x f_{X_n}(x) \, dx \quad \text{as} \quad \mu_n \to \infty \]

\[ = E[X_n] = \mu \]

Also \( \text{Var}(Y_n) \leq E[Y_n^2] \leq E[\mu^2] = \mu^2 < \infty \)
Let \( S_n = Y_1 + Y_2 + \ldots + Y_n \)

First we show \( \forall \varepsilon \lim_{n \to \infty} \left| \frac{S_n}{n} - \frac{S_n^*}{n} \right| \to 0 \)  \( \forall \varepsilon > 0 \)

i.e. \( \frac{S_n}{n} - \frac{S_n^*}{n} \to 0 \) almost surely!

Let \( A = \{ \omega \in S \mid \frac{S_n}{n}(\omega) - \frac{S_n^*}{n}(\omega) \to 0 \} \)

It suffices to show \( P(A) = 0 \). Let \( \omega \in A \)

Then for any \( \varepsilon > 0 \) there is \( N_{\varepsilon} \) such that

\[
\forall \omega \in S_{\varepsilon} \quad X_{n\omega}(\omega) \neq Y_{n\omega}(\omega)
\]

Otherwise, there is \( \omega \in S_{\varepsilon} \) fixed such that

\[
X_n(\omega) = Y_n(\omega) \quad \forall \omega \in S_{\varepsilon} \quad \text{so}
\]

\[
\left| \frac{S_n}{n} - \frac{S_n^*}{n} \right| \leq \left| \frac{X_{1\omega}(\omega) + \ldots + X_{n\omega}(\omega)}{n} \right| \to 0
\]

We showed

\[
A = \bigcap_{\varepsilon > 0} \bigcup_{\omega \in S} \{ \omega \in S \mid X_{n\omega}(\omega) \neq Y_{n\omega}(\omega) \}
\]
\[ P(A) = \lim_{n \to \infty} \sum_{u=1}^{\infty} P(|X_u| > u) \leq \sum_{u=1}^{\infty} P(1|X| > u) \]

Since \( X_u \) are identically distributed.

\[ \leq \sum_{u=1}^{\infty} P(1|X| > u) \cdot dx = \mathbb{E}[|X|] < \infty \]

\[ \Rightarrow \lim_{n \to \infty} \sum_{u=1}^{\infty} P(\cap uA_n) = 0 \] (since these is the limit of the previous series)

\[ \Rightarrow P(A) = P(\cap \cup uA_n) = \sum_{u=1}^{\infty} P(A_u) \]

It remains to show

\[ \underset{n \to \infty}{\text{PS law}} \left| \frac{S_n^x}{n} - \mu^x \right| = 0 \quad \text{if} \quad y = 1 \]

where

\[ \mu^x = \mathbb{E}\left[ \frac{S_n^x}{n} \right] = \frac{\sum_{i=1}^{n} \mu^x_i}{n} \] and
\[
\lim_{n \to \infty} P_n = \lim_{n \to \infty} \sum_{i=1}^{n} \mu_i \quad \text{Gromov-} \quad \lim_{n \to \infty} \rho_n = \rho
\]

Let \( \alpha > \frac{1}{2} \) and
\[
U_n = \lfloor \alpha n \rfloor = \text{largest integer less equal to } \alpha n
\]
\[
U_0 = 0
\]

We first show
\[
P \left( \lim_{n \to \infty} \left| \frac{\sum_{i=1}^{n} \mu_i - \rho n}{U_n} \right| = 0 \right) \quad y = 1
\]

As before let
\[
B = \{ \omega \in S \mid \left| \frac{\sum_{i=1}^{n} \mu_i - \rho n}{U_n} \right| > \varepsilon \}
\]

If \( \omega \in B \) then there is an \( \varepsilon > 0 \) such that for all \( n \geq m \) there is \( n > m \) satisfying
\[
\left| \frac{\sum_{i=1}^{n} \mu_i - \rho n}{U_n} \right| > \varepsilon
\]

i.e.
\[
B = \cap_{m=1}^{\infty} \cup_{n=m}^{\infty} \{ \omega \in S \mid \left| \frac{\sum_{i=1}^{n} \mu_i - \rho n}{U_n} \right| > \varepsilon \} \quad y
\]
\[ P(\beta_n) = P \left( \left| \frac{S_{\beta_n}}{\sqrt{\sigma_n}} - E \frac{S_{\beta_n}}{\sigma_n} \right| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \text{Var} \left( \frac{S_{\beta_n}}{\sqrt{\sigma_n}} \right) = \frac{1}{\sigma_n^2 \varepsilon^2} \text{Var} \left( S_{\beta_n} \right) \]

By independence

\[ \text{Var} \left( S_{\beta_n} \right) = \sum_{j=1}^{\sigma_n} \text{Var} \left( Y_j \right) \]

\[ \leq \sum_{j=1}^{\sigma_n} E \left[ V_{\beta_n}^2 \right] = \sum_{j=1}^{\sigma_n} E \left[ X_j^2 \cdot I_j (x_1) \right] \]

\[ \leq \sum_{j=1}^{\sigma_n} E \left[ X_j^2 \cdot I_j (x_1) \right] \]

\[ = E \left[ X_1^2 \cdot \sum_{j=1}^{\sigma_n} I_j (x_1) \right] \]

\[ = E \left[ X_1^2 \cdot \sum_{j=1}^{\sigma_n} I_j (x_1) \right] \]

\[ \leq E \left[ X_1^2 \cdot \sum_{j=1}^{\sigma_n} I_j (x_1) \right] \]

\[ = \sigma_n E \left[ X_1^2 \cdot I_{\beta_n} (x_1) \right] \]
So \( P(\beta_m) \leq \frac{1}{\varepsilon^2} \mathbb{E} \left[ x_i^2 \frac{I_{U_m}(x_i)}{U_m} \right] \) and

\[
\sum_{u=1}^{\infty} P(\beta_m) \leq \frac{1}{\varepsilon^2} \mathbb{E} \left[ x_i^2 \sum_{u=1}^{\infty} \frac{I_{U_m}(x_i)}{U_m} \right]
\]

For each \( x \neq 0 \) let \( H(x) \) be the index for which

\[
U_{H(x)-1} < |x| \leq U_H(x) = 1 \cdot 2^{H(x)} \leq 2^{H(x)}
\]

Note that \( I_{\beta_m}(x) = 0 \) for \( m < H(x) \) so:

\[
x^2 \sum_{u=1}^{\infty} \frac{I_{U_m}(x)}{U_m} \leq x^2 \sum_{u=1}^{H(x)} \frac{1}{U_m} \leq x^2 \sum_{u=1}^{\infty} \frac{1}{2^{U_m}}
\]

\[
= x^2 \cdot \frac{1}{2^{H(x)-1}} \cdot \frac{2}{2-1} \leq \frac{|x| \cdot 2}{2^{H(x)-1}}
\]

So \( x^2 \sum_{u=1}^{\infty} \frac{I_{U_m}(x)}{U_m} \leq \frac{2}{2-1} |x| \) for all \( x \in \mathbb{R} \)

\[
\Rightarrow \mathbb{E} \left[ \sum_{u=1}^{\infty} \frac{x_i^2 \cdot I_{U_m}(x_i)}{U_m} \right] \leq \frac{2}{2-1} \mathbb{E} \left[ |x_i| \right] < \infty
\]

\[
\Rightarrow \sum_{u=1}^{\infty} \mathbb{P}(\beta_m) < \infty
\]
as before the above shows

\[ P(B) = 0 \Rightarrow P(\lim_{n \to \infty} \left| \frac{S_n}{n} - \mu \right| = 0) = 1 \]

To show that for the entire sequence

\[ P(\lim_{n \to \infty} \left| \frac{S_n}{n} - \mu \right| = 0) = 1 \]

Note that \( S_n^* \) is increasing, due to \( Y_i \geq 0 \), so

\[ \frac{S_n}{\sqrt{n}} \leq \frac{S_m}{\sqrt{m}} \leq \frac{S_{m+1}}{\sqrt{m+1}} \]

Or

\[ \frac{1}{2} \frac{S_n}{\sqrt{n}} \leq \frac{S_m}{\sqrt{m}} \leq 2 \frac{S_m}{\sqrt{m+1}} \]

On a set of prob 1

Since \( \lambda > 1 \), arbitrarily we deduce

\[ P(\lim_{n \to \infty} \frac{S_n}{\sqrt{n}} = \mu) = 1 \text{ hence } P(\lim_{n \to \infty} \frac{S_n}{\sqrt{n}} = \mu) = 1 \]

\[ \therefore \]
Proof of Th 3. Let

\[ Y_m = \frac{X_m - \mu}{\sigma} \]

then \( Y_m \) are identically distributed, independent and

\[ \mathbb{E} [Y_m] = 0 \quad \mathbb{V}[Y_m] = \mathbb{E}[Y_m^2] = 1 \]

hence the characteristic functions satisfy:

\[ \varphi_{Y_m} = \varphi_Y \] and \( \varphi_Y \) together with

its first and second derivative exists, are continuous and:

\[ \varphi_{Y_1}(0) = 1 \]
\[ \varphi_{Y_1}'(0) = i \mathbb{E} [Y_1] = 0 \]
\[ \varphi_{Y_1}''(0) = i^2 \mathbb{E} [Y_1^2] = -1 \]

See Lecture 21.
Let
\[ Z_n = \frac{\sum_{i=1}^{n} Y_i - n \mu}{\sigma_1 \sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \]

As in the proof of Theorem 1, we will use the characteristic function:
\[
Z_n(t) = e^{i \frac{t}{\sqrt{n}} \sum_{i=1}^{n} Y_i} = \mathcal{F}_Y \left( \frac{t}{\sqrt{n}} \right)^n
\]

where I used independence of \( Y_i \) and the fact that they are identically distributed \( \Rightarrow Y_i \sim Y_1 \).

Let \( L_n(t) = \ln \mathcal{F}_Y \left( \frac{t}{\sqrt{n}} \right)^n = n \ln \mathcal{F}_Y \left( \frac{t}{\sqrt{n}} \right) \)

Applying twice L'Hospital we get
\[
\lim_{n \to \infty} L_n(t) = \lim_{n \to \infty} \ln \mathcal{F}_Y \left( \frac{t}{\sqrt{n}} \right)
\]
\[
= \lim_{n \to \infty} \frac{1}{n} \mathcal{F}_Y' \left( \frac{t}{\sqrt{n}} \right)
\]
\[
= \lim_{n \to \infty} \frac{2 t^{3/2}}{n^{3/2}} \mathcal{F}_Y \left( \frac{t}{\sqrt{n}} \right)
\]
\[- \frac{t}{2} \lim_{n \to \infty} \frac{\varphi_{Y_n}(\frac{t}{\sqrt{n}})}{\sqrt{n}} = \lim_{n \to \infty} \varphi_{X_n}(\frac{t}{\sqrt{n}}) \cdot t = -\frac{t^2}{2} \]

\[\lim_{n \to \infty} \frac{Y_n^2(t)}{2n} = 0 \]

where $Z$ is the standard normal random variable

\[f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\]

By Levy's Continuity Theorem (Lemma 1 pg 4)

\[F_{\frac{Y_n^2}{n}}(a) \xrightarrow{n \to \infty} F_Z(a) = \Phi(a) \text{ for all } a \in \mathbb{R} \]

(since $F_Z(a)$ is continuous)

\[\Rightarrow P\left( \frac{Y_1 + \cdots + Y_n}{\sqrt{n}} \leq a \right) \xrightarrow{n \to \infty} \Phi(a) \]

\[\Rightarrow P\left( \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}} \leq a \right) \xrightarrow{n \to \infty} \Phi(a) \]