Probability

§1 Basic Notion and Property.

Notation:

$\Omega = \{ \dots \}$. sample space (finite / infinite).

A, $A \in \Omega$. event.

$P(A) \in [0,1]$ probability of A happens.

C_n^k ; $\binom{n}{k}$. combination number (binomial coefficient).

$A_n^k = \frac{n!}{(n-k)!}$ permutation number

$\hat{A}_n^k = n^k$ permutation with repetition.

$\tilde{C}_n^k = C_{n+k-1}^k = C_{n+k-1}^{n-1}$ combination with repetition.

- discrete. probability.

$\Omega = \{w_1, \dots w_n\}$. $A \in \Omega$. $P(A) = \frac{|A|}{|\Omega|}$. $|\dots|$ - power (number of events)

<u>example</u>. $\Omega = \{10000 \leq n \leq 99999\}$. $A = \{n \bmod 3 = 0\}$. $P(A) = ?$

<u>digitize method</u>: $\boxed{a|b|c|d|e}$

for arbitrary b, c, d, e. $(b+c+d+e) \bmod 3$: 0 or 1 or 2.

a has. 3. possible cases. $a = 3.6.9$ $a = 2,5,8$ $a = 1,4,7$.

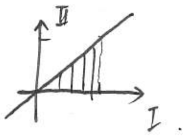
$P(A) = 3 \times 10^4 = 30000$

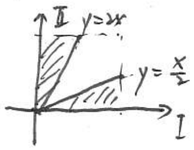
- continuous. probability.

$\Omega = \{$infinite elements$\}$. $A \in \Omega$. $P(A) = \frac{mes(A)}{mes(\Omega)}$ mes - measure.

<u>example</u>. $\Omega = [0, 10]$. choose. 2 numbers from $[0, 10]$. (I, II).

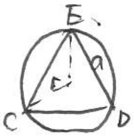
a). $P(I > II)$. b) P(one of number in 2. time more than other).





- Bertrand Paradox.

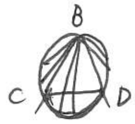
Q: Randomly select a chord within a circle



inscribed regular triangle. side $= a$. AB - arbitrary chord.

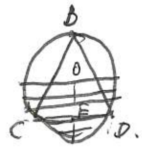
$P(AB > a) = ?$

I. method of random of chord



$|\Omega| = \odot$. $\quad P(AB > a) = \frac{|\overset{\frown}{CD}|}{|\odot|} = \frac{2\pi/3}{2\pi} = \frac{1}{3}$.

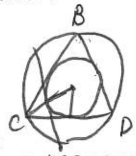
II. method of random radius. (choose point in the radius.)



max: $2R$.
min: 0.

$|\Omega| = R$. $\quad P(AB > a) = \frac{|OE|}{R} = \frac{\frac{R}{2}}{R} = \frac{1}{2}$

III. method of random centre. (choose point, construct chord centre in the point)



$|\Omega| = \text{disc}$. $\quad P(AB > a) = \frac{\pi r^2}{\pi R^2} = \frac{1}{4}$.

$P = \frac{1}{3}$ (measure of arc). $\frac{1}{2}$ (measure of length). $\frac{1}{4}$ (measure of area)

(why different results ?).

$\Rightarrow$ we need to well-defined the randomness.

§2. Axiom of Probability. (Kolmogorov).

$(\Omega, \mathbb{P}) \to (\Omega, \mathcal{A}, \mathbb{P})$. - probability space.

$\mathcal{A}$ - σ-algebra.

* 注意 deck 题 joker 也分颜色.

* $B - C = B \cap \bar{C}$.

Property.

(1). $\forall A \in \mathcal{A}$. $\quad P(A) \in [0,1]$. $\quad P(\Omega) = 1$. $\quad P(\emptyset) = 0$.

(2). $\forall A \in \mathcal{A}$. $\quad \bar{A} = \Omega \setminus A = \{\omega \in \Omega \mid \omega \notin A\}$.

$P(\bar{A}) = 1 - P(A)$

(3) $\forall A, B \in \mathcal{A}$: $A \cap B = \emptyset$. $\Rightarrow P(A \cup B) = P(A) + P(B)$.

$\forall \{A_i\}_{i=1}^{n} \in \mathcal{A}$: $\forall i \neq j$ $\quad A_i \cap A_j = \emptyset \Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

(4). $\forall A, B \in \mathcal{A}$: $A \subseteq B$. $\Rightarrow P(A) \leq P(B)$.

$\forall A, B \in \mathcal{A}$: $A \subseteq B \quad \Rightarrow \quad P(B \setminus A) = P(B) - P(A)$.

(5) $\forall A, B \in \mathcal{A}$. $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

* 在题目条件中，";"表示"and"，同时分割了其他描述性短语和下一个条件.

如 "at least A and B". - 表示至少 A 个，和正好 B 个

§3. Conditional Probability.

<u>Def.</u> the probability of event A when event B has occurred. $-P(A|B)$. ↙ syn. given that.

$$P(A|B) = \frac{P(A\cap B)}{P(B)}, \quad P(B)>0.$$

(if $P(B)=0$. $P(A|B)$ is undefined.)

<u>Property.</u>

(1) $P(\Omega|B)=1 \quad P(\emptyset|B)=0.$

(2). $A\subseteq C$. $\because P(A)\le P(C)$. $\Rightarrow \quad P(A|B)\le P(C|B)$.

(3). $P(\bar{A}|B) = 1-P(A|B)$.

(4). $P(A\overset{\text{union } \cup}{+}C|B) = P(A|B)+P(C|B)-P(A\cap C|B)$

(5) $A\subseteq C. \quad P(C-A|B)=P(C|B)-P(A|B)$.

Probability multiplication formula.

$$P(AB) \triangleq P(A\cap B) = P(B)\cdot P(A|B) = P(A)\cdot P(B|A).$$

$$P(A_1A_2\dots A_n) = P(A_1)P(A_2|A_1)\,P(A_3|A_1 * A_2) * \dots * P(A_n|A_1 * \dots * A_{n-1}).$$ $*$ - intersectA

<u>Independent Event.</u>

<u>Def.</u> the event A is independent from B: if $P(A|B) = P(A)$. ↙ probability of A given B.

similarly. $P(B|A)=P(B)$.

<u>Judge.</u> A.B are independent $\Leftrightarrow P(A\cap B) = P(A)\cdot P(B)$.

<u>Property.</u> (1). $P(A\cup B) = P(A)+P(B)-P(A)\cdot P(B)$.

(2) $P(A_1\cap A_2\cap\dots\cap A_n) = P(A_1)\times P(A_2)\times\dots\times P(A_n)$.

<u>Def.</u> (independent collectively). ↙ syn. jointly the events. $A_1, A_2, \dots A_n$ are independent. collectively if for any k. such that $1\le i_1 < i_2 < \dots < i_k \le n$: $P(A_{i_1}A_{i_2}\dots A_{i_k}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_k})$.

They are pairwise independent. if: $\forall i,j$. s.t. $1\le i<j\le n$. $P(A_i\cap A_j) = P(A_i)P(A_j)$

$\triangle$. pairwise independent $\underset{\Longleftarrow}{\not\Rightarrow}$ jointly / collectively independent.

e.g. a regular pyramid.  paint R.G.B. each of three face. in one color; the 4th face painted in 3 colors at once. ↙ 3面涂一色. 一面涂3色.

△只看一面，看到某色概率.

$P(R)=P(G)=P(B)=\frac{1}{2}$

$P(BG)=P(RG)=P(BR)=P(RGB)=\frac{1}{4}$.

$P(RGB)\neq P(B)\cdot P(G)\cdot P(R)$.

Total Probability Formula. (TPF).

Let $\{H_i\}_{i=1,2,\dots,n}$ be a complete system of events (i.e. $\forall i\neq j$ $H_i \cap H_j = \phi$ $\sum_{i=1}^{n} H_i = \Omega$. $P(H_i) > 0$ for any A). then we have $P(A) = \sum_{i=1}^{n} P(H_i) P(A|H_i)$

Pf: $P(A) = P(A\cdot\Omega) = P(A\cdot(\bigcup_{i=1} H_i))$

AHi mutually exclusive.

$= \sum P(AH_i) = \sum_{i} P(H_i) P(A|H_i)$.

△常用于分步问题. 首先要对最后一步前的状态构建 complete system of events

Bayes' Theorem.

Let $\{H_i\}_{i=1,2,\dots n}$ be a complete system of event. and some event A.

for all $k=1,2,\dots,n$. $P(H_k|A) = \dfrac{P(H_k)P(A|H_k)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)}$

Remark: if $P(B_iC) > 0$. $B_i \cap B_j = \phi$ $\forall i \neq j$. $AC \subseteq \bigcup_{k=1}^{n} B_k$

then $P(A|C) = \sum_{k=1}^{n} P(A|B_kC) P(B_k|C)$.

△$P(H_k)$为H_k的先验概率. $P(H_k|A)$是H_k的后验概率.
Bayes是通过A已发生的信息. 修正H_k的概率.

§ Binary Distribution.

Bernoulli Trail

Problem: Given some event $A \subset \Omega$. let the probability of success of A is p.
while the failure be $q = 1-p$. (independent p, q).

In n trial. the probability of obtaining k successes. $P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}$

We have (1). $P_n(k) > 0$ (2). $\sum_{k=0}^{n} P_n(k) = 1$.

lemma. For fixed n. the sequence $P_n(0), P_n(1), \dots, P_n(n)$. is unimodal (单峰的).

$$\begin{cases} P_n(k+1) \geq P_n(k) & \text{for } k \leq np-1+p. \\ P_n(k+1) < P_n(k) & \text{for } k > np-1+p. \end{cases}$$

coro. The maximum $P_n(k)$ is attained by $k \in [np-1+p, np+p]$. (k-most probable outcome)
若为整数. 有2个最大值点.

the pro. of success. number between m_1, m_2. $P_n(m_1, m_2) = \sum_{k=m_1}^{m_2} P_n(k) = \sum_{k=m_1}^{m_2} \binom{n}{k} p^k (1-p)^{n-k}$.

Poisson 's Theorem.

在固定时空中.稀有事件发生次数.

Let $np = \lambda = \text{const}$. For any m and const. λ: $\lim\limits_{\substack{n\to\infty \\ np=\lambda}} P_n'(m) = \dfrac{\lambda^m}{m!} e^{-\lambda}$.

<u>Note</u>. Let $p_\lambda(m) = \dfrac{\lambda^m}{m!} e^{-\lambda}$, then $\sum\limits_{m=0}^{\infty} \dfrac{\lambda^m}{m!} e^{-\lambda} = 1$.

($p_\lambda(m)$. λ 不指事次件数. 表示一类情况发生(如 $n\to\infty$, 稀有事件).).

此处的 $n\to\infty$ 并不严格. 对于"大数"即可使用).

Moivre - Laplace Local Limit Theorem.

Denote $x_n = \dfrac{m-np}{\sqrt{npq}}$, suppose $m\to\infty$, $n\to\infty$, the values x_n are bounded

(sufficiently $n^{\frac{1}{6}-\epsilon}$). Then $\sqrt{npq}\, P_n(m) \sim \dfrac{1}{\sqrt{2\pi}} e^{\frac{x_n^2}{2}}$

<u>Remark</u>: define Gaussian function $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,

the thm. can be expressed as: $P_n(m) \sim \dfrac{1}{\sqrt{npq}} \varphi(x)$

Pf: $n-m = n(1-p) + np - m = nq - x_n\sqrt{npq} \xrightarrow[n,m\to\infty]{} \infty$ (since x_n is bounded) (1).

by Stirling formula: $k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$

$$\sqrt{npq}\, P_n(m) = \frac{\sqrt{npq}\; n!}{m!\,(n-m)!} p^m q^{n-m} \sim \frac{\sqrt{npq}\,\sqrt{n}\, n^n}{\sqrt{2\pi}\sqrt{m}\, m^m \sqrt{n-m}\,(n-m)^{n-m}} \cdot p^m q^{n-m}$$

$$= \frac{1}{\sqrt{2\pi}} \left(\frac{np}{m}\right)^m \left(\frac{nq}{n-m}\right)^{n-m} \cdot \underbrace{\sqrt{\frac{np}{m}\frac{nq}{n-m}}}_{\to 1}$$

$$\sim \frac{1}{\sqrt{2\pi}} \exp\left(-m \ln\frac{m}{np} - (n-m)\ln\frac{n-m}{nq}\right).$$

$$\frac{m}{np} = \frac{np + x_n\sqrt{pqn}}{np} = 1 + \underbrace{\frac{x_n\sqrt{q}}{\sqrt{np}}}_{\to 0}, \quad \frac{n-m}{nq} = \frac{nq - x_n\sqrt{npq}}{nq} = 1 - \underbrace{\frac{x_n\sqrt{p}}{\sqrt{nq}}}_{\to 0}. \quad (2)$$

by Taylor extension. $\ln(1+z) = z - \dfrac{z^2}{2}(1+o(1)) \quad z\to 0$.

$$\sqrt{npq}\, P_n(m) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-m\left(\frac{x_n\sqrt{q}}{\sqrt{np}} - \frac{x_n^2 q}{2np}(1+o(1))\right)\right.$$

$$\left. - (n-m)\left(-\frac{x_n\sqrt{p}}{\sqrt{nq}} - \frac{x_n^2 p}{2nq}(1+o(1))\right)\right)$$

$$-x_n\left(\frac{m\sqrt{q}}{\sqrt{np}} - \frac{(n-m)\sqrt{p}}{\sqrt{nq}}\right) \overset{\text{use (2). replace } m \text{ and } n-m.}{=} -x_n\left[\sqrt{q}\left(\sqrt{np} + x_n\sqrt{q}\right) - \sqrt{p}\left(\sqrt{nq} - x_n\sqrt{p}\right)\right]$$

$$= -x_n\left(x_n q + x_n p\right) = -x_n^2 \quad (3)$$

$$\frac{x_n^2}{2}\left(\frac{mq}{np}(1+o(1)) + \frac{(n-m)p}{nq}(1+o(1))\right) = \frac{x_n^2}{2}\left(p+q + x_n\underbrace{\left(\frac{q^2-p^2}{\sqrt{npq}}\right)}_{\to 0}\right) \sim \frac{x_n^2}{2}.$$

$$\Rightarrow \sqrt{npq}\, P_n(m) \sim \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_n^2}{2}(1+o(1))\right) \sim \frac{1}{\sqrt{2\pi}} e^{-\frac{x_n^2}{2}}. \qquad \square$$

△ 标号的是对 m, n, p, q, x_n 的转化式，逐步地对原式进行代换.

Moirve-Laplace Integral Limit Theorem.

Assume $a_n = \frac{m_1 - np}{\sqrt{npq}}$, $b_n = \frac{m_2 - np}{\sqrt{npq}}$. Suppose that $m_1 \to \infty$, $n \to \infty$ and the values b_n are bounded. Then we have:

$$P_n(m_1, m_2) = P_n(m_1 \le m \le m_2) \simeq \frac{1}{\sqrt{2\pi}} \int_{a_n}^{b_n} e^{-\frac{x^2}{2}}\, dx.$$ ←常用于计算置信区间.

Remark 1. define $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt$. Φ - cumulative distribution func. of the normal distribution

we have $\Phi(-\infty) = 0$ $\Phi(+\infty) = 1$. $\Phi(0) = \frac{1}{2}$. $\Phi(-x) = 1 - \Phi(x)$

Then $P_n(m_1, m_2) \approx \Phi(b_n) - \Phi(a_n)$.

Remark 2: define $\Phi_0(x) = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}}\, dt$. Φ_0 - Laplace func.

we have $\Phi_0(0) = 0$ $\Phi_0(+\infty) = 1$. $\Phi_0(-x) = -\Phi_0(x)$.

Then $P_n(m_1, m_2) \approx \frac{1}{2}(\Phi_0(b_n) - \Phi_0(a_n))$.

Parametrization of the Bernoulli Trial.

Let's consider $X_i = 1$ if event A occurs in the i-th trial, otherwise $X_i = 0$. Then the number of successes is given by $S_n = \sum_{i=1}^{n} X_i$

Def. The value $\mu = \frac{S_n}{n}$ is called the relative frequency.

Lemma For any $\varepsilon > 0$, $P(|\mu - p| < \varepsilon) \xrightarrow[n\to\infty]{} 1$

§4. Random Variable.

A probability space: $(\Omega, \mathcal{F}, \mathbb{P})$. Ω - sample space. $\mathcal{F}$ - σ-algebra.
$\mathbb{P}$ - probability measure.

Def 4.1. (random variable). A random variable is a function that assigns a real number to an outcome $X: (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$. △样本空间上的实值函数.

i.e. $\forall B \in \mathcal{B}_{\mathbb{R}}$. $X^{-1}(B) = \{\omega \mid X(\omega) \in B\} \subseteq \mathcal{F}$. $\mathcal{B}_{\mathbb{R}}$ is Borel σ-algebra.

△此处 X^{-1} 没有 X 是单射要求. △ "$\in \mathcal{F}$" means "it's an event"

Def 4.2. The σ-algebra generated by a random variable X is the smallest σ-algebra: contains the preimage of every Borel set under X.

$\sigma(X) = \{X^{-1}(B) \mid B \in \mathcal{B}_{\mathbb{R}}\}$, where $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$.

若 P_X 在M下绝对连续，即为密度函数.

Def 4.3 (distribution). The distribution of a random variable X is the probability measure P_X on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ given as: $P_X(B) = \mathbb{P}(\{\omega : X(\omega) \in B\})$

Def 4.4. Random variable X and Y are identically distributed ($X \overset{d}{=} Y$) if $P_X = P_Y$. (note that X, Y can be given on different probability space).

Def 4.5. distribution function: $F_\xi(x) = P_\xi((-\infty, x)) = \mathbb{P}(\{\omega : \xi(\omega) < x\}) = P(\xi < x)$

Prop. 4.1. (Continuity of probability measure)

(1). Let a system of events s.t. $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_n \subseteq \cdots$

Then we have. $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \lim_{i \to \infty} \mathbb{P}(A_i)$

(2) Let a system of events s.t. $B_1 \supseteq B_2 \supseteq \cdots \supseteq B_n \supseteq \cdots$

Then we have $\mathbb{P}(\bigcap_{i=1}^{\infty} B_i) = \lim_{i \to \infty} \mathbb{P}(B_i)$.

Remark: The bijection function exists between $F_X \leftrightarrow P_X$ on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$

(1) $P_X([a,b]) = F_X(b) - F_X(a)$.

(2) $P_X((a,b)) = P_X(\bigcap_{c<b} (a,c]) = \lim_{c \to b-} P_X((a,c]) = F_X(b-) - F_X(a)$

<u>Example</u>. choose the best σ-algebra for a given random variable.

Let $\Omega = [-1,1]$, $\mathcal{F} = \mathcal{B}([-1,1])$. $\mathbb{P}: \mathcal{F} \to [0,1]$. s.t. $\mathbb{P}([a,b)) = \frac{b-a}{2}$.

Consider the random variable (1) $\xi(\omega) = |\omega|$ (2) $\eta(\omega) = \max(0, |\omega-\frac{1}{4}|-\frac{1}{2})$

Sol: find η^{-1} and ξ^{-1} △此处"原像"不要搞，不要求是单射

$\sigma(\xi) = \{(a,b) \cup (-b,-a) \mid (a,b) \subseteq [0,1]\} \cup \{0\}$. ← 由这类集合生成的 σ-algebra. 无非关于原点对称的集合.

$$\eta(\omega) = \begin{cases} 0 & \omega \in [\frac{1}{4}, \frac{3}{4}] \\ -\omega+\frac{1}{4} & \omega \in [-1, \frac{1}{4}] \\ \omega - \frac{3}{4}, & \omega \in [\frac{3}{4}, 1] \end{cases}$$

$\eta^{-1}(0) = [\frac{1}{4}, \frac{3}{4})$

$$\eta^{-1}((a,b)) = \begin{cases} (\frac{1}{4}-b, \frac{1}{4}-a) & (a,b) \subseteq (\frac{1}{4}, \frac{5}{4}) \\ (\frac{3}{4}+a, \frac{3}{4}+b) \cup (\frac{1}{4}-b, \frac{1}{4}-a), & (a,b) \subseteq (0, \frac{1}{4}) \end{cases}$$

$$\sigma(\eta) = (\frac{1}{4}, \frac{3}{4}) \cup \{(\frac{1}{4}-b, \frac{1}{4}-a) \mid (a,b) \subseteq (\frac{1}{4}, \frac{5}{4})\} \cup \{(\frac{3}{4}+a, \frac{3}{4}+b) \cup (\frac{1}{4}-b, \frac{1}{4}-a) \mid (a,b) \subseteq [0, \frac{1}{4})\}$$

<u>prop 4.2</u> (distribution func.)

(1) $F_\xi(\cdot) \nearrow$ (monotonic increasing)

(2) F_ξ is left cont. and it has the right limit.

(3) $\lim_{x\to+\infty} F_\xi(x) = 1$. $\lim_{x\to-\infty} F_\xi(x) = 0$.

Remark: From (2). $\lim_{x\to x_0^+} F_\xi(x) = P(\xi \le x_0) = F_\xi(x) + P(\xi = x_0)$. ↙对绝对连续分布. 有 $P(\xi = x_0) \equiv 0$.

Common Distribution: (Discrete)

1. Discrete distribution.

Ω - finite / countable $= \{\omega_1, \omega_2, \dots\}$. ; ξ - random variable with the value $\{x_i\}_{i\in\mathbb{R}}$.

$p_i = P(\xi = x_i)$, $\sum_i p_i = 1$.

The distribution function :

ξ	x_1	x_2	x_3	$\cdots$	x_n	$\cdots$
P	p_1	p_2	p_3	$\cdots$	p_n	$\cdots$

$$F_\xi(x) = \mathbb{P}(\xi < x) = \sum_{x_i < x} \mathbb{P}(\xi = x_i) = \sum_{x_i < x} p_i$$

2. Bernoulli distribution. ($Ber(p)$)

$X \in \{0,1\}$. $P(X=0) = q$. $P(X=1) = p$. $E(X) = p$. $D(X) = pq$

3. Poisson distribution. ($Pois(\lambda)$).

$X \in \mathbb{N}$. $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \dots$ $E(X) = \lambda$. $D(X) = \lambda$.

4. Binomial distribution ($Bin(n,p)$).

$X \in \{0, 1, 2, \dots, n\}$. $P(X=k) = \binom{n}{k} p^k (1-p)^k$. $k \in [0:n]$

$E(X) = np$. $D(X) = npq$.

5. Discrete Uniform distribution ($Unif\{1, \dots, n\}$)

$X \in \{1, \dots, n\}$. $p = n^{-1}$

$E(X) = \frac{n+1}{2}$ $D(X) = \frac{n^2-1}{2}$.

6. $Geom_1(p)$. X - trail until 1st. success. (include it)

$P(X=k) = p \cdot (1-p)^{k-1}$, $k = 1, 2, \dots$ $E(X) = \frac{1}{p}$. $D(X) = \frac{q}{p^2}$

7. $Geom_2(p)$ X - failure before 1st success. (not incl. it)

$P(X=k) = p(1-p)^k$. $k = 0, 1, 2, \dots$ $E(X) = \frac{q}{p}$. $D(X) = \frac{q}{p^2}$

Characteristics of discrete random variable.

Def 4.6. The mathematical expectation of the discrete random variable X. is

$$\mathbb{E}(X) = \sum_{i=1}^{n} x_i \, p(X = x_i) = \sum_{i=1}^{n} x_i p_i$$

similarly the 2nd and kth moment of a discrete random variable X.

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} x_i^2 \cdot p_i \qquad m_k = \mathbb{E}(X^k) = \sum_{i=1}^{n} x_i^k p_i$$

Def 4.7 The variance of a discrete random variable X is defined as the number:

$$\mathbb{D}(X) = \sum_{i=1}^{n} (x_i - \mathbb{E}(X))^2 \cdot p_i = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = m_2 - m_1^2 \geq 0.$$

△此处的"abs. cont."是 $\mathbb{P}$ abs. cont. w.r.t. μ. (即 $\mathbb{P} \ll \mu$). Radon - Nikodym thm.

$\exists f$, μ-integrable. $\forall B(\xi) \in \mathcal{B}_{\mathbb{R}}$, $F_\xi(x) = \mathbb{P}(\xi < x) = \int_B f \, d\mu$.

Characteristics of absolutely continuous distribution

Ω - cont. $P(\xi = k) = 0$. △对 abs. cont. 分布的随机变量. 考虑他在某一个取值的概率是无意义的.

Def 4.8 Probability density function of a random variable ξ. is non-negative function $f(y)$. s.t. $F_\xi(x) = \int_{-\infty}^{x} f(y) \, dy$. $F_\xi(x)$ - cumulative density func. CDF

PDF

prop. 4.3 (1) $\int_{-\infty}^{\infty} f(y) \, dy = 1$.

(2) $f(x) = F'(x)$ if the distribution func. is differentiable.

(3). $\mathbb{P}(a \leq \xi < b) = \int_a^b f(y) \, dy = F(b) - F(a)$.

Def 4.9. The mathematical expectation of random variable with abs. cont. distri.

$$E(\xi) = \int_{-\infty}^{\infty} x f(x) \, dx.$$

the 2nd moment $\mathbb{E}(\xi^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x) \, dx$.

the variance. $\mathbb{D}(\xi) = \mathbb{E}(\xi - \mathbb{E}(\xi))^2 = \mathbb{E}(\xi^2) - (\mathbb{E}(\xi))^2$

Remark. For a func. of random variable. $g(\xi)$.

$$\mathbb{E}(g(\xi)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

e.g. k-th moment $g(x) = x^k$ $\quad m_k = \mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x)\, dx.$

Common Continuous Distribution.

1. Uniform Distribution on $[0,1] \equiv \text{Unif}[0,1]$

在[0,1]之间为1. 其余地方为0.

$f(x) = 1 \cdot \text{Ind}_X([0,1])$. $\quad \mathbb{E}(X) = \frac{1}{2}$. $\quad D(X) = \frac{1}{12}$.

(for $f(x) = \frac{1}{b-a} \text{Ind}_X([a,b])$. - distribution $\text{Unif}[a,b]$. same result).

2. Standard Normal. Distribution. Norm(0,1).

$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $x \in \mathbb{R}$. $\quad F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt.$

$\mathbb{E}(X) = 0$. $\quad D(X) = 1$.

3. Exponential Distribution $Exp(\lambda)$.

$f(x) = \lambda e^{-\lambda x}$ $\quad F(x) = 1 - e^{-\lambda x}$. $\quad x \geq 0$, $\lambda > 0$.

$\mathbb{E}(X) = \frac{1}{\lambda}$

4. Standard Cauchy Distribution Cauchy(0,1)

$f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$.

$F(x) = \frac{1}{\pi} \arctan x + \frac{1}{2}$. $\quad x \in \mathbb{R}$.

$\mathbb{E}(X) = \infty$ (not exist!)

5. Normal distribution $N(\mu, \sigma^2)$.

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt.$$

The function of a random variable and its distribution

Let $(\Omega, \mathcal{F}, \mathbb{P})$. - probability space. ξ - random variable on the space.

Consider $\eta := g(\xi)$. $g(\cdot): \mathbb{R} \to \mathbb{R}$.

<u>Thm 4.3.</u> Let ξ be a random variable. $g: \mathbb{R} \to \mathbb{R}$. Borel func. ($\forall B \in \mathcal{B}_{\mathbb{R}}$. $g^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$)

Then $g(\xi)$ is a random variable.

<u>Pf:</u> $g(\xi): \Omega \to \mathbb{R}$. check $\forall$ Borel set it's preimage is an event. (i.e. $\in \mathcal{F}$).

$$g(\xi)^{-1}(B) = \{\omega \mid g(\xi(\omega)) \in B\} = \{\omega \mid \xi(\omega) \in g^{-1}(B)\} = \xi^{-1}(B_1).$$

$B_1 = g^{-1}(B)$ is a Borel set. ξ is random. variable thus $g(\xi)^{-1}(B) \in \mathcal{F}$. $\square$.

△ PDF of $\eta = g(\xi)$ not always exist (如 g 是 p.w. const func. η. 有离散分布.

<u>Thm 4.4.</u> Let ξ be a random variable. has CDF $F_\xi(x)$ and PDF $f_\xi(x)$.

Then the random variable $\eta = a\xi + b$ $(a \neq 0)$. has PDF $f_\eta(x) = \frac{1}{|a|} f_\xi\left(\frac{x-b}{a}\right)$.

<u>Pf:</u> i) $a>0$. $F_\eta(x) = P(a\xi + b < x) = P\left(\xi < \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right) = \int_{-\infty}^{(x-b)/a} f_\xi(t)\, dt.$

$$= \int_{-\infty}^{x} \frac{1}{a} f_\xi\left(\frac{y-b}{a}\right) dy.$$

ii) $a<0$. $F_\eta(x) = P(a\xi + b < x) = P\left(\xi > \frac{x-b}{a}\right) = \int_{\frac{x-b}{a}}^{+\infty} f_\xi(t)\, dt$

$$= \int_{x}^{+\infty} \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right) dy = \int_{-\infty}^{x} \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right) dy.$$

<u>Thm 4.5.</u> Let ξ have a probability density func. $f_\xi(x)$. and let $g: \mathbb{R} \to \mathbb{R}$. be a monotonic func. Then the random variable $\eta = g(\xi)$ has a PDF given by

$$f_\eta(x) = \left|\left(g^{-1}(x)\right)'\right| f_\xi(g^{-1}(x))$$

Remark: If the func. g is p.w. monotonic func. ($x^2 \in [-1,1]$ etc.).
the formula above can be given by $f_\eta(y) = \sum f_\xi(g_i^{-1}(y)) \cdot \left|(g_i^{-1}(y))'\right| = \sum \frac{f_\xi(g_i^{-1}(y))}{|g'(g_i^{-1}(y))|}$

△ "仅单调"需要导数? -> 仅有第一类间断点. 可以分别处理.

Coro 4.6. (1) $\xi \sim N(\mu, \sigma^2). \Rightarrow \frac{\xi-\mu}{\sigma} \sim N(0,1)$

(2) $\xi \sim \mathrm{Unif}[0,1] \Rightarrow a\xi+b \sim \mathrm{Unif}[b, a+b]$ for $a>0$.

(3) $\xi \sim \mathrm{Exp}(\alpha) \Rightarrow \alpha\xi \sim \mathrm{Exp}(1)$.

(4) $\xi \sim N(0,1) \Rightarrow \sigma\xi+\eta \sim N(\eta, \sigma^2)$

Quantile Transform. (Smirnov's method of modeling r.v.).

Thm 4.7. Let CDF $F(x) = F_\xi(x)$ be cont. Then the random variable $\eta = F(\xi)$ has a uniform distribution on the interval $[0,1]$. (i.e. $\eta \sim U_{0,1}$).

$F(x)$. CDF连续 $\Rightarrow \eta = F(\xi) \sim U_{0,1}$

Remark: The thm 4.7 can be used to generate random variables with a specified distribution using a uniformly distribution:

define $F^{-1}(x) = \inf\{t \mid F(t) \geq x\}$. (for any distribution func. F. not necess. cont.).

△用于生成随机变量.

Thm 4.8. Let $\eta \sim U_{0,1}$ and F be arbitrary distribution function. Then the random variable $\xi = F^{-1}(\eta)$ has the distribution func. F.

$\eta \sim U_{0,1}$ F. CDF $\Rightarrow \xi = F^{-1}(\eta)$ 的CDF为F.

Coro. 4.8.1 $\eta \sim U_{0,1}$. TFAS: 1). $-\frac{1}{a}\ln(1-\eta) \in E_a$. 2). $a+\sigma\tan(\pi\eta - \frac{\pi}{2}) \in C_{a,\sigma}$

3). $\Phi^{-1}_{\sigma,1}(\eta) \in N_{\sigma,1}$.

△ Thm 4.7 要求连续. (Why? $\exists F^{-1}(x) = x_0$. $x \in [F(x_0^-), F(x_0)]$, $\mathbb{P}(\eta \in [F(x_0^-), F(x_0)]) = \mathbb{P}(\xi = x_0) > 0$)

Thm 4.8 不要求连续 (Why? F^{-1} 取左端点. $F^{-1}(x) = \inf\{t \mid F(t) \geq x\}$

保证 $\mathbb{P}(\xi = x_0) = F(x_0) - F(x_0^-)$ 对 $\eta \in [F(x_0^-), F(x_0)]$. $F^{-1}(\eta) = x_0$.

Characteristics of random variable.

1. Raw moment of order k. (原始矩)

Def 4.10. The raw moment of order k of a random varible X is define as the expected value of its k-th power: $a_k = \mathbb{E}[X^k]$.

For d.r.v. with probability mass func. $p(x)$: $a_k = \sum_x x^k p(x)$

For cont. r.v. with probabitity density func. $f(x)$: $a_k = \int_{-\infty}^{\infty} x^k f(x)\, dx$.

2. Central moment of order k.

Def 4.11. The central moment of order k. of a random variable X is defined as.

$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k]$$

For d.r.v. $\mu_k = \sum_x (x - \mathbb{E}[X])^k p(x)$.

For. cont. r.v. $\mu_k = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^k f(x)\, dx$ $(\mathbb{E}[\text{const}] = \text{const})$

△一般不直接计算. $\mu_3 = \mathbb{E}[(X-\mathbb{E}[X])^3] = a_3 - 3a_1a_2 + 3a_1^3 - a_1^3$.

3. Mode. (众数).

Def 4.12. The mode of a random variable X. is the value at which its PMF ($p(x)$) or PDF ($f(x)$). reaches a maximum.

$$\text{mode}(X) = \arg\max_x P(X=x). \quad (\text{discrete})$$

$$\text{mode}(X) = \arg\max_x f(x) \quad (\text{continuous}).$$

→ argmax 表示 f 取到最大时 x 的值.

4. Quantile of level p. (分位数).

Def 4.13. $Q_p = F^{-1}(p) = \inf\{x \mid F(x) \geq p\}$. ($F(x)$ is CDF of X).

4*. Median.

Def. 4.13*. Median is Quatite of level $p = 0.5$.

$$F(x_{0.5}) = \int_{-\infty}^{x_{0.5}} f(x)\, dx = 0.5.$$

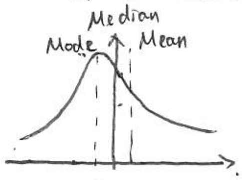
5. Skewness coefficient. (偏度系数)

Def 4.14. The skewness coefficient measures the asymmetry of pro. distri.

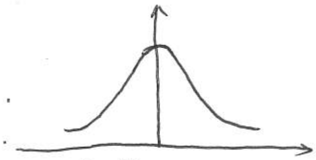
$$\gamma_1 = \frac{\mathbb{E}[(X-\mathbb{E}[X])^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{3/2}}$$

△ if $\gamma_1 > 0$. distribution skew to the right. (long right tail)

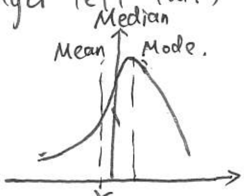
$\gamma_1 < 0$ distribution skew to the left (longer left tail)



$\gamma_1 < 0$



$\gamma_1 = 0$



$\gamma_1 > 0$

6. Excess kurtosis (峰度系数)

Def 4.15. The excess kurtosis measures the peakedness of a probability distribution

$$\gamma_2 = \frac{\mathbb{E}[(X-\mathbb{E}[X])^4]}{\sigma^4} - 3 = \boxed{\frac{\mu_4}{\sigma^4}} - 3$$

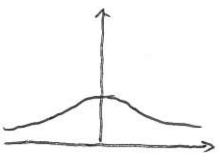
→ 最小值为1. 仅出现在 $\gamma_1 = 0$ 时. 对称两点分布.

if $\gamma_1 \neq 0$. $\frac{\mu_4}{\sigma^4} > 1$.

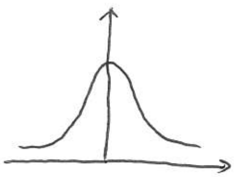
△ if $\gamma_2 = 0$. distribution has same kurtosis as Norm. (mesokurtic).

$\gamma_2 > 0$ distribution is more peaked. (leptokurtic).

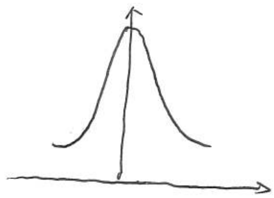
$\gamma_2 < 0$. distribution is flatter (platykurtic).



$\gamma_2 < 0$



$\gamma_2 = 0$



$\gamma_2 > 0$.

7. Probability-generating function. (discrete r.v. taking values in $\mathbb{Z}_+ \cup \{0\}$).

Def 4.16. $G_X(t) = \mathbb{E}(t^X) = \sum_{k=0}^{\infty} P(X=k)\, t^k$. $t \in [0,1]$

→ P这项与 t 无关. 可微. 连续 自然满足.

Prop 4.9 (1) $G_X(1) = 1$. $G_X(0) = P(X=0)$. (2) $G_X(t) \nearrow$ $t \in [0,1]$.

(3). $G_X(t)$ is convex. (4). if $\mathbb{E}(X) < \infty \Leftrightarrow G_X'(1) < \infty$ and $E(X) = G_X'(1)$.

Example.

(1). $Ber(p)$: $G_X(t) = \sum_{i=0}^{\infty} t^i\, \mathbb{P}(X=i) = tp + 1\cdot(1-p)$

(2) $Bin(n,p)$: $G_X(t) = \sum_{k=0}^{\infty} t^k\, \mathbb{P}(X=k) = \sum_{k=0}^{\infty} t^k C_n^k p^k (1-p)^{n-k} = (tp + (1-p))^n$

(3) $DUnif\{0,\dots,n\}$. $G_X(t) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n} t^i = \begin{cases} \frac{1}{n}\left(\frac{1}{1-t}\right). & |t| \neq 1. \\ \frac{1}{n}. & t = 1. \end{cases}$

(4) $Pois(\lambda)$ $G_X(t) = e^{-\lambda(1-t)}$.

8. Characteristic function.

<u>Def 4.17</u> $\varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x)\,dx$. $t \in \mathbb{R}$. `the expectation of complex exponential.

<u>Def 4.18</u> complex random variable are $X = X_1 + iX_2 : \Omega \to \mathbb{C}$. where X_1, X_2 - real r.v.

<u>Prop. 4.10</u> (1) $t \in \mathbb{R}$. $X \in \mathbb{R}$. $\Rightarrow |E(e^{itX})| \leq E|e^{itX}| = 1$. $\Rightarrow \varphi_X(t)$ exists.
(the characteristic func. is defined for all $t \in \mathbb{R}$).

(2). $\varphi_X(0) = 1$. $|\varphi_X(t)| \leq 1$.

(3). $\varphi_X(-t) = \overline{\varphi_X(t)}$.

(4). $\forall t$ $\varphi_X(t) \in \mathbb{R} \Leftrightarrow$ Distribution of X is symmetric $P_X = P_{-X}$.

(5). $\varphi_{aX+b}(t) = e^{itb} \varphi_X(at)$.

(6). $\varphi_X(t)$ uni. cont. on $\mathbb{R}$.
$\forall \varepsilon > 0.\ \exists \delta > 0 : \forall |h| < \delta, \forall t\ |\varphi(t+h) - \varphi(t)| < \varepsilon$ (i.e. $\sup_{t \in \mathbb{R}} |\varphi(t+h) - \varphi(t)| \to 0$).

(7). Non-negative definiteness.
$\forall n \in \mathbb{N}$. $\forall t_1, t_2, \dots, t_n$ $\forall c_1, \dots, c_n \in \mathbb{C}$. $\sum_{j,k=1}^{n} c_j \varphi(t_j - t_k) \overline{c_k} \geq 0$.

Pf: $\sum_{j,k=1}^{n} c_j \varphi(t_j - t_k) \overline{c_k} = \sum c_j \mathbb{E}(e^{it_jX} \cdot e^{-it_kX}) \overline{c_k} = \mathbb{E}(\sum c_j e^{it_jX} \cdot \overline{c_k e^{it_kX}})$
$= \mathbb{E}(\sum_{j=1} c_j e^{it_jX} \cdot \overline{\sum_{k=1} c_k \cdot e^{it_kX}}) = \mathbb{E}(S_n \cdot \overline{S_n}) = \mathbb{E}(Re^2 S_n + Im^2 S_n) \geq 0$.

<u>Thm 4.11.</u> (Bochner - Khinchin).

Any continuous non-negative definite func. $\varphi : \mathbb{R} \to \mathbb{C}$. s.t. $\varphi(0) = 1$.
is the characteristic function of some random variable.

9. Laplace Transform function (for cont. r.v. X).

<u>Def 4.19</u> $\mathcal{L}_X(s) = \mathbb{E}[e^{-sX}] = \int_0^{\infty} e^{-sx} f_X(x)\,dx$. $s > 0$.

<u>Prop. 4.12</u> (1) $L_{aX+b}(s) = e^{-bs} L_X(as)$. $a, b \in \mathbb{R}$.

(2). $L_X(0) = 1$ $\qquad L_X(s) \leq 1$.

(3) $\text{sign}(L_X^{(k)}(s)) = (-1)^k$

(4). $L_X^{(k)}(0) = (-1)^k \mathbb{E}(X^k)$.

(5). If $X \in \mathbb{Z}_{\geq 0}$. then $G_X(e^{-u}) = L_X(u)$. $u \geq 0$.

△ Characteristic function $\longrightarrow$ moment.

$\varphi'(0) = i\,E(X)$

$\varphi''(0) = -E(X^2)$.

$\cdots$

$\varphi^{(k)}(t) = i^k E(X^k e^{itX})$. $\qquad \varphi^{(k)}(0) = i^k E(X^k)$.

(holds. as long as $E(X^n)$ exists)

<u>Prop 4.13</u>. Let the first n-th moments exist, i.e. $E(|X|^n) < \infty$ $n \in \mathbb{N}$.

Then $\forall k = 1, \dots, n$ $\varphi^{(k)}(t) = i^k E(X^k e^{itX})$. and $E(X^k) = (-i)^k \varphi^{(k)}(0)$.

pf by induction: for $(k+1)$-th derivative

$$\frac{\varphi_X^{(k)}(t+h) - \varphi_X^{(k)}(t)}{h} = \frac{i^k}{h}\left(\mathbb{E}(X^k e^{i(t+h)X}) - \mathbb{E}(X^k e^{itX})\right) = \frac{i^k}{h}\mathbb{E}\left(X^k(e^{i(t+h)X} - e^{itX})\right)$$

by $\left|e^{iy} - 1 - iy - \frac{y^2}{2} + \frac{iy^3}{6} - \dots + \frac{i^k y^k}{k!}\right| \le \frac{|y|^{k+1}}{(k+1)!}$ $\quad$ ($k=0$. $|e^{iy}-1| \le |y|$).

Thus $|X^k e^{itX}(e^{ihX} - 1)| \le |X|^{k+1} \cdot |h|$

by Lebesgue dominated conv. thm. $\frac{i^k}{h}\mathbb{E}(X^k e^{itX}(e^{ihX} - 1)) \xrightarrow[h\to 0]{} i^{k+1}\mathbb{E}(X^{k+1} e^{itX})$.

期望(实质是积分)与极限换序.

<u>Prop 4.14</u> suppose. for some $k \in \mathbb{N}$. the derivative $\varphi_X^{(2k)}(0)$ exists. $\varphi_X(t) = \mathbb{E}(e^{itX})$.

Then $\mathbb{E}X^{2k} < \infty$ and $\mathbb{E}X^{2k} = (-1)\,\varphi_X^{(2k)}(0)$

△ 对奇数2k+1. $\varphi_X^{(2k+1)}(0)$ 存在 $\not\Rightarrow$ $\mathbb{E}X^{2k+1} < \infty$, 但 $\mathbb{E}X^{2k} < \infty$ 成立

<u>Thm 4.15</u> Let X be a random variable. $\varphi_X(t) = \mathbb{E}e^{itX}$, $P(X=a) = P(X=b) = 0$, $a<b$. Then $P(X \in [a,b]) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-iat} - e^{-ibt}}{it}\,\varphi_X(t)\,dt$.

<u>coro</u>. X, Y has same distribution $\Leftrightarrow$ $F_X = F_Y$ / $f_X = f_Y$ $\forall x$. $\Leftrightarrow$ $\varphi_X = \varphi_Y$.

10. Moment generating function.

<u>Def 4.20</u>. $M_X(t) = \mathbb{E}[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$. if it conv.

<u>Remark</u>: if $M_X(t)$ exists in some $U_\varepsilon(0)$. $\mathbb{E}[X^n] = M_X^{(n)}(0)$ (所有原点矩可通过导数获得).

§ 5. Random Vector. (multi. dim. random. vector)

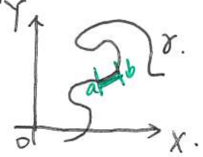
<u>Def 5.1.</u> (random vector). Let $(\Omega, \mathcal{F}, \mathbb{P})$. be a probability space. A random vector of dimension n is a mapping. $X : \Omega \to \mathbb{R}^n$.

$X(\omega) = (X_1(\omega), X_2(\omega), \dots, X_n(\omega))^T$. each X_i is r.v. i.e. meas. func. $X_i : \Omega \to \mathbb{R}$.

<u>Def 5.2.</u> (singular cont. distri.).: $\exists A \subseteq \mathbb{R}^n$: $\mu(A) = 0$ (→ Lebesgue meas.), $P_X(A) = 1$ (>0). (若>0. 一定∃$A' \supseteq A$. $P_X(A') = 1$.)

("概率"集中在更低维度的区域内).

e.g. in $\mathbb{R}^2$. (X, Y). distribute (concentrated on) a curve γ in $\mathbb{R}^2$.



△ 这是多维中介于 discrete / abs. cont. 之间的情况.

对 abs. cont. 取线段 ab. $\mathbb{P} = 0$. (面积 $= 0$)

对 singular cont. $\mathbb{P} = \frac{|ab|}{|\gamma|}$

<u>Def 5.3.</u> (Joint distribution func.).

The joint distribution func. of random vector $\vec{X} = (X_1, \dots X_n)^T$ is

$$F_{\vec{X}}(x_1, \dots x_n) = \mathbb{P}(X_1 < x_1, \dots X_n < x_n).$$

<u>Prop. 5.1</u> (Prop. for JCDF). (for simplicity. (ξ_1, ξ_2))

(1). For any x_1, x_2. $0 \le F_{\xi_1, \xi_2}(x_1, x_2) \le 1$.

(2). $F_{\xi_1, \xi_2}(x_1, x_2)$. is non-decreasing in each vector (x_1, x_2). → $x_1 \uparrow$ $F \uparrow$, $x_2 \uparrow$ $F \uparrow$.

(3). $\forall i$ $\lim\limits_{x_i \to -\infty} F_{\xi_1, \dots, \xi_n}(x_1, x_2, \dots x_n) = 0$.

the (double) limit exists. $\lim\limits_{x_1 \to +\infty} \lim\limits_{x_2 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = 1$.

(4). $F_{\xi_1, \xi_2}(x_1, x_2)$ is left-cont. in each coordinate vector (x_1, x_2).

(5). derive "marginal distribution func." of ξ_1 and ξ_2.

$$\lim_{x_1 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_2}(x_2). \qquad \lim_{x_2 \to +\infty} F_{\xi_1, \xi_2}(x_1, x_2) = F_{\xi_1}(x_1)$$

(6) Non-negative. if $\forall a < b$ $\forall c < d$

$$P(a < X \le b, c < Y \le d) = F(b,d) - F(a,d) - F(b,c) + F(a,c) \ge 0.$$

△ 性质 1,2,3,4,6 可构成等价条件、满足(1)~(6) 一定∃(X, Y). $(X, Y) \sim F_{\xi_1, \xi_2}$

△ 在 check 性质时，不考虑 5. 没有可 check 的.

Def 5.4. (1) discrete case - Joint Probability Mass Function. (JPMF).

$$P_X(x_1, \cdots, x_n) = \mathbb{P}(X_1 = x_1, \dots, X_n = x_n). \quad (\sum P_X(x_1 \dots x_n) = 1 \text{ - normalization})$$

(2) Abs. cont. case - Joint Probability Density Function. (f_X).

$$\exists f_X \text{ s.t. } F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_X(t_1, \dots, t_n)\, dt_1 \dots dt_n.$$

Prop. 5.2 (1) $f_X(\cdot) \geq 0$.

(abs. cont.) (2). $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(t_1, \dots, t_n)\, dt_1 \dots dt_n = 1.$

(3). $f_X(x_1 \dots x_n) = \dfrac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n).$

Def 5.5. (Marginal distribution)

The distribution of a single component / a subset of components.

(1) discrete case $P_{X_1}(x_1) = \sum_{(x_2, \dots x_n)} P_X(x_1, x_2, \dots x_n)$

(2) Abs. cont. case: $f_{X_1}(x_1) = \int_{\mathbb{R}^{n-1}} f_X(x_1, x_2, \dots, x_n)\, dx_2, \dots dx_n$

Def 5.6. (Independence) Random variables $X_1, \dots X_n$ are jointly independent if

$$\forall B_1, \dots B_d \in \mathcal{B}(\mathbb{R}). \ \mathbb{P}(X_1 \in B_1, \dots X_d \in B_d) = P(X_1 \in B_1) \times \dots \times P(X_d \in B_d)$$

(1) discrete: $\mathbb{P}(X_1 = x_1 \dots X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$

(2) abs. cont. $f_{X_1, \dots X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Def 5.7 (Conditional Distribution)

(1) For discrete X_1, X_2. $\mathbb{P}(X_1 = x_1 \mid X_2 = x_2) = \dfrac{\mathbb{P}(X_1 = x_1, X_2 = x_2)}{\mathbb{P}(X_2 = x_2)}.$

(2) For abs cont. (conditional density). if f_X exists, then the conditional density of X_1 given X_2 is $f_{X_1|X_2}(x_1 | x_2) = \dfrac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}$

Def 5.8 (Moment)

(1) For discrete $E(g(X, Y)) = \sum_{i=1,\dots\infty; j=1,\dots\infty} g(x_i, y_j)\, P(X = x_i, Y = y_j).$

(2) For abs. cont. $E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\, f_{X,Y}(x, y)\, dx\, dy.$

Def 5.9. (Moment of a Random Vector)

(1) Expectation Vector. $\mathbb{E}(X) = (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])^T$

(2) Covariance Matrix. $\Sigma = Cov(X) = \mathbb{E}([X - \mathbb{E}(X)][X - \mathbb{E}(X)]^T)$.

it's symmetric and positive semi-definite. ($\forall \vec{\alpha} \in \mathbb{R}^n$. $\langle \vec{\alpha}, \Sigma\vec{\alpha} \rangle \geq 0$)

· diagonal: variance of components. · Off-diagonal: covariance $Cov.(X_i, X_j)$

Pf: of p. semi-definite. $\langle \vec{\alpha}, \Sigma\vec{\alpha} \rangle = \sum_{i,j} \alpha_i Cov(X_i, X_j) \alpha_j = \sum_{i=1}^{n} \alpha_i \left(\sum_{j=1}^{n} \alpha_j \, cov(X_i, X_j) \right)$.

$= cov.\left(\sum_{i=1}^{n} \alpha_i X_i, \sum_{j=1}^{n} \alpha_j X_j \right) \overset{i \Leftrightarrow j}{=} D\left(\sum \alpha_i X_i \right) \geq 0$.

Def 5.10 (Linear Transformation)

Let A be $m \times n$ matrix. $X \in \mathbb{R}^n$. then $Y = AX + b$ is also a random vector.

with. $\mathbb{E}[Y] = A\,\mathbb{E}[X] + b$ $\quad Cov(Y) = A\, cov(X)\, A^T$

Covariation and Correlation.

Def 5.11. (covariance) $cov(X, Y) = \mathbb{E}((X - \mathbb{E}X)(Y - \mathbb{E}Y)) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$.

Remark: (1) X, Y are uncorrelated if $cov(X, Y) = 0$ (i.e. $\mathbb{E}(XY) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$).

(2) uncorrelated $\not\Leftrightarrow$ independent. (counter e.g. $X \sim Unif[-1,1]$. $Y = X^2$).

Prop 5.3. (1) symmetry $cov(X, Y) = cov(Y, X)$.

(2) $D(X) = cov(X, X)$.

(3). Linearity. $cou(aX + b, Y) = a\, cov(X, Y)$.

(4). $D(X \pm Y) = D(X) + D(Y) \pm 2cov(X, Y)$

(5) $cov(X, Y) \leq \sqrt{D(X) \cdot D(Y)}$.

Def 5.12. (correlation) $\quad corr(X, Y) = \rho_{X,Y} = \dfrac{Cov.(X, Y)}{\sqrt{D(X) D(Y)}}$

Prop 5.4. (1). $corr(X, Y) = corr(Y, X)$.

(2) $Corr.(\cdot, \cdot) \in [-1, 1]$.

measure the strength of linear dependence between variable.

(3). $corr(aX + b, Y) = sgn(a) \cdot corr(X, Y)$.

(4) $|\rho_{X,Y}| = 0$. $\Rightarrow \exists\, a, b \in \mathbb{R}: Y = aX + b$.

lemma. $|\rho_{X,Y}| = 1 \Rightarrow \exists\, a, b \in \mathbb{R}: Y = aX + b$ (linear dependence)

(5) X, Y uncorrelated $\Leftrightarrow \rho_{X,Y} = 0$.

Special Multivariate Distribution.

(1) Let X_i - independent identically distributed random variable with $F_X(\cdot)$

Consider sort(X_i) : $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$; the distribution of $X_{(i)}$. $\forall i \in 1, \dots n$.

$$F_{X_{(i)}}(t) = \sum_{i \le j \le n}^{n} C_n^j (F(t))^j (1-F(t))^{n-j}. \qquad f_{X_{(i)}}(t) = \frac{n!}{(i-1)!(n-i)!} [F(t)]^{i-1} [1-F(t)]^{n-i} f(x)$$

至少有 i 个 $X < t$.

在PDF推导中是恰好某个分布在$(t, t+\Delta t)$内.

△ 一般用2个角度推PDF和CDF. 求导时很难, 猜定相消.

(2). For independent X, Y.
$$\begin{cases} F_{max}(t) = P(\max(X,Y) < t) = F_X(t) \cdot F_Y(t) \\ F_{min}(t) = 1-(1-F_X(t))(1-F_Y(t)) \end{cases}$$

(3). Multivariate Normal Distribution. $X \sim N(\mu, \Sigma)$.

$\mu \in \mathbb{R}^n$ - mean vector. $\Sigma \in \mathbb{R}^{n \times n}$ - covariance matrix.

PDF: $$f(\vec{x}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})\right).$$

<u>property</u>: (1) $(X, Y) \sim N(\vec{\mu}, \Sigma)$: X, Y uncorrelated $\Leftrightarrow$ X, Y independent.

← 注意这仅是2维结论

(2) if $\vec{X} \sim N(\vec{\mu}, \Sigma) \Rightarrow A\vec{X}+b \sim N(A\vec{\mu}+b, A\Sigma A^T)$. A - size $m \times n$.

$1 \le m \le n$ 即可 (Jm为所求分布的维数).

For $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho_{x,y})$.

$$f_{x,y}(X,Y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

△ 注意 property (2) 的运用. 在已知 Σ. μ 的情况下. (Σ 的对角为 X_i 方差).

例: 令 $A = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}^T$ 可求 $x_1 + x_2$ 分布.

令 $A = \begin{pmatrix} 3 & 2 \\ -2 & 0 \\ 0 & 1 \end{pmatrix}^T$ 可求 $(3X_1 - 2X_2, 2X_1 + X_3)$ 分布.

△ 若 $X_i \sim N(\mu_i, \sigma_i^2)$. X_i 独立. $\sum X_i \sim N(\sum \mu_i, \sum \sigma_i^2)$

(正态分布对独立变量具有可加性).

Transformation of random variable.

§ Unidimensional. Function

Let some random vector $\mathbb{X}$ is given by PDF. $f_{\mathbb{X}}(\cdot)$. ← vector function.

If $Y = g(\vec{X})$. then $F_Y(y) = P(g(\vec{X}) \le y)$.

express this as a multivariate integral: $F_Y(y) = \int_{g(\vec{x}) \le y} f_{\mathbb{X}}(\vec{x})\, d\vec{x}$.

The distribution of the sum of r.v. (convolution).

For random variable X and Y. with JPDF $f_{X,Y}(x,y)$; consider $Z = X + Y$.

1) $X.Y$ independent: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$.

$$f_Z(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx.$$

2). $X.Y$ not independent.

$$F_Z(z) = \iint_{x+y<z} f(x,y)\, dx\, dy = \iint_{-\infty}^{z-x} f(x,y)\, dx\, dy = \iint_{-\infty}^{z} f(x, t-x)\, dt\, dx = \iint_{-\infty}^{z} f(t-y, y)\, dt\, dy$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)\, dy = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\, dx.$$

generalized convolution formula.

3). $X.Y$ discrete (without PDF).

$$P(Z = z) = \sum_{k=-\infty}^{+\infty} P(X = z-k) \cdot P(Y = k).$$

The distribution of the quotient of r.v.

For r.v. $X.Y$, with. JPDF $f_{X,Y}$; consider the quotient $Z = \frac{X}{Y}$.

$$F_Z(z) = \mathbb{P}\left(\frac{X}{Y} < z\right) = \int_0^{\infty} \int_{-\infty}^{zx} p(x,y)\, dy\, dx + \int_{-\infty}^{0} \int_{zx}^{\infty} p(x,y)\, dy\, dx$$

if $X.Y$ independent. $F_Z(z) = \int_0^{\infty} F_Y(xz) f_X(x)\, dx + \int_{-\infty}^{0} (1 - F_Y(xz)) f_X(x)\, dx.$

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_0^{\infty} x f_X(x) f_Y(xz)\, dx - \int_{-\infty}^{0} x f_X(x) f_Y(xz)\, dx.$$

if $X.Y$ not independent. $f_Z(z) = \int_0^{\infty} x f_{X,Y}(x, xz)\, dx - \int_{-\infty}^{0} x f_{X,Y}(x, xz)\, dx.$

8 Multidimensional Function

Let $\vec{X} = (X_1, X_2, \dots, X_n)$ be a random vector in $\mathbb{R}^n$. and let $g: \mathbb{R}^n \to \mathbb{R}^n$. be meas. func. Then the image $\vec{Y} = g(\vec{X}) \in \mathbb{R}^m$ is also a random vector.

If $\vec{X}$ has JPDF $f_{\vec{X}}(x_1, x_2, \dots, x_n)$. and if g is bijective and differentiable with inverse g^{-1}, then the density of Y is given by the change of variables formula:

$$f_Y(y_1, y_2, \dots, y_m) = f_{\vec{X}}(g^{-1}(\vec{y})) \cdot |\det J_{g^{-1}}(y)|.$$

Reminder: $J_g(\vec{x}) = \begin{bmatrix} \frac{\partial Y_1}{\partial X_1} & \cdots & \frac{\partial Y_1}{\partial X_n} \\ \vdots & & \\ \frac{\partial Y_n}{\partial X_1} & \cdots & \frac{\partial Y_n}{\partial X_n} \end{bmatrix}$

△例：二维情况：作变量代换 $\begin{cases} x = x(u,v) \\ y = y(u,v) \end{cases}$ $\quad J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$

$$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v)) \cdot |J|.$$

Conditional mathematical expectation.

<u>Def.</u> X - r.v. $E(X) < \infty$. and $\mathcal{F}$ - σ-algebra. then $Z := E(X|\mathcal{F})$. is such r.v.:

(1) Z is $\mathcal{F}$-measurable

(2) $\forall B \in \mathcal{F}$ $\quad E(X \cdot \mathbb{1}_B) = E(Z \cdot \mathbb{1}_B)$. ← B的特征函数. $1_B = \begin{cases} 1, & x \in B. \\ 0, & x \notin B. \end{cases}$

<u>Lemma</u> $E(X|\mathcal{F})$ ~~is~~ exists and is unique

<u>Lemma.</u> $E(X|Y) =: g(Y)$.

discrete case: $\quad g(y) = \dfrac{E(X \cdot 1(Y=y))}{P(Y=y)} = \sum x_i (P(X=x_i | Y=y))$.

abs. cont. case: $\quad g(y) = \int_{R} x \dfrac{f_{X,Y}(x,y)}{f_Y(y)} dx$

<u>Property</u>

(1) $E(\text{const} | \mathcal{F}) = \text{const}$.

(2) $E(aX|\mathcal{F}) = aE(X|\mathcal{F})$.

(3) $E(X+Y|\mathcal{F}) = E(X|\mathcal{F}) + E(Y|\mathcal{F})$.

(4). $E(E(X|\mathcal{F})) = E(X)$.

(5) If $X \in \mathcal{F}$. $E(X|\mathcal{F}) = X$.

(6). If X is independent of $\mathcal{F}$. $\quad E(X|\mathcal{F}) = X$.

(7). If $X \leq Y$. then $E(X|\mathcal{F}) \leq E(Y|\mathcal{F})$.

(8). If $E|X| < \infty$. then $|E(X|\mathcal{F})| \leq E(|X| \,|\, \mathcal{F})$.

(9). If Y is $\mathcal{F}$-measurable. then $E(XY|\mathcal{F}) = Y \cdot E(X|\mathcal{F})$.

(10). If $\mathfrak{A} \subset \mathfrak{B} \subset \mathcal{F}$. then $E(\xi|\mathfrak{A}) = E(E(\xi|\mathfrak{B})|\mathfrak{A})$.

<u>Remark</u>: conditional expectation w.r.t. set B:

$$E(\mathbb{1}_A | B) = \frac{E(\mathbb{1}_A \cdot \mathbb{1}_B)}{P(B)}. \qquad E(\xi | B) = \frac{E(\xi \cdot \mathbb{1}_B)}{P(B)}$$

conditional expectation w.r.t. σ-algebra $\mathfrak{A}$.; $\forall B \in \mathfrak{A}$. $B = \bigcup_{k=1}^{m} B_k$

$$\int_B E(X|\mathfrak{A})\, dP = \sum_{k=1}^{m} E(X|B_k) \cdot P(B_k) = \sum_{k=1}^{m} E(X \cdot \mathbb{1}_{B_k}) = \int_B X\, dP.$$

Probability inequality for r.v.

1. Jensen's inequality

X - r.v. $\mathbb{E}X < \infty$. $\varphi(x)$ - convex. $\varphi(\mathbb{E}(X)) \leq \mathbb{E}(\varphi(X))$.

<u>Remark</u>: (1) φ - convex. "$\geq$"

(2) For conditional E. correct too.

<u>e.g.</u> (1) $\varphi = x^2$. $(E(X))^2 \leq E(X^2)$ .. (i.e. $D(X) \geq 0$).

2. Markov's inequality.

X - r.v. $X \geq 0$. $\mathbb{E}X < \infty$. $\forall a > 0$. $P(X \geq a) \leq \dfrac{E(X)}{a}$.

<u>Coro</u> (1) $\varphi(x) \geq 0$. increasing and $\mathbb{E}[\varphi(X)] < \infty$. then

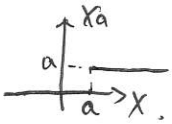
$\forall a > 0$. $P(X \geq a) = P(\varphi(X) \geq \varphi(a)) \leq \dfrac{\mathbb{E}[\varphi(X)]}{\varphi(a)}$

(2) if $X \in \mathbb{R}$. others same. $P(|X| \geq a) \leq \dfrac{E(|X|)}{a}$.

3. Chebyshev's Inequality.

X - r.v. $\mu = \mathbb{E}(X) < \infty$. $\sigma^2 = \mathbb{D}(X)$ $\forall \varepsilon > 0$. $P(|X - \mu| \geq \varepsilon) \leq \dfrac{\sigma^2}{\varepsilon^2}$

<u>Pf</u>: 2). $X_a := \begin{cases} a; & X \geq a > 0 \\ 0, & \text{other} \end{cases}$



$E(X_a) = a \cdot P(X \geq a)$

$P(X \geq a) = \dfrac{EX_a}{a} \leq \dfrac{EX}{a}$

1). discrete.: $\varphi(\sum x_i p_i) \leq \sum p_i \varphi(x_i)$ $\quad \sum p_i = 1$.

abs. cont. $\varphi(\int x f(x)\, dx) \leq \int f(x) \varphi(x)\, dx$. $\quad \int f(x)\, dx = 1$.

4. Markov. (for sum of r.v.).

$S_n = X_1 + X_2 + \dots + X_n$. ($X_i$ - r.v. not necessarily to be independent) $S_n \geq 0$ almost sure

$\forall a > 0$. $\mathbb{P}(S_n \geq a) \leq \frac{\mathbb{E}(S_n)}{a}$

5. Chebyshev. (for sum of r.v.)

S_n. same as above and $\mathbb{E}(S_n) < \infty$.

$\forall \varepsilon > 0$. $\mathbb{P}(|S_n - \mathbb{E}[S_n]| \geq \varepsilon) \leq \frac{D(S_n)}{\varepsilon^2}$

For independent r.v. we have $D(S_n) = \sum_{i=1}^{n} D(X_i)$

(with not independence $D(S_n) = \sum D(X_i) + 2\sum_{1 \leq i < j \leq n} Cov.(X_i, X_j)$).

6. Conditional Math. Expectation of sum of r.v.

X_i - i.i.d. r.v. $\sigma_s = \sigma(X_1, X_2, \dots X_s)$. $S_k = \sum_{i=1}^{k} X_i$ $EX_i = a$.

$$E(S_k | \sigma_s) = \begin{cases} E(X_1 + \dots + X_k) \mid \sigma(X_1 + \dots + X_k + X_{k+1} + \dots + X_s)) = S_k. & k \leq s. \\ E(X_1 + \dots + X_s + \dots + X_k) \mid \sigma(X_1, \dots X_s)) = S_s + a(k-s)\ & k > s. \end{cases}$$

Type of Convergence of r.v.

$\{X_n\}$ - sequence of r.v. X - r.v.

(1). Almost sure converge

$X_n \xrightarrow{a.s.} X$: $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

(2) Converge in probability.

$X_n \xrightarrow{\mathbb{P}} X$: $\forall \varepsilon > 0$. $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$.

(3) Converge in L^p (Mean convergence)

$X_n \xrightarrow{L^p} X$: $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$. $p \geq 1$.

(4). Converge in Distribution. (Weak conv.)

$X_n \xrightarrow{d} X$: $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all cont. point of F_X.

$$
\begin{tikzcd}
(1)^{a.s.} \arrow[r] & (2)^{\mathbb{P}} \arrow[d] \\
(3)_{L^p} \arrow[ur] & (4)_d
\end{tikzcd}
$$

△除了给出的"→"其余在一般情况均不成立.

Borel-Cantelli Lemma. (for events).

Let (A_n) - sequence of events. in prob. space $(\Omega, \mathcal{F}, \mathbb{P})$.

(1). First Borel-Cantelli lemma.

If. $\sum_{n=1}^{\infty} P(A_n) < +\infty$. then $\mathbb{P}(A_n$ occurs infinitely often$) = 0$.

(i.e. $\mathbb{P}(\limsup_{n\to\infty} A_n) = 0$. $\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$.) ← 无论n多大, 总有$A_m (m>n)$发生.

(2) Second Borel-Cantelli lemma

If $\sum_{n=1}^{\infty} P(A_n) = +\infty$. and A_n are independent.

then $\mathbb{P}(A_n$ occurs infty often$) = 1$.

Borel - Cantelli Lemma. (for a.s. conv.).

$\{X_n\}$ - sequence of r.v. X - r.v. $\forall \varepsilon > 0$. $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) < +\infty$ or $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| < \varepsilon) = +\infty$

Then $X_n \xrightarrow[n\to+\infty]{a.s} X$.

Pf: by event case.

△注意."$<\varepsilon$"或许更符合极限/收敛定义.但实际情景中"$\geq\varepsilon$"才是更常用的.
因为可以直接应用 Chebyshev.不等式.

define $A_n(\varepsilon) = \{\omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}$.

if $\sum \mathbb{P}(A_n) < +\infty \Rightarrow \mathbb{P}(\limsup_{n\to\infty} |X_n(\omega) - X(\omega)| \geq \varepsilon) = 0 \Rightarrow \mathbb{P}(\lim_{n\to\infty} |X_n(\omega) - X(\omega)| \geq \varepsilon) = 0$

$\Rightarrow \mathbb{P}(\lim_{n\to\infty} |X_n(\omega) - X(\omega)| < \varepsilon) = 1 \Rightarrow \mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

△ the converse not true in general.

If $\forall \varepsilon > 0$. $\sum \mathbb{P}(|X_n - X| \geq \varepsilon) = +\infty$. and $A_n(\varepsilon)$ independent $\Rightarrow \mathbb{P}(|X_n - X| \geq \varepsilon$ infty often$) \doteq 1$.

$\Rightarrow X_n \overset{a.s.}{\not\to} X$.

△大数定律的目标.都是描述样本均值 $\frac{1}{n}\sum X_i$ 随 $n\to\infty$ 趋近于总体均值 $\mu = \mathbb{E}[X_i]$.但收敛方式/初始条件各不相同.

Law of Large Number

(Khinchin / Khintchine 辛钦.叠咔擦咔擦~)

1. Hinčin's Law of Large Number. (Weak) - 弱大数定律的最一般形式

$X_1, \dots, X_n$ - i.i.d. r.v. $\mathbb{E}[X_i] = \mu$. $D(X_i) = \sigma^2$ $(\mu. \sigma < \infty)$.

→ 这个条件可以删去.(注意期望$<\infty \overset{\Leftarrow}{\not\Rightarrow}$ 方差$<\infty$).
辛钦的创新即将"有限方差"的假设.
弱化为"有限期望"

$\forall \varepsilon > 0$. $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \varepsilon) \xrightarrow[n\to\infty]{} 0$

(equivalently. $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} \mu$).

2. Lévy's Law of Large Number. (Weak)

$X_1, X_2, X_3 \cdots$ i.d. r.v. (not independent necessity!). $\mathbb{E}[X_i] = \mu$. $\exists C$. s.t. $\mathbb{E}[X_i^2] \leq C$.

Then $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\mathbb{P}} \mu$.

3. Kolmogorov's Strong Law of Large Number

I) $\{b_n\}$. numberical sequence. with $b_n > 0$. and $b_n \nearrow$. $X_1, X_2, \dots$ independent r.v. (not necess. identical!). s.t. $\mathbb{E}X_i^2 < \infty$ (thus. variance exist).

Then. $\sum_n \frac{DX_n}{b_n^2} < \infty$, which implies. $\frac{S_n - \mathbb{E}S_n}{b_n} \xrightarrow[n\to\infty]{a.s.} 0$. $\forall \varepsilon > 0. \mathbb{P}(|\frac{1}{n}\sum X_i - \mathbb{E}X_i| \leq \varepsilon) = 1$

$\Uparrow$

II) $\{X_n\}$ - i.i.d. r.v. s.t. $\mathbb{E}|X_1| < \infty$. Then $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{a.s.} \mathbb{E}X_1$. with probability 1.

△强弱区间体现在:弱仅说明.对足够大 n^*. $\frac{1}{n}\sum X_i \to \mathbb{E}X_1$. 不保证 $\forall n > n^*$. $\frac{1}{n}\sum X_i$ 均停在 $\mathbb{E}X_1$ 附近
强说明. $|\frac{1}{n}\sum X_i - \mathbb{E}X_i|$ 不能无限次(即概率为0).离开0.

强 - 长期平均行为. 弱 - 总体趋势.

Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. r.v. $\mathbb{E}[X_i] = \mu$. $\mathbb{D}[X] = \sigma^2 > 0$. $S_n = \sum_{k=1}^{n} X_k$

normalized sum.

We have. $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to+\infty]{d} N(0,1)$.

△ σ 是单个 r.v. 的标准差. (如对 Bernoulli trial. 应取 $\sqrt{pq}$ 而非 $\sqrt{npq}$.

△ 用于估算: $P(S_n \in (a,b)) \approx \Phi\left(\dfrac{b - E(S_n)}{\sqrt{D(S_n)}}\right) - \Phi\left(\dfrac{a - E(S_n)}{\sqrt{D(S_n)}}\right)$

$\Phi(\cdot)$ - distribution func. of $N(0,1)$

Complentary (of LLN)

Chevbyshev. WLLN.

$\{X_n\}$ - jointly unrelated. . r.v. if $\exists C$. for all i. $D[X_i] \le C$.

$\mathrm{Cov}(X_i, X_j) = 0. \forall i \ne j$. 对应 $D(\frac{1}{n}\sum X_i) = \frac{1}{n^2}\sum DX_i$. 弱于 independence

Then $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

Markov. WLLN.

$\{X_n\}$ - r.v. $\frac{1}{n^2} D\left(\sum_{i=1}^{n} X_i\right) \to 0$. Then. $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{P} \frac{1}{n}\sum_{i=1}^{n} E(X_i)$

△ 注意在这2个大数定律中. 独立/同分布 假设均是可以被舍去的.