MATH 2431
All-in-one Summary

CHEN, Yixiang

Based on MATH2431 lectured by BAO, Zhigang in 23-24Spring

1 Events and their Probabilities

- **Sample Space**: Given an experiment, the sample space Ω is the set of all outcomes ω.
- **Event**: A subset of Ω.
- **Field**: Any collection $\mathcal{F}$ of subsets of Ω is called a field if it satisfies the following properties:
  - $\emptyset \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
  - If $A \in \mathcal{F}$, then $A^{\complement} \in \mathcal{F}$, i.e. $\mathcal{F}$ is closed under complement.
  - If $A, B \in \mathcal{F}$, then $A \cup B \in \mathcal{F}$, i.e. $\mathcal{F}$ is closed under finite union.
- **σ-Field**: Any collection $\mathcal{F}$ of events in Ω is called a σ-field if it satisfies the following properties:
  - $\emptyset \subset \mathcal{F}$ and $\Omega \subset \mathcal{F}$.
  - If $A \subset \mathcal{F}$, then $A^{\complement} \subset \mathcal{F}$.
  - If $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$, i.e. $\mathcal{F}$ is closed under countable union.
- **Probability Measure**: Given a measurable space $(\Omega, \mathcal{F})$, a probability on $(\Omega, \mathcal{F})$ is a function $\mathbb{P} : \mathcal{F} \to [0, 1]$, satisfying:
  - $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.
  - If $A_1, A_2, \ldots \in \mathcal{F}$ are disjoint, then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.
- **Inclusion-Exclusion Formula**: If $A, B \in \mathcal{F}$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- **General Measure**: Given $(\Omega, \mathcal{F})$, a measure is a set function $\mu : \mathcal{F} \to [0, \infty]$, such that
  - $\mu(\emptyset) = 0$.
  - it satisfies countable additivity.
- **Set Limit**: Given $A_1, A_2, \ldots \in \mathcal{F}$, $\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for infinitely many } n\}$ and $\liminf_{n\to\infty} = \bigcup_{n=1}^{\infty} \cap_{n=m}^{\infty} A_n = \{\omega \in \Omega : \omega \in A_n \text{ for all but finitely many } n\}$.

- **Set Convergence**: (A_n) converges if $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$.

- **Continuity of Probability Measure**: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and (A_n) in $\mathcal{F}$ such that $\lim_{n\to\infty} A_n = A$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A)$.

- **Conditional Probability**: If $\mathbb{P}(B) > 0$, then the conditional probability that A occurs given that B has occurred is $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}$.

- **Law of Total Probability**: Let $B_1, \ldots, B_n$ be a partition of Ω. Suppose that $\mathbb{P}(B_i) > 0$ for all i. Then, $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$.

- **Independence**: Let $A, B \in \mathcal{F}$. We say A is independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Let $\bigcup_{i\in I}\{A_i\} \subset \mathcal{F}$, where I is not necessarily countable. $\bigcup_{i\in I}\{A_i\}$ is called mutually independent if $\mathbb{P}\left(\bigcap_{j\in J} A_j\right) = \prod_{j\in J} \mathbb{P}\left(A_j\right)$, for every $J \subset I$ and pairwise independent if $\mathbb{P}\left(A_{i_1} \cap A_{i_2}\right) = \mathbb{P}\left(A_{i_1}\right)\mathbb{P}\left(A_{i_2}\right)$, for every $\{i_1, i_2\} \subset I$.

- $A \amalg B \iff A \amalg B^{\complement}$.

- If $\{A, B, C\}$ is independent, then $A \amalg (B \cup C)$ and $A \amalg (B \cap C)$.

- **Lemma**: If $(\mathcal{F}_i)_{i\in I}$ is a system of σ-algebra on Ω, then $\bigcap_{i\in I} \mathcal{F}_i$ is a σ-field.

- **Generated σ-Field**: Let $A \subset 2^{\Omega}$. The σ-field generated by A is $\sigma(A) = \bigcap_{\mathcal{G}\in X} \mathcal{G}$, where $X = \{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-field such that } A \subset \mathcal{G}\}$.

- **Product Space**: Given $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$, $(\Omega, \mathcal{F}, \mathbb{P})$ is a product space if
  - $\Omega = \Omega_1 \times \Omega_2$.
  - $\mathcal{F} = \sigma\left(\mathcal{F}_1 \times \mathcal{F}_2\right)$.
  - $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$.

2 Random Variables and their Distributions

- **Random Variable**: A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$, for all intervals $B \subset \mathbb{R}$. In this case, X is $\mathcal{F}$-measurable.

- **Borel Set**: A set is a Borel Set if it can be formed from open sets through the operations of countable union, countable intersection and relative complement.

- **Borel σ-Field of $\mathbb{R}$**: Any σ-field that is a collection of Borel sets ia called a Borel σ-field, i.e. σ-field $\mathcal{B}(\mathbb{R})$ generated by all open sets in $\mathbb{R}$.

- $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{X} (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{Q})$, where $\mathbb{Q} : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is indeed a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\mathbb{Q} = \mathbb{P} \circ X^{-1}$.

- **Distribution Function**: $F_X(x) = \mathbb{P} \circ X^{-1}\left((-\infty, x]\right)$.

- **Lemma**: A cumulative distribution function has the following properties:
  - $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$.
  - If $x < y$, then $F(x) \le F(y)$.
  - F is right-continuous.
- **Discrete Random Variable**: X is discrete if it takes values in some countable subset $\{x_1, x_2, \dots\}$ of $\mathbb{R}$.
- **Probability Mass Function**: $f : \mathbb{R} \longrightarrow [0,1]$, where $f(x) = \mathbb{P} \circ X^{-1}(\{x\})$ and the cumulative distribution function is given by $F(x) = \sum_{u\le x} f(u)$. Hence, $f(x) = F(x) - F(x^-)$.
- **Continuous Random Variable**: X is continuous if its distribution function $F(x)$ can be written as $F(x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$, for every $x \in \mathbb{R}$, for some integrable function $f : \mathbb{R} \longrightarrow [0,1]$.
- **Random Vector in $\mathbb{R}^2$**: $\vec{X} = (X_1, X_2) : \Omega \longrightarrow \mathbb{R}^2$ is a random vector if $\vec{X}^{-1}(D) = \{\omega : \Omega : \vec{X}(\omega) = (X_1(\omega), X_2(\omega) \in D\} \in \mathcal{F}$, for every $D \in \mathcal{B}(\mathbb{R}^2)$, i.e. $(\Omega, \mathcal{F}) \xrightarrow{\vec{X}} (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. Equivalently, $\vec{X} = (X_1, X_2) : \Omega \longrightarrow \mathbb{R}^2$ is a random vector if both $X_1, X_2 : \Omega \longrightarrow \mathbb{R}$ are random variables. $(\Omega, \mathcal{F}, \mathbb{P}) \xrightarrow{\vec{X}=(X_1,X_2)} \left(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \mathbb{P}\circ\vec{X}^{-1}\right)$.
- **Jointly Distribution Function**: $F_{X_1,X_2}(x_1,x_2) = \mathbb{P}\circ\vec{X}^{-1}\left((-\infty,x_1]\times(-\infty,x_2]\right)$.
- **Lemma**: The jointly distribution funcion has the following properties:
  - $\lim_{x,y\to-\infty} F_{X,Y}(x,y) = 0$ and $\lim_{x,y\to+\infty} F_{X,Y}(x,y) = 1$.
  - If $x_1 \le x_2$ and $y_1 \le y_2$, then $F_{X,Y}(x_1,y_1) \le F_{X,Y}(x_2,y_2)$.
  - Continuity from above.
- **Marginal Distribution**: $F_X(x) = \mathbb{P}\circ X^{-1}\left((-\infty,x]\right) = \lim_{y\to+\infty} F_{X,Y}(x,y)$.
- **Jointly Discrete Random Variables**: $X, Y : \Omega \longrightarrow \mathbb{R}$ are jointly discrete if (X,Y) takes values in some countable subset of $\mathbb{R}^2$ only. The jointly probability mass function $f : \mathbb{R}^2 \longrightarrow [0,1]$ is given by $f_{X,Y}(x,y) = \mathbb{P}\circ(X,Y)^{-1}(\{x,y\})$. And hence $F_{X,Y}(x,y) = \sum_{u\le x, v\le y} f_{X,Y}(u,v)$.
- **Jointly Continuous Random Variables**: $X, Y : \Omega \longrightarrow \mathbb{R}$ are jointly continuous if $F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\mathrm{d}v\mathrm{d}u$, for every $x, y \in \mathbb{R}$, for some integrable function $f : \mathbb{R}^2 \longrightarrow [0,+\infty]$ called a jointly probability density function, which is not a probability measure. More generally, $\mathbb{P}\circ(X,Y)^{-1}(B) = \iint_B f_{X,Y}(u,v)\mathrm{d}v\mathrm{d}u$.

3 Discrete Random Variable

Assume X, Y are discrete random variables.

- **Independence of Randon Variables**: $X, Y : \Omega \to \mathbb{R}$ are independent if $\mathbb{P}(X \in E, Y \in F) = \mathbb{P}(X \in E)\,\mathbb{P}(Y \in F)$, for every $E, F \in \mathcal{B}(\mathbb{R})$. Equivalently, $X \amalg Y$ if $F_{X,Y}(x,y) = F_X(x)F_Y(y)$, for every $x, y \in \mathbb{R}$. Or equivalently, $X \amalg Y$ if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, for every $x, y \in \mathbb{R}$. Let $X_1, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables. They are mutually independent if any of the following is true:
  - $\mathbb{P} \circ (X_1, \ldots, X_n)^{-1}(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mathbb{P} \circ X_i^{-1}(A_i)$, for every $A_i \in \mathcal{B}(\mathbb{R})$.
  - $F_{X_1,\ldots,X_n} = \prod_i F_{X_i}$.
  - $f_{X_1,\ldots,X_n} = \prod_i f_{X_i}$,

  The independence of events is a special case of the independence of random variables, i.e. $A \amalg B \iff \mathbf{1}_A \amalg \mathbf{1}_B$. Also, $X \amalg Y \iff X^{-1}(E) \amalg Y^{-1}(F)$, for every $E, F \in \mathcal{B}(\mathbb{R})$.
- **σ-Field generated by Random Variable**: $\sigma(X) = \{X^{-1}(E) : E \in \mathcal{B}(\mathbb{R})\} \subset \mathcal{F}$.
- **Independence of two σ-Fields**: Let $\mathcal{H}, \mathcal{G} \subset \mathcal{F}$ be two σ-fields. $\mathcal{H} \amalg \mathcal{G}$ if $A \amalg B$ for every $A \in \mathcal{H}$ and every $B \in \mathcal{G}$.
- **Theorem**: If $X \amalg Y$ and $g, h : \mathbb{R} \to \mathbb{R}$ such that $g(X)$ and $h(Y)$ are still random variables, then $g(X) \amalg h(Y)$.
- **Expectation**: The expectation $\mathbb{E}X$ is defined as $\mathbb{E}X = \sum_i x_i f_X(x_i) = \sum_{x : f_X(x) > 0} x f_X(x)$, if the above sum is absolutely convergent.
- **Lemma**: If $g : \mathbb{R} \to \mathbb{R}$ such that $g(X)$ is still a random variable, then $\sum y f_Y(y) = \mathbb{E}g(X) = \sum_x g(x) f_X(x)$. If $X, Y : \Omega \to \mathbb{R}$ are two jointly discrete random variables and $g : \mathbb{R}^2 \to \mathbb{R}$ such that $g(X, Y)$ is still a random variable, then $\mathbb{E}g(X,Y) = \sum_{x,y} g(x,y) f_{X,Y}(x,y)$.
- **Moments and Central Moments**: $\mathbb{E}X^k$ and $\mathbb{E}(X - \mathbb{E}X)^k$.
- **Properties of $\mathbb{E}(\cdot)$**:
  - If $x \geq 0$, then $\mathbb{E}X \geq 0$.
  - If $a, b \in \mathbb{R}$, $\mathbb{E}(aX + bY) = a\mathbb{E}X + b\mathbb{E}Y$.
- **Uncorrelated Relation**: X and Y are uncorrelated if $\mathbb{E}XY = \mathbb{E}X\mathbb{E}Y$.
- **Properties of Variance**:
  - $\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$.
  - $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, if X, Y are uncorrelated.
- **Covariance**: $\mathrm{Cov}(X,Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$. In general, $\mathrm{Var}(X+Y) = \mathrm{Var}\,X + \mathrm{Var}\,Y + 2\mathrm{Cov}(X,Y)$. More generally, $\mathrm{Var}\left(\sum_i X_i\right) = \sum_i \mathrm{Var}\,X_i + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$.

- **Conditional Distribution**: The conditional distribution of Y given $X = x$ is given by $\mathbb{P}(Y \in \cdot \mid X = x) : \mathcal{B}(\mathbb{R}) \to [0,1]$. The conditional distribution function of Y given $X = x$ is given by $F_{Y|X}(\cdot \mid x) = \mathbb{P}(Y \leq \cdot \mid X = x) : \mathbb{R} \to [0,1]$. The conditional mass function is given by $f_{Y|X}(\cdot \mid x) = \mathbb{P}(Y = \cdot \mid X = x)$. We define the above only when $\mathbb{P}(X = x) > 0$.

- **Conditional Expectation given an Event**: The conditional expectation of Y given $X = x$ is given by $\mathbb{E}(Y \mid X = x) = \sum_y y f_{Y|X}(y \mid x) = \psi(x)$, which is a function of x.

- **Conditional Expectation given a Random Variable**: The conditional expectation of Y given X is given by $\mathbb{E}(Y \mid X) = \psi(X)$, which is a random variable and is a function of X.

- **Law of Total Expectation**: Let $\psi(X) = \mathbb{E}(Y \mid X)$. Then, $\mathbb{E}\psi(X) = \sum_x \psi(x) f_X(x) = \sum_x \mathbb{E}(Y \mid X = x)\, \mathbb{P}(X = x) = \mathbb{E}Y$.

- **Theorem**: Let $\psi(X) = \mathbb{E}(Y \mid X)$. Then, $\mathbb{E}(\psi(X) g(X)) = \mathbb{E}(Y g(X))$.

- **Sum of Random Variable**: Assume $X \amalg Y$ and X, Y have joint probability mass function $f_{X,Y}(x,y)$. Then, $f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x) = (f_X * f_Y)(z)$.

4 Continuous Random Variables

Assume that X, Y are two continuous random variables.

- **Independence**: X, Y are independent if any of the following is true:
  - $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\, \mathbb{P}(Y \in B)$, for every $A, B \in \mathcal{B}(\mathbb{R})$.
  - $F_{X,Y}(x,y) = F_X(x) F_Y(y)$, for every $x, y \in \mathbb{R}$.
  - $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, for every $x, y \in \mathbb{R}$.

- **Expectation**: $\mathbb{E}X = \int x f_X(x) \mathrm{d}x$, only when $\int |x|\, f_X(x) \mathrm{d}x$ converges.

- **Tail Sum Formula**: If $X \geq 0$ has distribution function $F_X(x)$, then $\mathbb{E}X = \int_0^{+\infty} (1 - F_X(x))\, \mathrm{d}x$.

- **Inverse Transform Sampling**: $Y = G^{-1}(U)$ has the distribution function $G(x)$.

- **Conditional Distribution**: The conditional distribution of Y given $X = x$ is given by $\mathbb{P}(Y \leq y \mid X = x) = \int_{-\infty}^{y} \frac{f_{X,Y}(x,v)}{f_X(x)} \mathrm{d}v$. More generally, $\mathbb{P}(Y \in A \mid X = x) = \int_A \frac{f_{X,Y}(x,v)}{f_X(x)} \mathrm{d}v$. Hence, the conditional probability density function is given by $f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$. They are defined only when $f_X(x) > 0$.

- **Conditional Expectation**:
  - Given an event $X = x$, $\mathbb{E}(Y \mid X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y \mid x) \mathrm{d}y = \psi(x)$.
  - Given a random variable X, $\mathbb{E}(Y \mid X) = \psi(X)$.

- **Law of Total Expectation**: $\mathbb{E}Y = \mathbb{E}(\psi(x))$ and $\mathbb{E}(Y(g(x))) = \mathbb{E}(\psi(x) g(x))$.

- **Distribution of $g(X)$**: Let $Y = g(X)$ be a continuous random variable. $F_Y(y) = \int_{g^{-1}((-\infty,y])} f_X(x)\mathrm{d}x$. Hence, $f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\int_{g^{-1}((-\infty,y])} f_X(x)\mathrm{d}x$. If $g(x)$ is strictly monotone and differentiable, $f_Y(y) = \begin{cases} f_X\left(g^{-1}(y)\right)\cdot\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|, & \text{if } y = g(x) \text{ for some } x \\ 0, & \text{otherwise.}\end{cases}$ Consider multivariable cases. Suppose (X, Y) are jointly continuous with $f_{X,Y}$. $U = g(X, Y)$, $V = g(X, Y)$ are jointly continuous. Assume that X, Y can be uniquely solved from U, V, i.e. there exist unique functions a, b such that $X = a(U, V)$, $Y = b(U, V)$, and that the function g, h are differentiable and the Jacobian determinant does not equal to 0. Then,

$$\begin{aligned} f_{U,V}(u,v) &= f_{X,Y}(x,y)\frac{1}{|\det(\mathbf{J}(x,y))|} \\ &= \begin{cases} f_{X,Y}\left(a(u,v), b(u,v)\right)\frac{1}{|\det(\mathbf{J}(x,y))|}, & \text{if } (u,v) = (g(x,y), h(x,y)) \text{ for some } x, y \\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

  where

$$\mathbf{J}(x,y) = \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{pmatrix}.$$

- **Sum of Continuous Random Variables**: $F_{X+Y}(z) = \int\left(\int_{-\infty}^{z} f_{X,Y}(v-y, y)\mathrm{d}v\right)\mathrm{d}y$. Hence, $f_{X+Y}(z) = \int f_{X,Y}(z-y, y)\mathrm{d}y = \int f_{X,Y}(x, z-x)\mathrm{d}x$. If $X \amalg Y$, then $f_{X+Y}(z) = (f_X * f_Y)(z)$.

5 Generating Function

- **Generating Function**: For any sequence $\{a_n\}$, the generating function is given by

$$G_a(s) = \sum_{i=0}^{\infty} a_i s^i,$$

  if it converges.

- **Probability Generating Function**: The probability generating function of a random variable X with nonnegative value is given by

$$G_X(s) = \mathbb{E}s^X = \sum_{i=0}^{\infty} f_X(i)s^i.$$

  –

$$f_X(i) = \frac{\mathrm{d}^i}{i!\mathrm{d}s^i}G_X(s)\bigg|_{s=0}.$$

  – Generally,

$$\mathbb{E}s^X = \int s^x \mathrm{d}F_X(x).$$

- **Moment and Probability Generating Function**: If X has probability generating function G_X, then
  - $\mathbb{E}X = \lim_{s\uparrow 1} G'(s) := G'(1)$,
  - $E\left(X(X-1)\cdots(X-k+1)\right) = \lim_{s\uparrow 1} G^{(k)}(s) := G^{(k)}(1)$.
- **Sum of Independent Random Variables**: If X and Y are independent,

$$G_{X+Y} = G_X \cdot G_Y.$$

- **Sum of Random Number of Random Variables**: Assume that $\{X_i\}$ is i.i.d. random variables and $N \geq 0$ is a random variable independent of all X_i's. Let $T = \sum_{i=1}^{N} X_i$. Then,

$$G_T = G_N \circ G_X.$$

  - The sum of a Poisson number of independent Bernoulli random variables is still Poisson.
- **Joint Probability Generating Function**: Suppose X_1, X_2 are nonnegative, integer-valued and jointly discrete random variables. Then, the jointly probability generating function is given by

$$G_{X_1,X_2}(s_1,s_2) = \mathbb{E}s_1^{X_1}s_2^{X_2} = \sum_i \sum_j s_1^i s_2^j f_{X_1,X_2}(i,j).$$

  -
$$f_{X_1,X_2}(i,j) = \frac{\partial^{i+j}}{i!j!\partial s_1^i \partial s_2^j} G_{X_1,X_2}(s_1,s_2)\bigg|_{(s_1,s_2)=(0,0)}.$$

  - If X_1 and X_2 are independent, then

$$G_{X_1,X_2}(s_1,s_2) = G_{X_1}(s_1)G_{X_2}(s_2).$$

- **Moment Generating Function**:

$$M_X(t) = \mathbb{E}e^{tX} = \int e^{tx}\mathrm{d}F_X(x).$$

  -
$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\int e^{tx}\mathrm{d}F_X(x)\bigg|_{t=0} = \int x^k e^{tx}\mathrm{d}F_X(x)\bigg|_{t=0} = \mathbb{E}X^k.$$

- **Characteristic Function**:

$$\phi_X(t) = \mathbb{E}e^{itX} = \int e^{itx}\mathrm{d}F_X(x).$$

  -
$$\phi^{(k)}(0) = i^k \int x^k \mathrm{d}F_X(x).$$

- Taylor Expansion:
$$\phi_X(s) = \sum_{j=0}^{k} \frac{\mathbb{E}X^j}{j!}(is)^j + o\left(s^k\right).$$
- If $\phi^{(k)}(0)$ exists, then
  * $\mathbb{E}|X|^k < \infty$, if k is even,
  * $\mathbb{E}|X|^{k-1} < \infty$, if k is odd.
- If $\mathbb{E}|X|^k < \infty$, then $\phi^{(k)}(0)$ exists.
- If X and Y are independent, then
$$\phi_{X+Y} = \phi_X \cdot \phi_Y.$$
- X and Y are independent if and only if
$$\phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t).$$
- If $X \sim \mathrm{Be}(p)$, then
$$\phi_X(t) = q + pe^{it}.$$
- If $X \sim N\left(\mu, \sigma^2\right)$, then
$$\phi_X(t) = \exp\left(i\mu t - \frac{1}{2}\sigma^2 t^2\right).$$
- Cummulant Generating Function: $\log \phi_X(t)$.

• **Inversion Theorem**: If X is a continuous random variable with p.d.f. f and c.f. ϕ, then
$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\mathrm{d}t,$$
at every point x where f is differentiable.
More generally, let $\overline{F}(x) = \frac{1}{2}\left(F(x) + F(x^-)\right)$.
$$\overline{F}(b) - \overline{F}(a) = \lim_{N\to\infty}\int_{-N}^{-N} \frac{e^{-iat} - e^{-ibt}}{2\pi it}\phi(t)\mathrm{d}t.$$

6 Convergence of Random Variables

• **Weak Convergence**: Let $\{F_n\}$ be a sequence of distribution functions. F_n converges weakly to the distribution function F, denoted by $F_n \to F$, if $F_n(x) \to F(x)$ pointwise at each point x where F is continuous.

• **Convergence in Distribution**: Let $\{X_n\}$ be a family of random variables with distribution function $\{F_n\}$. X_n converges to X in distribution, denote by $X_n \xrightarrow{D} X$ if $F_n \to F$.

- **Vague Convergence**: Let $\{F_n\}$ be a sequence of distribution functions. If $F_n(x) \to G(x)$ at all continuities of G where G is not necessarily a distribution function, then F_n converges to G vaguely, denote by $F_n \xrightarrow{v} G$.

- **Weak Law of Large Number**: Let $\{X_n\}$ be i.i.d. random variables. Assume that $\mathbb{E}|X_1| < \infty$ and $\mathbb{E}X_1 = \mu$. Let $S_n = \sum_{i=1}^n X_i$. Then,
$$\frac{1}{n}S_n \xrightarrow{D} \mu.$$

- **Central Limit Theorem**: Let $\{X_n\}$ be i.i.d. random variables with $\mathbb{E}|X_i|^2 < \infty$. Let $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2$ and $S_n = \sum_{i=1}^n X_i$. Then.
$$\frac{1}{\sigma}\sqrt{n}\left(\frac{1}{n}S_n - \mu\right) = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1).$$
More generally, let $\{X_n\}$ be independent random variables with $\mathbb{E}X_i = 0$, $\operatorname{Var} X_i = \sigma_i^2$, $\mathbb{E}|X_i|^3 < \infty$ such that
$$\frac{1}{\sigma^3(n)}\sum_{i=1}^n \mathbb{E}|X_i|^3 \to 0,$$
as $n \to \infty$, where $\sigma^2(n) = \sum_{i=1}^n \sigma_i^2$. Then,
$$\frac{S_n}{\sigma(n)} \xrightarrow{D} N(0,1).$$

Let $\{X_n\}$ be a sequence of random variables. Let $S_n = \sum_{i=1}^n X_i$.

- **Almost Sure Convergence**: X_n converges to X almost surely (a.s.), denote by $X_n \xrightarrow{\text{a.s.}} X$ if
$$\mathbb{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}) = \mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

- **Convergence in r-th Mean**: X_n converges to X in r-th mean, denote by $X_n \xrightarrow{r} X$ if
$$\mathbb{E}|X_n - X|^r \to 0 \iff \lim_{n\to\infty} \mathbb{E}|X_n - X|^r = 0,$$
where $r \geq 1$.

- **Convergence in Probability**: X_n converges to X in probability, denote by $X_n \xrightarrow{\mathbb{P}} X$ if for every $\varepsilon > 0$,
$$\mathbb{P}(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}) \to 0 \iff \lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0.$$

- **Implications**:
  - $\left(X_n \xrightarrow{\mathbb{P}} X\right) \Rightarrow \left(X_n \xrightarrow{D} X\right)$.
  - $\left(X_n \xrightarrow{\text{a.s.}} X\right) \Rightarrow \left(X_n \xrightarrow{\mathbb{P}} X\right)$.

- $\left(X_n \xrightarrow{D} X\right) \Rightarrow \left(X_n \xrightarrow{\mathbb{P}} X\right)$.
- $\left(X_n \xrightarrow{r} X\right) \Rightarrow \left(X_n \xrightarrow{s} X\right)$, if $r \geq s \geq 1$.
- No other implications hold in general.

- **Markov Inequality**: Let X be any random variable. Let ϕ be any nonnegative increasing function on $[0, \infty)$. Then,
$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}\phi(|X|)}{\phi(a)},$$
for every $a > 0$.

- **Lyapunov's Inequality**: Let Z be any random variable.
$$\left(\mathbb{E}|Z|^s\right)^{\frac{1}{s}} \leq \left(\mathbb{E}|Z|^r\right)^{\frac{1}{r}},$$
for every $r \geq s > 0$.

Let $A_n(\varepsilon) := \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$ and $B_m(\varepsilon) := \bigcup_{n=m}^{\infty} A_n(\varepsilon)$.

- **Theorem 1**: $X_n \xrightarrow{\text{a.s.}} X$ if and only if
$$\lim_{m\to\infty} \mathbb{P}(B_m(\varepsilon)) = \mathbb{P}\left(\bigcap_{m=1}^{\infty} B_m(\varepsilon)\right) = \mathbb{P}\left(\bigcap_{m=1}^{\infty}\bigcup_{n=m}^{\infty} A_n(\varepsilon)\right) = \mathbb{P}\left(\limsup_{n\to\infty} A_n(\varepsilon)\right) = \mathbb{P}(A_n(\varepsilon) \text{ i.o.}) = 0,$$
for every $\varepsilon > 0$.

- **Theorem 2**: $X_n \xrightarrow{\text{a.s.}} X$ if
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n(\varepsilon)) < \infty,$$
for every $\varepsilon > 0$.
  - Note that
$$\sum_{n=1}^{\infty} \mathbb{P}(A_n(\varepsilon)) < \infty \text{ implies } \lim_{m\to\infty} \sum_{n=m}^{\infty} \mathbb{P}(A_n(\varepsilon)) = 0.$$

- **Parial Converse Statements**:
  - If $X_n \xrightarrow{D} c$, where c is a constant, then $X_n \xrightarrow{\mathbb{P}} c$.
  - If $X_n \xrightarrow{\mathbb{P}} X$ and $\mathbb{P}(|X_n| \leq k) = 1$ for all n uniformly with some constant $k > 0$, then $X_n \xrightarrow{r} X$, for every $r \geq 1$.

- **L^2-Weak Law of Large Number**: Assume that $\{X_n\}$ is uncorrelated random variables with $\mathbb{E}X_i = \mu$ and Var $X_i \leq C$ for every i. Then,
$$\frac{S_n}{n} \xrightarrow{2} \mu,$$
and consequently,
$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mu.$$

- **Weak Law of Large Number for Triangular Array**: Let $\{X_{n,j}\}_{j=1,\dots,n}$ be a triangular array of random variables. Let $S_n = \sum_{i=1}^n X_{n,i}$, $\mathbb{E}S_n = \mu_n$ and $\operatorname{Var} S_n = \sigma_n^2$. If

$$\frac{\sigma_n^2}{b_n^2} \to 0,$$

for some sequence $\{b_n\}$, then

$$\frac{S_n - \mu_n}{b_n} \xrightarrow{\mathbb{P}} 0.$$

7 Borel-Cantelli Lemma and 0-1 Law

Let $\{A_n\}$ be a sequence of events in $(\Omega, \mathcal{F})$. Let $\{X_n\}$ be a sequence of random variables. Let $S_n = \sum_{i=1}^n X_i$.

- $$\limsup_{n\to\infty} A_n = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = A_n \text{ i.o..}$$

- **Borel-Cantelli Theorem**: For any sequence of $\{A_n\}$ in $\mathcal{F}$,
  - $\mathbb{P}(A_n \text{ i.o.}) = 0$ if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$,
  - $\mathbb{P}(A_n \text{ i.o.}) = 1$ if $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ and $\{A_n\}$ is independent.
    * If $\{A_n\}$ is pairwise independent and $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, then

$$\frac{\sum_{m=1}^n \mathbf{1}_{A_m}}{\sum_{m=1}^n \mathbb{P}(A_m)} \xrightarrow{\text{a.s.}} 1.$$

- **Theorem**: $X_n \xrightarrow{\mathbb{P}} X$ if and only if for every subsequence $X_{n(m)}$, there exists a further subsequence such that

$$X_{n(m_k)} \xrightarrow{\text{a.s.}} X.$$

- **Strong Law of Large Number**: Let $\{X_n\}$ be i.i.d. random variables with $\mathbb{E}X_i = \mu$ and $\mathbb{E}X_i^4 < \infty$ (which is not necessary), then

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu.$$

- **Tail Sum Formula**:

$$\mathbb{E}|X| = \int_0^{\infty} \mathbb{P}(|X| > t)\, \mathrm{d}t.$$

- **Glivenko-Cantelli Theorem**: Let $\{X_n\}$ be i.i.d. samples of X with distribution function F. Let

$$F_N(x) := \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(X_i \le x).$$

Then,

$$\sup_x |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0.$$

- **Borel 0-1 Law**: Let $\{A_n\}$, not necessarily independent, be events in $\mathcal{F}$. Let $\mathcal{A} := \sigma(A_1, A_2, \ldots)$. Suppose that
  - $A \in \mathcal{A}$, i.e. A is a combination of A_i's and $A_i^{\complement}$'s,
  - A is independent of any finite collection of $\{A_n\}$.

  Then,
  $$\mathbb{P}(A) = 0 \text{ or } 1.$$

- **Kolmogorov 0-1 Law**: Let $\{X_n\}$ be independent random variables on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{H}_n := \sigma\{X_{n+1}, X_{n+2}, \ldots\} = \sigma\left\{\bigcup_{i=m+1}^{\infty} \sigma(X_i)\right\}$. Let $\mathcal{H}_\infty = \bigcap_{n=1}^{\infty} \mathcal{H}_n$, the tail σ-field. If $E \in \mathcal{H}_\infty$,
  - E is expressed in terms of $X_i^{-1}(B_i)$'s where $B_i \in \mathcal{B}(\mathbb{R})$, but is independent of any collection of $X_i^{-1}(B_i)$'s,
  - $\mathbb{P}(E) = 0$ or 1.

- **Tail Function**: $Y : \Omega \to \mathbb{R} \cup [-\infty, \infty]$ is a tail function if it is $\mathcal{H}_\infty$-measurble, i.e. it is a function of $X_1, X_2, \ldots$ and is independent of any finite collection of X_i's.

- **Theorem**: If Y is a tail function of independent random variable sequence $\{X_n\}$, then
  $$\mathbb{P}(Y = k) = 1,$$
  for some $k \in [-\infty, \infty]$.
  - Let $\{X_n\}$ be independent. Then,
  $$\mathbb{P}\left(\lim_{n\to\infty} \frac{S_n}{n} \text{ exists}\right) = 0 \text{ or } 1.$$
  - For any random power series, the radius of convergence equals to a constant almost surely.