STA257 Probability and Statistics I

Gun Ho Jang

Update on June 13, 2018

Note: This note is prepared for STA257. There might be numerous fault arguments/statements/typos. If you spot one, please contact the instructor or you may look up references which may contain errors too.

Mode of Convergence

Definition 42. A sequence of random variables X_n *converges to* X *in distribution* ($X_n \xrightarrow{d} X$) if $P(X_n \le x) \to P(X \le x)$ as $n \to \infty$ for any x with $P(X = x) = 0$. A sequence of random variables X_n *converges to* X *in probability* ($X_n \xrightarrow{p} X$) if, for any $\epsilon > 0$, $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$. A sequence of random variables X_n *converges to* X *almost surely* ($X_n \xrightarrow{a.s.} X$) if $P(\limsup_{n\to\infty} |X_n - X| = 0) = 1$. A sequence of random variables X_n *converges to* X *in* L^p ($X_n \xrightarrow{L^p} X$) for $p > 0$ if $\mathbb{E}(|X_n - X|^p) \to 0$ as $n \to \infty$.

In the above convergences, all random variables are converging except convergence in distribution. The convergence in distribution indicates distribution functions of random variables are converging instead of random variables.

The definition of almost sure convergence $X_n \xrightarrow{a.s.} X$ contains two properties X_n converges and the limit is X with probability one, or, $P(\lim_{n\to\infty} X_n$ exists and $\lim_{n\to\infty} X_n = X) = 1$.

Implications

Theorem 65. (a) $X_n \to X$ a.s. $\Longrightarrow X_n \to X$ in probability.
(b) $X_n \to X$ in $L^p \Longrightarrow X_n \to X$ in probability.
(c) $X_n \to X$ in probability $\Longrightarrow X_n \to X$ in distribution.

Proof. (a) Fix $\epsilon > 0$. Note that $\lim_{n\to\infty} X_n = X$ a.s. implies $\limsup_{n\to\infty} |X_n - X| = 0$ a.s. Hence

$$0 = P(\limsup_{n\to\infty} |X_n - X| > \epsilon) = P(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{|X_n - X| > \epsilon\}) = \lim_{m\to\infty} P(\bigcup_{n=m}^{\infty} \{|X_n - X| > \epsilon\}) \ge \lim_{m\to\infty} P(|X_m - X| > \epsilon).$$

(b) Fix $\epsilon > 0$. The probability $P(|X_n - X| > \epsilon)$ converges to

$$P(|X_n - X| > \epsilon) = \mathbb{E}[1(|X_n - X| > \epsilon)] \le \mathbb{E}\Big[\frac{1}{\epsilon^p}|X_n - X|^p 1(|X_n - X| > \epsilon)\Big] \le \frac{1}{\epsilon^p}\mathbb{E}[|X_n - X|^p] \to 0.$$

(c) Note that $P(X_n \le x) = P(X_n \le x, X \le x+\epsilon) + P(X_n \le x, X > x+\epsilon) \le P(X \le x+\epsilon) + P(|X_n - X| > \epsilon)$. Similarly $P(X \le x - \epsilon) = P(X \le x - \epsilon, X_n \le x) + P(X \le x - \epsilon, X_n > x) \le P(X_n \le x) + P(|X_n - X| > \epsilon)$.

Hence

$$P(X \le x - \epsilon) \le \liminf_{n\to\infty} P(X_n \le x) \le \limsup_{n\to\infty} P(X_n \le x) \le P(X \le x + \epsilon).$$

For any point x with $P(X = x)$, by taking ϵ small enough, we get $P(X_n \le x) \to P(X \le x)$, that is, $X_n \to X$ in distribution. $\square$

Example 90. Let $U \sim \text{uniform}(0,1)$.

- Let $X_n = 1(U \in [0,1/n])$. Then $X_n \to 0$ in probability, a.s. and in L^p for $p > 0$.
  Take $\epsilon \in (0,1)$. $P(|X_n - 0| > \epsilon) = P(X_n > \epsilon) = P(U \le 1/n) = 1/n \to 0$. $\limsup X_n = \limsup 1(U \in [0,1/n]) = 0$. $\mathbb{E}[|X_n - 0|^p] = \mathbb{E}[X_n^p] = \mathbb{E}[X_n] = \mathbb{E}[1(U \le 1/n)] = 1/n \to 0$.

- Let $Y_n = n1(U \in [0,1/n])$. Then $Y_n \to 0$ in probability, a.s. but not in L^p for $p \ge 1$.
  Take $\epsilon \in (0,1)$. $P(|Y_n - 0| > \epsilon) = P(Y_n > \epsilon) = P(U \le 1/n) = 1/n \to 0$. $\limsup Y_n = \limsup n1(U \in [0,1/n]) = 0$. $\mathbb{E}[|Y_n - 0|^p] = \mathbb{E}[Y_n^p] = \mathbb{E}[n^p 1(U \le 1/n)] = n^p(1/n) = n^{p-1}$ which diverges to ∞ if $p > 1$ and converges to 1 if $p = 0$. Hence Y_n does not converge to 0 in L^p for $p \ge 1$.

- Let $Z_n = 1(U \in [a_n, b_n))$ where $n = 2^k + m$ with $0 \le m < 2^k$, $a_n = m/2^k$ and $b_n = (m+1)/2^k$. Then $Z_n \to 0$ in probability and in L^p for $p > 0$ but not a.s. because $\limsup_{n\to\infty} Z_n = 1$.
  Take $\epsilon \in (0,1)$. $P(|Z_n - 0| > \epsilon) = P(Z_n > \epsilon) = 2^{-k_n} \to 0$ where $k_n = \lfloor \log_2(n) \rfloor$. $\mathbb{E}[|Z_n - 0|^p] = \mathbb{E}[Z_n^p] = \mathbb{E}[Z_n] = 2^{-k_n} \to 0$. $\limsup Z_n = 1$. Hence $P(\lim Z_n = 0) = 0$ and Z_n does not converge to 0 a.s.

- Let $W_n = U$ if n is odd and $W_n = 1 - U$ if n is even. Then $W_n \to U$ in distribution but not in probability.
  Note $P(W_n \le x) = x$ for any n and $0 < x < 1$. But $P(|W_n - W_{n-1}| > \epsilon) = P(|2U - 1| > \epsilon) = \max(0, 1 - 2\epsilon)$ implies W_n does not converge in probability.

Theorem 66. If a sequence of random variables X_n converges to X in probability, then there exists a subsequence n_k such that X_{n_k} converges to X a.s.

Proof. Let $n_0 = 0$. Sequentially take $n_k > n_{k-1}$ such that $P(|X_n - X| > 2^{-k}\}) < 2^{-k}$ for all $n \ge n_k$. Then $\{\lim_{k\to\infty} X_{n_k} \ne X\} \subset \bigcap_{m=1}^{\infty} \bigcup_{k \ge m} B_k$ where $B_k = \{|X_{n_k} - X| > 2^{-k}\}$. So we get

$$P(\{\lim_{k\to\infty} X_{n_k} \ne X\}) \le \lim_{m\to\infty} P(\bigcup_{k\ge m} B_k) \le \lim_{m\to\infty} \sum_{k \ge m} P(B_k) \le \lim_{m\to\infty} \sum_{k\ge m} 2^{-k} = \lim_{m\to\infty} 2^{1-m} = 0.$$

Hence the theorem follows. $\square$

Theorem 67. A sequence x_n of real numbers converges to x if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $x_{n_{k_l}}$ converges to x.

Proof. Sufficiency ($\Longrightarrow$) is obvious. Necessity ($\Longleftarrow$). If x_n does not converge to x, then the sequence $|x_n - x|$ does not converge to 0. Then there exists a $\delta > 0$ and a subsequence n_k such that $|x_{n_k} - x| > \delta$. However, from the assumption, there exists a further sequence n_{k_l} such that $x_{n_{k_l}} \to x$, i.e., $|x_{n_{k_l}} - x| \to 0$. Two statements contradicts. Thus x_n converges to x. $\square$

Theorem 68. A sequence of random variables X_n converges to X in probability if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $X_{n_{k_l}}$ converges to X a.s.

Proof. Necessity part ($\Longleftarrow$) is direct from Theorem 65.
Sufficiency ($\Longrightarrow$). Note that $X_n \xrightarrow{p} X$ implies $X_{n_k} \xrightarrow{p} X$. By applying Theorem 66, there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \xrightarrow{a.s.} X$ as $l \to \infty$. □

Example 91. Suppose $X_n \xrightarrow{p} X$, $|X_n| \le Y$ with $\mathbb{E}(Y) < \infty$. For any subsequence n_k, there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \xrightarrow{a.s.} X$. By applying the dominated convergence theorem, $\mathbb{E}[X_{n_{k_l}}] \to \mathbb{E}[X]$. Theorem 67 implies $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Exercise 33. Prove the generalized dominated convergence theorem with $X_n \to X$ in probability.

Exercise 34. Show that $X_n \to X$ in L^p if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \to X$ in L^p and a.s. Note. L^p is a vector space equipped with a topology.

Example 92. Suppose $X_n \to X$ and $Y_n \to Y$ a.s. Then $X_n + Y_n \to X + Y$ a.s. because $P(\lim_{n\to\infty}(X_n + Y_n) \ne X + Y) \le P(\lim_{n\to\infty} X_n \ne X) + P(\lim_{n\to\infty} Y_n \ne Y) = 0$. Similarly, $X_n Y_n \to XY$ a.s.

Example 93. Suppose $X_n \to X$, $Y_n \to Y$ in probability. For any subsequence n_k, there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \to X$ and $Y_{n_{k_l}} \to Y$ a.s. Hence $X_{n_{k_l}} + Y_{n_{k_l}} \to X + Y$ and $X_{n_{k_l}} Y_{n_{k_l}} \to XY$ a.s. Hence $X_n + Y_n \to X + Y$ and $X_n Y_n \to XY$ in probability.

Theorem 69. (a) If $X_n \xrightarrow{d} c$ where c is a constant, then $X_n \xrightarrow{p} c$.
(b) If $X_n \xrightarrow{p} X$ and $P(|X_n| \le M) = 1$ for some $M > 0$, then $X_n \xrightarrow{L^p} X$ for any $p > 0$.

Proof. (a) Fix $\epsilon > 0$, $P(|X_n - c| > \epsilon) \le P(X_n \le c - \epsilon) + 1 - P(X_n \le c + \epsilon) \to 0$ since $P(X_n \le x) \to 0$ for any $x < c$ and $P(X_n \le x) \to 1$ for any $x > c$.
(b) Note that $P(|X_n| \le M) = 1$ and $X_n \xrightarrow{p} X$ implies $P(|X| \le M) = 1$ and $|X_n - X| \le 2M$ for all n. Thus $|X_n - X|^p \le (2M)^p$ and $|X_n - X|^p \xrightarrow{p} 0$. Then for any $\epsilon > 0$,

$$\begin{aligned}\mathbb{E}[|X_n - X|^p] &= \mathbb{E}[|X_n - X|^p 1(|X_n - X| \le \epsilon)] + \mathbb{E}[|X_n - X|^p 1(|X_n - X| > \epsilon)] \\ &\le \epsilon^p + (2M)^p \mathbb{E}[1(|X_n - X| > \epsilon)] = \epsilon^p + (2M)^p P(|X_n - X| > \epsilon).\end{aligned}$$

Hence $\limsup_{n\to\infty} \mathbb{E}[|X_n - X|^p] \le \epsilon^p + (2M)^p \limsup_{n\to\infty} P(|X_n - X| > \epsilon) = \epsilon^p$ and again $\limsup_{n\to\infty} \mathbb{E}[|X_n - X|^p] \le \inf_{\epsilon>0} \epsilon^p = 0$. Therefore $\mathbb{E}[|X_n - X|^p] \to 0$. □

Theorem 70. Let X be a random variable with $P(X = x) = 0$ for all x and F be the distribution function of X. Then $F(X) \sim \text{uniform}(0, 1)$ and $F^{-1}(U) \sim X$ for any $U \sim \text{uniform}(0, 1)$.

Proof. The random variable X does not have any point mass from the assumption. Hence $F(x-) = \lim_{y<x: y\to x} F(y) = P(X < x) = P(X \le x) - P(X = x) = P(X \le x) = F(x)$ implies F is continuous.

Let $V = F(X)$ for simplicity. For any $v \in (0, 1)$, there exists x_v such that $F(x_v) = v$. Then $F_V(v) = P(V \le v) = P(F(X) \le v) = P(X \le x_v) = F(x_v) = v$, that is, $V \sim \text{uniform}(0, 1)$.

Let $Y = F^{-1}(U)$. For any x, $P(Y \le x) = P(F^{-1}(U) \le x) = P(F(F^{-1}(U)) \le F(x)) = P(U \le F(x)) = F(x) = P(X \le x)$. Hence $Y = F^{-1}(U)$ and X have the same distribution. □

Theorem 71 (Skorokhod's representation theorem)**.** If $X_n \xrightarrow{d} X$, then there exist random variables $Y, Y_1, Y_2, \ldots$ in a probability space such that
(a) X_n and Y_n have the same distribution as well as X and Y have the same distribution,
(b) $Y_n \xrightarrow{a.s.} Y$.

The below proof requires a bit of mathematics and you may skip this proof.

Proof. For simplicity, let $X_0 = X$. Let F_n be the distribution function of X_n for $n = 0, 1, 2, \ldots$. Consider functions $Y_n(u) = \inf\{x : F_n(x) \geq u\}$ for $n = 0, 1, 2, \ldots$. For a uniform random variable $U \sim \text{uniform}(0, 1)$, define random variables $Y_n = Y_n(U)$ for $n \geq 0$. Note that (a) $u \leq F_n(x)$ if and only if $Y_n(u) \leq x$, (b) $Y_n(\cdot)$ is non-decreasing, (c) $u \leq F_n(Y_n(u))$. Thus $P(Y \leq y) = P(Y(U) \leq y) = P(U \leq F_n(y)) = F_n(y) = P(X_n \leq y)$ which implies X_n and Y_n have the same distribution. Similarly, X and Y have the same distribution.

For any $x < y$, the event $x < Y(U) \leq y$ is equivalent to $F(x) < U \leq F(y)$ also $x < Y_n(U) \leq y$ is equivalent to $F_n(x) < U \leq F_n(y)$. If $P(Y = y) = 0 = P(Y = x)$, then $F_n(x) \to F(x$ and $F_n(y) \to F(y)$. Hence Let $h(F, u) = \inf\{x : F(x) \geq u\}$. then $h(F, \cdot)$ is non-decreasing. Take y so that $P(Y = y = 0)$. Let $u = F(y)$. Then there exists a unique u such that. Let $u = \max\{v : Y(v) = y\}$

Still $Y_n \xrightarrow{a.s.} Y$ should be proved, that is, $Y_n(u) \to Y(u)$ almost surely. For any $u \in (0, 1)$ and $\epsilon > 0$, let $y = Y(u)$. Then pick an x so that $y - \epsilon < x < y$ and $P(Y = x) = 0$. Since $F_n(x) \to F(x)$, there exists $N > 0$ such that $|F_n(x) - F(x)| < (F(y) - F(x))/2$ for all $n \geq N$. Then $F_n(x) < F(x) + (F(y) - F(x))/2 < F(y-) \leq u$. Hence $Y_n(u) > x$ for all $n \geq N$ which implies $Y(u) - \epsilon = y - \epsilon \leq \liminf_{n\to\infty} Y_n(u)$. By taking $\epsilon > 0$ arbitrarily small, $Y(u) \leq \liminf_{n\to\infty} Y_n(u)$.

For any $v \in (F(y), 1)$ and $\epsilon > 0$, there exists $z > y$ such that $Y(v) < z < Y(v) + \epsilon$ with $P(Y = z) = 0$. Then for sufficiently large n, $|F_n(z) - F(z)| < (F(z) - F(y))/2$ which implies $F_n(z) > (F(y) + F(z))/2 > u$. Hence $Y_n(u) < z < Y(v) + \epsilon$. Send n to infinity and ϵ to zero to obtain $\limsup_{n\to\infty} Y_n(u) \leq Y(v)$ for any $v > F(y) \geq u$. Hence $Y_n(u) \to Y(u)$ as long as $\lim_{v\searrow u} Y(v) = Y(u)$. Since Y is non-decreasing, there are at most countably many discontinuity points, say D. Then $P(Y \in D) = P(U \in Y^{-1}(D)) = 0$ because $Y^{-1}(D)$ is at most countable. Hence $Y_n \xrightarrow{a.s.} Y$. □

Note. Roughly speaking, Skorohkod's representation theorem can be interpreted as, for a given $U \sim \text{uniform}(0, 1)$, new random variables $Y_n = F_n^{-1}(U) \sim F_n \sim X_n$ converges almost surely to $Y = F^{-1}(U)$ where F_n is the distribution function of X_n.

Theorem 72 (Continuous mapping theorem)**.** Let g be a continuous function.
(a) $X_n \xrightarrow{a.s.} X$ implies $g(X_n) \xrightarrow{a.s.} g(X)$.
(b) $X_n \xrightarrow{p} X$ implies $g(X_n) \xrightarrow{p} g(X)$.
(c) $X_n \xrightarrow{d} X$ implies $g(X_n) \xrightarrow{d} g(X)$.

Proof. Recall that g is continuous if $g(x_n) \to g(x)$ as long as $x_n \to x$.
(a) $P(\limsup_{n\to\infty} |g(X_n) - g(X)| > 0) \leq P(\limsup_{n\to\infty} |X_n - X| > 0) = 0$.
(b) For any subsequence n_k, $X_{n_k} \xrightarrow{p} X$ and hence there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \xrightarrow{a.s.} X$. Then by part (a), $g(X_{n_{k_l}}) \xrightarrow{a.s.} g(X)$. Theorem 66 implies $g(X_n) \xrightarrow{p} g(X)$.
(c) From Skorokhod's representation theorem, there exist $Y, Y_1, Y_2, \ldots$ such that $P(X \leq x) = P(Y \leq x)$, $P(X_n \leq x) = P(Y_n \leq x)$ for all x and $Y_n \xrightarrow{a.s.} Y$. By part (a), $g(Y_n) \xrightarrow{a.s.} g(Y)$ which implies $g(Y_n) \xrightarrow{d} g(Y)$. Then $P(g(X_n) \leq x) = P(g(Y_n) \leq x) \to P(g(Y) \leq x) = P(g(X) \leq x)$ for any x with $P(g(X) = x) = 0$. Hence $g(X_n) \xrightarrow{d} g(X)$. □

Basic L^1 Convergence

Lemma 73. If $Y \geq 0$ and $\mathbb{E}(Y) < \infty$, then for any $\epsilon > 0$ there exists $M > 0$ such that $\mathbb{E}[Y1(Y > M)] < \epsilon$.

Proof. Suppose $\mathbb{E}[Y1(Y > y)]$ does not converge to 0. Then there exists an increasing sequence y_n such that $\mathbb{E}[Y1(Y > y_n]) \to c$ where $c > 0$. The convergence implies there exists $n_0 > 0$ such that $\mathbb{E}[Y1(Y > y_n)] \geq$

$2c/3$ for all $n \geq n_0$. For any $k \geq 1$, we take n_k sequentially increasing. Since $\mathbb{E}[Y1(Y > n_{k-1})] > 2c/3$, there exists $n_k > n_{k-1}$ such that $\mathbb{E}[Y1(y_{n_{k-1}} < Y \leq y_n)] \geq c/3$ for all $n \geq n_k$. Then

$$\mathbb{E}[Y] \geq \sum_{k=1}^{\infty} \mathbb{E}[Y1(y_{n_{k-1}} < Y \leq Y_{n_k})] \geq \sum_{k=1}^{\infty} \frac{c}{3} = \infty.$$

Which contradicts to the assumption $\mathbb{E}(Y) < \infty$. Thus $\limsup_{y\to\infty} \mathbb{E}[Y1(Y > y)] = 0$. □

Exercise 35. Prove that $nP(X > n) \to 0$ as $n \to \infty$ if $\mathbb{E}(|X|) < \infty$.

Lemma 74. Suppose a random variable Y has a finite absolute expectation, that is, $\mathbb{E}(|Y|) < \infty$ and a sequence A_n of events satisfy $P(A_n) \to 0$. Then $\mathbb{E}(Y1_{A_n}) \to 0$ where 1_A is an indicator function of the event A.

Proof. Fix $\epsilon > 0$. From the finite expectation assumption, there exists $M > 0$ such that $\mathbb{E}[|Y|1(|Y| > M)] < \epsilon/2$ by Lemma 73. There exists $N > 0$ such that $P(A_n) < \epsilon/(2M)$ for all $n \geq N$. Then for any $n \geq N$,

$$\begin{aligned}|\mathbb{E}[Y1_{A_n}]| &\leq \mathbb{E}[|Y|1_{A_n}] = \mathbb{E}[|Y|1(|Y| > M)1_{A_n}] + \mathbb{E}[|Y|1(|Y| \leq M)1_{A_n}] \\ &\leq \mathbb{E}[|Y|1(|Y| > M)] + \mathbb{E}[M1_{A_n}] \leq \epsilon/2 + MP(A_n) \leq \epsilon/2 + M\epsilon/(2M) \\ &\leq \epsilon.\end{aligned}$$

The arbitrariness of $\epsilon > 0$ implies $|\mathbb{E}[Y1(Y \in A_n)]| \to 0$ and the lemma holds. □

Theorem 75 (Dominated Convergence Theorem)**.** Suppose that $X_n \to X$ in probability, $|X_n| \leq Y$ and $\mathbb{E}(Y) < \infty$. Then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Proof. We rather prove $\mathbb{E}(|X_n - X|) \to 0$. Which implies the theorem via the triangle inequality.

Fix $\epsilon > 0$. From $|X_n| \leq Y$, we get $|X| \leq Y$ and hence $|X_n - X| \leq 2Y$. The convergence $X_n \xrightarrow{p} X$ implies $P(|X_n - X| > \epsilon/2) \to 0$.

$$\begin{aligned}\mathbb{E}(|X_n - X|) &= \mathbb{E}[|X_n - X|1(|X_n - X| \leq \epsilon/2)] + \mathbb{E}[|X_n - X|1(|X_n - X| > \epsilon/2)] \\ &\leq \mathbb{E}[\epsilon/2 1(|X_n - X| \leq \epsilon/2)] + \mathbb{E}[2Y1(|X_n - X| > \epsilon/2)]\end{aligned}$$

From Lemma 74, $\mathbb{E}[2Y1(|X_n - X| > \epsilon/2)] \to 0$. Hence there exists $N > 0$, such that $\mathbb{E}[2Y1(|X_n - X| > \epsilon/2)] < \epsilon/2$ for all $n \geq N$.

$$\leq \epsilon/2 + \epsilon/2 \leq \epsilon.$$

By taking $\epsilon > 0$ arbitrarily small, the result $\mathbb{E}(|X_n - X|) \to 0$ is obtained. □

Theorem 76 (Monotone Convergence Theorem)**.** Let X_n be non-negative non-decreasing random variables. Suppose $X = \lim_{n\to\infty} X_n$ is finite a.s. Then $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$.

Proof. Apply DCT for $0 \leq X_n \leq X$ and $\mathbb{E}(X) < \infty$.
Second Proof using MCT in integration: Since $X_n \to X$ a.s., $f_n(x) := P(X_n > x) \to P(X > x) =: f(x)$ as long as $P(X = x) = 0$. Hence $f_n \to f$ a.e. and $f_n \nearrow f$. Using the monotone convergence theorem of

integral we get

$$\mathbb{E}(X_n) = \int_0^\infty P(X_n > x)\,dx = \int_0^\infty f_n(x)\,dx \nearrow \int_0^\infty f(x)\,dx = \int_0^\infty P(X > x)\,dx = \mathbb{E}(X).$$

Thus the theorem follows. $\square$

Example 94. Suppose $X_n \geq 0$ with $\sum_{n=1}^\infty \mathbb{E}(X_n) < \infty$. Let $Y_n = X_1 + \cdots + X_n$. Then Y_n converges to $Y = \sum_{n=1}^\infty X_n$ a.s. By the MCT, $\sum_{n=1}^\infty \mathbb{E}(X_n) = \lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}(X_k) = \lim_{n\to\infty} \mathbb{E}(Y_n) \to \mathbb{E}(Y) = \mathbb{E}(\sum_{n=1}^\infty X_n)$.

Theorem 77 (Fatou's lemma)**.** Let $X_1, X_2, \ldots$ be a sequence of non-negative random variables. Then

$$\mathbb{E}[\liminf_{n\to\infty} X_n] \leq \liminf_{n\to\infty} \mathbb{E}(X_n).$$

Proof. Let $Y_n = \inf_{m\geq n} X_m$ so that $\liminf_{n\to\infty} X_n = \lim_{n\to\infty} \inf_{m\geq n} X_m = \lim_{n\to\infty} Y_n$. Obviously Y_n is non-decreasing. Also $\mathbb{E}(Y_n) = \mathbb{E}[\inf_{m\geq n} X_m] \leq \mathbb{E}[X_m]$ for all $m \geq n$ implies $\mathbb{E}(Y_n) \leq \inf_{m\geq n} \mathbb{E}(X_m)$. Using the monotone convergence theorem implies

$$\mathbb{E}[\liminf_{n\to\infty} X_n] = \mathbb{E}[\lim_{n\to\infty} Y_n] = \lim_{n\to\infty} \mathbb{E}(Y_n) \leq \lim_{n\to\infty} \inf_{m\geq n} \mathbb{E}(X_m) = \liminf_{n\to\infty} \mathbb{E}(X_n).$$

$\square$

Theorem 78 (Dominated convergence theorem in classical sense)**.** Suppose $X_n \to X$ a.s. and $|X_n| \leq Y$ with $\mathbb{E}(Y) < \infty$. Then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Classical Proof. Note $Y + X_n \geq 0$ and $Y + X_n \to Y - X$ a.s. By Fatou's lemma, $\mathbb{E}(Y+X) = \mathbb{E}[\liminf_{n\to\infty}(Y + X_n)] \leq \liminf_{n\to\infty} \mathbb{E}(Y + X_n) = \mathbb{E}(Y) + \liminf_{n\to\infty} \mathbb{E}(X_n)$ which implies $\mathbb{E}(X) \leq \liminf_{n\to\infty} \mathbb{E}(X_n)$. Similarly, $Y - X_n \geq 0$ with $Y - X_n \to Y - X$ a.s. Hence $\mathbb{E}(Y - X) \leq \liminf_{n\to\infty} \mathbb{E}(Y - X_n) = \mathbb{E}(Y) - \limsup_{n\to\infty} \mathbb{E}(X_n)$. Hence we get

$$\mathbb{E}(X) \leq \liminf_{n\to\infty} \mathbb{E}(X_n) \leq \limsup_{n\to\infty} \mathbb{E}(X_n) \leq \mathbb{E}(X)$$

which implies $\mathbb{E}(X_n) \to \mathbb{E}(X)$ as $n \to \infty$. $\square$

Example 95. Suppose random variables X_n satisfy $\sum_{n=1}^\infty \mathbb{E}(|X_n|) < \infty$. Let $Y = |X_1| + |X_2| + \cdots = \sum_{n=1}^\infty |X_n|$. Then $|X_n| \leq Y$ and $\mathbb{E}(Y) = \sum_{n=1}^\infty \mathbb{E}(|X_n|) < \infty$. By DCT, $X_1 + X_2 + \cdots \to X$ a.s. and $\mathbb{E}(\sum_{n=1}^\infty X_n) = \sum_{n=1}^\infty \mathbb{E}(X_n)$.

Example 96. Suppose $\mathbb{E}(|X|^r) < \infty$. Let $X_n = |X| 1(|X| \geq n)$. Then $X_n \to 0$ a.s. and $|X_n| \leq |X|$. Which implies $X_n^r \to 0$ a.s. and $|X_n^r| \leq |X|^r$. By DCT, $\mathbb{E}(X_n^r) \to 0$. Then $n^r P(|X| \geq n) \leq \mathbb{E}[X_n^r] \to 0$.

Exercise 36. Show the next theorem.

Theorem (Generalized Dominated Convergence Theorem)**.** If all X, Y, X_n, Y_n have finite absolute expectation, $|X_n| \leq Y_n$ for all n, $X_n \xrightarrow{p} X$, $Y_n \xrightarrow{p} Y$, and $\mathbb{E}(Y_n) \to \mathbb{E}(Y)$, then $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Example 97. Suppose $X_n \xrightarrow{p} X$, $|X_n| \leq Y$ with $\mathbb{E}(Y) < \infty$. For any subsequence n_k, there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \xrightarrow{a.s.} X$. By applying the dominated convergence theorem, $\mathbb{E}[X_{n_{k_l}}] \to \mathbb{E}[X]$. Theorem 67 implies $\mathbb{E}(X_n) \to \mathbb{E}(X)$.

Exercise 37. Prove the generalized dominated convergence theorem with $X_n \to X$ in probability.

Exercise 38. Show that $X_n \to X$ in L^p if and only if for any subsequence n_k there exists a further subsequence n_{k_l} such that $X_{n_{k_l}} \to X$ in L^p and a.s. Note. L^p is a vector space equipped with a topology.

Convergence in distribution

Theorem 79. $X_n \xrightarrow{d} X$ if and only if $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for any bounded continuous function g.

Proof. **Sufficiency ($\Longrightarrow$).** Take Skorokhod's representation theorem, say Y and $Y_1, Y_2, \ldots$. Let $M = \sup_x |g(x)| < \infty$. Hence $|g(Y_n)| \le M < \infty$. The dominated convergence theorem implies $\mathbb{E}[g(X_n)] = \mathbb{E}[g(Y_n)] \to \mathbb{E}[g(Y)] = \mathbb{E}[g(X)]$.
Necessity ($\Longleftarrow$). For $y < z$, define a continuous function $h_{y,z}$ by $h_{y,z}(x) = 1$ if $x \le y$, $h_{y,z}(x) = 0$ if $x > z$, and $h_{y,z}(x) = (z-x)/(z-y)$ so that $h_{y,z}$ is continuous and bounded like $0 \le 1(x \le y) \le h_{y,z}(x) \le 1(x \le z) \le 1$. From $\mathbb{E}[h_{y,z}(X_n)] \to \mathbb{E}[h_{y,z}(X)]$ and $P(X_n \le y) = \mathbb{E}[1(X_n \le y)] \le \mathbb{E}[h_{y,z}(X_n)] \le \mathbb{E}[1(X_n \le z)] = P(X_n \le z)$, we get $\limsup_{n\to\infty} P(X_n \le y) \le \limsup_{n\to\infty} \mathbb{E}[h_{y,z}(X_n)] = \mathbb{E}[h_{y,z}(X)] = \liminf_{n\to\infty} \mathbb{E}[h_{y,z}(X_n)] = \liminf_{n\to\infty} P(X_n \le z)$. Pick x so that $P(X = x) = 0$. For any $\epsilon > 0$, $\mathbb{E}[h_{x-\epsilon,x}(X)] \le \liminf_{n\to\infty} P(X_n \le x) \le \limsup_{n\to\infty} P(X_n \le x) \le \mathbb{E}[h_{x,x+\epsilon}(X)]$. Hence the limit of $P(X_n \le x)$ exists because $\inf_{\epsilon>0}\{\mathbb{E}[h_{x-\epsilon,x}(X)] - \mathbb{E}[h_{x,x+\epsilon}(X)]\} \le \inf_{\epsilon>0} P(x - \epsilon \le X \le x + \epsilon) = F_X(y+) - F_X(y-) = F_X(y) - F_X(y) = 0$. Hence $X_n \xrightarrow{d} X$. □

Theorem 80. $X_n \xrightarrow{d} X$ if and only if $\mathrm{chf}_{X_n}(t) \to \mathrm{chf}_X(t)$.

Proof. The sufficiency ($\Longrightarrow$) is direct from Theorem 79.
The necessity ($\Longleftarrow$) requires tedious rigorous steps. A sketch is given below using inversion formula. Fix $a < b$. Define $h_n = [(e^{-iat} - e^{-ibt})/(it)]\mathrm{chf}_{X_n}(t)$ is continuous, bounded by $b - a$ and converges to $h = [(e^{-iat} - e^{-ibt})/(it)]\mathrm{chf}_X(t)$. Hence

$$\lim_{n\to\infty}[P(a < X_n < b) + \{P(X_n = a) + P(X_n = b)\}/2] = \lim_{n\to\infty}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} h_n(t)\,dt$$

change the order of limit and apply dominated convergence theorem

$$= \lim_{T\to\infty}\lim_{n\to\infty}\frac{1}{2\pi}\int_{-T}^{T} h_n(t)\,dt = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T} h(t)\,dt = P(a < X < b) + \{P(X = a) + P(X = b)\}/2.$$

Roughly speaking, by taking $a < b$ so that $P(X = a) = P(X = b) = 0$, the convergence $P(a < X_n \le b) \to P(a < X \le b)$ is obtained as well as $X_n \xrightarrow{d} X$. □

Theorem 81. If $X_n \xrightarrow{d} X$, then $aX_n + b \xrightarrow{d} aX + b$ for any $a, b \in \mathbb{R}$.

Proof. Proof I: If $a = 0$, then $aX_n + b \equiv b \equiv aX + b$. Assume either $a > 0$ or $a < 0$. For any x so that $P(X = x) = 0$, if $a > 0$, $P(aX_n + b \le ax + b) = P(X_n \le x) \to P(X \le x) = P(aX + b \le ax + b)$, if $a < 0$, then $P(aX_n + b \le ax + b) = P(X_n \ge x) = 1 - P(X_n < x) \to 1 - P(X \le x) = P(aX + b \le ax + b)$ where $P(X = x) = 0$ is used.
Proof II: $\mathrm{chf}_{aX_n+b}(t) = \mathbb{E}[e^{it(aX_n+b)}] = e^{itb}\mathbb{E}[e^{i(ta)X_n}] = e^{itb}\mathrm{chf}_{X_n}(ta) \to e^{itb}\mathrm{chf}_X(ta) = \mathbb{E}[e^{itaX+itb}] = \mathbb{E}[e^{it(aX+b)}] = \mathrm{chf}_{aX+b}(t)$. Hence $aX_n + b \xrightarrow{d} aX + b$ as $n \to \infty$. □

Theorem 82 (Slutsky's lemma)**.** Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$ for a constant c.
(a) $X_n + Y_n \xrightarrow{d} X + c$,
(b) $X_n Y_n \xrightarrow{d} Xc$,
(c) $X_n/Y_n \xrightarrow{d} X/c$ if $c \neq 0$.

Proof. **A tedious proof:** (a) Fix x so that $P(X = x) = 0$. Note that, for any $\epsilon > 0$ with $P(|X-x| = \epsilon) = 0$), if $X_n + Y_n \le x + c$, then $|Y_n - c| \ge \epsilon$ or $X_n \le x + \epsilon$. Thus $P(X_n + Y_n \le x + c) \le P(|Y_n - c| \ge \epsilon) + P(X_n \le x + \epsilon) \to P(X \le x + \epsilon)$. Similarly, $X_n \le x - \epsilon$ and $|Y_n - c| < \epsilon$ implies $X_n + Y_n \le x + c$. Hence $P(X_n + Y_n \le x + c) \ge P(X_n \le x - \epsilon, |Y_n - c| < \epsilon) \ge P(X_n \le x - \epsilon) - P(|Y_n - c| \ge \epsilon) \to P(X \le x - \epsilon)$. In sum, $\limsup_{n\to\infty} |P(X_n + Y_n \le x + c) - P(X \le x)| \le P(x - \epsilon < X \le x + \epsilon) \to 0$ as $\epsilon \to 0$. Finally $X_n + Y_n \xrightarrow{d} X + c$.
(b) Fix $\epsilon > 0$, there exists $M > 0$ so that $P(|X| > M) < \epsilon$ and $P(|X| = M) = 0$. Then $P(|X_n(Y_n - c)| > \epsilon) \le P(|X_n| > M) + P(|Y_n - c| > \epsilon/M) \to P(|X| > M) < \epsilon$. Hence $X_n(Y_n - c) \xrightarrow{p} 0$. Also $cX_n \xrightarrow{d} cX$. By part (a), $X_n Y_n = cX_n + X_n(Y_n - c) \xrightarrow{d} cX$.
(c) The continuous mapping theorem implies $1/Y_n \xrightarrow{d} 1/c$. Apply (b) to obtain $X_n/Y_n = X_n(1/Y_n) \xrightarrow{d} X/c$.
An elegant proof: Note that $(X_n, Y_n) \xrightarrow{d} (X, c)$ because for any $\epsilon > 0$, $P(X_n \le x, Y_n < c - \epsilon) \le P(Y_n < c - \epsilon) = 0$ and $P(X_n \le x, Y_n \le c + \epsilon) = P(X_n \le x) - P(X_n \le x, Y_n > c + \epsilon) \to P(X \le x)$ for all x with $P(X = x) = 0$. Three maps $(x, y) \mapsto x + y$, $(x, y) \mapsto xy$ and $(x, y) \mapsto x/y$ are continuous. The continuous mapping theorem implies the results. ☐

Law of Large Numbers

Example 98 (Weak law of large numbers)**.** Let $X_1, \ldots, X_n$ be an i.i.d. (independent and identically distributed) with mean μ and finite variance σ^2. Then the sample mean $\overline{X}_n = (X_1 + \cdots + X_n)/n$ has mean $\mathbb{E}(\overline{X}_n) = \mu$ and variance $\mathbb{V}\mathrm{ar}(\overline{X}_n) = \mathbb{V}\mathrm{ar}(X_1)/n = \sigma^2/n$. Chebychev's inequality implies, for any $\epsilon > 0$,

$$P(|\overline{X}_n - \mu| > \epsilon) = P(|\overline{X}_n - \mu| > (\epsilon/\sigma)\sigma) \le \mathbb{V}\mathrm{ar}(\overline{X}_n)/(\epsilon/\sigma)^2 = \sigma^2/(n\epsilon^2) \to 0.$$

In other words, $\overline{X}_n$ converges to the mean μ in probability as n increases.

Theorem 83. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}(X_n^2) < \infty$. For $\mu = \mathbb{E}(X_1)$, $\overline{X}_n = (X_1 + \cdots + X_n)/n \longrightarrow \mu$ almost surely and in L^2.

Proof. Note that $\mathbb{E}[\overline{X}_n] = \mu$ and $\mathbb{E}[(\overline{X}_n - \mu)^2] = \mathbb{V}\mathrm{ar}(\overline{X}_n) = \mathbb{V}\mathrm{ar}(X_1 + \cdots + X_n)/n^2 = n\mathbb{V}\mathrm{ar}(X_1)/n^2 = \mathbb{V}\mathrm{ar}(X_1)/n \to 0$ implies L^2 convergence.
Claim: Let Y_n be nonnegative i.i.d. random variables with $\mathbb{E}(Y_n^2) < \infty$. Then $V_n/n \to \mu_y$ where $V_n = Y_1 + \cdots + Y_n$ and $\mu_y = \mathbb{E}(Y_n)$.
Let $n_k = k^2$ and $\sigma_y^2 = \mathbb{V}\mathrm{ar}(Y_1)$. Then, for $\epsilon > 0$, $P(|V_{n_k}/n_k - \mu_y| > \epsilon) \le \epsilon^{-2}\mathbb{V}\mathrm{ar}(V_{n_k}/n_k) = \epsilon^{-2}\mathbb{V}\mathrm{ar}(Y_1)/n_k = \epsilon^{-2}\sigma_y^2/k^2$. Hence $\sum_{k=1}^\infty P(|V_{n_k}/n_k - \mu_y| > \epsilon) \le \epsilon^{-2}\sigma_y^2 \sum_{k=1}^\infty 1/k^2 < \infty$ implies $\limsup_{k\to\infty} |V_{n_k}/n_k - \mu_y| \le \epsilon$ almost surely. By taking $\epsilon \to 0$, $\limsup_{k\to\infty} |V_{n_k}/n_k - \mu_y| = 0$ almost surely that is equivalent to $V_{n_k}/n_k \longrightarrow \mu_y$ almost surely. For any n, there exists k such that $k^2 \le n \le (k+1)^2$. Then

$$\frac{k^2}{(k+1)^2}\frac{V_{k^2}}{k^2} = \frac{V_{k^2}}{(k+1)^2} \le \frac{V_n}{n} \le \frac{V_{(k+1)^2}}{k^2} = \frac{V_{(k+1)^2}}{(k+1)^2}\frac{(k+1)^2}{k^2}.$$

As $n \to \infty$, $(k/(k+1))^2 \to 1$ and $V_{k^2}/k^2 \to \mu_y$ a.s. Hence $V_n/n \longrightarrow \mu_y$ almost surely.

Recall that $X_n = X_{n,+} - X_{n,-}$ where $X_{n,+} = \max(0, X_n)$ and $X_{n,-} = \max(0, -X_n)$. Let $S_n = X_{1,+} + \cdots + X_{n,+}$ and $T_n = X_{1,-} + \cdots + X_{n,-}$. Then $\overline{X}_n = (X_1 + \cdots + X_n)/n = (X_{1,+} - X_{1,-} + \cdots + X_{n,+} - X_{n-})/n = S_n/n - T_n/n \xrightarrow{a.s.} \mathbb{E}[X_{1,+}] - \mathbb{E}[X_{1,-}] = \mathbb{E}[X_{1,+} - X_{1,-}] = \mathbb{E}(X_1)$. □

Theorem 84 (Weak law of large numbers)**.** Let X_n be i.i.d. with $\mathbb{E}(|X_n|) < \infty$. Then $\overline{X}_n \longrightarrow \mathbb{E}(X_1)$ in probability.

Proof. Let $\mu = \mathbb{E}(X_1)$. Recall $\mathrm{chf}_{X_1}(t) = 1 + i\mu t + o(|t|)$. Note that $\mathrm{chf}_{\overline{X}_n}(t) = \mathbb{E}[\exp(it\overline{X}_n)] = \mathbb{E}[\exp(it(X_1 + \cdots + X_n)/n)] = \mathbb{E}[\exp(itX_1/n)] \cdots \mathbb{E}[\exp(itX_n/n)] = \{\mathbb{E}[\exp(i(t/n)X_1)]\}^n = \{\mathrm{chf}_{X_1}(t/n)\}^n = (1 + i\mu(t/n) + o(|t/n|))^n = \exp(n \log(1 + i\mu(t/n) + o(|t/n|))) = \exp(n[i\mu(t/n) + o(|t/n|) + o(|i\mu(t/n)|)]) = \exp(it + o(|t|)) \to \exp(it)$ which is the characteristic function of constant μ. Hence $\overline{X}_n \xrightarrow{d} \mu$. Thus $\overline{X}_n \xrightarrow{p} \mu$. □

Exercise 39. Let X_n be i.i.d. with $\mathbb{E}(|X_n|) < \infty$. Show that $\overline{X}_n \to \mathbb{E}(|X_1|)$ in L^1.

Theorem 85 (Strong law of large numbers)**.** Let X_n be i.i.d. with $\mathbb{E}(|X_n|) < \infty$. Then $\overline{X}_n \longrightarrow \mathbb{E}(X_1)$ almost surely.

A proof of strong law of large numbers is beyond our scope. A sketch of proof is as follows. Define $Y_n = X_n 1(|X_n| \le n)$. Then $Y_n = X_n$ almost surely using $\sum_n P(Y_n \neq X_n) = \sum_n P(|X_n| > n) \le \mathbb{E}(|X_1|) < \infty$. Take $n_k = \lfloor \alpha^k \rfloor$ for a $\alpha > 1$. Then, for $T_n = Y_1 + \cdots + Y_n$, $\sum_k P(|(T_{n_k} - \mathbb{E}(T_{n_k}))/n_k| > \epsilon) \le \epsilon^{-2} \sum_k \mathbb{V}\mathrm{ar}(T_{n_k})/n_k^2 < \infty$ implies $(T_{n_k} - \mathbb{E}(T_{n_k}))/n_k \to 0$ almost surely. Using $\mathbb{E}(T_{n_k})/n_k \to \mathbb{E}(X_1)$, $T_{n_k}/n_k \to \mathbb{E}(X_1)$ almost surely. Then apply similar method to Theorem 83 to obtain $T_n/n \to \mathbb{E}(X_1)$ almost surely. Since the $X_n = Y_n$ almost surely, $\overline{X}_n/n \to \mathbb{E}(X_1)$ almost surely.

Central Limit Theorem

Central limit theorem was found long ago for binomial cases which is called de Moivre-Laplace theorem.

Theorem 86. For k around np, the binomial probability is approximated by

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Big(- \frac{(k-np)^2}{2np(1-p)} \Big).$$

Proof. Note that $k/n \approx p$, $(n-k)/n \approx 1-p$ and let $z_n = (k-np)/\sqrt{n}$ or $k = np + z_n\sqrt{n}$. Then

$$\begin{aligned}
&\log\Big[\binom{n}{k} p^k (1-p)^{n-k}\Big] = \log\Big[\frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}\Big] \\
&\approx \frac{1}{2}\log 2\pi + (n + \frac{1}{2})\log n - n - \Big[\frac{1}{2}\log 2\pi + (k + \frac{1}{2})\log k - k + \frac{1}{2}\log 2\pi + (n - k + \frac{1}{2})\log(n-k) - (n-k)\Big] \\
&\qquad + k\log(p) + (n-k)\log(1-p) \\
&= -\frac{1}{2}\log 2\pi\frac{k(n-k)}{n} - k\log(k/n) - (n-k)\log(1-k/n) + k\log p + (n-k)\log(1-p) \\
&= -\frac{1}{2}\log 2\pi\frac{k(n-k)}{n} - k\log\Big(1 + \frac{z_n}{p\sqrt{n}}\Big) - (n-k)\log\Big(1 - \frac{z_n}{(1-p)\sqrt{n}}\Big)
\end{aligned}$$

using a Taylor expansion of log given by $\log(1-z) = -[z + z^2/2 + O(|z|^3)]$

$$
\begin{aligned}
&= -\frac{1}{2}\log 2\pi n\frac{k}{n}(1-\frac{k}{n}) - k\log\Big(1+\frac{z_n}{p\sqrt{n}}\Big) - (n-k)\log\Big(1-\frac{z_n}{(1-p)\sqrt{n}}\Big)\\
&= -\frac{1}{2}\log 2\pi np(1-p)(1+O_p(\frac{1}{n^{1/2}})) - k\Big(\frac{z_n}{p\sqrt{n}} - \frac{z_n^2}{2p^2n} + O_p(\frac{1}{n^{3/2}})\Big) + (n-k)\Big(\frac{z_n}{(1-p)\sqrt{n}} + \frac{z_n^2}{2(1-p)^2n} + O_p(\frac{1}{n^{3/2}})\Big)\\
&= -\frac{1}{2}\log 2\pi np(1-p) + \frac{z_n}{\sqrt{n}}\Big(\frac{n-k}{1-p} - \frac{k}{n}\Big) + \frac{z_n^2}{2n}\Big(\frac{k}{p^2} + \frac{n-k}{(1-p)^2}\Big) + O_p(\frac{1}{n^{1/2}})\\
&= -\frac{1}{2}\log 2\pi np(1-p) + \frac{z_n}{\sqrt{n}}\Big(-\frac{z_n\sqrt{n}}{p(1-p)}\Big) + \frac{z_n^2}{2n}\Big(\frac{n}{p(1-p)} + O_p(\sqrt{n})\Big) + O_p(\frac{1}{n^{1/2}})\\
&= -\frac{1}{2}\log 2\pi np(1-p) - \frac{z_n^2}{2p(1-p)} + O_p(\frac{1}{n^{1/2}})\\
&= -\frac{1}{2}\log 2\pi np(1-p) - \frac{(k-np)^2}{2np(1-p)} + O_p(\frac{1}{n^{1/2}}).
\end{aligned}
$$

□

When $X_n \sim \text{binomial}(n,p)$, define $Z_n = (X_n - np)/\sqrt{np(1-p)}$. Then for any $a < b$

$$
\begin{aligned}
P(a < Z_n < b) &= P(np + a\sqrt{np(1-p)} < X_n < np + b\sqrt{np(1-p)})\\
&= \sum_{k: np+a\sqrt{np(1-p)} < k < np+b\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k}\\
&\approx \sum_{k: np+a\sqrt{np(1-p)} < k < np+b\sqrt{np(1-p)}} \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Big(-\frac{(k-np)^2}{2np(1-p)}\Big)\\
&\approx \int_{np+a\sqrt{np(1-p)}}^{np+b\sqrt{np(1-p)}} \frac{1}{\sqrt{2\pi np(1-p)}} \exp\Big(-\frac{(k-np)^2}{2np(1-p)}\Big)\, dk
\end{aligned}
$$

let $z = (k-np)/\sqrt{np(1-p)}$

$$
= \int_a^b \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{z^2}{2}\Big)\, dz.
$$

Hence $Z_n \xrightarrow{d} N(0,1)$ which is an earliest version of central limit theorem. Actually this proof showed a sequence of densities converges to the standard normal density which is stronger than convergence in distribution.

Theorem 87 (Lévy's Central Limit Theorem)**.** Let X_n be i.i.d. with $\mu = \mathbb{E}(X_n)$ and $\sigma^2 = \mathbb{V}\text{ar}(X_n)$. Then $\sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$.

Proof. Let $Y_j = (X_j - \mu)/\sigma$ so that Y_n are i.i.d. with mean zero and variance 1. The characteristic function of Y_j satisfies

$$
\text{chf}_{Y_j}(t) = 1 + i\cdot 0\cdot t - 1^2\cdot t^2/2 + o(t^2) = 1 - t^2/2 + o(t^2).
$$

Let $Z_n = \sqrt{n}(\overline{X}_n - \mu)/\sigma = \sqrt{n}\overline{Y}_n$ and its characteristic function is

$$
\begin{aligned}
\mathrm{chf}_{Z_n}(t) &= \mathbb{E}[e^{itZ_n}] = \mathbb{E}[\exp(it\sqrt{n}\overline{Y}_n)] = \{\mathbb{E}[\exp(itY_1/\sqrt{n})]\}^n = \{\mathrm{chf}_{Y_1}(t/\sqrt{n})\}^n = \Big\{1 - \frac{1}{2}\big(\frac{t}{\sqrt{n}}\big)^2 + o\big(\frac{t^2}{n}\big)\Big\}^n \\
&= \exp\Big[n\log\Big(1 - \frac{t^2}{2n} + o\big(\frac{t^2}{n}\big)\Big)\Big] = \exp\Big[-n\Big\{\frac{t^2}{2n} + o\big(\frac{t^2}{n}\big) + \frac{1}{2}\Big(\frac{t^2}{2n} + o\big(\frac{t^2}{n}\big)\Big)^2 + O\Big(\big(\frac{t^2}{n} + o\big(\frac{t^2}{n}\big)\big)^3\Big)\Big\}\Big] \\
&= \exp\Big(-\frac{t^2}{2} + o(t^2)\Big).
\end{aligned}
$$

Hence $Z_n \xrightarrow{d} N(0,1)$. □

Example 99. Let $X_n \sim i.i.d.$ Poisson(μ). Then $\mathbb{E}(X_n) = \mu$ and $\mathbb{V}\mathrm{ar}(X_n) = \mu$. The Lévy's central limit theorem implies $(X_1 + \cdots + X_n - n\mu)/\sqrt{n\mu} \xrightarrow{d} N(0,1)$.

Generally, if $Y_n \sim$ Poisson(μ_n) with $\mu_n \to \infty$, then $(Y_n - \mu_n)/\sqrt{\mu_n} \xrightarrow{d} N(0,1)$.

For a sequence of independent Poisson random variables $Z_n \sim$ Poisson(μ_n). If $s_n^2 = \mu_1 + \cdots + \mu_n \to \infty$, then $(Z_1 + \cdots + Z_n - s_n^2)/s_n \xrightarrow{d} N(0,1)$.

Example 100. There is an annual marathon in a town. Every year around 1000 people participated in. In history, 40% of them were women. What is the probability of more than 450 women participate in this year when the number of participant is 1000.

Let W be the number of women participant. Then $W \sim$ binomial$(1000, 0.4)$. Since binomial is sum of independent Bernoulli, there exists W_i's such that $W = W_1 + \cdots + W_n$ where $n = 1000$ and $W_i \sim i.i.d.$ Bernoulli(0.4). The central limit theorem implies

$$P(W > 450) = P((W - 400)/\sqrt{1000 \times 0.4 \cdot 0.6} > (450 - 400)/\sqrt{240}) \approx P(N(0,1) > 3.227) = 0.000624$$

The standard normal probabilities are often approximated as follows

$$P(Z > z) \approx \frac{1}{z}\phi(z) = \frac{1}{z}\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$$

where $z > 0$ and $\phi(z) = \frac{1}{z}\frac{1}{\sqrt{2\pi}}\exp(-z^2/2)$. In the above example $\phi(3.227)/3.227 = 0.000677$ which is very close to the real probability 0.000624.

Theorem 88 (δ-method)**.** Suppose $X_1, X_2, \ldots$ is a sequence of random variables and a_n is a sequence of positive real numbers diverging to infinity. If $a_n(X_n - \mu) \xrightarrow{d} Z$ for some random variable Z and a constant μ, then for any continuously differentiable function g, $a_n(g(X_n) - g(\mu)) \xrightarrow{d} g'(\mu)Z$.

Proof. Note that $a_n(X_n - \mu) \xrightarrow{d} Z$ implies $X_n \xrightarrow{p} \mu$. By Taylor expansion, $g(X_n) - g(\mu) = g'(\mu)(X_n - \mu) + o(|X_n - \mu|)$. Hence $a_n(g(X_n) - g(\mu)) = g'(\mu)a_n(X_n - \mu) + o(|a_n(X_n - \mu)|) \xrightarrow{d} g'(\mu)Z$. □

Example 101. Let $X_n \sim i.i.d.$ Exponential(λ). Then $\mathbb{E}[X_n] = 1/\lambda$ and $\mathbb{V}\mathrm{ar}(X_n) = 1/\lambda^2$. Using the strong law of large numbers, $\bar{X}_n = (X_1 + \cdots + X_n)/n \xrightarrow{a.s.} 1/\lambda$. By the central limit theorem, $\sqrt{n}(\bar{X}_n - 1/\lambda)/(1/\lambda^2)^{1/2} = \lambda\sqrt{n}(\bar{X}_n - 1/\lambda) \xrightarrow{d} N(0,1)$. Apply δ-method for $g(x) = 1/x$ to obtain

$$\lambda\sqrt{n}(1/\bar{X}_n - \lambda) \xrightarrow{d} -\lambda^2 N(0,1) \sim N(0, \lambda^4).$$

Finally $\sqrt{n}(1/\bar{X}_n - \lambda) \xrightarrow{d} N(0, \lambda^2)$ by Slutsky's lemma.

Example 102. Let $X_n \sim i.i.d.$ uniform$(0, \theta)$. Then $P(n(\theta - X_{(n)}) > x) = P(X_{(n)} \leq \theta - x/n) = [P(X_1 \leq \theta - x/n)]^n = ((\theta - x/n)/\theta)^n = (1 - x/(n\theta))^n \to \exp(-x/\theta)$. Hence $n(\theta - X_{(n)}) \xrightarrow{d}$ Exponential$(1/\theta)$. Since the limit distribution is a Gaussian distribution, it is called a *non-central limit theorem*.

Exercise 40. Two independent and identically distributed random variables X and Y satisfies that $(X + Y)/\sqrt{2}$ and X have the same distribution. Assume X has variance 1. Show that X has a normal distribution. Find the mean of X. [Hint: central limit theorem.]

Exercise 41. Assume $X_1, X_2, \ldots \sim i.i.d.$ uniform$(-\theta, \theta)$ for some $\theta > 0$. Show that $X_{(n)} = \max(X_1, \ldots, X_n)$ converges to θ almost surely. Prove that $X_{(1)} = \min(X_1, \ldots, X_n)$ converges to $-\theta$ almost surely. Show that $n(X_{(1)} + X_{(n)})$ converges in distribution. Specify the convergent distribution.

Exercises. (Ri) 2.78, 2.81, 2.82, 2.83, 2.87, 3.77, 3.81, 3.82, 3.89, 3.90, 3.94, 3.98, 3.99; (RM) 5.4.1, 5.4.2, 5.4.3, 5.4.4, 5.4.6, 5.4.7, 5.4.11, 5.4.13, 5.4.19, 5.4.21, 5.4.28, 5.4.29.