Lecture 26

Let us recall the following from previous lecture.

Remark 0.1 *Let $X_n, n \geq 1, X$ be random variables. Then following relations holds.*

- $X_n \to X$ *a.s.* $\Rightarrow X_n \to X$ *in Probability .*

  Recall that

$$\left\{ \lim_{n\to\infty} X_n = X \right\} \;=\; \bigcap_{\varepsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\}.$$

  Now

$$\begin{aligned} P\Big(\bigcap_{\varepsilon>0} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \;=\; 1 \;&\Rightarrow\; P\Big(\bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \;=\; 1 \\ &\quad\; \text{for all } \varepsilon > 0 \\ &\Rightarrow\; \lim_{n\to\infty} P\Big(\bigcap_{m=n}^{\infty} \{|X_m - X| \leq \varepsilon\} \Big) \;=\; 1 \\ &\quad\; \text{for all } \varepsilon > 0 \\ &\Rightarrow\; \lim_{n\to\infty} P\big(|X_n - X| \leq \varepsilon\big) \;=\; 1 \\ &\quad\; \text{for all } \varepsilon > 0. \end{aligned}$$

  This implies convergence in probability.

- *Converse is not true. i.e. convergence in probability need not imply convergence a.s.*

  For example, consider $X_n, n \geq 1,\; X_n \sim$ Bernoulli $(\frac{1}{n})$ and are independent.

  For $\varepsilon > 0$,

$$P(|X_n| > \varepsilon) = P\big(X_n = 1\big) = \frac{1}{n} \to 0 \text{ as } n \to \infty.$$

  Hence $X_n \to 0$ in probability.

  Now note that
  $\displaystyle\sum_{n=1}^{\infty} P\big(X_n = 1\big) \;=\; \infty$. *Hence using Borel-Cantelli lemma, it follows that $P\big(X_n = 1 \text{ i.o}\big) = 1$. Hence X_n doen't converge to 0 a.s.*

Note that

$$\sum_{n=1}^{\infty} P\big(X_{n^2} = 1\big) \;=\; \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

Hence using Borel-Cantelli lemma, we have

$$P\big(X_{n^2} = 1 \text{ i.o.}\, n\big) = 0.$$

i.e.

$$P\Big(\bigcup_{n} \bigcap_{m \geq n} \{X_{m^2} = 0\}\Big) \;=\; 1.$$

Hence

$$P(\lim_{n\to\infty} X_{n^2} = 0) \;=\; 1.$$

i.e., $X_{n^2} \to 0$ *a.s. as* $n \to \infty$. *So there is a subsequence along which convergence is a.s. This is an illustration for the following result.*

- $X_n \to X$ *in probability* $\Rightarrow$ *along a subsequence* $X_n \to X$ *a.s.*

- $X_n \to X$ *in* m^{th} *moment* $\Rightarrow$ $X_n \to X$ *in Probability.*

  Using Markov inequality we get

$$P\big(|X_n - X| > \varepsilon\big) \;\leq\; \frac{1}{\varepsilon^m} E|X_n - X|^m.$$

  Since

$$\lim_{n\to\infty} E|X_n - X|^m \;=\; 0, \text{ we get } \lim_{n\to\infty} P\big(|X_n - X| > \varepsilon\big) = 0, \text{ for all } \varepsilon > 0.$$

  This completes the proof.

- $X_n \to X$ *in probability* $\Rightarrow$ $X_n \to X$ *in distribution.*

  The proof below uses the technique[1]

[1] Give a random variable Z with finite mean, and $A \in \mathcal{F}$, an event,

$$EZ = E[ZI_A] + E[ZI_{A^c}].$$

For $\varepsilon > 0$, consider

$$
\begin{aligned}
|\Phi_{X_n}(t) - \Phi_X(t)| &= |E[e^{itX_n} - e^{itX}]| \\
&\leq E|e^{it(X_n - X)} - 1| \;=\; E\sqrt{2(1 - \cos t(X_n - X))} \\
&= 2E\Big|\sin\frac{t(X_n - X)}{2}\Big| \\
&= 2E\Big[\Big|\sin\frac{t(X_n - X)}{2}\Big| I_{\{|X_n - X| \leq \varepsilon\}}\Big] \\
&\quad +2E\Big[\Big|\sin\frac{t(X_n - X)}{2}\Big| I_{\{|X_n - X| > \varepsilon\}}\Big] \\
&\leq |t|\, E\big[|X_n - X)|\, I_{\{|X_n - X| \leq \varepsilon\}}\big] \;+ P\big(|X_n - X| > \varepsilon\big) \\
&\leq |t|\,\varepsilon + P\big(|X_n - X| > \varepsilon\big).
\end{aligned}
$$

Hence

$$
\lim_{n\to\infty} |\Phi_{X_n}(t) - \Phi_X(t)| \;\leq\; |t|\,\varepsilon, \text{ for all } \varepsilon > 0.
$$

i.e.,

$$
X_n \to X \text{ in probability } \;\Rightarrow\; \Phi_{X_n}(t) \to \Phi_X(t), \text{ for all } t \in \mathbb{R}.
$$

Hence by the continuity theorem (see the Section on Characteristic functions), $F_{X_n}(x) \to F_X(x)$ at all continuity points of F.

Following is a useful technical result for the rest of the chapter. The result given below relate the event on a.s. convergence of ranodm variables with limsup or liminf of suitable events and hence Borel-Cantell lemma becomes a natural tool to show a.s. convergence of random variables.

Lemma 0.1 *Let $X_n, X, n \geq 1$ be random variables such that $P(\{|X_n - X| \geq \varepsilon \text{ i.o}\}) = 0$ for all $\varepsilon > 0$. Then $X_n \to X$ a.s.*

Proof: For $\varepsilon > 0$, consider

$$
\begin{aligned}
P\Big(\liminf_{n\to\infty}\{|X_n - X| < \varepsilon\}\Big) &= 1 - P\Big(\limsup_{n\to\infty}\{|X_n - X| \geq \varepsilon\}\Big) \\
&= 1.
\end{aligned}
$$

i.e.

$$
P\Big(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}\{|X_m - X| < \varepsilon\}\Big) = \; 1 \text{ for all } \varepsilon > 0.
$$

Therefore

$$
P\Big(\bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{|X_m - X| < \frac{1}{k}\}\Big) \;=\; \lim_{k\to\infty} P\Big(\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{|X_m - X| < \frac{1}{k}\}\Big) = 1.
$$

Since

$$\left\{\lim_{n\to\infty} X_n = X\right\} \ = \ \bigcap_{k=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{|X_m - X| < \frac{1}{k}\}$$

we get $X_n \to X$ a.s. $\square$

0.0.1 Limit theorems

In this section, we look at limit theorems. As told earlier, we only consider special cases but are good enough for many well known situations such as random walks (simple symmetric). We first discuss law of large numbers in the weak form.

Theorem 0.1 *(Weak law of large numbers) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and finite variance σ^2. Then $\frac{S_n}{n}$ converges in probability to μ. i.e., for each $\epsilon > 0$,*

$$\lim_{n\to\infty} P\Big(\Big|\frac{S_n}{n} - \mu\Big| \geq \epsilon\Big) \ = \ 0\,,$$

where $S_n \ = \ X_1 + \cdots + X_n$.

Proof: Note that

$$E\Big(\frac{S_n}{n}\Big) \ = \ \mu, \ E\Big(\frac{S_n}{n} - \mu\Big)^2 \ = \ \frac{\sigma^2}{n}\,.$$

Hence applying Chebyshev's inequality to $\frac{S_n}{n}$ we get

$$P\Big(\Big|\frac{S_n}{n} - \mu\Big| \geq \epsilon\Big) \ \leq \ \frac{\sigma^2}{n\epsilon^2}\,.$$

Therefore

$$\lim_{n\to\infty} P\Big(\Big|\frac{S_n}{n} - \mu\Big| \geq \epsilon\Big) \ = \ 0\,.$$

$\square$

Jacob Bernoulli proved weak law of large numbers when X_n's are i.i.d. Bernoulli(p) random variables. As we have seen earlier that convergence in probability only guarantee a.s convergence along sub sequence. So convergence of the averages of S_n to its mean is not completely satisfactory. What one need is a.s. convergence instead of convergence in probability. In the next theorem, we will do it for Bernoulli case.

Theorem 0.2 *(Strong law -Bernoulli case) Let $X_n, n \geq 1$ be a sequence of i.i.d. Bernoulli $(\frac{1}{2})$ random variables. Then $\frac{S_n}{n} \to \frac{1}{2}$ a.s.*

Proof: Using Hoeffing's inequality, for each $\varepsilon > 0$, we have

$$P\Big(\Big|\frac{S_n}{n} - \frac{1}{2}\Big| \geq \varepsilon\Big) \leq 2e^{-2n\varepsilon^2}.$$

Hence

$$\sum_{n=1}^{\infty} P\Big(\Big|\frac{S_n}{n} - \frac{1}{2}\Big| \geq \varepsilon\Big) \leq 2\sum_{n=1}^{\infty}(e^{-2\varepsilon^2})^n < \infty.$$

Therefore by using Borel-Cantelli lemma, it follows that

$$P\Big(\Big|\frac{S_n}{n} - \frac{1}{2}\Big| \geq \varepsilon \text{ i.o.}\Big) \;=\; 0.$$

Hence using Lemma 0.1, we have $\frac{S_n}{n} \to \frac{1}{2}$ a.s. $\square$.

The above result was only intended at illustrating the use inequalities like Hoeffding or Chernoff. One can prove the above SLLN(Strong law of large numbers) when $X_1, X_2, \cdots$ are i.i.d. with $EX_1 = \mu$ and satisfies $P(a \leq X \leq b) = 1$ for some $a < b$ by using the following Hoeffding's inequality (by replaing t by nt in the inequality).

Lemma 0.2 *(Hoeffding's inequality-a general form) Let $X_1, X_2, \cdots$ be i.i.d. with mean μ and satisfies $P(a \leq X \leq b) = 1$ for some $a < b$. Then*

$$P\big(|S_n - n\mu| \geq t\big) \;\leq\; 2e^{-\frac{2t^2}{n(b-a)}}.$$

Now we will prove strong law without any structural condition on the distribution of X_n's.

Theorem 0.3 *(Strong law of large numbers) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean. Then*

$$\lim_{n\to\infty} \frac{S_n}{n} \;\to\; \mu \text{ a.s.}.$$

Proof: We will only give the proof of the case, when X_1 has finite fourth moment.

$$
\begin{aligned}
(S_n - n\mu)^4 &= \Big(\sum_{i=1}^n (X_i - \mu)\Big)^4 \\
&= \sum_{i=1}^n (X_i - \mu)^4 \,+\, 4\sum_{i\neq j}^n (X_i-\mu)^3(X_j-\mu) \\
&\quad + 6\sum_{i\neq j}^n (X_i-\mu)^2(X_j-\mu)^2 \\
&\quad + 12 \sum_{i,j,k\ \mathrm{distinct}}^n (X_i-\mu)^2(X_j-\mu)(X_k-\mu) \\
&\quad + 24 \sum_{i,j,k,l\ \mathrm{distinct}}^n (X_i-\mu)(X_j-\mu)(X_k-\mu)(X_l-\mu)
\end{aligned}
$$

[Note we used above number of permutations of four objects (need not be distinct)]

Hence using the fact that $(X_i - \mu)$, $(X_j - \mu)$, $(X_k - \mu)$, $(X_l - \mu)$ are independent for i, j, k, l distinct, we get

$$
E(S_n - n\mu)^4 \;=\; \sum_{i=1}^n E(X_i-\mu)^4 \,+\, 6\sum_{i\neq j}^n E(X_i-\mu)^2 E(X_j-\mu)^2 \,. \tag{0.1}
$$

Since X_1, $X_2, \ldots$ are identically distributed, we get

$$
\begin{aligned}
E(S_n - n\mu)^4 &= nE(X_1-\mu)^4 \,+\, 3n(n-1)E(X_1-\mu)^2E(X_2-\mu)^2 \\
&\le nK \,+\, 3n(n-1)K\,,
\end{aligned} \tag{0.2}
$$

where $K = E(X_1 - \mu)^4$. Hence

$$
E\Big[\frac{S_n}{n} - \mu\Big]^4 \;\le\; \frac{K}{n^3} \,+\, \frac{3K}{n^2}\,. \tag{0.3}
$$

Therefore for each $\varepsilon > 0$,

$$
\begin{aligned}
P\Big(\Big|\frac{S_n}{n} - \mu\Big| \ge \varepsilon\Big) &\le \frac{1}{\varepsilon^4}E\Big[\frac{S_n}{n} - \mu\Big]^4 \\
&\le \frac{K}{\varepsilon^4}\Big(\frac{1}{n^3} + \frac{3}{n^2}\Big).
\end{aligned}
$$

Hence

$$
\sum_{n=1}^\infty P\Big(\Big|\frac{S_n}{n} - \mu\Big| \ge \varepsilon\Big) < \infty.
$$

Now using Borel-Cantelli lemma it follows that for each $\varepsilon > 0$,

$$P\Big(\Big|\frac{S_n}{n} - \mu\Big| \geq \varepsilon \text{ i.o.}\Big) \;=\; 0.$$

Hence $\frac{S_n}{n}$ converges to μ a.s. This completes the proof. $\square$

As an application we show that any continuous function can be approximated by Bernstein polynomials.

Example 0.1 *Let $f : [0,\ 1] \to \mathbb{R}$ be a continuous function. Consider the Bernstein polynomials*

$$B_n(x) \;=\; \sum_{k=0}^{n} f(\frac{k}{n}) \binom{n}{k} \, x^k (1-x)^{n-k},\ 0 \leq x \leq 1\,.$$

Fix $x \in (0,1)$. Let $X_1, X_2, \ldots$ be independent and identically distributed Bernoulli(x) random variables. Then using strong law of large numbers, we have

$$\frac{S_n}{n} \to x \text{ a.s. as } n \to \infty\,.$$

Now note that S_n is Binomial (n,x) random variable. Hence

$$B_n(x) \;=\; E\Big[f\Big(\frac{S_n}{n}\Big)\Big]\,.$$

Set

$$Y_n \;=\; f\Big(\frac{S_n}{n}\Big)\,.$$

Then $Y_n \to f(x)$ a.s. as $n \to \infty$ and $|Y_n| \leq K$ where K such that $-K \leq f(x) \leq K$. Here we use the fact that every continuous function defined on $[0,\ 1]$ is bounded. Now apply the dominated convergence theorem (Theorem 6.0.31), we get

$$\lim_{n\to\infty} EY_n \;=\; f(x)\,.$$

i.e.

$$\lim_{n\to\infty} B_n(x) \;=\; f(x),\ 0 < x < 1\,.$$

For $x = 0, x = 1$, the proof follows by observing that $B_n(0) = f(0)$ and $B_n(1) = f(1)$.

Now we will give an alternate proof in fact we prove here the uniform convergence. A part of the proof given below is by Chebychev for the Bernoulli's *(weak) law of large numbers.*

For $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - y| \leq \delta \ \Rightarrow \ |f(x) - f(y)| \leq \varepsilon.$$

Also let K denote the maximum of f in $[0,\ 1]$. Now consider

$$\begin{aligned}
|B_n(x) - f(x)| &\leq E\Big|f(\frac{S_n}{n}) - f(x)\Big| \\
&= E\Big[\Big|f(\frac{S_n}{n}) - f(x)|I\Big\{|\frac{S_n}{n} - x\Big| \leq \delta\Big\}\Big] \\
&\quad +E\Big[\Big|f(\frac{S_n}{n}) - f(x)|I\Big\{|\frac{S_n}{n} - x\Big| > \delta\Big\}\Big] \\
&\leq \varepsilon + 2KP\Big(\Big|\frac{S_n}{n} - x\Big| > \delta\Big) \\
&\leq \varepsilon + 2K\frac{E\big[|\frac{S_n}{n} - x|^2\big]}{\delta^2} \\
&= \varepsilon + 2K\frac{E\big[|S_n - nx|^2\big]}{n^2\delta^2} \\
&= \varepsilon + 2K\frac{nx(1-x)}{n^2\delta^2} \\
&\leq \varepsilon + K\frac{1}{2n\delta^2}.
\end{aligned}$$

Now given $\varepsilon > 0$, choose n_0 such that $K\frac{1}{2n\delta^2} \leq \varepsilon$ for all $n \geq n_0$. Hence we get

$$|B_n(x) - f(x)| \leq 2\varepsilon \ \forall\ n \geq n_0,\ 0 \leq x \leq 1.$$

This proves that $B_n(x) \to f(x)$ uniformly in x.

Theorem 0.4 *(Central limit theorem) Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables, each having finite mean μ and finite non zero variance σ^2. Then*

$$\lim_{n\to\infty} P\Big[\frac{S_n - n\mu}{\sigma\sqrt{n}} \ \leq \ x\Big] \ = \ N(x),\ x \in \mathbb{R}\,,$$

where $N(\cdot)$ is the standard normal distribution function.

Proof:

Set

$$\bar{S}_n \ = \ \frac{S_n - n\mu}{\sigma\sqrt{n}}\,.$$

For $t \geq 0$,

$$\begin{aligned} \Phi_{\bar{S}_n}(t) &= e^{-in\mu\frac{t}{\sigma\sqrt{n}}}\Phi_{S_n}(\frac{t}{\sigma\sqrt{n}}) \\ &= e^{-in\mu\frac{t}{\sigma\sqrt{n}}}(\Phi_{X_1}(\frac{t}{\sigma\sqrt{n}}))^n , \end{aligned} \tag{0.4}$$

where the second equality uses the fact that $X_1, X_2, \ldots$ are independent and identically distributed.

Hence, from (0.4), we get

$$\Phi_{\bar{S}_n}(t) = e^{n[\ln(\Phi_{X_1}(\frac{t}{\sigma\sqrt{n}})) - i\mu(\frac{t}{\sigma\sqrt{n}})]} . \tag{0.5}$$

Hence for $t \neq 0$,

$$\lim_{n\to\infty} n[\ln(\Phi_{X_1}(\frac{t}{\sigma\sqrt{n}})) - i\mu(\frac{t}{\sigma\sqrt{n}})] = \frac{t^2}{\sigma^2}\lim_{n\to\infty}\frac{\ln(\Phi_{X_1}(\frac{t}{\sigma\sqrt{n}})) - i\mu(\frac{t}{\sigma\sqrt{n}})}{(\frac{t}{\sigma\sqrt{n}})^2} \tag{0.6}$$

Consider

$$\begin{aligned} \lim_{n\to\infty}\frac{\ln(\Phi_{X_1}(\frac{t}{\sigma\sqrt{n}})) - i\mu(\frac{t}{\sigma\sqrt{n}})}{(\frac{t}{\sigma\sqrt{n}})^2} &= \lim_{x\to 0}\frac{\ln(\Phi_{X_1}(x)) - i\mu x}{x^2} \\ &= \lim_{x\to 0}\frac{\Phi_{X_1}^{(1)}(x) - i\mu\Phi_{X_1}(x)}{2x\Phi_{X_1}(x)} \\ &= \frac{\Phi_{X_1}^{(2)}(0) - i\mu\Phi_{X_1}^{(1)}(0)}{2} = -\frac{\sigma^2}{2} . \end{aligned} \tag{0.7}$$

The last two equalities follows from l'Hospital's rule in view of the properties of characteristic functions. Combining (0.5), (0.5) and (0.5), we get, for each $t \neq 0$

$$\lim_{n\to\infty}\Phi_{\bar{S}_n}(t) = e^{-\frac{t^2}{2}} .$$

For $t = 0$, the above limit follows easily. Hence for each $t \in \mathbb{R}$, we have

$$\lim_{n\to\infty}\Phi_{\bar{S}_n}(t) = e^{-\frac{t^2}{2}} . \tag{0.8}$$

Hence using Theorem 7.0.38, we have from (0.8), we complete the proof.

As a corollary to Central limit theorem, one can get DeMoivre-Laplace limit theorem.

Theorem 0.5 *Let S_n be Bernoulli (n, p) random variable. Then*

$$\lim_{n\to\infty} P\Big(\frac{S_n - np}{\sqrt{np(1-p)}} \ \leq \ x \Big) \ = \ N(x), \ x \in \mathbb{R} \, .$$

Proof:

Observe that if $X_1, X_2 \dots$ are independent and identically distributed Binomial (p) random variables, then $X_1 + \dots X_n$ is Bernoulli (n, p) random variable. Now apply Central limit theorem, we get DeMoivre-Laplace limit theorem.

One can use central limit theorem to get the normal approximation formula. Under the hypothesis of central limit theorem, we have

$$\lim_{n\to\infty} \Big[P\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \ \leq x \Big) - N(x) \Big] \ = 0.$$

This imnplies

$$\lim_{n\to\infty} \Big[P\big(S_n \leq x \big) - N\Big(\frac{x - n\mu}{\sigma\sqrt{n}} \Big) \Big] \ = 0. \tag{0.9}$$

The above we write as

$$P\big(S_n \leq x \big) \approx N\Big(\frac{x - n\mu}{\sigma\sqrt{n}} \Big)$$

for sufficiently large values of n.

Here $\approx$ means "approximately equal to". To use the above normal approximation precisely, one need to know the rate of convergence in the central limit theorem, that is the content of the next theorem.

Theorem 0.6 *(Berry-Esseen Theorem) Let $X_1, X_2, \cdots$ be i.i.d. random variables such that X_1 has finite first three moments. Then*

$$\sup_{x\in\mathbb{R}} \Big| P\Big(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \Big) - N(x) \Big| \leq \ \frac{3\rho}{\sigma\sqrt{n}},$$

where $\mu = EX_1, \sigma^2 = \mathrm{Var}(X_1)$ and $\rho = E|X_1|^3$.

Example 0.2 *Let $X_1, X_2, \dots$ be an independent and identically distributed Bernoulli(p) random variables. Then using (0.9), we get for x a non negative integer,*

$$\begin{aligned} P\big(S_n = x \big) \ &= \ P\Big(x - \frac{1}{2} < S_n \leq x + \frac{1}{2} \Big) \\ &\approx \ N\Big(\frac{x + (1/2) - n\mu}{\sigma\sqrt{n}} \Big) \ - \ N\Big(\frac{x - (1/2) - n\mu}{\sigma\sqrt{n}} \Big) \, , \end{aligned}$$

where $\mu \ = \ p, \ \sigma^2 \ = \ p(1-p)$.

BEST OF LUCK