STAT/MTHE 353: 6 – Convergence of Random Variables and Limit Theorems

T. Linder

Queen's University

Winter 2017

Markov and Chebyshev Inequalities

Recall that a random variable X is called nonnegative if $P(X \geq 0) = 1$.

Theorem 1 (Markov's inequality)

Let X be a nonnegative random variable with mean $E(X)$. Then for any $t > 0$

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Proof: Assume X is continuous with pdf f. Then $f(x) = 0$ if $x < 0$, so

$$\begin{aligned} E(X) &= \int_0^\infty x f(x)\,dx \geq \int_t^\infty x f(x)\,dx \geq t \int_t^\infty f(x)\,dx \\ &= tP(X \geq t) \end{aligned}$$

If X is discrete, replace the integrals with sums... $\square$

Example: Suppose X is nonnegative and $P(X > 10) = 1/5$; show that $E(X) \geq 2$. Also, Markov's inequality for $|X|$.

Theorem 2 (Chebyshev's inequality)

Let X be a random variable with finite variance $\mathrm{Var}(X)$. Then for any $t > 0$

$$P\big(|X - E(X)| \geq t\big) \leq \frac{\mathrm{Var}(X)}{t^2}$$

Proof: Apply Markov's inequality to the nonnegative random variable $Y = |X - E(X)|^2$

$$\begin{aligned} P(|X - E(X)| \geq t) &= P(|X - E(X)|^2 \geq t^2) \leq \frac{E\big(|X - E(X)|^2\big)}{t^2} \\ &= \frac{\mathrm{Var}(X)}{t^2} \end{aligned}$$ $\square$

Example: Chebyshev with $t = k\sqrt{\mathrm{Var}(X)}$ ...

The following result often gives a much sharper bound if the MGF of X is finite in some interval around zero.

Theorem 3 (Chernoff's bound)

Let X be a random variable with MGF $M_X(t)$. Then for any $a \in \mathbb{R}$

$$P(X \geq a) \leq \min_{t>0} e^{-at} M_X(t)$$

Proof: Fix $t > 0$. Then we have

$$\begin{aligned} P(X \geq a) &= P(tX \geq ta) = P\big(e^{tX} \geq e^{ta}\big) \\ &\leq \frac{E\big(e^{tX}\big)}{e^{ta}} \quad \text{(Markov's inequality)} \\ &= e^{-ta} M_X(t) \end{aligned}$$

Since this holds for all $t > 0$ such that $M_X(t) < \infty$, it must hold for the t minimizing the upper bound. $\square$

Example: Suppose $X \sim N(0,1)$. Apply Chernoff's bound to upper bound $P(X \geq a)$ for $a > 0$ and compare with the bounds obtained from Chebyshev's and Markov's inequalities.

Convergence of Random Variables

A *probability space* is a triple $(\Omega, \mathcal{A}, P)$, where Ω is a *sample space*, $\mathcal{A}$ is a collection of subsets of Ω called *events*, and P is a probability measure on $\mathcal{A}$. In particular, the set of events $\mathcal{A}$ satisfies

1) Ω is an event
2) If $A \subset \Omega$ is an event, then A^c is also an event
3) If $A_1, A_2, A_3, \ldots$ are events, then so is $\bigcup_{n=1}^{\infty} A_n$.

P is a function from the collection of events $\mathcal{A}$ to $[0,1]$ which satisfies the *axioms of probability*:

1) $P(A) \geq 0$ for all events $A \in \mathcal{A}$.
2) $P(\Omega) = 1$.
3) If $A_1, A_2, A_3, \ldots$ are mutually exclusive events (i.e., $A_i \cap A_j = \emptyset$ for all $i \neq j$), then $P\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} P(A_i)$.

- Recall that a random variable X is a function $X : \Omega \to \mathbb{R}$ that maps any point ω in the sample space Ω to a real number $X(\omega)$.
- A random variable X must satisfy the following: any subset A of Ω in the form
$$A = \{\omega \in \Omega : X(\omega) \in B\}$$
for any "reasonable" $B \subset \mathbb{R}$ is an *event*. For example B can be any set obtained by a countable union and intersection of intervals.
- Recall that a sequence of real numbers $\{x_n\}_{n=1}^{\infty}$ is said to converge to a limit $x \in \mathbb{R}$ (notation: $x_n \to x$) if for any $\epsilon > 0$ there exists N such that $|x_n - x| < \epsilon$ for all $n \geq N$.

Definitions (Modes of convergence) Let $\{X_n\} = X_1, X_2, X_3, \ldots$ be a sequence of random variables defined in a probability space $(\Omega, \mathcal{A}, P)$.

(i) We say that $\{X_n\}$ converges to a random variable X *almost surely* (notation: $X_n \xrightarrow{a.s.} X$) if
$$P(\{\omega : X_n(\omega) \to X(\omega)\}) = 1$$

(i) We say that $\{X_n\}$ converges to a random variable X *in probability* (notation: $X_n \xrightarrow{P} X$) if for any $\epsilon > 0$ we have
$$P\big(|X_n - X| > \epsilon\big) \to 0 \quad \text{as} \quad n \to \infty$$

(iii) If $r > 0$ we say that $\{X_n\}$ converges to a random variable X *in rth mean* (notation: $X_n \xrightarrow{r.m.} X$) if
$$E\big(|X_n - X|^r\big) \to 0 \quad \text{as} \quad n \to \infty$$

(iv) Let F_n and F denote the cdfs of X_n and X, respectively. We say that $\{X_n\}$ converges to a random variable X *in distribution* (notation: $X_n \xrightarrow{d} X$) if
$$F_n(x) \to F(x) \quad \text{as} \quad n \to \infty$$
for any x such that F is continuous at x.

Remarks:

- A very important case of convergence in rth mean is when $r = 2$. In this case $E\big(|X_n - X|^2\big) \to 0$ as $n \to \infty$ and we say that $\{X_n\}$ converges in *mean square* to X.
- Almost sure convergence is often called convergence with *probability 1*.
- Almost sure convergence, convergence in rth mean, and convergence in probability all state that X_n is eventually *close* to X (in different senses) as n increases.
- In contrast, convergence in distribution is only a statement about the closeness of the *distribution* of X_n to that of X for large n.

Example: Sequence $\{X_n\}$ such that $F_n(x) = F(x)$ for all $x \in \mathbb{R}$ and $n = 1, 2, \ldots$, but $P(|X_n - X| > 1/2) = 1$ for all n ...

Theorem 4

The following implications hold:

$$\begin{tikzcd}
\boxed{X_n \xrightarrow{a.s.} X} \arrow[dr, Rightarrow] & & \\
& \boxed{X_n \xrightarrow{P} X} \arrow[r, Rightarrow] & \boxed{X_n \xrightarrow{d} X} \\
\boxed{X_n \xrightarrow{r.m.} X} \arrow[ur, Rightarrow] & &
\end{tikzcd}$$

Remark: We will show that in general, $X_n \xrightarrow{a.s.} X$ does not imply that $X_n \xrightarrow{r.m.} X$, and also that $X_n \xrightarrow{r.m.} X$ does not imply $X_n \xrightarrow{a.s.} X$.

Theorem 5

Convergence in rth mean implies convergence in probability; i.e., if $X_n \xrightarrow{r.m.} X$, then $X_n \xrightarrow{P} X$.

Proof: Assume $X_n \xrightarrow{r.m.} X$ for some $r > 0$; i.e., $E(|X_n - X|^r) \to 0$ as $n \to \infty$. Then for any $\epsilon > 0$,

$$\begin{aligned} P(|X_n - X| > \epsilon) &= P(|X_n - X|^r > \epsilon^r) \\ &\le \frac{E(|X_n - X|^r)}{\epsilon^r} \to 0 \text{ as } n \to \infty \end{aligned}$$

where the inequality follows from Markov's inequality applied to the nonnegative random variable $|X_n - X|^r$. □

Theorem 6

Almost sure convergence implies convergence in probability; i.e., if $X_n \xrightarrow{a.s.} X$, then $X_n \xrightarrow{P} X$.

Proof: Assume $X_n \xrightarrow{a.s.} X$. We want to show that for any $\epsilon > 0$ we have $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. Thus, defining the event

$$A_n(\epsilon) = \big\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\big\}$$

we want to show that $P\big(A_n(\epsilon)\big) \to 0$ for any $\epsilon > 0$.

Define

$$A(\epsilon) = \big\{\omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ for infinitely many } n\big\}$$

Then for any $\omega \in A(\epsilon)$ we have that $X_n(\omega) \not\to X(\omega)$. Since $X_n \xrightarrow{a.s.} X$ we must have that

$$P\big(A(\epsilon)\big) = 0$$

Proof cont'd: Now define the event

$$B_n(\epsilon) = \big\{\omega : |X_m(\omega) - X(\omega)| > \epsilon \text{ for some } m \geq n\big\}$$

Notice that $B_1(\epsilon) \supset B_2(\epsilon) \supset \cdots \supset B_{n-1}(\epsilon) \supset B_n(\epsilon) \supset \cdots$; i.e., $\big\{B_n(\epsilon)\big\}$ is a *decreasing sequence* of events, satisfying

$$A(\epsilon) = \bigcap_{n=1}^{\infty} B_n(\epsilon)$$

Therefore by the *continuity of probability*

$$P\big(A(\epsilon)\big) = P\bigg(\bigcap_{n=1}^{\infty} B_n(\epsilon)\bigg) = \lim_{n\to\infty} P\big(B_n(\epsilon)\big)$$

But since $P\big(A(\epsilon)\big) = 0$, we obtain that $\lim_{n\to\infty} P\big(B_n(\epsilon)\big) = 0$.

Proof cont'd: We clearly have

$$A_n(\epsilon) \subset B_n(\epsilon) \quad \text{for all } n$$

so $P\big(A_n(\epsilon)\big) \leq P\big(B_n(\epsilon)\big)$, so $P\big(A_n(\epsilon)\big) \to 0$ as $n \to \infty$. We obtain

$$\lim_{n\to\infty} P\big(|X_n - X| > \epsilon\big) = 0$$

for any $\epsilon > 0$ as claimed. □

Example: ($X_n \xrightarrow{P} X$ does not imply $X_n \xrightarrow{a.s.} X$.) Let $X_1, X_2, \ldots$ be independent random variables with distribution

$$P(X_n = 0) = 1 - \frac{1}{n}, \quad P(X_n = 1) = \frac{1}{n}$$

and show that there is a random variable X such that $X_n \xrightarrow{P} X$, but $P(\{\omega : X_n(\omega) \to X(\omega)\}) = 0$.

Solution: ...

Example: ($X_n \xrightarrow{r.m.} X$ does not imply $X_n \xrightarrow{a.s.} X$.) Use the previous example...

Lemma 7 (Sufficient condition for a.s. convergence)

Suppose $\sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$ *for all* $\epsilon > 0$. *Then* $X_n \xrightarrow{a.s.} X$.

Proof: Recall the event

$$A(\epsilon) = \big\{\omega : |X_n(\omega) - X(\omega)| > \epsilon \text{ for infinitely many } n\big\}$$

Then

$$A(\epsilon)^c = \big\{\omega : \text{ there exists } N \text{ such that } |X_n(\omega) - X(\omega)| \leq \epsilon \text{ for all } n \geq N\big\}$$

For $\epsilon = 1/k$ we obtain a *decreasing sequence of events* $\big\{A(1/k)^c\big\}_{k=1}^{\infty}$. Let

$$C = \bigcap_{k=1}^{\infty} A(1/k)^c$$

and notice that if $\omega \in C$, then $X_n(\omega) \to X(\omega)$ as $n \to \infty$. Thus if $P(C) = 1$, then $X_n \xrightarrow{a.s.} X$.

Proof cont'd: But from the continuity of probability

$$P(C) = P\Big(\bigcap_{k=1}^{\infty} A(1/k)^c\Big) = \lim_{k\to\infty} P\big(A(1/k)^c\big)$$

so if $P\big(A(1/k)^c\big) = 1$ for all k, then $P(C) = 1$ and we obtain $X_n \xrightarrow{a.s.} X$. Thus $P\big(A(\epsilon)\big) = 0$ for all $\epsilon > 0$ implies $X_n \xrightarrow{a.s.} X$.

As before, let $A_n(\epsilon) = \big\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\big\}$ and

$$B_n(\epsilon) = \big\{\omega : |X_m(\omega) - X(\omega)| > \epsilon \text{ for some } m \ge n\big\}$$

We have seen in the proof of Theorem 6 that

$$P(A(\epsilon)) = \lim_{n\to\infty} P(B_n(\epsilon))$$

Thus if we can show that $\lim_{n\to\infty} P(B_n(\epsilon)) = 0$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$.

Proof cont'd: Note that

$$B_n(\epsilon) = \bigcup_{m=n}^{\infty} A_m(\epsilon)$$

and that the condition $\sum_{n=1}^{\infty} P\big(A_n(\epsilon)\big) = \sum_{n=1}^{\infty} P(|X_n - X| > \epsilon) < \infty$ implies

$$\lim_{n\to\infty} \sum_{m=n}^{\infty} P\big(A_n(\epsilon)\big) = 0$$

We obtain

$$\begin{aligned} P\big(B_n(\epsilon)\big) &= P\Big(\bigcup_{m=n}^{\infty} A_m(\epsilon)\Big) \\ &\qquad \text{(union bound)} \\ &\le \sum_{m=n}^{\infty} P\big(A_m(\epsilon)\big) \to 0 \text{ as } n \to \infty \end{aligned}$$

so $\lim_{n\to\infty} P(B_n(\epsilon)) = 0$ for all $\epsilon > 0$. Thus $X_n \xrightarrow{a.s.} X$. □

Example: ($X_n \xrightarrow{a.s.} X$ does not imply $X_n \xrightarrow{r.m.} X$.) Let $X_1, X_2, \ldots$ be random variables with marginal distributions given by

$$P(X_n = n^3) = \frac{1}{n^2}, \quad P(X_n = 0) = 1 - \frac{1}{n^2}$$

Show that there is a random variable X such that $X_n \xrightarrow{a.s.} X$, but $X_n \not\xrightarrow{r.m.} X$ if $r \ge 2/3$.

Theorem 8

Convergence in probability implies convergence in distribution; i.e., if $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.

Proof: Let F_n and F be the cdfs of X_n and X, respectively and let $x \in \mathbb{R}$ be such that F is continuous at x. We want to show that $X_n \xrightarrow{P} X$ implies $F_n(x) \to F(x)$ as $n \to \infty$.

For any given $\epsilon > 0$ we have

$$\begin{aligned} F_n(x) &= P(X_n \le x) \\ &= P\big(X_n \le x,\, X \le x+\epsilon\big) + P\big(X_n \le x,\, X > x+\epsilon\big) \\ &\le P\big(X \le x+\epsilon\big) + P\big(|X_n - X| > \epsilon\big) \\ &= F(x+\epsilon) + P\big(|X_n - X| > \epsilon\big) \end{aligned}$$

Proof cont'd: Similarly,

$$\begin{aligned} F(x-\epsilon) &= P(X \le x-\epsilon) \\ &= P\big(X \le x-\epsilon,\, X_n \le x\big) + P\big(X \le x-\epsilon,\, X_n > x\big) \\ &\le P\big(X_n \le x\big) + P\big(|X_n - X| > \epsilon\big) \\ &= F_n(x) + P\big(|X_n - X| > \epsilon\big) \end{aligned}$$

We obtain

$$F(x-\epsilon) - P\big(|X_n - X| > \epsilon\big) \le F_n(x) \le F(x+\epsilon) + P\big(|X_n - X| > \epsilon\big)$$

Since $X_n \xrightarrow{P} X$, we have $P\big(|X_n - X| > \epsilon\big) \to 0$ as $n \to \infty$. Choosing $N(\epsilon)$ large enough so that $P\big(|X_n - X| > \epsilon\big) < \epsilon$ for all $n \ge N(\epsilon)$, we obtain

$$F(x-\epsilon) - \epsilon \le F_n(x) \le F(x+\epsilon) + \epsilon$$

for all $n \ge N(\epsilon)$. Since F is continuous at x, letting $\epsilon \to 0$ we obtain $\lim_{n\to\infty} F_n(x) = F(x)$. $\square$

If X is a constant random variable, then the converse also holds:

Theorem 9

Let $c \in \mathbb{R}$. If $X_n \xrightarrow{d} c$, then $X_n \xrightarrow{P} c$.

Proof: For X with $P(X = c)$ let F_n and F be as before and note that

$$F(x) = \begin{cases} 0 & \text{if } x < c \\ 1 & \text{if } x \ge c \end{cases}$$

Then F is continuous at all $x \ne c$, so $F_n(x) \to F(x)$ as $n \to \infty$ for all $x \ne c$. For any $\epsilon > 0$

$$\begin{aligned} P\big(|X_n - c| > \epsilon\big) &= P(X_n < c-\epsilon) + P(X_n > c+\epsilon) \\ &\le P(X_n \le c-\epsilon) + P(X_n > c+\epsilon) \\ &= F_n(c-\epsilon) + 1 - F_n(c+\epsilon) \end{aligned}$$

Since $F_n(c-\epsilon) \to 0$ and $F_n(c+\epsilon) \to 1$ as $n \to \infty$, we obtain $P\big(|X_n - c| > \epsilon\big) \to 0$ as $n \to \infty$. $\square$

Laws of Large Numbers

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables with finite mean $\mu = E(X_i)$ and variance $\sigma^2 = \mathrm{Var}(X_i)$. The *sample mean* $\bar{X}_n$ is defined by

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Theorem 10 (Weak law of large numbers)

We have $\bar{X}_n \xrightarrow{P} \mu$; i.e., for all $\epsilon > 0$

$$\lim_{n\to\infty} P\big(|\bar{X}_n - \mu| > \epsilon\big) = 0$$

Proof: Since the X_i are independent,

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

Since $E(\bar{X}_n) = \mu$, Chebyshev's inequality implies for any $\epsilon > 0$

$$P\big(|\bar{X}_n - \mu| > \epsilon\big) \le \frac{\mathrm{Var}(\bar{X}_n)}{\epsilon^2} = \frac{\sigma}{\epsilon^2 n}$$

Thus $P\big(|\bar{X}_n - \mu| > \epsilon\big) \to 0$ as $n \to \infty$ for any $\epsilon > 0$. $\square$

Theorem 11 (Strong law of large numbers)

If $X_1, X_2, \ldots$ is an i.i.d. sequence with finite mean $\mu = E(X_i)$ and variance $\mathrm{Var}(X_i) = \sigma^2$, then $\bar{X}_n \xrightarrow{a.s.} \mu$; i.e.,

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1$$

Proof: First we show that the subsequence $\bar{X}_{1^2}, \bar{X}_{2^2}, \bar{X}_{3^2}, \ldots$ converges a.s. to μ.

Let $S_n = \sum_{i=1}^n X_i$. Then $E(S_{n^2}) = n^2\mu$ and $\mathrm{Var}(S_{n^2}) = n^2\sigma^2$. For any $\epsilon > 0$ we have by Chebyshev's inequality

$$\begin{aligned} P\big(|\bar{X}_{n^2} - \mu| > \epsilon\big) &= P\left(\Big|\frac{1}{n^2}S_{n^2} - \mu\Big| > \epsilon\right) = P\big(|S_{n^2} - n^2\mu| > n^2\epsilon\big) \\ &\le \frac{\mathrm{Var}(S_{n^2})}{n^4\epsilon^2} = \frac{\sigma^2}{n^2\epsilon^2} \end{aligned}$$

Thus $\sum_{n=1}^{\infty} P\big(|\bar{X}_{n^2} - \mu| > \epsilon\big) < \infty$ and Lemma 7 gives $\bar{X}_{n^2} \xrightarrow{a.s.} \mu$.

Proof cont'd: Next suppose that $X_i \ge 0$ for all i. Then $S_n(\omega) = X_1(\omega) + \cdots + X_n(\omega)$ is a nondecreasing sequence. For any n there is a unique integer i_n such that $i_n^2 \le n < (i_n+1)^2$. Thus

$$S_{i_n^2} \le S_n \le S_{(i_n+1)^2} \quad \Longrightarrow \quad \frac{S_{i_n^2}}{(i_n+1)^2} \le \frac{S_n}{n} \le \frac{S_{(i_n+1)^2}}{i_n^2}$$

This is equivalent to

$$\left(\frac{i_n}{i_n+1}\right)^2 \frac{S_{i_n^2}}{i_n^2} \le \frac{S_n}{n} \le \left(\frac{i_n+1}{i_n}\right)^2 \frac{S_{(i_n+1)^2}}{(i_n+1)^2}$$

i.e.

$$\left(\frac{i_n}{i_n+1}\right)^2 \bar{X}_{i_n^2} \le \bar{X}_n \le \left(\frac{i_n+1}{i_n}\right)^2 \bar{X}_{(i_n+1)^2}$$

Letting $A = \{\omega : \bar{X}_{n^2}(\omega) \to \mu\}$, we know that $P(A) = 1$. Since $i_n/(i_n+1) \to 1$ and $(i_n+1)/i_n \to 1$ as $n \to \infty$, and for all $\omega \in A$, $\bar{X}_{i_n^2}(\omega) \to \mu$ and $\bar{X}_{(i_n+1)^2}(\omega) \to \mu$ as $n \to \infty$, we obtain

$$\bar{X}_n(\omega) \to \mu \text{ for all } \omega \in A$$

Proof cont'd: Now we remove the restriction $X_i \ge 0$. Define

$$X_i^+ = \max(X_i, 0), \qquad X_i^- = \max(-X_i, 0)$$

and note that $X_i = X_i^+ - X_i^-$ and $X_i^+ \ge 0$, $X_i^- \ge 0$. Letting $\mu^+ = E(X_i^+)$, $\mu^- = E(X_i^-)$ and

$$A_1 = \left\{\omega : \frac{1}{n}\sum_{i=1}^n X_i^+(\omega) \to \mu^+\right\}, \qquad A_2 = \left\{\omega : \frac{1}{n}\sum_{i=1}^n X_i^-(\omega) \to \mu^-\right\}$$

we know that $P(A_1) = P(A_2) = 1$. Thus $P(A_1 \cap A_2) = 1$. But for all $\omega \in A_1 \cap A_2$ we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i(\omega) &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^+(\omega) - \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^-(\omega) \\ &= \mu^+ - \mu^- = \mu \end{aligned}$$

We conclude that $\frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} \mu$. $\square$

Remark: The condition $\mathrm{Var}(X_i) < \infty$ is not needed. The strong law of large numbers (SLLN) holds for any i.i.d. sequence $X_1, X_2, \ldots$ with finite mean $\mu = E(X_1)$.

Example: Simple random walk ...

Example: (Single server queue) Customers arrive one-by-one at a service station.

- The time between the arrival of the $(i-1)$th and ith customer is Y_i, and $Y_1, Y_2, \ldots$ are i.i.d. nonnegative random variables with finite mean $E(Y_1)$ and finite variance.
- The time needed to service the ith customer is U_i, and $U_1, U_2, \ldots$ are i.i.d. nonnegative random variables with finite mean $E(U_1)$ and finite variance.

Show that if $E(U_1) < E(Y_1)$, then (after the first customer arrives) the queue will eventually become empty with probability 1.

Central Limit Theorem

- If $X_1, X_2, \ldots$ are Bernoulli(p) random variables, then $S_n = X_1 + \cdots + X_n$ is a Binomial(n, p) random variable with mean $E(S_n) = np$ and variance $\mathrm{Var}(S_n) = np(1-p)$.
- Recall the De Moivre-Laplace theorem:

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) = \Phi(x)$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ is the cdf of a $N(0,1)$ random variable X.

- Since $\Phi(x)$ is continuous at every x, the above is equivalent to

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} X$$

where $\mu = E(X_1) = p$ and $\sigma = \sqrt{\mathrm{Var}(X_1)} = p(1-p)$.

The De Moivre-Laplace theorem is a special case of the following general result:

Theorem 12 (Central Limit Theorem)

Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean μ and finite variance σ^2. Then for $S_n = X_1 + \cdots + X_n$ we have

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} X$$

where $X \sim N(0,1)$.

Remark: Note that $\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$. The SLLN implies that with probability 1, $\bar{X}_n - \mu \to 0$ as $n \to \infty$ The central limit theorem (CLT) tells us something about the speed at which $\bar{X}_n - \mu$ converges to zero.

To prove the CLT we will assume that each X_i has MGF $M_{X_i}(t)$ that is defined in an open interval around zero. The key to the proof is the following result which we won't prove:

Theorem 13 (Levy continuity theorem for MGF)

Assume $Z_1, Z_2, \ldots$ are random variables such that the MGF $M_{Z_n}(t)$ is defined for all $t \in (-a, a)$ for some $a > 0$ and all $n = 1, 2, \ldots$ Suppose X is a random variable with MGF $M_X(t)$ defined for $t \in (-a, a)$. If

$$\lim_{n\to\infty} M_{Z_n}(t) = M_X(t) \quad \text{for all } t \in (-a, a)$$

then $Z_n \xrightarrow{d} X$.

Proof of CLT: Let $Y_n = X_n - \mu$. Then $E(Y_n) = 0$ and $\mathrm{Var}(Y_n) = \sigma^2$. Note that

$$\frac{\sum_{i=1}^{n} Y_i}{\sigma\sqrt{n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}} = Z_n$$

Letting $M_Y(t)$ denote the (common) moment generating function of the Y_i, we have by the independence of $Y_1, Y_2, \ldots$

$$M_{Z_n}(t) = M_Y\left(\frac{t}{\sigma\sqrt{n}}\right)^n$$

If $X \sim N(0,1)$, then $M_X(t) = e^{t^2/2}$. Thus if we show that $M_{Z_n}(t) \to e^{t^2/2}$, or equivalently

$$\lim_{n\to\infty} \ln\big(M_{Z_n}(t)\big) = \frac{t^2}{2}$$

then Levy's continuity theorem implies the CLT.

Proof cont'd: Let $h = \dfrac{t}{\sigma\sqrt{n}}$. Then

$$\begin{aligned}
\lim_{n\to\infty} \ln\big(M_{Z_n}(t)\big) &= \lim_{n\to\infty} \ln\left(M_Y\left(\frac{t}{\sigma\sqrt{n}}\right)^n\right) \\
&= \lim_{n\to\infty} n \ln\left(M_Y\left(\frac{t}{\sigma\sqrt{n}}\right)\right) \\
&= \frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{\ln M_Y(h)}{h^2}
\end{aligned}$$

It can be shown that $M_Y(h)$ and all its derivatives are continuous at $h = 0$. Since $M_Y(0) = E(e^{0\cdot Y}) = 1$, the above limit is indeterminate. Applying l'Hospital's rule, we consider the limit

$$\frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{M_Y'(h)/M_Y(h)}{2h} = \frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{M_Y'(h)}{2hM_Y(h)}$$

Again, since $M'(0) = E(Y) = 0$, the limit is indeterminate.

Proof cont'd: Apply l'Hospital's rule again:

$$\begin{aligned}
\lim_{n\to\infty} \ln\big(M_{Z_n}(t)\big) &= \frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{\ln M_Y(h)}{h^2} \\
&= \frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{M_Y'(h)}{2hM_Y(h)} \\
&= \frac{t^2}{\sigma^2} \lim_{h\to 0} \frac{M_Y''(h)}{2M_Y(h) + 2hM_Y'(h)} \\
&= \frac{t^2}{\sigma^2} \cdot \frac{M_Y''(0)}{2M_Y(0)} = \frac{t^2}{\sigma^2} \cdot \frac{\sigma^2}{2} \\
&= \frac{t^2}{2}
\end{aligned}$$

Thus $M_{Z_n}(t) \to e^{t^2/2}$ and therefore $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} X$, where $X \sim N(0,1)$. $\square$