Review of Probability

January 31, 2002

1 Probability space and random variable

Probability theory has its own vocabulary; this is due partly to the circumstances of its development and partly to the fact that the central problems, motivation, ideas and techniques of Probability are distinctively its own.

Definition: A measure space $(\Omega, \mathcal{F}, \mathsf{P})$ is a **probability space** if $\mathsf{P}(\Omega) = 1$. Ω is the collection of all possible outcomes, and is called the **sample space**. Each member in the σ-algebra $\mathcal{F}$ is an **event**.

Defintion: A measurable function $X : \Omega \to \mathsf{R}$ is called a **random variable**. Clearly then, $f(X)$ is also a random variable if $f : \mathsf{R} \to \mathsf{R}$ is a measurable function.

Definition: The **expectation** of a random variable X is defined as $\mathsf{E}X \doteq \int_\Omega X \, d\mathsf{P}$ if the integral is well-defined. If $X \in \mathsf{L}^2(\mathsf{P})$, then we define the **variance** of X as

$$\mathrm{Var}X \doteq \mathsf{E}\big[X - \mathsf{E}X\big]^2 = \mathsf{E}\big(X^2\big) - \big(\mathsf{E}X\big)^2$$

The square-root of the variance $\sqrt{\mathrm{Var}X}$ is called the **standard deviation**.

Definition: The p-th **absolute moment** of X is defined as $\mathsf{E}\left(|X|^p\right) = \int_\Omega |X|^p \, d\mathsf{P}$.

Example (Bernoulli random variable): A ramdom variable X with

$$\mathsf{P}(X = 1) = p, \quad \mathsf{P}(X = 0) = q; \qquad p + q = 1.$$

We have $\mathsf{E}X = p$ and $\mathrm{Var}X = pq$.

Example (Binomial random variable): A ramdom variable S_n with

$$\mathsf{P}(S_n = k) = \binom{n}{k} p^k q^{n-k}; \qquad k = 0, 1, \cdots, n.$$

We have $\mathsf{E}S_n = np$ and $\mathrm{Var}S_n = npq$.

Example (Geometric random variable): A ramdom variable T with

$$\mathsf{P}(T = k) = q^{k-1}p; \qquad k = 1, 2 \cdots.$$

We have $\mathsf{E}T = \frac{1}{p}$ and $\mathrm{Var}T = \frac{q}{p^2}$.

Example (Poisson random variable): A ramdom variable X with

$$\mathsf{P}(X = k) = e^{-\lambda k}\frac{\lambda^k}{k!}; \qquad k = 0, 1, \cdots. \;\; \lambda > 0$$

We have $\mathsf{E}X = \lambda$ and $\mathrm{Var}X = \lambda$.

Exercise: Show that a Poisson distribution can be approximated by Binomial distributions with "n large and p small. That is, let $p_n = \frac{\lambda}{n}$, show

$$\binom{n}{k} p_n^k (1-p_n)^{n-k} - e^{-\lambda}\frac{\lambda^k}{k!} \longrightarrow 0, \quad \forall\, k \in \mathsf{N}_0.$$

Definition: Suppose X is a random variable. For every set $E \in \mathcal{B}(\mathsf{R})$, define

$$\mu_X(E) \doteq \mathsf{P}\{\omega;\ X(\omega) \in E\} = \mathsf{P}(X \in E).$$

It is not difficult to verify that μ_X define a *probability measure* on space $(\mathsf{R}, \mathcal{B}(\mathsf{R}))$. We call μ_X the **probability measure induced by** X.

Definition: Function $F(x) \doteq \mathsf{P}(X \le x) = \mu_X\{(-\infty, x]\}$ is said to be the **cumulative distribution function** of X. It is non-decreasing, right-continuous, with

$$F(\infty) = 1, \quad F(-\infty) = 0.$$

Any function $F : \mathsf{R} \to [0,1]$ with these properties are called a **probability distribution function**. If there is a non-negative function f such that $F(x) = \int_{-\infty}^{x} f(z)\,dz$, then f is called the **probability density function** of X.

Example (uniform random variable): With $-\infty < a < b < \infty$, $f(x) = \frac{1}{b-a}1_{[a,b]}(x)$ the density of X, we have $\mathsf{E}X = \frac{b+a}{2}$, $\mathrm{Var}X = \frac{1}{2}(b-a)^2$.

Example (exponential random variable): With $\lambda > 0$, $f(x) = \lambda e^{-\lambda x}1_{[0,\infty)}(x)$ the density of X, we have $\mathsf{E}X = \frac{1}{\lambda}$, $\mathrm{Var}X = \frac{1}{\lambda^2}$.

Example (normal random variable): With $\mu \in \mathsf{R}, \sigma > 0$, $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ the density of X, we have $\mathsf{E}X = \mu$, $\mathrm{Var}X = \sigma^2$.

Below is a collection of exercises.

Exercise: For every random variable X and measurable function $h : \mathsf{R} \to \mathsf{R}$, we have

$$\mathsf{E}h(X) = \int_{\mathsf{R}} h(x)\,d\mu_X(x),$$

whenever either side is well-defined.

Proof. Ehe equality holds when $h = 1_E$ for some $E \in \mathcal{B}(\mathsf{R})$, and thus holds for non-negative simple functions. Using MCT, the equality holds for any non-negative measurable function h. Note $h = h^+ - h^-$ in general. We complete proof. $\square$

Exercise: For any non-negative random variable X, we have

$$\mathsf{E}X = \int_0^\infty \mathsf{P}(X \geq x)\, dx = \int_0^\infty \mathsf{P}(X > x)\, dx = \int_0^\infty \big[1 - F(x)\big]\, dx$$

Proof. By definition,

$$\begin{aligned}\mathsf{E}X &= \int_\Omega X(\omega)\, d\mathsf{P}(\omega) = \int_\Omega \int_{\mathsf{R}+} 1_{[0,X(\omega)]}(x)\, dx\, d\mathsf{P}(\omega) = \int_{\mathsf{R}+} \int_\Omega 1_{[0,X(\omega)]}(x)\, dP(\omega)\, dx \\ &= \int_{\mathsf{R}+} \mathsf{P}(X \geq x)\, dx.\end{aligned}$$

We can also replace $1_{[0,X(\omega)]}(x)$ by $1_{\big[0,X(\omega)\big)}(x)$ to obtain the second equality. □

Exercise: For any given probability distribution function $F : \mathsf{R} \to [0,1]$, consider its right- and left-continuous inverse

$$\begin{aligned}X^+(\omega) &\doteq \inf\{x \in \mathsf{R};\ F(x) > \omega\} = \sup\{x \in \mathsf{R};\ F(x) \leq \omega\}\,; \quad \forall\, 0 \leq \omega \leq 1. \\ X^-(\omega) &\doteq \inf\{x \in \mathsf{R};\ F(x) \geq \omega\} = \sup\{x \in \mathsf{R};\ F(x) < \omega\}\,; \quad \forall\, 0 \leq \omega \leq 1.\end{aligned}$$

Then the result mapping $X^\pm : \Omega \to \mathsf{R}$ on the probability space $(\Omega, \mathcal{F}, \mathsf{P}) = ([0,1], \mathcal{B}[0,1], \lambda)$ are random variables with probability distribution functions $F_{X^\pm} = F$.

Proof. It is easily checked that for any $\omega \in [0,1]$, $x \in \mathsf{R}$, we have

$$\omega \leq F(x) \quad \Leftrightarrow \quad X^-(\omega) \leq x.$$

Therefore $\mathsf{P}(X^- \leq x) = \lambda\big(\{\omega;\ \omega \leq F(x)\}\big) = F(x)$. It suffices now to show that $\mathsf{P}(X^+ \neq X^-) = 0$. Since $X^+ \geq X^-$, we have

$$\{X^+ \neq X^-\} = \cup_{q \in \mathsf{Q}}\{X^- \leq q < X^+\};$$

here Q is the set of all rational numbers. But for every $q \in \mathsf{Q}$, we have

$$\mathsf{P}(X^- \leq q < X^+) = \mathsf{P}(X^- \leq q) - \mathsf{P}(X^+ \leq q) \leq F(q) - \mathsf{P}\big(\{\omega; \omega < F(q)\}\big) = 0.$$

This completes the proof. □

2 Independence

Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$.

Definition: A collection of events $\{E_\alpha\}_{\alpha \in I} \subseteq \mathcal{F}$ is said to be **independent** if

$$\mathsf{P}\left(\cap_{j=1}^n E_{\alpha_j}\right) = \prod_{j=1}^n \mathsf{P}\left(E_{\alpha_j}\right) \quad \text{for every } n \in \mathsf{N} \text{ and } \{\alpha_1, \cdots, \alpha_n\} \subset I.$$

Exercise: If $\{E_\alpha\}_{\alpha \in I}$ is a collection of independent events, then so is $\{F_\alpha\}_{\alpha \in I}$ where $F_\alpha = E_\alpha$ or $F_\alpha = E_\alpha^c$.

Broel-Cantelli Lemma: For any collection of events $\{E_n\}_{n\in\mathsf{N}} \subseteq \mathcal{F}$ with $\sum_n \mathsf{P}(E_n) < \infty$,

$$\mathsf{P}(E_n, \text{i.o.}) = 0; \qquad \text{where } \{E_n, \text{i.o.}\} = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} E_m.$$

Conversely, if $\{E_n\}_{n\in\mathsf{N}}$ are independent events and $\sum_n \mathsf{P}(E_n) = \infty$, we have

$$\mathsf{P}(E_n, \text{i.o.}) = 1.$$

Proof. Note that $\mathsf{P}(\cup_{m=n}^{\infty} E_m) \leq \sum_{m=n}^{\infty} \mathsf{P}(E_m) \to 0$ as $n \to \infty$, we have $\mathsf{P}(E_n, \text{i.o.}) = 0$. Conversely, note that

$$\mathsf{P}\left(\{\cup_{m=n}^{\infty} E_m\}^c\right) = \mathsf{P}(\cap_{m=n}^{\infty} E_m^c) \leq \mathsf{P}(\cap_{m=n}^{N} E_m^c) = \prod_{m=n}^{N} \mathsf{P}(E_m^c)$$

by independence, for all N. However, for every $x \geq 0$, $1 - x \leq e^{-x}$, it follows that

$$\mathsf{P}\left(\{\cup_{m=n}^{\infty} E_m\}^c\right) \leq \prod_{m=n}^{N} e^{-\mathsf{P}(E_m)} = e^{-\sum_{m=n}^{N} \mathsf{P}(E_m)} \quad \longrightarrow \quad 0, \quad \text{as } N \to \infty.$$

This completes proof. □

Definition: A family of random variables $\{X_\alpha\}_{\alpha\in I}$ are **independent** if the corresponding events $E_\alpha = \{X_\alpha \in B_\alpha\}$, $\alpha \in I$ are independent for all $B_\alpha \in \mathcal{B}(\mathsf{R})$.

Remark: If $\{X_\alpha\}$ are independent random variables, so are $\{f_\alpha(X_\alpha)\}_{\alpha\in I}$; here $f_\alpha : \mathsf{R} \to \mathsf{R}$ is measurable, $\forall \alpha \in I$.

Remark: Suppose $X_1, X_2, \cdots, X_n$ are independent random variables. Show that

$$\mu_{X_1,\cdots,X_n} = \otimes_{j=1}^{n} \mu_{X_j};$$

Here μ_{X_j} is the probability measure induced on R by X_j, and $\mu_{X_1,\cdots,X_n}$ is the probability measure induced on R^n by $(X_1, \cdots, X_n)$. Indeed, these two measures agree on rectangles, thanks to independence. Therefore, they are the same measure (why?)

Proposition: Suppose $X_1, X_2, \cdots, X_n$ are independent random variables.

1. If $X_i \in \mathsf{L}^1(\mathsf{P})$, then $\prod_{j=1}^{n} X_j \in \mathsf{L}^1$ and $\mathsf{E}\left(\prod_{j=1}^{n} X_j\right) = \prod_{j=1}^{n} \mathsf{E} X_j$.
2. If $X_i \in \mathsf{L}^2(\mathsf{P})$, then $\mathrm{Var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} \mathrm{Var} X_j$.

Proof. It follows from Tonelli theorem that

$$\mathsf{E}\left(\prod_{j=1}^{n} |X_j|\right) = \int_{\mathsf{R}^n} \prod_{j=1}^{n} |x_j| \, d\mu_{X_1,\cdots,X_n} = \int_{\mathsf{R}^n} \prod_{j=1}^{n} |x_j| \, d\left(\otimes \mu_{X_j}\right) = \prod_{j=1}^{n} \int_{\mathsf{R}} |x_j| \, d\mu_{X_j} < \infty.$$

Then 1 follows from Fubini. Claim 2 is a direct consequence of 1. □

Exercise: **Weistrass Approximation Theorem**. For any continuous function $f : [0,1] \to \mathsf{R}$, there exists a sequence of polynomials

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}, \qquad 0 \le x \le 1,\ n \in \mathsf{N},$$

such that $B_n \to f$ uniformly as $n \to \infty$.

Proof. Let $\{X_n\}_{n\in\mathsf{N}}$ be a sequence of iid (independent, identically distributed) Bernoulli random variables with

$$\mathsf{P}(X_j = 1) = p = 1 - \mathsf{P}(X_j = 0).$$

It follows that

$$B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathsf{P}(S_n = k) = \mathsf{E} f\left(\frac{S_n}{n}\right).$$

It follows that

$$\begin{aligned} |B_n(p) - f(p)| &= \left| \mathsf{E} f\left(\frac{S_n}{n}\right) - f(p) \right| \le \mathsf{E} \left| f\left(\frac{S_n}{n}\right) - f(p) \right| \\ &\le \mathsf{E}\left(\left| f\left(\frac{S_n}{n}\right) - f(p) \right| 1_{\{|\frac{S_n}{n} - p| \le \delta\}} \right) + \mathsf{E}\left(\left| f\left(\frac{S_n}{n}\right) - f(p) \right| 1_{\{|\frac{S_n}{n} - p| > \delta\}} \right) \end{aligned}$$

for every $\delta > 0$. However, f is continuous on $[0,1]$, it must be bounded, say by K, and uniformly continuous on $[0,1]$. For an arbitrary $\epsilon > 0$, choose $\delta > 0$ such that

$$|f(x) - f(y)| \le \epsilon \qquad \text{whenever} \quad |x - y| \le \delta, \ \ \forall\, x, y \in [0,1].$$

It follows that, for every $p \in [0,1]$,

$$|B_n(p) - f(p)| \le \epsilon + 2K\mathsf{P}\left(\left| \frac{S_n}{n} - p \right| \ge \delta \right) \le \epsilon + 2K \frac{\mathsf{E}\left(\frac{S_n}{n} - p\right)^2}{\delta^2} = \epsilon + 2K \frac{pq}{n\delta^2} \le \frac{K}{2\delta^2} \cdot \frac{1}{n} + \epsilon \le 2\epsilon$$

for sufficiently large n. This completes the proof. □

3 Uniform integrability

A family of random varaibles $\{X_\alpha\}_{\alpha\in I}$ is said to be **uniformly integrable** if

$$\sup_\alpha \int_\Omega |X_\alpha| \cdot 1_{\{|X_\alpha| \ge \lambda\}}\, d\mathsf{P} \quad \longrightarrow \quad 0, \quad \text{as } \ \lambda \to \infty.$$

Proposition: A family of random varaibles $\{X_\alpha\}_{\alpha\in I}$ is uniformly integrable if and only if

1. $\sup_\alpha \mathsf{E}|X_\alpha| < \infty$
2. for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\sup_\alpha \int_A |X_\alpha|\, d\mathsf{P} < \epsilon, \qquad \forall\ A \in \mathcal{F},\ \mathsf{P}(A) < \delta.$$

Proof. "$\Leftarrow$": Let $K = \sup_\alpha \mathsf{E}|X_\alpha|$. For every $\epsilon > 0$, we have

$$\mathsf{P}(|X_\alpha| \geq \lambda) \leq \frac{K}{\lambda} < \delta, \quad \forall\, \alpha \in I,$$

for any $\lambda > \frac{K}{\delta}$. It follows that

$$\sup_\alpha \int_\Omega |X_\alpha| \cdot 1_{\{|X_\alpha| \geq \lambda\}} \, d\mathsf{P} < \epsilon$$

by assumption.

"$\Rightarrow$": For any $\epsilon > 0$, choose λ such that

$$\sup_\alpha \int_\Omega |X_\alpha| \cdot 1_{\{|X_\alpha| \geq \lambda\}} \, d\mathsf{P} \leq \epsilon,$$

which implies that

$$\mathsf{E}|X_\alpha| \leq \lambda + \int_\Omega |X_\alpha| \cdot 1_{\{|X_\alpha| \geq \lambda\}} \, d\mathsf{P} \leq \lambda + \epsilon, \quad \forall\, \alpha \in I.$$

Moreover, let $\delta = \frac{\epsilon}{\lambda}$, then for any $A \in \mathcal{F}$, $\mathsf{P}(A) < \delta$, we have

$$\int_A |X_\alpha| \, d\mathsf{P} = \int_{A \cap \{|X_\alpha| \geq \lambda\}} |X_\alpha| \, d\mathsf{P} + \int_{A \cap \{|X_\alpha| < \lambda\}} |X_\alpha| \, d\mathsf{P} \leq \int_{\{|X_\alpha| \geq \lambda\}} |X_\alpha| \, d\mathsf{P} + \lambda \mathsf{P}(A) \leq \epsilon + \epsilon = 2\epsilon.$$

This completes the proof. □

Exercise: Show that each of the following conditions is a sufficient condition for uniformly integrability.

1. $|X_\alpha| \leq X$ for some $X \in \mathsf{L}^1$.
2. $\sup_\alpha \mathsf{E}h(|X_\alpha|) < \infty$ for some function $h : [0, \infty) \to [0, \infty)$ such that

$$\frac{h(x)}{x} \text{ is non-decreasing, and } \lim_{x \to \infty} \frac{h(x)}{x} = \infty.$$

In particular, if $\sup_\alpha \mathsf{E}|X_\alpha|^p < \infty$ for some $p > 1$ then $\{X_\alpha\}$ are uniformly integrable.

Absoulte continuity of integral: Suppose $X \in \mathsf{L}^1$. Then for an arbitrary $\epsilon > 0$, there exists a $\delta > 0$ such that $\int_A |X| \, d\mathsf{P} < \epsilon$ whenever $\mathsf{P}(A) < \delta$. This is a direct consequence by taking $X_\alpha \equiv X$.

Proposition: Suppose $\{X_n\}_{n \in \mathsf{N}}$ is a sequence of integrable random variables such that $X_n \xrightarrow{P} X$. Then the following conditions are equivalent.

1. $\{X_n\}_{n \in \mathsf{N}}$ is uniformly integrable.
2. $X_n \to X$ in L^1.
3. $\mathsf{E}|X_n| \to \mathsf{E}|X|$ and $\mathsf{E}|X| < \infty$.

Remark: Suppose $X_n \to X$ in L^1. It follows that $\{X_n\}_{n \in \mathsf{N}}$ is uniformly inegrable.

Proof. "$1 \Rightarrow 2$": There exists a subsequence, say X_n' such that $X_n' \to X$ almost surely. It follows from Fatou Lemma $\mathsf{E}|X| \le \liminf_n \mathsf{E}|X_n'| \le \sup_n |X_n'| < \infty$, or $X \in \mathsf{L}^1$. Therefore,

$$\mathsf{E}|X_n - X| = \int_{\{|X_n - X| \le \epsilon\}} |X_n - X|\, d\mathsf{P} + \int_{\{|X_n - X| \ge \epsilon\}} |X_n - X|\, d\mathsf{P} \le \epsilon + \int_{\{|X_n - X| \ge \epsilon\}} |X_n - X|\, d\mathsf{P}, \quad \forall\, \epsilon > 0.$$

However, by assumption $\mathsf{P}(|X_n - X| \ge \epsilon) \to 0$, we have

$$\int_{\{|X_n - X| \ge \epsilon\}} |X_n - X|\, d\mathsf{P} \quad \to \quad 0,$$

since that the sequence $\{X_n - X\}_{n \in \mathsf{N}}$ is also uniformly integrable (why?). For n large enough, we should have

$$\mathsf{E}|X_n - X| \le \epsilon + \epsilon = 2\epsilon.$$

"$2 \Rightarrow 3$": This claim is trivial, thanks to the inequality $|x| + |y| \ge |x + y|$.

"$3 \Rightarrow 1$": Since $|X_n| \xrightarrow{P} |X|$ (why?), we can assume X_n and X are all non-negative. We first show that $X_n \to X$ in L^1. Indeed, $(X - X_n)^+ \le X$, and $(X - X_n)^+ \xrightarrow{P} 0$, it follows from DCT that $\lim_n \mathsf{E}(X - X_n)^+ = 0$, which implies

$$\mathsf{E}(X - X_n)^- = \mathsf{E}(X - X_n)^+ - \mathsf{E}(X - X_n) \quad \to \quad 0,$$

which in turn implies

$$\lim_n \mathsf{E}|X_n - X| = 0.$$

In order to show the uniform integrability of $\{X_n\}_{n \in \mathsf{N}}$, note that $\sup_n \mathsf{E}|X_n| < \infty$ is trivial. Moreover,

$$\int_A |X_n|\, d\mathsf{P} \le \int_A |X|\, d\mathsf{P} + \int_A |X - X_n|\, d\mathsf{P} \le \int_A |X|\, d\mathsf{P} + \mathsf{E}|X - X_n|, \quad \forall\, A \in \mathcal{F}.$$

The uniform integrability follows readily, thanks to the absoulte continuity of integral, and the fact that $\lim_n \mathsf{E}|X_n - X| = 0$. □

4 Law of large Numbers and Central Limit Theorem

The strong law of large numbers (SLLN) and central limit theorem (CLT) are two most important results in classical probability theory.

Strong Law of Large Numbers: Suppose $\{X_n\}_{n \in \mathsf{N}}$ is a sequence of iid random variables and let $S_n \doteq \sum_{j=1}^n X_j$. Then

$$\frac{S_n}{n} \quad \longrightarrow \quad \mathsf{E}X_1 \quad \text{alomst surely}, \qquad \text{as} \;\; n \to \infty,$$

whenever $\mathsf{E}X_1$ is well-defined.

We will later present a proof of SLLN using martingale convergence theory.

Central Limit Theorem (Classical): Suppose that $\{X_n\}_{n\in\mathsf{N}}$ is a sequence of iid random variables with $\mu = \mathsf{E}X_j$ and $\sigma^2 = \mathrm{Var}X_j \in (0,\infty)$. Let $S_n \doteq \sum_{j=1}^n X_j$. Then for every pair of real numbers x,

$$\mathsf{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right] \quad\longrightarrow\quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz.$$

The proof of CLT can be achieved with the help of characteristic functions. We also have the following extension, which deals with independent but not necessarily identically distributed random variables.

Central Limit Theorem (Lindeberg): Suppose that $\{X_n\}_{n\in\mathsf{N}}$ is a sequence of independent random variables with $\mu_j \doteq \mathsf{E}X_j$ and $\sigma_j^2 = \mathrm{Var}X_j \in (0,\infty)$. Let $S_n \doteq \sum_{j=1}^n X_j$ and $B_n \doteq \sum_{j=1}^n \sigma_j^2$. If for any $\epsilon > 0$,

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{j=1}^{n} \mathsf{E}\left[(X_j - \mu_j)^2 \cdot 1_{\{|X_j - \mu_j| \ge \epsilon B_n\}}\right] = 0,$$

then for every pair of real numbers x,

$$\mathsf{P}\left[\frac{S_n - \sum_{j=1}^n \mu_j}{B_n} \le x\right] \quad\longrightarrow\quad \Phi(x).$$

Exercise: Show that the Lindeberg condition is satisfied if either of the following conditions holds.

1. $\{X_n\}_{n\in\mathsf{N}}$ is a sequence of iid random variables.
2. $|X_n| \le K$ for all $n \in \mathsf{N}$; that is, $|X_n|$ is uniformly bounded, and $B_n^2 \to \infty$.
3. (Lyapounov condition) for some $\delta > 0$,

$$\lim_{n\to\infty} \frac{1}{B_n^{2+\delta}} \sum_{j=1}^{n} \mathsf{E}|X_j - \mu_j|^{2+\delta} = 0.$$

The rate of convergence of CLT is obtained in the following theorem.

Theorem (Berry & Esseen): Let $\{X_n\}_{n\in\mathsf{N}}$ is a sequence of iid random variables with $\mu = \mathsf{E}X_j$, $\sigma^2 = \mathrm{Var}X_j$ and $\mathsf{E}|X_j|^3 < \infty$. Then

$$\sup_{x\in\mathsf{R}} \left| \mathsf{P}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le x\right) - \Phi(x) \right| \le \frac{C \cdot \mathsf{E}|X_1|^3}{\sigma^3\sqrt{n}}$$

for some universal constant C.

Remark: It can be shown that $\frac{1}{\sqrt{2\pi}} \le C \le 0.8$.