Probability Theory. by Prof. Jieliang Hong.

Notes taken by Kai Chen.

Week 1.1 Lecture 1.

Theorem 1.1 (The basic principle of counting).

Suppose there are 2 experiments.

If exp. 1 can result in m possible outcomes and exp. 2 can result in n possible outcomes, then together there are $m \times n$ possible outcomes.

Ex. How many functions defined on $\{1, 2, \cdots, n\}$ are possible if each functional value is 0 or 1? 2^n.

Def. Permutation (排列) means the different ordered arrangment of objects.

Ex. Choose 4 students consecutively without replacement (reoccurrence?) from 80 stu.s

$80 \times 79 \times 78 \times 77$ possible outcomes.

Theorem 1.2. Suppose we have n objects. then

there are $n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 1$ permutations. (factorial.)

Ex. How many different arrangements can be formed from the letters BOY?

$3!$.

BABY?

$4!/2!$

Theorem 1.3. Suppose there are n objects.

of which n_1 are alike, n_2 are alike, $\cdots$, n_r are alike.

Then there are $\dfrac{n!}{n_1! n_2! \cdots n_r!}$ different permutations.

Ex. How about PEPPER? $\dfrac{6!}{3! \cdot 2!}$.

Def. Combination:

Choose 4 students from 80 students.

$$\binom{80}{4} = \frac{80!}{76!\,4!}.$$

Theorem 1.4. If we choose r objects from a total of n different objects at a time, then the # of possible combinations is $\binom{n}{r}$,

where $\binom{n}{r} = \dfrac{n!}{(n-r)!\,r!} \left(= \binom{n}{n-r} \right)$.

Ex. A student has to sell 2 books from 6 math, 7 science and 4 economics.

How many choices are possible if

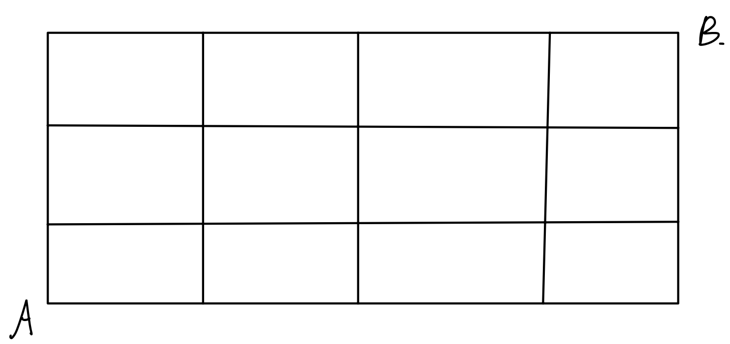
(a) both books are to be on the same objects.

$$\binom{6}{2} + \binom{7}{2} + \binom{4}{2}.$$

(b). the books are to be on different objects.

$$\binom{6}{1}\binom{7}{1} + \binom{6}{1}\binom{4}{1} + \binom{7}{1}\binom{4}{1}.$$

Ex.



from A to B

3 upwards.

4 rightwards.

$\binom{7}{3}$ choices.

Theorem 1.5 (Binomial Theorem)

For any positive integer $n \geq 1$,

$0^0 = e^{0\ln 0} = e^0 = 1.$

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

$$= \binom{n}{0} x^0 y^{n-0} + \binom{n}{1} x^1 y^{n-1} + \cdots + \binom{n}{n} x^n y^{n-n}.$$

Proof by induction.

① (Basis step.) The case for $n=1$ holds.

② (Inductive step.) Assume that $n=k$ holds for some $k \geq 1$.

Then $n=k+1$ holds.

By induction, the case holds for all $n \geq 1$.

Pf of theorem 1.5.

① The case holds for $n=1$.

② Assume that

$$(x+y)^k = \sum_{j=0}^{k} \binom{k}{j} x^j y^{k-j}.$$

we need to prove that

$$(x+y)^{k+1} = \sum_{i=0}^{k+1} \binom{k+1}{i} x^i y^{k+1-i}.$$

Notice that

$$(x+y)^{k+1} = (x+y)\cdot(x+y)^k = (x+y)\cdot\left(\sum_{j=0}^{k} \binom{k}{j} x^j y^{k-j}\right)$$

$$= \sum_{j=0}^{k} \binom{k}{j} x^{j+1} y^{k-j} + \sum_{j=0}^{k} \binom{k}{j} x^j y^{k+1-j}.$$

$$= \sum_{i=1}^{k+1} \binom{k}{i-1} x^i y^{k-i+1} + \sum_{i=0}^{k} \binom{k}{i} x^i y^{k+1-i}$$

$$= \binom{k}{0} x^0 y^{k+1} + \sum_{i=1}^{k} \left[\binom{k}{i-1} + \binom{k}{i}\right] x^i y^{k+1-i} + \binom{k}{k} x^{k+1} y^0$$

$$= \binom{k+1}{0} x^0 y^{k+1} + \sum_{i=1}^{k} \binom{k+1}{i} x^i y^{k+1-i} + \binom{k+1}{k+1} x^{k+1} y^0$$

$$= \sum_{i=0}^{k+1} \binom{k+1}{i} x^i y^{k+1-i}.$$

By induction, the case holds for all $n \geq 1$. $\square$.

$$\underbrace{\binom{k+1}{i}}_{i \text{ out of } (k+1)} = \underbrace{\binom{k}{i}}_{i \text{ out of } k} + \underbrace{\binom{k}{i-1}}_{(i-1) \text{ out of } k}$$

Week 2. Lecture 2.

(Probability Space) $\begin{cases} \text{Sample Space.} \\ \text{Events.} \\ \text{Probability Measure.} \end{cases}$

1. The sample space S is the set of all possible outcomes of an experiment.

Eg 1. If I toss a coin. $S = \{\text{head}, \text{tail}\}$.

Eg 2. Choose 1 student uniformly at random in a class of 80 students. Do this repeatedly for 4 times

$S = \{ (x_1, x_2, x_3, x_4).\ 1 \le x_i \le 80, \quad i = 1, 2, 3, 4 \}$

2. An event is a subset of the sample space S, denoted $E \subset S$

3. Set operation: Let E, F be sets.

(a). union $E \cup F = \{x : x \in E \text{ or } x \in F\}$.

(b). intersection $EF = E \cap F = \{x : x \in E \text{ and } x \in F\}$.

(c). complement. $E^c = \{x : x \notin E\}$.

(d). difference. $E - F = E \cap F^c = \{x : x \in E \text{ and } x \notin F\}$

* Venn diagram.

Let $E_1, E_2, \cdots$ be a sequence of events in S.

Define $\bigcup_{n=1}^{\infty} E_n = \{x :$ there exists some $n \ge 1$. s.t. $x \in E_n\}$.

$= \{x : \exists n \ge 1, \text{ s.t } x \in E_n\}$.

$\bigcap_{n=1}^{\infty} E_n = \{x : x \in E_n \text{ for all } n\}$.

Theorem. 2.1. (DeMorgan's Law) For each $n \ge 1$,

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c. \quad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c.$$

This also holds if $n = \infty$.

// [Negation of a statement : $\exists n \geq 1,\ x \in E_n$.

$\forall n \geq 1,\ x \notin E_n$.]

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c.$$

Proof: ① : $\forall x \in LHS$. $x \notin \bigcup_{i=1}^{n} E_i \Rightarrow \forall n \geq 1,\ x \notin E_n$

$\Rightarrow \forall n \geq 1,\ x \in E_n^c \Rightarrow x \in RHS.$

② Similar.

For any $x \in \bigcap_{n=1}^{\infty} E_n^c$, we have $\forall n \geq 1,\ x \in E_n^c$.

$\Rightarrow \forall n \geq 1,\ x \notin E_n \Rightarrow x \notin \bigcup_{n=1}^{\infty} E_n \Rightarrow x \in LHS$. $\square$.

4. Axiom of Probability : Let S be a sample space.

For each event E, the probability $P(E)$ satisfies:

(i) $0 \leq P(E) \leq 1$. (ii). $P(S) = 1$.

(iii) For any sequence of mutually exclusive events $E_1, E_2, \ldots$

(i.e. $E_i \cap E_j = \emptyset,\ \forall i \neq j$). $P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$.

5. If all outcomes in S are equally likely to occur, then for any event E,

$$P(E) = \frac{\#\text{ of outcomes in } E}{\#\text{ of outcomes in } S}.$$

Eg 3. Calculate the probability that at least two students are born on the same day in a class of 80 stus.

ans: $$\frac{365^{80} - \binom{365}{80} 80!}{365^{80}} \approx 99.991033\%.$$

Eg 4. Pocker.

hearts ♡ diamonds ◇ spades ♤ club ♧

A = Ace. J = Jack. Q = Queen. K = king.

[5 cards]. Straight 顺子. $\begin{cases} A2345 \\ \vdots \\ 10JQKA \end{cases}$ ~~JQKA2~~ X.

Flush. 同花. Straight Flush. 同花顺.

Royal Flush 皇家同花顺. (10 J Q K A).

Eg 5. If it is assumed that all $\binom{52}{5}$ poker hands are equally likely.
扑克牌手牌.
what is the probability of being dealt with

(a) a flush. $$\frac{\binom{4}{1}\binom{13}{5}}{\binom{52}{5}} \approx 0.198\%.$$

(b) a straight. $$\frac{\binom{10}{1} \times 4^5}{\binom{52}{5}} \approx 0.394\%.$$

(c). straight flush. $$\frac{10 \times 4}{\binom{52}{5}} \approx 0.00154\%.$$

(d). a royal flush. $\dfrac{1\times 4}{\binom{52}{5}} \approx 0.000154\%$

$P(E\cup F) = P(E)+P(F)$ if $E\cap F=\phi$.
$P(E\cup F) = P(E\cup F \bigcup_{n=1}^{\infty}\phi)$
$= P(E)+P(F)+\sum_{n=1}^{\infty}P(\phi)$
$= P(E)+P(F)$. ✓.

6. Some propositions for the probability.

(a). $P(E) = 1 - P(E^c)$.

Proof: $\underline{P(S) = P(S\cup \bigcup_{n=1}^{\infty}\phi) = P(S)+\sum_{n=1}^{\infty}P(\phi)}$.

$\Rightarrow \underline{0 = \sum_{n=1}^{\infty}P(\phi) \Rightarrow P(\phi)=0}$.

So $\underline{1 = P(S) = P(E\cup E^c \bigcup_{n=1}^{\infty}\phi)}$

$= P(E)+P(E^c)+\sum_{n=1}^{\infty}P(\phi) = P(E)+P(E^c)$. $\square$.

(b). If $E\subset F$, then $P(E)\le P(F)$.

Proof: $P(F) = P((F-E)\cup E) = \underbrace{P(F-E)}_{\geq 0} + P(E) \geq P(E)$. $\square$.

(c). $P(E\cup F) = P(E)+P(F)-P(E\cap F)$

Proof: $LHS = P(E\cup F) = P\Big((E-F)\cup(F-E)\cup(E\cap F)\Big) = P(E-F)+P(F-E)+P(E\cap F)$.

$RHS = P((E-F)\cup(E\cap F)) + P((F-E)\cup(E\cap F)) - P(E\cap F)$

$= P(E-F) + \cancel{P(E\cap F)} + P(F-E) + P(E\cap F) - \cancel{P(E\cap F)}$

$= LHS$. $\square$.

(d). Inclusion - Exclusion Identity

$$P\Big(\bigcup_{i=1}^{n}E_i\Big) = \sum_{i=1}^{n}P(E_i) - \sum_{i_1<i_2}P(E_{i_1}\cap E_{i_2}) + \sum_{i_1<i_2<i_3}P(E_{i_1}\cap E_{i_2}\cap E_{i_3})$$

$$- \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^{n} E_i\right).$$

Proof by induction.

Eg 5. Let $S=[0,+\infty)$. $\boxed{\text{Dirac mass.}}$

Define $P(E)=\begin{cases} 1, & \text{if } 0\in E. \\ 0, & \text{if } 0\notin E. \end{cases}$

Then, P satisfies the axiom of probability.

(a) $0\le P(E)\le 1$. (b). $P(S)=1$.

(c). Let $E_1, \cdots$ be a sequence of mutually exclusive events,

then ① If $\forall n\ge 1$, $0\notin E_n$, then

$$0\notin \bigcup_{n=1}^{\infty} E_n \Rightarrow P\left(\bigcup_{n=1}^{\infty} E_n\right)=0.$$

and $$\sum_{n=1}^{\infty} P(E_n)=0.$$

② If $\exists n_0\ge 1$, $0\in E_{n_0}$, then

$\forall m\ne n_0$, $0\notin E_m$. So $P\left(\bigcup_{n=1}^{\infty} E_n\right)=1$.

$$1=\sum_{n=1}^{\infty} P(E_n)=P(E_{n_0})+\sum_{m\ne n_0} P(E_m).$$

$$=1+0.$$

Week 3. Lecture 3.

Conditional Probability and Independence.

1. For two events E, F s.t. $P(E) > 0$, the conditional probability that F occurs given that E has occured is denoted by $P(F|E) = \dfrac{P(F \cap E)}{P(E)}$.

Eg 1. A class of 80 stu.s of which 10 are ♀. Choose 1 stu. uniformly at random. Repeat it twice. Given that $\underbrace{\text{at least one of them is male}}_{E}$. Find the probability that $\underbrace{\text{the other one is ♀}}_{F}$. the two stu.s can be the same person.

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{\frac{2 \times 10 \times 70}{80 \times 80}}{\frac{2 \times 10 \times 70}{80 \times 80} + \frac{70 \times 70}{80 \times 80}} = \frac{2 \times 10 \times 70}{\underbrace{2 \times 10 \times 70 + 70 \times 70}_{\text{reduced sample space}}}$$

reduced sample space.

($E \cap$ Sample space).

Thm 3.1 If each outcome of a finite sample space is equally likely, then we may compute conditional prob. of the form $P(F|E)$ by using E as the reduced sample space.

Eg. 2. In the pocker game, 52 cards are equally distributed to A, B, C, D. If A & B have a total # of 8 diamonds, what is the prob. that C has 3 diamonds.

$A+B$: 26 cards. $\to$ 8 ◇, $\to$ 18

C: 13 cards $\to$ 3 ◇, $\to$ 10.

D: rest. 13 cards $\to$ ◇ + 11.

Sol. 1. E: $(A+B)$ have 8 ◇

F: C has 3 ◇

Then $P(F|E) = \dfrac{P(F \cap E)}{P(E)} = \dfrac{\dfrac{\binom{13}{8}\binom{39}{18}\binom{5}{3}\binom{21}{10}\binom{2}{2}\binom{11}{11}}{\binom{52}{26}\binom{26}{13}\binom{13}{13}}}{\dfrac{\binom{13}{8}\binom{39}{18}}{\binom{52}{26}}} \approx 0.33.$

$A+B$: 26 cards $\to$ 8 ◇, $\to$ 18.

Sol. 2. $C+D = 5$ ◇ $+ 21$.

$P(F|E) = \dfrac{\binom{5}{3}\binom{21}{10}}{\binom{26}{13}} \approx 0.33.$

2.

Thm 3.2. (Multiplication Rule) For events E, F, we have $P(E \cap F) = P(E) \cdot P(F|E)$

More generally, $P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \times P(E_2|E_1) \times P(E_3|E_1 \cap E_2) \times \cdots \times P(E_n|E_1 \cap \cdots \cap E_n)$.

$= P(E_1) \cdot P(E_2 \cap \cdots \cap E_n | E_1) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3 \cap \cdots \cap E_n | E_1 \cap E_2) = \cdots\cdots$

Eg. 3.

5 white
7 black

Each time a ball is selected, its colour is noted and it is returned along with 2 other balls of the same colour.

Compute the following prob.

(a) the first 2 balls are B & the next 2 are white.

$(B_1 B_2 W_1 W_2)$

$P(B_1 B_2 W_1 W_2) = P(B_1) P(B_2|B_1) P(W_1|B_1 \cap B_2) P(W_2|B_1 \cap B_2 \cap W_1)$

$= \frac{7}{12} \times \frac{9}{14} \times \frac{5}{16} \times \frac{7}{18}$.

(b) of the first 4 balls, exactly 2 are black.

$P(W_1 W_2 B_1 B_2) = \frac{5}{12} \times \frac{7}{14} \times \frac{7}{16} \times \frac{9}{18}$. $= P(B_1 B_2 W_1 W_2)$

Similarly, the $\binom{4}{2} = 6$ numbers of prob. are the same.

Prob. $= \binom{4}{2} \times \frac{5 \times 7 \times 7 \times 9}{12 \times 14 \times 16 \times 18}$.

3. For two event E, F, we say E and F are independent

if $P(E \cap F) = P(E) \times P(F)$ (in other words, $P(F) = P(F|E)$)

Otherwise, they are dependent.

For a sequence of events $E_1, \cdots, E_n$,

they are independent iff $P(E_{i_1} \cap \cdots \cap E_{i_k}) = \prod_{j=1}^{k} P(E_{i_j})$

$\forall\ i_1 < i_2 < \cdots < i_k$, where $k \geq 1$.

4. **(Total Probability Formula)** Let $A_1, \cdots, A_n$ be mutually exclusive with

$S = \bigcup_{k=1}^{n} A_k$. Then, $\forall$ event B,

$= \bigsqcup_{k=1}^{n} A_k$.

$$P(B) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i).$$

Proof: $P(B) = P(B \cap S) = P\left(\bigcup_{i=1}^{n} B \cap A_i\right)$

$$= \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)\,P(A_i) \qquad \square.$$

(Bayes' Theorem) Let $A_1, \cdots, A_n$ be mutually exclusive so that $S = \bigsqcup_{i=1}^{n} A_i$.

Then $\forall$ event B, $P(A_j|B) = \dfrac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$. $\dfrac{P(A_j \cap B)}{P(B)} = \dfrac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)}$.

Proof: $LHS = \dfrac{P(A_j \cap B)}{P(B)} = \dfrac{P(B|A_j)\,P(A_j)}{\text{Total Prob.}} = \dfrac{P(B|A_j)\,P(A_j)}{\sum_{i=1}^{n} P(B|A_i)\,P(A_i)} = RHS \qquad \square.$

Eg 4. $\boxed{\begin{matrix} 5 & W \\ 9 & B \end{matrix}}$ A fair die is rolled and that number of balls are randomly chose from the urn.

(a). $P(\underline{\text{all of the balls selected are white}}) \triangleq A$.

$$= P(A) = \sum_{i=1}^{6} P(A|i)\,P(i) = \sum_{i=1}^{5} P(A|i)\,P(i)$$

$$= \sum_{i=1}^{5} \frac{\binom{5}{i}}{\binom{14}{i}} \times \frac{1}{6}.$$

(b). What is the conditional prob. that $\underline{\text{the die landed on } 3}$ $\triangleq A$.

if $\underline{\text{all the balls selected are white}}$ $\triangleq B$?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\overbrace{P(B|A)}^{\binom{5}{3}/\binom{14}{3}}\ \overbrace{P(A)}^{=\frac{1}{6}}}{\sum_{i=1}^{5} \frac{\binom{5}{i}}{\binom{14}{i}} \times \frac{1}{6}}$$

$P(A \cap B) = P(A|B)\,P(B)$

$= P(B|A)\,P(A)$.

Week 5. Lecture 4.

Discrete Random Variables.

1. **Def**: A random variable $X: S \to \mathbb{R}$.

   If X can take on at most a countable number of possible values is called a discrete random variable.

   Eg1. 80 stu. = 70 male + 10 female.

   Choose 1 stu. randomly for 4 times.

   $X \triangleq \#$ of male stu. chosen. Then X is a discrete R.V. taking values in $\{0, 1, 2, 3, 4\}$.

   Furthermore, $\forall k \in \{0, 1, 2, 3, 4\}$. $P(X=k) = \binom{4}{k} \left(\frac{7}{8}\right)^k \left(\frac{1}{8}\right)^{4-k}$.

2. For a discrete R.V. X,

   define: probability mass function (p.m.f) $p(m)$ of X by

   $$p(m) = P(X=m).$$

3. A R.V. X is said to be a Bernoulli random variable with parameter $p \in [0,1]$

   if $P(X=0) = 1-p$, $P(X=1) = p$.

   We say $\boxed{X \sim \text{Bernoulli}(p)}$.

4. If we toss a coin independently for n times and let $X = \#$ of heads coming up,

   二项式的. 二项的.

   then X is a binomial random variable with parameter (n, p).

   denoted by $\boxed{X \sim \text{Bin}(n,p)}$.

   Take p.m.f of $\text{Bin}(n,p)$ is:

   $$\forall k \in \{0, 1, \cdots, n\}, \quad P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

   One can check that: $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^k = 1$. (binomial theorem)

Eg. 2. Alice is in a class of 80 students.

After 100 independent trials. we count X as the # of times when Alice is picked.

Then $X \sim Bin(100, \frac{1}{80})$.

indicator function.

Remark. If $X_i = 1_{\{\text{Alice is picked at } i\text{-th trial.}\}}$

(indicator)

$= \begin{cases} 1, & \text{yes} \\ 0, & \text{no.} \end{cases}$

then, $X = \sum_{i=1}^{100} X_i$ where $X_i \sim$ Bernoulli (p).

5. Let $X \sim Bin(n, \frac{\lambda}{n})$ for some $\lambda > 0$.

Then, $P(X=k) = \binom{n}{k} (\frac{\lambda}{n})^k (1-\frac{\lambda}{n})^{n-k}$, $\forall\ 0 \le k \le n$.

Fix any $k \ge 0$, we will let $n \to \infty$,

$$P(X=k) = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} (1-\frac{\lambda}{n})^n (1-\frac{\lambda}{n})^{-k}$$

$$= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} (1-\frac{\lambda}{n})^n (1-\frac{\lambda}{n})^{-k}$$

$$\to \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^k \cdot e^{-\lambda}}{k!} \quad (n \to \infty).$$

Rmk. $(1-\frac{\lambda}{n})^n = e^{\ln(1-\frac{\lambda}{n}) \cdot n} \approx e^{(-\frac{\lambda}{n}) \cdot n} = e^{-\lambda}$.

$\ln(1+x) \approx x$ if $x \approx 0$.

We call the limiting distribution by $\boxed{\text{Poisson}(\lambda)}$ and its p.m.f is

$P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $\forall k \ge 0$.

6. There is a coin having probability $p \in (0,1)$ of coming up heads.

Toss the coin until it shows up heads for the first time.

Let X = # of tosses needed.

Then $\boxed{X \sim \text{Geometric}(p)}$, the p.m.f of which is

$$P(X=k) = (1-p)^{k-1}p, \quad \forall k \geq 1.$$

Remark. In Statistics, $Y \sim \text{Geometric}(p)$ if

$$P(Y=k) = (1-p)^k p, \quad \forall k \geq 0. \qquad X \overset{d}{=} Y+1.$$

Eg 3. If we toss the coin n times, and X represents the difference between # of heads and # of tails. Find the p.m.f of X.

U. V.

Solution: ① If n is odd, the all possible values are $\{n, n-2, n-4, \cdots, 3, 1\}$.

$U+V=n$. ① $U-V=k$ ② $V-U=k$.

$(n = U+V, \ X = |U-V| \text{ is odd.})$ √.

$$P(X=k) = \binom{n}{\frac{n+k}{2}} p^{\frac{n+k}{2}} (1-p)^{\frac{n-k}{2}} + \binom{n}{\frac{n-k}{2}} p^{\frac{n-k}{2}} (1-p)^{\frac{n+k}{2}}, \quad \forall k \in \{n, n-2, \cdots, 3, 1\}.$$

√.

② If n is even, $X \in \{0, 2, 4, \cdots, n\}$. √. $U+V=n$. ① $U-V=k$ ② $V-U=k$.

$$\begin{cases} P(X=k) = \text{the same} & \forall k \in \{2, 4, \cdots, n\}. \\ P(X=0) = \binom{n}{\frac{n}{2}} p^{\frac{n}{2}} (1-p)^{\frac{n}{2}}. \end{cases}$$

Example 5: At all times, a pipe-smoking mathematician carries 2 matchboxes-1 in his left-hand pocket and 1 in his right-hand pocket. Both matchboxes initially contained 100 matches. Each time he needs a match, he is equally likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of his matchboxes is empty. What is the probability that there are exactly k matches, k= 0,1,2, … , 100 in the other box?

101st. R.

$$\underbrace{\cdots\cdots\cdots\cdots}_{100R + (100-k)L} \ \_$$

$$P(X=k) = \binom{200-k}{100} \left(\tfrac{1}{2}\right)^{100} \left(\tfrac{1}{2}\right)^{100-k} \cdot \tfrac{1}{2}.$$

108 cards = {1,2,...,108}.
Pick one card uniformly at random.
Record the humber. Return the card. Repeat until we collect all the 108 numbers. (Coupon Collector Problem).

What is the average number of trials needed?

7. For a discrete R.V., the expectation of X is defined by

$$EX = \sum_{m=0}^{\infty} m \cdot P(X=m).$$

Continuous Random Variable.

1. Def A non-negative function $f: (-\infty, \infty) \to [0, \infty)$ is called a probability density function (p.d.f) if

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Then a R.V. X is called a continuous random variable if

$\exists$ a p.d.f f s.t. $\forall\, a, b \in \mathbb{R}$, $P(a \le X \le b) = \int_a^b f(x)\, dx$.

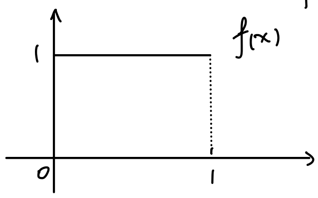
Rmk. Let $b=a$ to get $P(X=a) = P(a \le X \le a) = \int_a^a f(x)\, dx = 0.$

Week 5. Lecture 5.

1. A random variable $\boxed{X \sim \text{Uniform}(\alpha, \beta)}$

if the p.d.f of X is

$$f(x) = \frac{1}{\beta-\alpha} \mathbb{1}_{(\alpha,\beta)}(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \alpha < x < \beta. \\ 0. & \text{otherwise.} \end{cases}$$



Indicator func. $\mathbb{1}_A(x)$, $x \in \mathbb{R}$.

$$\mathbb{1}_A(x) = \begin{cases} 1. & x \in A \\ 0. & x \notin A. \end{cases}$$

Eg 1. Let $X \sim \text{Uniform}(1,5)$ Find $P(X > 3.5)$

Recall. $P(a \le x \le b) = \int_a^b f(x)\,dx$. $\forall a, b \in \mathbb{R}$. (including $a = -\infty$ and $b = -\infty$)

☆ 单点值密度不影响总概率.

Solution: $P(X > 3.5) = P(X \geq 3.5) = \lim_{b \to \infty} P(3.5 \le x \le b)$. $= \int_{3.5}^{+\infty} f(x)\,dx$.

Use $f(x) = \frac{1}{5-1} \mathbb{1}_{(1,5)}(x)$ to see

$$P(X > 3.5) = \int_{3.5}^{5} \frac{1}{4}\,dx = \frac{3}{8}.$$

2. We say X is an exponential random variable with parameter $\lambda > 0$ if

the p.d.f is $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0. \\ 0, & x < 0. \end{cases}$ $\boxed{X \sim \text{Exp}(\lambda)}$

$P(X \geq a) = e^{-\lambda \cdot a}$.

Def. We say a non-negative R.V. X is memoryless if

$$P(X > t+s \mid X > t) = P(X > s), \ \forall t, s > 0.$$

$$= P(X - t > s \mid X > t).$$

i.e. The distribution of $X-t$ conditioning on $X > t$ is equal to X.

If $X \sim \text{Exp}(\lambda)$, then $\forall t > 0$,

$$P(X > t) = \int_t^{\infty} \lambda e^{-\lambda x}\,dx = e^{-\lambda t}$$

Therefore, $LHS = \frac{P(X > t+s,\ X > t)}{P(X > t)} = \frac{P(X > t+s)}{P(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s}$.

RHS $= P(X>s) = e^{-\lambda s} =$ LHS.

Claim. If X is memoryless, then $X \sim Exp(\lambda)$ for some $\lambda>0$.

Sketch of proof: LHS $= \dfrac{P(X>t+s)}{P(X>t)} =$ RHS $= P(X>s)$ $\forall t,s>0$.

$\Rightarrow \forall t,s>0$ $P(X>t+s) = P(X>t)\,P(X>s)$

$\Rightarrow$ $\ln P(X>t+s) = \ln P(X>t) + \ln P(X>s)$.

Let $g(t) = \ln P(X>t)$

$g(t+s) = g(t) + g(s)$

$\Rightarrow$ $g(t) = ct+d$ for some c,d.

by $g(0)=0$. $g(\infty) = -\infty$, $g(t) = -\lambda t$. for some $\lambda>0$.

Eg 2. Let X be a continuous R.V. with p.d.f $f(x) = \begin{cases} \lambda e^{-\frac{x}{100}}, & x\geq 0 \\ 0, & x<0 \end{cases}$

(a) $P(50 \leq X \leq 150)$

Solution 1: Since $1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} \lambda e^{-\frac{x}{100}}dx = \lambda \cdot 100$, $\lambda = \frac{1}{100}$. $X \sim Exp(\frac{1}{100})$

Then, $P(50\leq X\leq 150) = \int_{50}^{150} \frac{1}{100} e^{-\frac{x}{100}}dx = \left(-e^{-\frac{x}{100}}\right)\Big|_{x=50}^{x=150}$.

$= e^{-\frac{1}{2}} - e^{-\frac{3}{2}}$.

Solution 2: $P(50\leq X\leq 150) = P(X\geq 50) - P(X\geq 150) = e^{-\frac{1}{100}\cdot 50} - e^{-\frac{1}{100}\cdot 150}$.

(b) $P(X\leq 100) = 1 - P(X\geq 100) = 1 - e^{-\frac{1}{100}\cdot 100} = 1-e^{-1}$.

3. We say $\boxed{X \sim Gamma(\alpha,\lambda)}$ if the p.d.f is

$$f(x) = \begin{cases} \lambda \cdot e^{-\lambda x} \cdot \dfrac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}. & x\geq 0 \\ 0. & x<0 \end{cases}$$

where $\alpha, \lambda > 0$, and the gamma function is $\Gamma(\alpha) = \int_0^\infty e^{-y} y^{\alpha-1} dy$.

If $\alpha = n \in \mathbb{N}$, then $\Gamma(n) = \int_0^\infty e^{-y} y^{n-1} dy = \int_0^\infty y^{n-1} d(-e^{-y}) = (-y^{n-1}e^{-y})\Big|_{y=0}^{y=\infty} - \int_0^\infty (-e^{-y}) d(y^{n-1})$

$$= (n-1)\int_0^\infty e^{-y} y^{n-2} dy = \cdots = (n-1)! \int_0^\infty e^{-y} dy = (n-1)!.$$

$\Rightarrow \Gamma(n) = (n-1)!$

In fact, if $X_1, \cdots, X_n$ are independent $\text{Exp}(\lambda)$, then $X_1 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$

4. We say $\boxed{X \sim N(\mu, \sigma^2)}$ is a normal random variable if the density is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

$\mu = \int_{-\infty}^\infty x f(x) dx$

$\sigma^2 = \int_{-\infty}^\infty (x-\mu)^2 f(x) dx$

Check, $1 \overset{?}{=} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$

$$\underline{\underline{y = x-\mu}} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy$$

$$\underline{\underline{\frac{y}{\sigma} = t}} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}\cancel{\sigma}} e^{-\frac{1}{2}t^2} \cancel{\sigma} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}t^2} dt.$$

$$X \sim N(\mu, \sigma^2).$$
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$
$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

The proof is reduced to $\sqrt{2\pi} \overset{?}{=} \int_{-\infty}^\infty e^{-\frac{1}{2}t^2} dt \overset{\text{def}}{=} I \quad \Leftrightarrow I^2 = 2\pi.$

Notice $I^2 = \left[\int_{-\infty}^\infty e^{-\frac{1}{2}t^2} dt\right]^2$

$$= \int_{-\infty}^\infty e^{-\frac{x^2}{2}} dx \cdot \int_{-\infty}^\infty e^{-\frac{y^2}{2}} dy$$

$$= \int_{-\infty}^\infty \int_{-\infty}^\infty e^{-\frac{x^2+y^2}{2}} dx dy.$$

$$\underline{\underline{\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}}} \int_0^\infty r dr \int_0^{2\pi} d\theta \cdot e^{-\frac{r^2}{2}} = 2\pi \int_0^\infty r e^{-\frac{r^2}{2}} dr = 2\pi \left(-e^{-\frac{r^2}{2}}\right)\Big|_0^\infty = 2\pi.$$

When $\mu=0$, $\sigma^2=1$. we say $X \sim N(0,1)$ is a standard normal R.V.

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

5. For a continuous R.V. X, the expectation of X is defined by

$$EX = E(X) = E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx. \quad \left(\text{similar to } \sum_{x} x \cdot P(X=x) \right).$$

For any function g, we have

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx.$$

<u>Recall</u>: for a discrete R.V. X, the expectation is $EX = \sum_{m=-\infty}^{\infty} m\, P(X=m).$

For any function g, $E[g(X)] = \sum_{m=-\infty}^{\infty} g(m) \cdot P(X=m).$

6. The <u>variance</u> of X is given by

$$Var(X) \overset{def}{=} E\left[(X-EX)^2\right] \underset{\text{Prop.}}{=} E(X^2) - (EX)^2$$

7. <u>Prop.</u> (of expectation)

Let X, Y be 2 random variables.

(a). If $X \geq 0$, the $EX \geq 0$.

(b). If $c \in \mathbb{R}$, then $E(cX) = c\,EX.$

(c). $E[X+Y] = EX + EY.$

<u>Note</u>. $Var(X) = E\left[(X-\underset{\uparrow\ \text{const}}{EX})^2\right] = E\left[X^2 - 2EX \cdot X + (EX)^2\right]$

$= E[X^2] - E[2EX \cdot X] + (EX)^2$

$= E(X^2) - 2(EX)^2 + (EX)^2 = E(X^2) - (EX)^2.$

Eg 3. $X \sim Unif(0,1)$. $f(x) = \begin{cases} 1, & x \in (0,1) \\ 0, & \text{otherwise.} \end{cases}$

$EX = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 1 \, dx = \frac{1}{2}.$

$EX^2 = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 \cdot 1 \, dx = \frac{1}{3}. \Rightarrow Var(X) = E(X^2) - (EX)^2$

$= \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}.$

Week 6. Lecture 6.

1. $EX = \sum_m m \cdot P(X=m).$

$E[g(X)] = \sum_m m \, P(g(X)=m) = \sum_n g(n) \, P(X=n).$

$\downarrow$ cts. (continuous).

$\int_{-\infty}^{\infty} g(x) f(x) dx.$

Prop. (of expectation).

(i) $\forall c \in \mathbb{R}$, $E(c) = c$. $g(X) \equiv c$.

$E(c) = \sum_m c \, P(X=m) = c$

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

(ii). $\forall a, b \in \mathbb{R}$, $E(aX+b) = aEX + b$.

Let $g(x) = ax+b$.

$E[aX+b] = E[g(X)]$

$= \int_{-\infty}^{\infty} g(x) f(x) dx$

$= \int_{-\infty}^{\infty} (ax+b) f(x) dx = a \int_{-\infty}^{\infty} x f(x) dx + b \int_{-\infty}^{\infty} f(x) dx$

$= a E(X) + b.$

Eg1. If $X = Exp(\lambda)$

then $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx \overset{t=\lambda x}{=\!=} \int_0^{\infty} t e^{-t} \frac{1}{\lambda} dt = \frac{1}{\lambda}\int_0^{\infty} t e^{-t} dt.$

$= \frac{1}{\lambda}\Gamma(2) = \frac{1}{\lambda} 1! = \frac{1}{\lambda}.$ where $\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt.$ Or. $\boxed{\int_0^{+\infty} t^n e^{-t} dt = n!}$.

Next, $Var(X) = E(X^2) - (EX)^2$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx \overset{t=\lambda x}{=\!=} \frac{1}{\lambda^2}\int_0^{\infty} t^2 e^{-t} dt = \frac{1}{\lambda^2}\cdot 2! = \frac{2}{\lambda^2}.$$

$$Var(X) = E(X^2) - (EX)^2 = \frac{1}{\lambda^2}.$$

Eg2. $X \sim Gamma(\alpha, \lambda)$ with $f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}$, $x > 0$.

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} dx \overset{t=\lambda x}{=\!=} \int_0^{\infty} t e^{-t} \frac{t^{\alpha-1}}{\Gamma(\alpha)} \frac{1}{\lambda} dt.$$

$$= \frac{1}{\lambda\Gamma(\alpha)} \int_0^{\infty} t^{\alpha} e^{-t} dt = \frac{1}{\lambda\Gamma(\alpha)} \Gamma(\alpha+1) = \frac{1}{\lambda}\frac{\Gamma(\alpha+1)}{\Gamma(\alpha)}.$$

Notice that if $\alpha = n \in \mathbb{N}$. then $E(X) = \frac{1}{\lambda}\frac{\Gamma(n+1)}{\Gamma(n)} = \frac{1}{\lambda}\frac{n!}{(n-1)!} = n\frac{1}{\lambda}.$

Next, $Var(X) = E(X^2) - (EX)^2.$

$$E(X^2) = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} dx \overset{t=\lambda x}{=\!=} \int_0^{+\infty} \frac{t^2}{\lambda} e^{-t} \frac{t^{\alpha-1}}{\Gamma(\alpha)} \frac{1}{\lambda} dt.$$

$$= \frac{1}{\lambda^2\Gamma(\alpha)} \int_0^{\infty} t^{\alpha+1} e^{-t} dt = \frac{1}{\lambda^2}\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)}$$

$$Var(X) = \frac{1}{\lambda^2}\frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} - \frac{1}{\lambda^2}\frac{\Gamma^2(\alpha+1)}{\Gamma^2(\alpha)}.$$

If $\alpha = n \in \mathbb{N}$. $Var(X) = \frac{1}{\lambda^2}\frac{(n+1)!}{(n-1)!} - \frac{1}{\lambda^2}n^2 = \frac{1}{\lambda^2}\cdot n.$

Eg 3. $X \sim N(\mu, \alpha^2)$.

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

$$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \alpha^2. \quad \boxed{\text{(Check Quiz 5)}}$$

Eg 4. $X \sim$ Bernoulli (p).

$$E(X) = \sum_m m \cdot P(X=m) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = p.$$

$$E(X^2) = \sum_m m^2 \cdot P(X=m) = 0^2 \cdot P(X=0) + 1^2 \cdot P(X=1) = p.$$

$$Var(X) = p - p^2 = p(1-p)$$

Eg 5. $X \sim$ Bin. (n, p).

$E(X) = np$.

$Var(X) = n \cdot p(1-p)$

$$E(X) = \sum_{m=0}^{n} m \cdot P(X=m) = \sum_{m=0}^{n} m \cdot \binom{n}{m} p^m (1-p)^{n-m}$$

$$= \sum_{m=0}^{n} \frac{n! \cdot m}{m! \cdot (n-m)!} p^m (1-p)^{n-m}.$$

$$= \underline{\sum_{m=1}^{n} \frac{n!}{(m-1)!\,(n-m)!} p^m (1-p)^{n-m}}.$$

$$= n \cdot p \sum_{m=1}^{n} \frac{(n-1)!}{(m-1)!\,(n-m)!} p^{m-1} (1-p)^{n-m}.$$

$$= n \cdot p \sum_{k=0}^{n-1} \frac{(n-1)!}{k!\,(n-1-k)!} p^k (1-p)^{n-1-k}.$$

$$= n \cdot p \,(p + (1-p))^{n-1} = n \cdot p.$$

$$E(X^2) = \sum_{k=0}^{n} k^2 P(X=k) = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k}.$$

$$= \sum_{k=0}^{n} \frac{n!\, k}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k}.$$

$$= \sum_{k=1}^{n} \frac{n!\,(k-1)}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k} + \underline{\sum_{k=1}^{n} \frac{n!}{(k-1)!\,(n-k)!} p^k (1-p)^{n-k}}$$

$$= \sum_{k=2}^{n} \frac{n!}{(k-2)!\,(n-k)!} p^k (1-p)^{n-k} + np.$$

$$= p^2 \cdot n(n-1) \sum_{k=2}^{n} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2}(1-p)^{n-k} + n\cdot p.$$

$$= p^2 \cdot (n^2-n) \cdot (p+(1-p))^{n-2} + n\cdot p = p^2\cdot(n^2-n) + n\cdot p.$$

$$Var(X) = E(X^2) - (EX)^2 = p^2(n^2-n) + n\cdot p - n^2p^2 = n\cdot(p-p^2)$$

$$= n\cdot p(1-p).$$

<u>Eg 6.</u> $X \sim$ Poisson (λ).

i.e. $P(X=k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k \geq 0$.

$$E(X) = \sum_{k=0}^{\infty} k\cdot e^{-\lambda}\frac{\lambda^k}{k!}$$

Note: $\boxed{e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!},\ \forall x \in \mathbb{R}.}$

$$= \lambda\cdot e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots\cdots$

$$= \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \lambda\cdot e^{-\lambda}e^{\lambda} = \lambda.$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty}(k-1)e^{-\lambda}\frac{\lambda^k}{(k-1)!} + \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-1)!}$$

$$= \sum_{k=2}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-2)!} + \lambda = \lambda^2 + \lambda.$$

$$Var(X) = E(X^2) - (EX)^2 = \lambda.$$

<u>Eg 7.</u> $X \sim$ Geo (p).

$P(X=k) = (1-p)^{k-1}p$. $\forall k \geq 1$.

$$EX = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{\infty} k(1-p)^{k-1}$$

Let $f(x) = \sum_{k=1}^{\infty} k x^{k-1}$ $(x \in (0,1))$.

$\boxed{\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.}$

$F(x) = \sum_{k=0}^{\infty} x^k$. $F'(x) = \sum_{k=1}^{\infty} kx^{k-1} = f(x)$.

$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots\cdots$

$F(x) = \frac{1}{1-x}$ $F'(x) = \frac{1}{(1-x)^2} = f(x)$. $\Rightarrow \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}.$

$$EX = p\cdot\sum_{k=1}^{\infty} k(1-p)^{k-1} = p\cdot\frac{1}{p^2} = \frac{1}{p}.$$

$$E(X^2) = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p$$

$$= p\cdot\sum_{k=1}^{\infty} k\cdot k(1-p)^{k-1} = p\cdot\sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1} - p\sum_{k=1}^{\infty} k(1-p)^{k-1}.$$

$$= p\cdot\sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1} - \frac{1}{p}.$$

Let $g(x) = \sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1}$, $x\in(0,1)$.

$G(x) = \sum_{k=0}^{\infty} x^{k+1}$, $x\in(0,1)$. $G'(x) = \sum_{k=0}^{\infty}(k+1)x^k$

$$G''(x) = \sum_{k=0}^{\infty}(k+1)k\,x^{k-1} = \sum_{k=1}^{\infty}(k+1)k\,x^{k-1} = g(x).$$

Note that $G(x) = x\cdot\frac{1}{1-x}$. $G'(x) = \frac{1-x+x}{(1-x)^2} = \frac{1}{(1-x)^2}$. $G''(x) = \frac{2}{(1-x)^3}$.

$$\sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1} = \frac{2}{p^3}.$$

$$Var(X) = p\cdot\frac{2}{p^3} - \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{1}{p^2} - \frac{1}{p}.$$

2. For random variable X, the cumulative distribution function (c.d.f) of X is

$F(b) = F_X(b) = P(X\le b)$. $b\in\mathbb{R}$.

If X is a continuous R.V. then $F(b) = \int_{-\infty}^{b} f(x)\,dx$. $\Rightarrow F'(b) = f(b)$.

If X is a discrete R.V. then $F(b) = \sum_{m=-\infty}^{[b]} P(X=m)$.

$$P(X<b) = P\left(\bigcup_{n=1}^{\infty} X\le b-\tfrac{1}{n}\right) = \lim_{n\to\infty} P(X\le b-\tfrac{1}{n}) = \lim_{n\to\infty} F(b-\tfrac{1}{n}).$$

continuity of probability

Thm. If $A_n \subseteq A_{n+1}$ $\forall n \geq 1$. then $P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$.

Proof: $P(\bigcup_{n=1}^{\infty} A_n) = P(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cup \cdots\cdots \cup (A_{n+1} \setminus A_n) \cup \cdots\cdots)$.

$$= \sum_{n=1}^{\infty} P(A_n \setminus A_{n-1}) = \sum_{n=1}^{\infty} P(A_n) - P(A_{n-1}) = \lim_{n\to\infty} P(A_n). \qquad \square.$$

$A_0 \triangleq \phi$.

Week 6 Lecture 7.

The c.d.f of a R.V. X is $F(b) = F_X(b) = P(X \leq b)$ $\forall b \in \mathbb{R}$

1. **Prop.** (of c.d.f).

(a) F is a nondecreasing function. i.e. $\forall a<b.$ $F(a) \leq F(b)$

(b). $\lim_{b\to\infty} F(b) = 1$, $\lim_{b\to-\infty} F(b) = 0$.

Proof: $\lim_{b\to\infty} F(b) = \lim_{b\to\infty} P(X \leq b)$

$$= \lim_{n\to\infty} P(X \leq n) = P(\bigcup_{n=1}^{\infty} X \leq n) = P(S) = 1.$$

$\lim_{b\to-\infty} F(b) = \lim_{b\to-\infty} P(X \leq b)$

$$= \lim_{n\to\infty} P(X \leq -n) = P(\bigcap_{n=1}^{\infty} X \leq -n) = P(\phi) = 0.$$

$A_n \uparrow$. $(A_n \subseteq A_{n+1})$. $\Rightarrow \lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$.

$B_n \downarrow$ $(B_n \supseteq B_{n+1})$. $\Rightarrow \lim_{n\to\infty} P(B_n) = P(\bigcap_{n=1}^{\infty} B_n)$.

(c). F is right continuous i.e. $\forall b \in \mathbb{R}$, $\forall b_n \searrow b$ $(n\to\infty)$, $\lim_{n\to\infty} F(b_n) = F(b)$.

Proof: $\lim_{n\to\infty} P(X\le b_n) = P(\bigcap_{n=1}^{\infty} X\le b_n) = P(X\le b)$.

$\bigcap_{n=1}^{\infty} X\le b_n = X\le b$. $\quad b_n \downarrow b$.

"$\subseteq$": If $X>b$, $\exists\, b_N$ s.t. $X>b_N$. $\quad$ ↯.

$\Rightarrow X\le b$.

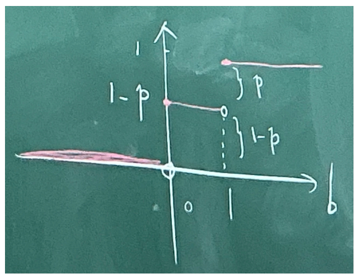
"$\supseteq$": Trivial.

(d). F has left limits. $\forall b\in\mathbb{R}$, $\forall a_n \uparrow b$. $\lim_{n\to\infty} F(a_n) = F(b^-) = P(X<b)$.

Cádlág: Right continuous & left limits.

Eg 1. If $X\sim$ Bernoulli (p). then

$$F_X(b) = \begin{cases} 1, & b\ge 1 \\ 1-p, & 0\le b<1 \\ 0, & b<0. \end{cases}$$



$F_X(b) = P(X\le b)$.

2. Function of a Random Variable.

Theorem 6.1 If $X\sim N(\mu,\sigma^2)$, then

$$Y = aX+b \sim N(a\mu+b,\, a^2\sigma^2),\quad a,b\in\mathbb{R}.$$

Proof: $\forall t\in\mathbb{R}$, $P(Y\le t) \overset{?}{=} \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi a^2\sigma^2}} \frac{[y-(a\mu+b)]^2}{e^{-2a^2\sigma^2}}\, dy$.

$$= P(aX+b\le t)$$

(a>0)

$$= P\left(X\le \frac{t-b}{a}\right) = \int_{-\infty}^{\frac{t-b}{a}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx.$$

$$\overset{y=ax+b}{=\!=\!=} \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}}\, d\left(\frac{y-b}{a}\right).$$

$$= \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{a} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}}\, dy. \qquad \square.$$

$\left[\text{If } a<0, \quad P\left(X \geq \frac{t-b}{a}\right) = \int_{\frac{t-b}{a}}^{\infty} f(x)\, dx \overset{y=ax+b}{=\!=\!=} \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{-a} e^{-\frac{(y-a\mu-b)^2}{2a^2\sigma^2}} dy. \right]$

$|a|$

It follows that if $X \sim N(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma} \sim N(0,1)$.

3. <u>Eg 2.</u> If $X \sim N(0,1)$, find the c.d.f and p.d.f of $Y = X^2$.

<u>Solution:</u> $\forall t>0$, $F(t) = P(Y \leq t) = P(X^2 \leq t)$

$$= P(-\sqrt{t} \leq X \leq \sqrt{t}) = \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

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

Take derivative to get

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} \cdot \frac{1}{2}\frac{1}{\sqrt{t}} - \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} \cdot \left(-\frac{1}{2}\frac{1}{\sqrt{t}}\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} \cdot \frac{1}{\sqrt{t}}. \qquad (t>0)$$

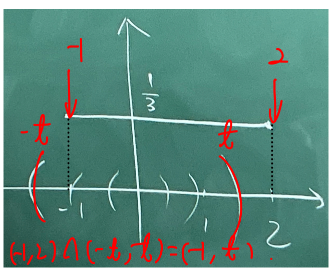
If $t \leq 0$, $f(t) = 0$.

<u>Eg 3.</u> Let $X \sim Unif(-1,2)$. Find the c.d.f and p.d.f of $Y = |X|$.

<u>Solution.</u> $\forall t>0$, $F_Y(t) = P(Y \leq t) = P(|X| \leq t)$

$$= P(-t \leq X \leq t)$$

$$= \int_{-t}^{t} f_X(x)\, dx.$$

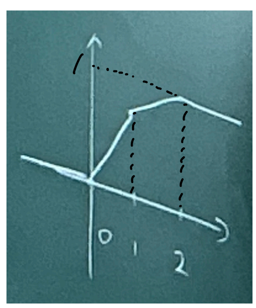


1°. $0 < t \leq 1$. $\quad F(t) = \int_{-t}^{t} \frac{1}{3} dx = \frac{2}{3} t.$

2° $1 < t \leq 2$. $\quad F(t) = \int_{-1}^{t} \frac{1}{3} dx = \frac{1}{3} t + \frac{1}{3}.$

3° $t > 2$. $\quad F(t) = \int_{-1}^{2} \frac{1}{3} dx = 1.$

4° $t \leq 0$, $\quad F(t) = 0.$



The p.d.f is obtained by taking derivative.

$$f(t) = \begin{cases} 0, & t \le 0. \\ \frac{2}{3}, & 0 < t \le 1. \\ \frac{1}{3}, & 1 < t \le 2. \\ 1, & t > 2 \end{cases}$$

<u>Theorem 6.2</u> Let X be a continuous R.V. with p.d.f $f_X(x)$. Suppose $g(x)$ is a strictly monotonic (increasing or decreasing), differentiable function.

Then $Y = g(X)$ has a p.d.f.

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & \text{if } y = g(x) \text{ for some } X. \\ 0. & \text{if } y \ne g(x),\ \forall x. \end{cases}$$

where $g^{-1}(y)$ is defined to be the X s.t. $g(x) = y$.

<u>Proof</u>: $\forall y \in \mathbb{R}$. $F_Y(y) = P(Y \le y) = P(g(X) \le y)$

Assume g is increasing. Then $g(X) \le y \iff X \le g^{-1}(y)$.

So $F_Y(y) = P(X \le g^{-1}(y)). = F_X(g^{-1}(y))$.

Take dirivative to get $\overset{1^\circ\ y = g(x)}{f_Y(y)} = F_X'(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$.

$$= f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y).$$

2° If $y > g(x)$. $\forall x \in \mathbb{R}$, then $F_Y(y) = 1. \Rightarrow f_Y(y) = 0.$

3° If $y < g(x)$ $\forall x \in \mathbb{R}$, then $F_Y(y) = 0. \Rightarrow f_Y(y) = 0.$ $\square$

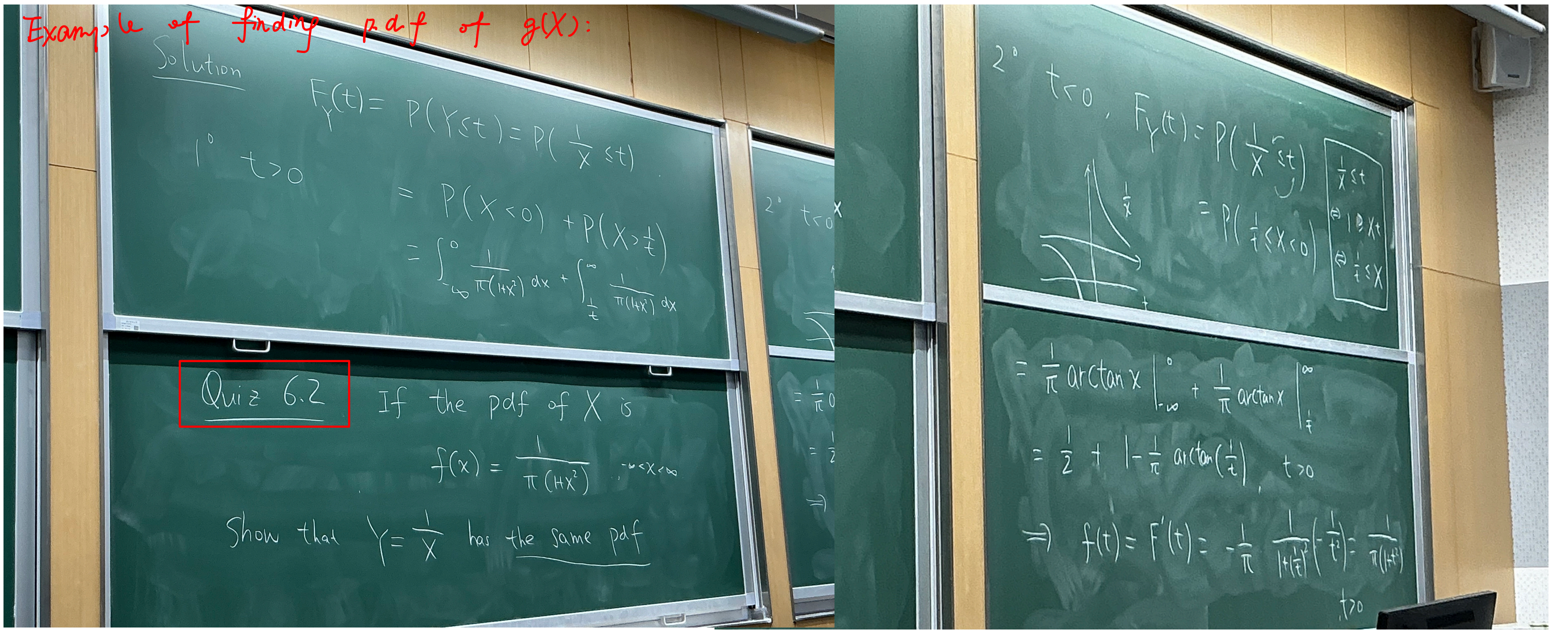
Steps for finding p.d.f of $g(X)$.

1° Find the c.d.f of $Y = g(X)$.

2° Differentiate to find the density.

3°. Specify in what region the result holds.

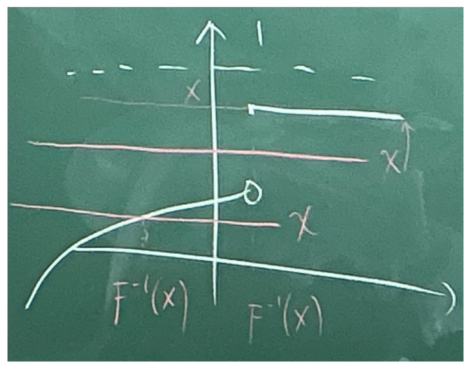
Example of finding pdf of $g(X)$:



Theorem 6.3. Let $F(x)$ be the c.d.f of any R.V. X.

Define for each $x \in (0,1)$, $F^{-1}(x) = \sup\{y \in \mathbb{R} : F(y) < x\}$.

Let U be the uniform $(0,1)$. Then the distribution of $F^{-1}(U)$ is the same as X.



Proof. Let $\Omega = (0, 1)$, $\mathcal{F}$ = the Borel sets, and P = Lebesgue measure. If $\omega \in (0, 1)$, let

$$X(\omega) = \sup\{y : F(y) < \omega\}$$

Once we show that

$$(*) \qquad \{\omega : X(\omega) \le x\} = \{\omega : \omega \le F(x)\}$$

the desired result follows immediately since $P(\omega : \omega \le F(x)) = F(x)$. (Recall P is Lebesgue measure.) To check ($\star$), we observe that if $\omega \le F(x)$ then $X(\omega) \le x$, since $x \notin \{y : F(y) < \omega\}$. On the other hand if $\omega > F(x)$, then since F is right continuous, there is an $\epsilon > 0$ so that $F(x + \epsilon) < \omega$ and $X(\omega) \ge x + \epsilon > x$. ∎

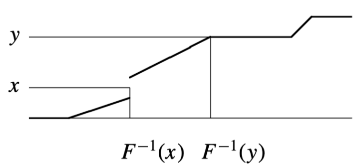


Figure 1.4. Picture of the inverse defined in the proof of Theorem 1.2.2.

Even though F may not be 1-1 and onto, we will call X the inverse of F and denote it by F^{-1}. The scheme in the proof of Theorem 1.2.2 is useful in generating random variables on a computer. Standard algorithms generate random variables U with a uniform distribution; then one applies the inverse of the distribution function defined in Theorem 1.2.2 to get a random variable $F^{-1}(U)$ with distribution function F.

Week 7. Lecture 8.

1. Joint distribution function:

For any R.V.s X, Y, the joint cumulative distribution function of X and Y is defined by

$$F(a,b) = P(X \le a, Y \le b), \quad \forall a, b \in \mathbb{R}.$$

Notice that

$$P(X \le a) = \lim_{b \to \infty} P(X \le a, Y \le b)$$

$$= F(a, \infty).$$

$$P(Y \le b) = F(\infty, b).$$

2. When X, Y are both discrete R.V.s with p.m.f is given by P_X, P_Y.

i.e. $P_X(m) = P(X=m)$.

The joint probability is given by

$$p(x,y) = P(X=x, Y=y).$$

In this case, $\underset{P(X \le a, Y \le b)}{\underset{\parallel}{F(a,b)}} = \sum_{m=-\infty}^{[a]} \sum_{n=-\infty}^{[b]} \underset{p(m,n).}{\underset{\parallel}{P(X=m, Y=n)}}$

3. Independent R.V.s

We say X, Y are independent if $\forall A, B \subseteq \mathbb{R}$,

$$P(X \in A, Y \in B) = P(X \in A) P(Y \in B).$$

Prop. So the two discrete R.V.s X, Y are independent

iff $\forall x, y \in \mathbb{R}$,

$$p(x,y) = P_X(x) P_Y(y).$$

$$P(X=x, Y=y) = P(X=x) P(Y=y).$$

Proof: "$\Rightarrow$": Let $A=x$, $B=y$.

$$P(X=x,\ Y=y) = P(X=x)\,P(Y=y).$$

"$\Leftarrow$": $\forall A, B \subseteq \mathbb{R}$, $P(X\in A, Y\in B)$

$$= \sum_{x\in A}\sum_{y\in B} P(X=x,\ Y=y).$$

$$= \sum_{x\in A}\sum_{y\in B}\Big(P(X=x)\,P(Y=y)\Big).$$

$$= \sum_{x\in A}\Big(P(X=x)\sum_{y\in B} P(Y=y)\Big).$$

$$= \Big(\sum_{x\in A} P(X=x)\Big)\Big(\sum_{y\in B} P(Y=y)\Big).$$

$$= P(X\in A)\ P(X\in B). \quad \square.$$

4. We say X, Y are <u>jointly continuous</u> if there exists a function $f(x,y)$ s.t. $\forall C \subseteq \mathbb{R}^2$,

$$P((X,Y)\in C) = \int_{(x,y)\in C} f(x,y)\,dx\,dy.$$

$\Leftrightarrow$ $\forall A, B \subseteq \mathbb{R}$

$$P(X\in A,\ Y\in B) = \int_A dx \int_B dy\ f(x,y),$$

The function $f(x,y)$ is called the <u>joint p.d.f</u> of X and Y.

The <u>joint c.d.f</u> is then given by

$$F(a,b) = P\big(X\le a,\ Y\le a\big) = \int_{-\infty}^{a} dx \int_{-\infty}^{b} dy\ f(x,y).$$

$$\Rightarrow f(a,b) = \frac{\partial^2}{\partial a\,\partial b} F(a,b).$$

边缘的. 边际的.

(marginal p.d.f).

The <u>marginal probability density function</u> of X and Y are

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx.$$

X, Y are <u>independent</u> iff

$$f(x,y) = f_X(x)\, f_Y(y).$$

$$\Updownarrow$$

$$P(X \approx x, Y \approx y) = P(X \approx x)\, P(Y \approx y).$$

5. For any joint p.m.f $p(x,y)$ or joint p.d.f $f(x,y)$.

we have $\forall$ function $g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$E[g(X,Y)] = \sum_m \sum_n g(m,n)\, p(m,n).$$

<u>Eg.</u> $g(X,Y) = 1_{X \in A}\, 1_{Y \in B}$.

$$E[g(X,Y)] = E\left[1_{X\in A,\, Y\in B}\right] = P(X \in A, Y \in B).$$

(Here we use $P(A) = 1 \cdot P(1_A = 1) + 0 \cdot P(1_A = 0) = E[1_A]$).

$$= \sum_{m\in A} \sum_{n \in B} p(m,n) = \sum_m \sum_n p(m,n)\, 1_{m\in A}\, 1_{n \in B} = \sum_m \sum_n p(m,n)\, g(m,n).$$

or $$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\,dy.$$

<u>Eg 1.</u> The joint p.d.f is

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x, y < +\infty. \\ 0. & \text{otherwise.} \end{cases}$$

One can check $\int_0^\infty \int_0^\infty f(x,y)\,dx\,dy = 1$.

(a) $$P(X \le 1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{(x \le 1)}\, f(x,y)\, dx\,dy$$

$$\parallel$$

$$E(1_{X \le 1}) = \int_0^1 dx \int_0^\infty dy\; 2e^{-x}e^{-2y}$$

$$= \int_0^1 e^{-x}dx \int_0^\infty 2e^{-2y}dy = (1-e^{-1})\cdot 1. = 1-e^{-1}.$$

(b) $$P(X>1, Y<1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{(x>1,\, y<1)}\, f(x,y)\,dx\,dy$$

$\parallel$

$$E\left(1_{(X>1,\, Y<1)}\right) \quad = \int_1^\infty dx \int_{-\infty}^1 dy\; 2e^{-x}e^{-2y}$$

$$= \int_1^\infty e^{-x}dx \int_{-\infty}^1 2e^{-2y}dy$$

$$= e^{-1}\cdot(1-e^{-2}).$$

(c) $$P(X<Y) = E\left(1_{X<Y}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{x<y}\, f(x,y)\,dx\,dy.$$

$$= \int_0^\infty dx \int_x^\infty dy\; 2e^{-x}e^{-2y}$$

$$= \int_0^\infty e^{-x}dx \int_x^\infty 2e^{-2y}dy$$

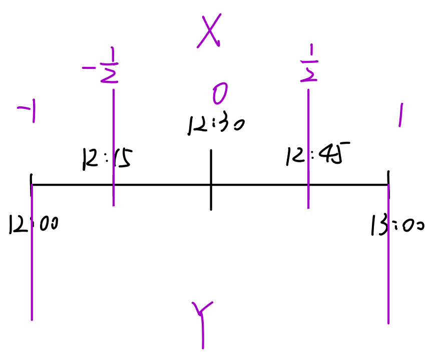
$$= \int_0^\infty e^{-x}e^{-2x}dx = \frac{1}{3}.$$

Eg 2. A man and a woman $\longrightarrow$ 12:30 P.M.

Assume the man Unif(12:15, 12:45)

the woman Unif(12:00, 1:00) independent.

(a) P(the man arrives first)

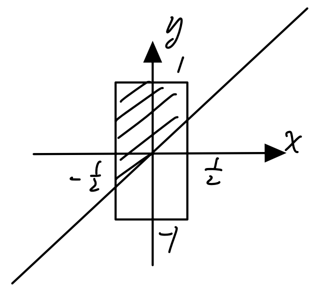


Solution: $X \sim Unif\left(-\frac{1}{2}, \frac{1}{2}\right)$

$Y \sim Unif(-1, 1).$

$$P(X<Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 1_{x<y}\, f(x,y)\,dx\,dy$$

where $f(x,y) = f_X(x)\, f_Y(y) = \begin{cases} 1\cdot\frac{1}{2}, & -\frac{1}{2}<x<\frac{1}{2},\ -1<y<1. \\ 0. & \text{otherwise} \end{cases}$

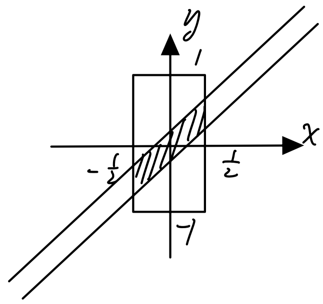


$$= \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} dx \int_{x}^{1} dy = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} (1-x)\, dx = \frac{1}{2}[1-0] = \frac{1}{2}.$$

(b) Find the probability that the first to arrive waits no longer than 5 minutes.

$$P\left(|X-Y| < \frac{5}{30}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbb{1}_{|x-y|<\frac{1}{6}}\, f(x,y)\, dx\, dy$$

$$= \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} dx \int_{x-\frac{1}{6}}^{x+\frac{1}{6}} dy = \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{3}\, dx = \frac{1}{6}.$$



6. The bivariate normal distribution

.... multivariate ....

The joint p.d.f of bivariate normal distribution (X, Y) is given by

$$f(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right\}$$

$$(X,Y) \sim N(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2; \rho)$$

$$\Rightarrow \quad X \sim N(\mu_x, \sigma_x^2).$$

$$Y \sim N(\mu_y, \sigma_y^2).$$

相关系数.

7. The covariance of X, Y is

$$\mathrm{Cov}(X,Y) \stackrel{\text{def}}{=} E\left[(X-EX)(Y-EY)\right].$$

$$\Rightarrow \mathrm{Cov}(X,X) = E\left[(X-EX)^2\right] = \mathrm{Var}(X).$$

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \text{correlation of } X, Y.$$

If we let $\vec{x} = (x, y)$

$\vec{\mu} = (\mu_x, \mu_y).$

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}, \quad \Sigma^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_x^2} & \frac{-\rho}{\sigma_x\sigma_y} \\ \frac{-\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}.$$

So $\quad f(x,y) = \dfrac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left\{-\frac{1}{2}(\vec{x}-\vec{\mu})\,\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right\}.$

$$(x-\mu_x,\ y-\mu_y)\,\Sigma^{-1}\begin{pmatrix} x-\mu_x \\ y-\mu_y \end{pmatrix} = \left(\frac{x-\mu_x}{\sigma_x} - \rho\frac{y-\mu_y}{\sigma_y}\right)^2 + (1-\rho^2)\left(\frac{y-\mu_y}{\sigma_y}\right)^2 \geq 0.$$

Week 8 Lecture 9

$\vec{x} = (x,y).\quad \vec{\mu} = (\mu_x, \mu_y).$

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} = \begin{array}{c|cc} & X & Y \\ \hline X & Cov(X,X) & Cov(X,Y) \\ Y & Cov(Y,X) & Cov(Y,Y) \end{array}$$

$\vec{x} = (x_1, x_2, x_3)$

$\vec{\mu} = (\mu_1, \mu_2, \mu_3).$

$$\Sigma = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline X_1 & & & \\ X_2 & & Cov & \\ X_3 & & & \end{array}.$$

$$f(\vec{x}) = \frac{1}{2\pi\sqrt{\det(\Sigma)}} \exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})^T\right].$$

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\,\frac{y-\mu_y}{\sigma_y}\right]\right\}.$$

$X \sim N(\mu_x, \sigma_x^2)$

$Y \sim N(\mu_y, \sigma_y^2) \qquad f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \dfrac{1}{\sqrt{2\pi\sigma_x^2}}\, e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$

$$f_X(x) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y-\mu_y}{\sigma_y} - \rho\frac{x-\mu_x}{\sigma_x}\right)^2 + (1-\rho^2)\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]\right\} dy.$$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\, e^{-\frac{1}{2}\frac{(x-\mu_x)^2}{\sigma_x^2}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_y}{\sigma_y} - \rho\frac{x-\mu_x}{\sigma_x}\right)^2} dy.$$

Notice that $\displaystyle\int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{y-\mu_y}{\sigma_y} - \rho\frac{x-\mu_x}{\sigma_x}\right)^2} dy$

$$\underline{\underline{t = \frac{y-\mu_y}{\sigma_y} - \rho\frac{x-\mu_x}{\sigma_x}}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}t^2} d(\sigma_y t) = \sigma_y \int_{-\infty}^{+\infty} e^{-\frac{1}{2(1-\rho^2)}t^2} dt.$$

By $\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2(1-\rho^2)}} dt = \sqrt{2\pi(1-\rho^2)}.$ $= \sigma_y\sqrt{2\pi(1-\rho^2)}$

$$\left(1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} dt. \right).$$

$$f_X(x) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \sigma_y\sqrt{1-\rho^2}\sqrt{2\pi} = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}.$$

If $\rho=0$. then $f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \cdot e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} = f_X(x) f_Y(y).$

So X and Y are independent.

If X, Y are indep. then

$$\mathrm{Cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E(X-\mu_X)\cdot E(Y-\mu_Y) = 0 \Rightarrow \rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = 0$$

Thus, for $(X,Y) \sim N(\mu_x, \mu_y; \sigma_x^2, \sigma_y^2; \rho)$

X, Y independent $\Leftrightarrow \rho = 0$

1. Sum of indep. R.V.s.

Let X, Y be independent continuous R.V.s.

Then $f(x,y) = f_{XY}(x,y) = f_X(x) f_Y(y).$

Let $Z = X+Y$, for the p.d.f of Z

we have $F_Z(z) = P(Z \le z) = P(X+Y \le z)$

$$E[g(X,Y)] = \iint g(x,y) f(x,y)\, dx\, dy.$$

$$\Rightarrow \quad P(X+Y \le z) = \iint 1_{X+Y\le z} f(x,y)\, dx\, dy$$

$$\| $$

$$E(1_{X+Y\le z}) \quad = \int_{-\infty}^{\infty} dy \int_{-\infty}^{z-y} dx \; f(x,y)$$

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

$$= \int_{-\infty}^{\infty} f_Y(y)\, dy \; \underbrace{\int_{-\infty}^{z-y} f_X(x)\, dx.}_{= F_X(z-y).}$$

$$\Rightarrow F_Z(z) = \int_{-\infty}^{\infty} F_X(z-y)\, f_Y(y)\, dy.$$

Differentiate to obtain

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

$$\triangleq f_X * f_Y(z).$$ Convolution. 卷积.

Eg 1. $X, Y \sim Exp(\lambda)$. indep.

Compute the p.d.f of $X+Y$.

Solution: $f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y)\, dy$

$$= \int_{-\infty}^{+\infty} f_X(z-y)\, f_Y(y) \cdot 1_{z-y>0} \; 1_{y>0} \; dy$$

$$= \int_0^z \lambda e^{-\lambda(z-y)} \cdot \lambda e^{\cancel{-\lambda y}}\, dy$$

$$= \lambda^2 e^{-\lambda z} z = \lambda e^{-\lambda z} \cdot \frac{(\lambda z)^1}{\Gamma(1)} \sim Gamma(2, \lambda).$$

<u>Eg. 2.</u> $X \sim N(\mu_x, \sigma_x^2)$

$Y \sim N(\mu_y, \sigma_y^2)$. indep.

$\Rightarrow X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$

<u>Solution:</u> $f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{1}{2}\left(\frac{z-y-\mu_x}{\sigma_x}\right)^2} \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} dy.$$

$$= \frac{1}{2\pi\sigma_x\sigma_y} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\frac{z-y-\mu_x}{\sigma_x}\right)^2 - \frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\} dy$$

"–"

Note that

$$\frac{\sigma_x^2(y-\mu_y)^2 + \sigma_y^2(y-z+\mu_x)^2}{2\sigma_x^2\sigma_y^2}$$

$$= \frac{1}{2\sigma_x^2\sigma_y^2}\left[(\sigma_x^2+\sigma_y^2)y^2 - 2y(\sigma_x^2\mu_y + \sigma_y^2(z-\mu_x)) + \sigma_x^2\mu_y^2 + \sigma_y^2(z-\mu_x)^2\right]$$

$$= \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left[y^2 - 2y\left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}\mu_y + \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}(z-\mu_x)\right) + \frac{\sigma_x^2\mu_y^2+\sigma_y^2(z-\mu_x)^2}{\sigma_x^2+\sigma_y^2}\right].$$

$$= \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left[y - \left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}\mu_y + \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}(z-\mu_x)\right)\right]^2$$

$$\underline{- \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_y^2}\mu_y + \frac{\sigma_y^2}{\sigma_x^2+\sigma_y^2}(z-\mu_x)\right)^2 + \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\frac{\sigma_x^2\mu_y^2+\sigma_y^2(z-\mu_x)^2}{\sigma_x^2+\sigma_y^2}} \quad (*)$$

$$(*) = \frac{\sigma_x^2+\sigma_y^2}{2\sigma_x^2\sigma_y^2}\left[\frac{1}{(\sigma_x^2+\sigma_y^2)^2}\left(\cancel{-\sigma_x^4\mu_y^2} - \cancel{\sigma_y^4(z-\mu_x)^2} + \cancel{\sigma_x^4\mu_y^2} + \sigma_x^2\sigma_y^2(z-\mu_x)^2 + \sigma_x^2\sigma_y^2\mu_y^2\right.\right.$$

$$+ \cancel{a_y^4 (z-\mu_x)^2} \Big) - 2\mu_y (z-\mu_x) \Big].$$

$$= \frac{1}{2a_x^2 a_y^2} \frac{1}{(a_x^2+a_y^2)} \qquad a_x^2 a_y^2 [\mu_y - (z-\mu_x)]^2 = \frac{(z-\mu_x-\mu_y)^2}{2(a_x^2+a_y^2)}.$$

Thus, $$f_{X+Y}(z) = \frac{1}{2\pi a_x a_y} \int_{-\infty}^{+\infty} \exp\left\{ -\frac{a_x^2+a_y^2}{2a_x^2 a_y^2} \Big[\underbrace{y - \Big(\frac{a_x^2}{a_x^2+a_y^2}\mu_y + \frac{a_y^2}{a_x^2+a_y^2}(z-\mu_x) \Big)}_{= t.} \Big]^2 \cancel{+} - \frac{(z-\mu_x-\mu_y)^2}{2(a_x^2+a_y^2)} \right\} dy.$$

$$= \frac{1}{2\pi a_x a_y} e^{-\frac{(z-\mu_x-\mu_y)^2}{2(a_x^2+a_y^2)}} \cdot \int_{-\infty}^{+\infty} e^{-\frac{a_x^2+a_y^2}{2a_x^2 a_y^2} t^2} dt.$$

$$\left[1 = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{t^2}{2a^2}} dt. \right]$$

$$= \frac{1}{2\pi a_x a_y} e^{-\frac{(z-\mu_x-\mu_y)^2}{2(a_x^2+a_y^2)}} \cdot \sqrt{2\pi \frac{a_x^2 a_y^2}{a_x^2+a_y^2}}$$

$$= \frac{1}{\sqrt{2\pi(a_x^2+a_y^2)}} e^{-\frac{(z-\mu_x-\mu_y)^2}{2(a_x^2+a_y^2)}} \sim N(\mu_x+\mu_y, a_x^2+a_y^2).$$

Upshot: Remember: $X \sim N(\mu, a^2)$

$$f(x) = \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(x-\mu)^2}{2a^2}}, \quad -\infty < x < +\infty.$$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi a^2}} e^{-\frac{(x-\mu)^2}{2a^2}} = 1.$$

Eg. 3. If $X \sim \text{Bernoulli}(\frac{1}{2})$
$Y \sim N(0,1)$. indep.

Find p.d.f of $Z = X+Y$.

$$P(X+Y \le z) = \frac{1}{2} P(X+Y \le z \mid X=0) + \frac{1}{2} P(X+Y \le z \mid X=1)$$

$$= \frac{1}{2} P(Y \le z) + \frac{1}{2} P(Y \le z-1).$$

$$= \frac{1}{2}\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}dx + \frac{1}{2}\int_{-\infty}^{z-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}dx.$$

$$f(z) = \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} + \frac{1}{2}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-1)^2}.$$

Week 9. Lecture 10.

1. Sum of independent discrete R.V.s.

Eg 1 If $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$, indep. then $X+Y \sim \text{Poisson}(\lambda_1+\lambda_2)$

Proof: $\forall n \geq 0$, $P(X+Y=n) \overset{?}{=} e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}$.

[Total probability] $= \sum_{k=0}^{n} P(X+Y=n \mid X=k)\, P(X=k)$

$$= \sum_{k=0}^{n} P(Y=n-k \mid X=k)\, P(X=k)$$

$$= \sum_{k=0}^{n} P(Y=n-k)\, P(X=k)$$

$$= \sum_{k=0}^{n} e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}\, e^{-\lambda_1}\frac{\lambda_1^k}{k!}$$

$$= e^{-(\lambda_1+\lambda_2)}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}\lambda_1^k\lambda_2^{n-k}$$

$$= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^n}{n!}. \qquad \square.$$

Eg 2. If $X \sim Bin(n,p)$, $Y \sim Bin(m,p)$. indep. then $X+Y \sim Bin(n+m, p)$.

Proof: $\forall\, 0 \leq N \leq n+m$. $P(X+Y=N) \overset{?}{=} \binom{n+m}{N} p^N (1-p)^{n+m-N}$.

$$= \sum_{k=0}^{\infty} P(X+Y=N \mid X=k)\, P(X=k)$$

$$= \sum_{k=0}^{\infty} P(Y=N-k \mid X=k)\, P(X=k)$$

$N-k \leq m$. $\quad k \geq N-m$.

$N \wedge n = \min\{N,n\}$. $\quad = \sum_{k=0}^{N\wedge n} P(Y=N-k)\, P(X=k)$.

$m \wedge n = \max\{m,n\}$.

1° $0 \leq N \leq n \wedge m$

$$\text{原式} = \sum_{k=0}^{N}\binom{m}{N-k} p^{N-k}(1-p)^{m-N+k}\binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{N} \binom{m}{N-k}\binom{n}{k} p^N (1-p)^{m+n-N}$$

$$= p^N (1-p)^{m+n-N} \sum_{k=0}^{N} \binom{m}{N-k}\binom{n}{k}$$

$$= p^N (1-p)^{m+N-N} \binom{m+n}{N}.$$

$2°$ $n \vee m \le N \le m+n$.

$$原式 = \sum_{k=0}^{n} \binom{m}{N-k}\binom{n}{k} p^N (1-p)^{n+m-N}$$

$$= \binom{m+n}{N} p^N (1-p)^{n+m-N}.$$

$3°$. $n \wedge m < N < n \vee m$. similar.

2. If X, Y are discrete R.V.s, we define the <u>conditional probability mass function of X given $Y=y$</u>.

$$p_{X|Y}(x|y) = P(X=x \mid Y=y)$$

$$= \frac{P(X=x, Y=y)}{P(Y=y)} = \frac{p(x,y)}{p_Y(y)}.$$

$P(Y=y)$ >0.

The <u>conditional cumulative distribution function</u>

$$F_{X|Y}(x|y) = P(X \le x \mid Y=y)$$

$$= \sum_{m \le x} P(X=m \mid Y=y).$$

If X, Y indep. then $p_{X|Y}(x|y) = p_X(x)$. $F_{X|Y}(x|y) = F_X(x)$.

$$\left[P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A) \text{ if } A, B \text{ indep.} \right].$$

<u>Eg 3.</u> Suppose the $p(x,y)$ of (X,Y) is $p(0,0)=0.4$ $p(0,1)=0.2$

$p(1,0)=0.1$ $p(1,1)=0.3$.

Find the conditional distribution of X given $Y=1$.

Solution: $P(X=0 \mid Y=1) = \frac{P(0,1)}{P(0,1)+P(1,1)} = \frac{0.2}{0.5} = \frac{2}{5}$.

$P(X=1 \mid Y=1) = \frac{P(1,1)}{P(0,1)+P(1,1)} = \frac{0.3}{0.5} = \frac{3}{5}$.

or: $P(X=0 \mid Y=1) = \frac{P(X=0, Y=1)}{P(Y=1)} = \frac{P(0,1)}{P(0,1)+P(1,1)} = \frac{2}{5}$.

Eg 4. If X and Y are indep. R.V.s with $X \sim \text{Poisson}(\lambda_1)$, $Y \sim \text{Poisson}(\lambda_2)$.

Calculate the conditional distribution of X given $X+Y=n$.

Solution: $P(X=k \mid X+Y=n) \quad \forall\, 0 \le k \le n$.

$$= \frac{P(X=k,\ X+Y=n)}{P(X+Y=n)} = \frac{P(X=k,\ Y=n-k)}{P(X+Y=n)} = \frac{P(X=k)\ P(Y=n-k)}{P(X+Y=n)}$$

$$= \frac{e^{-\lambda_1}\frac{\lambda_1^k}{k!}\ e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}}{e^{-\lambda_1-\lambda_2}\frac{(\lambda_1+\lambda_2)^n}{n!}} = \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.$$

$$= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k} \sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right).$$

4. If X and Y have joint pdf $f(x,y)$, then the conditional p.d.f of X given $Y=y$ is $f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$.

The condition c.d.f of X given $Y=y$ is

$$F_{X|Y}(a|y) = P(X \le a \mid Y=y)$$
$$= \int_{-\infty}^{a} f_{X|Y}(x|y)\, dx.$$

5. Total probability formula:

Let $A_1, \cdots, A_n$ be mutually exclusive and $S = \bigcup_{k=1}^{\infty} A_k$. Then

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)\ P(A_i).$$

The continuous version:

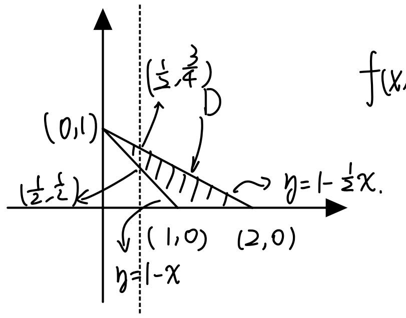
Let $f_Y(y)$ be the p.d.f of Y. Then $\forall$ event B.

$$P(B)=\int_{-\infty}^{+\infty} P(B\mid Y=y)\, f_Y(y)\, dy.$$

$\forall$ R.V. X we have

$$E(X)=\int_{-\infty}^{\infty} E(X\mid Y=y)\, f_Y(y)\, dy.$$

Eg 6. Let (X,Y) be Uniform on D



$$f(x,y)=\begin{cases} 2. & (x,y)\in D. \\ 0. & (x,y)\notin D. \end{cases}$$

(a) Find $P(X\le \frac{1}{2}\mid Y=y)$, $y\in(0,1)$. $=\begin{cases} 0, & 0<y<\frac{1}{2} \\ 1, & \frac{3}{4}<y<1. \end{cases}$

$$P(X\le \tfrac{1}{2}\mid Y=y)=\int_{-\infty}^{\frac{1}{2}} f_{X|Y}(x\mid y)\, dx.$$

Given $Y=y\in(\frac{1}{2},\frac{3}{4})$, $X\sim \mathrm{Unif}(1-y,\ 2(1-y))$.

$$P(X\le\tfrac{1}{2}\mid Y=y)=\frac{\frac{1}{2}-(1-y)}{2(1-y)-(1-y)}=\frac{y-\frac{1}{2}}{1-y}.$$

or:

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)\,dx=\int_{1-y}^{2(1-y)} 2\,dx=2(1-y)$$

$$P(X\le\tfrac{1}{2}\mid Y=y)=\int_{-\infty}^{\frac{1}{2}} f_{X|Y}(x\mid y)\,dx. \qquad \text{Use } f(x,y)=\begin{cases} 2. & 1-y\le x\le 2(1-y). \\ 0. & \text{otherwise.}\end{cases}$$

1°. $0<y<\frac{1}{2}$. $x\ge 1-y>\frac{1}{2}$. $\Downarrow$. 0.

2° $y>\frac{3}{4}$. $1-y<2(1-y)<\frac{1}{2}$.

$$\int_{-\infty}^{\frac{1}{2}} \frac{f(x,y)}{2(1-y)}\,dx=\int_{1-y}^{2(1-y)} \frac{2}{2(1-y)}\,dx=1.$$

3°. $\frac{1}{2}<y<\frac{3}{4}$. $P(X\le\frac{1}{2}\mid Y=y)=\int_{1-y}^{\frac{1}{2}} \frac{2}{2(1-y)}\,dx=\frac{-\frac{1}{2}+y}{1-y}.$

(b). $f_Y(y) = 2(1-y)$, $0<y<1$.

$$P(X \le \tfrac{1}{2}) = \int_{-\infty}^{+\infty} P(X \le \tfrac{1}{2} \mid Y=y) \cdot f_Y(y)\, dy ?$$

$$LHS = P(X \le \tfrac{1}{2}) = \frac{\frac{1}{2} \times \frac{1}{4} \times \frac{1}{2}}{\frac{1}{2} \times 1 \times 1} = \frac{1}{8}.$$

$$RHS = \int_{\frac{1}{2}}^{\frac{3}{4}} \frac{y-\frac{1}{2}}{1-y}\, 2(1-y)\, dy + \int_{\frac{3}{4}}^{1} 1 \cdot 2(1-y)\, dy = 2\int_0^{\frac{1}{4}} t\, dt + 2\int_0^{\frac{1}{4}} t\, dt = \frac{1}{16} + \frac{1}{16} = \frac{1}{8}.$$

Week 10. Lecture 11.

Continuous version of total probability fumula

$$P(B) = \int_{-\infty}^{+\infty} \underbrace{P(B \mid Y=y)}_{= \frac{P(B \cap \{Y=y\})}{P(Y=y)}} f_Y(y)\, dy.$$

$P(B \mid Y)$ is a R.V. of Y.
$\quad = g(Y)$.

$$P(B) = E\left[P(B \mid Y)\right] = \int_{-\infty}^{+\infty} P(B \mid Y=y) f_Y(y)\, dy.$$

1. Recall the Bivariate Normal Distribution

$$f(x,y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho \frac{x-\mu_x}{\sigma_x} \frac{y-\mu_y}{\sigma_y}\right]}.$$

$X \sim N(\mu_x, \sigma_x^2)$ $\quad$ $Y \sim N(\mu_y, \sigma_y^2)$

Given $Y=y$, the conditional pdf of X is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho \frac{x-\mu_x}{\sigma_x} \frac{y-\mu_y}{\sigma_y}\right]}}{\frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-\frac{(y-\mu_y)^2}{2\sigma_y^2}}}$$

$$= \frac{1}{\sqrt{2\pi \sigma_x^2 (1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \rho^2 \left(\frac{y-\mu_y}{\sigma_y}\right)^2 - 2\rho \frac{x-\mu_x}{\sigma_x} \frac{y-\mu_y}{\sigma_y}\right]}$$

$$= \frac{1}{\sqrt{2\pi\sigma_x^2(1-\rho^2)}} e^{-\frac{1}{2\sigma_x^2(1-\rho^2)}\left[(x-\mu_x) - \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)\right]^2}.$$

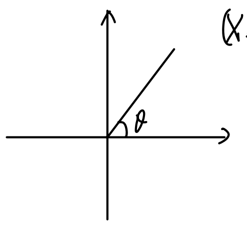
$$\sim N\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y),\ \sigma_x^2(1-\rho^2)\right).$$

If $\rho=0$, then $X\sim N(\mu_x, \sigma_x^2)$, $Y\sim N(\mu_y, \sigma_y^2)$ $\Rightarrow$ X, Y independent.

Also, we have "$\Leftarrow$". Thus, "$\Leftrightarrow$".

2. Suppose (X,Y) are independent $N(0,1)$.

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$



$(X,Y) \to (R, \theta)$, $R \in [0,+\infty)$, $\theta \in [0, 2\pi)$.

$$\begin{cases} x = R\cos\theta \\ y = R\sin\theta \end{cases}$$

Consider (X,Y) jointly continuous R.V. with joint p.d.f $f(x,y)$.

Define: $\begin{cases} U = g(X,Y) \\ V = h(X,Y) \end{cases}$ Let $K = \{(x,y)\in\mathbb{R}^2,\ f(x,y)>0\}$.

Set $G = \{(g(x,y), h(x,y)) \in \mathbb{R}^2,\ (x,y)\in K\} = \{(U,V)\}$.

$K = \mathbb{R}^2$ $G = [0,\infty)\times[0,2\pi)$ bijective

Solve $\begin{cases} u = g(x,y) \\ v = h(x,y) \end{cases}$ to get $\begin{cases} x = q(u,v) \\ y = r(u,v) \end{cases}$.

Let $J(u,v) = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} = \begin{vmatrix} \frac{\partial q(u,v)}{\partial u} & \frac{\partial q(u,v)}{\partial v} \\ \frac{\partial r(u,v)}{\partial u} & \frac{\partial r(u,v)}{\partial v} \end{vmatrix}.$

Theorem The joint p.d.f of $(U, V) = (g(X,Y), h(X,Y))$

is $f_{UV}(u,v) = f_{XY}(q(u,v), r(u,v)) \, |J(u,v)|$.

$$\left[dx\,dy = du\,dv\,|J(u,v)| \right].$$

Proof: For any function $W: \mathbb{R}^2 \to \mathbb{R}$.

$$E[W(U,V)] = E[W(g(X,Y), h(X,Y))].$$

$$= \iint_{(x,y) \in K} W(g(x,y), h(x,y)) \cdot f_{XY}(x,y)\, dx\, dy.$$

Use a change of variable by $\begin{cases} u = g(x,y) \\ v = h(x,y). \end{cases}$ to get

$$= \iint_{(u,v) \in G} W(u,v) \cdot f_{XY}(q(u,v), r(u,v)) \, |J(u,v)| \, du\, dv.$$

On the other hand.

$$E[W(U,V)] = \iint_{(u,v) \in G} W(u,v)\, f_{UV}(u,v)\, du\, dv$$

Since W is arbitrary, $f_{UV}(u,v) = f_{XY}(q(u,v), r(u,v)) \, |J(u,v)|$. almost surely.

Eg. $(X,Y) \to (R, \theta)$

$$\begin{cases} X = R\cos\theta \leftarrow q(u,v). \\ Y = R\sin\theta \leftarrow r(u,v). \end{cases} \qquad f_{XY}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}.$$

$$J(r,\theta) = R.$$

So $f_{R\theta}(R,\theta) = f_{XY}(R\cos\theta, R\sin\theta) \cdot R$

$$= \frac{1}{2\pi} e^{-\frac{1}{2}R^2} \cdot R$$

$$f_R(R) = \frac{1}{2\pi} \cdot e^{-\frac{R^2}{2}} \cdot R \cdot \int_0^{2\pi} d\theta = R \cdot e^{-\frac{R^2}{2}}, \quad R \geq 0.$$

$$f_\theta(\theta) = \int_0^\infty \frac{1}{2\pi} e^{-\frac{R^2}{2}} \cdot R\, dR = \frac{1}{2\pi} \qquad \theta \in [0, 2\pi).$$

$\Rightarrow \theta \sim \text{Unif}(0, 2\pi)$. Further, R, θ independent.

4. (Box-Muller Algorithm)

Let U_1, U_2 be independent Unif(0,1).

Note that $F_R(r) = \int_0^r R e^{-\frac{R^2}{2}} dR = \underset{\substack{\| \\ u}}{1 - e^{-\frac{r^2}{2}}}, \quad r > 0.$

Then $F_R^{-1}(u) = \sqrt{-2\ln(1-u)}, \quad u \in (0,1)$.

By letting $V_1 = \sqrt{-2\ln(1-U_1)}$ or $\sqrt{-2\ln(U_1)}$, we get $V_1 \overset{d}{=} R$. 分布一样.

$V_2 = 2\pi U_2 \sim \text{Unif}(0, 2\pi) \Rightarrow V_2 \overset{d}{=} \theta$.

Then $(V_1, V_2) \overset{d}{=} (R, \theta)$

So $\begin{cases} X = V_1 \cos V_2 \\ Y = V_1 \sin V_2 \end{cases}$ ~ independent $X, Y \sim N(0,1)$.

$$F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

$\Rightarrow X = F_X^{-1}(U)$. $U \sim \text{Unif}(0,1)$.

<u>Expectation</u>: For a discrete R.V. X, the expectation of X is

$$EX = \sum_{m=-\infty}^{\infty} m P(X=m) \text{ given } E|X| < +\infty.$$

$$E|X| = \sum_{m=-\infty}^{\infty} |m| P(X=m) < \infty.$$

$$Var(X) = E[(X-EX)^2] = \sum_{m=-\infty}^{\infty} (m-EX)^2 P(X=m).$$

For a continuous R.V. X, the expectation of X is

$$EX = \int_{-\infty}^{+\infty} x f(x)\, dx. \quad \text{given}$$

$$E|X| = \int_{-\infty}^{+\infty} |x| f(x)\, dx. < +\infty.$$

Week 10 Lecture 12

$(U, V) = (g(X, Y), h(X, Y))$.

$$\Rightarrow f_{UV}(u, v) = f_{XY}(q(u, v), r(u, v))\, |J(u, v)|.$$

$$\begin{cases} u = g(x, y) \\ v = h(x, y) \end{cases} \Rightarrow \begin{cases} x = q(u, v) \\ y = r(u, v) \end{cases}$$

Eg 1. $X \sim \text{Gamma}(\alpha, \lambda)$, $Y \sim \text{Gamma}(\beta, \lambda)$, Independent.

Find the joint p.d.f of $\left(\frac{X}{X+Y}, X+Y\right)$.

Solution: $\begin{cases} U = \frac{X}{X+Y} \\ V = X+Y \end{cases} \Rightarrow \begin{cases} X = UV \\ Y = V - UV \end{cases}$

$$J(u, v) = \begin{vmatrix} \frac{\partial X}{\partial u} & \frac{\partial X}{\partial v} \\ \frac{\partial Y}{\partial u} & \frac{\partial Y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ -v & 1-u \end{vmatrix} = v.$$

By independence,

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

$$= \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \cdot \lambda e^{-\lambda y} \frac{(\lambda y)^{\beta-1}}{\Gamma(\beta)}.$$

So $f_{UV}(u, v) = f_{XY}(uv, v - uv)\, |v|$.

$$= \lambda e^{-\lambda uv} \frac{(\lambda uv)^{\alpha-1}}{\Gamma(\alpha)}\, \lambda e^{-\lambda(v-uv)} \frac{(v-uv)^{\beta-1}}{\Gamma(\beta)} \cdot v.$$

$$= \lambda^{\alpha+\beta} e^{-\lambda v} \cdot u^{\alpha-1} (1-u)^{\beta-1}\, v^{\alpha+\beta-1} \frac{1}{\Gamma(\alpha)\Gamma(\beta)}.$$

$$= \underbrace{\frac{\lambda^{\alpha+\beta} v^{\alpha+\beta-1} e^{-\lambda v}}{\Gamma(\alpha+\beta)}}_{\sim \text{Gamma}(\alpha+\beta,\lambda)} \cdot \underbrace{u^{\alpha-1}(1-u)^{\beta-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\text{Beta}(\alpha,\beta)}.$$

<u>Eg2.</u> Let $X_1, X_2, \cdots, X_n \sim \text{Exp}(\lambda)$. Independent.

Define $Y_k = \sum_{i=1}^{k} X_i$, $1 \le k \le n$.

(a) Find the joint p.d.f of $(Y_1, Y_2, \cdots, Y_n)$.

<u>Solution:</u> Define
$$\begin{cases} y_1 = x_1 \\ y_2 = x_1 + x_2 \\ \vdots \\ y_n = x_1 + x_2 + \cdots + x_n \end{cases} \Rightarrow \begin{cases} x_1 = y_1 \\ x_2 = y_2 - y_1 \\ \vdots \\ x_n = y_n - y_{n-1} \end{cases}$$

$$J(y_1 \cdots y_n) = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{vmatrix} = 1.$$

$$f_{(X_1,\cdots,X_n)}(x_1,\cdots,x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$
$$= \lambda e^{-\lambda x_1} \cdots \lambda e^{-\lambda x_n}$$
$$= \lambda^n e^{-\lambda(x_1 + \cdots + x_n)}.$$

$$\Rightarrow f_{(X_1,\cdots,X_n)}(y_1, y_2 - y_1, \cdots, y_n - y_{n-1}) = \lambda^n e^{-\lambda y_n}. \qquad 0 < y_1 < y_2 < \cdots < y_n.$$

So the joint p.d.f $= \lambda^n e^{-\lambda y_n} \cdot \mathbb{1}_{\{0 < y_1 < y_2 < \cdots < y_n\}}$.

(b). Find the marginal p.d.f of Y_n.

$$f_{Y_n}(y_n) = \underbrace{\int \cdots \int}_{n-1} f(y_1, \cdots, y_n)\, dy_1\, dy_2 \cdots dy_{n-1}$$

$$= \lambda^n e^{-\lambda y_n} \underbrace{\int_0^{y_n} dy_{n-1} \int_0^{y_{n-1}} dy_{n-2} \cdots \int_0^{y_2} dy_1}$$

$\mathbb{1}_{\{0 < y_1 < y_2 < \cdots < y_n\}}$.

$$= \lambda^n e^{-\lambda y_n} \int_0^{y_n} \frac{y_{n-1}^{n-2}}{(n-2)!} dy_{n-1} = \lambda^n e^{-\lambda y_n} \frac{y_n^{n-1}}{(n-1)!} \sim \Gamma(n, \lambda).$$

2. Order Statistics : Let $X_1, \cdots, X_n$ be i.i.d (independent, identically distributed) R.V.s with a common p.d.f f and c.d.f F.

Define $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ to be the ordered sequence of $X_1, \cdots, X_n$.

Remark : $P(X_i = X_j) = 0, \ \forall i \neq j$. $\quad P(X_i - X_j = 0) = 0$.

Theorem. The joint p.d.f of $X_{(1)}, \cdots, X_{(n)}$ is

$$f_{X_{(1)}, \cdots, X_{(n)}}(x_1, \cdots, x_n) = n! \, f(x_1) f(x_2) \cdots f(x_n) \cdot 1_{x_1 < x_2 < \cdots < x_n}.$$

Heuristics: $LHS \approx P(X_{(1)} = x_1, \cdots, X_{(n)} = x_n)$

$x_1 < \cdots < x_n$ and $\{x_1, \cdots, x_n\} = \{X_1, \cdots, X_n\}$.

$X_1 = x_1, \cdots, X_n = x_n$. $\qquad X_1 = x_n, \cdots, X_n = x_1$.

$$P(X_1 = x_1, \cdots, X_n = x_n) = \prod_{k=1}^n P(X_k = x_k)$$

$$\approx \prod_{k=1}^n f(x_k).$$

$$P(X_1 = x_n, \cdots, X_n = x_1) = \prod_{k=1}^n P(X_k = x_{n-k+1}) \approx \prod_{k=1}^n f(x_{n-k+1}).$$

$$= n! \, f(x_1) \cdots f(x_n).$$

Proof using Infitesimal Method:

$\forall \varepsilon > 0$, Note that

$$P\left(X_{(1)} \in (x_1 - \tfrac{\varepsilon}{2}, x_1 + \tfrac{\varepsilon}{2}), \cdots, X_{(n)} \in (x_n - \tfrac{\varepsilon}{2}, x_n + \tfrac{\varepsilon}{2})\right). \quad \forall x_1 < \cdots < x_n.$$

$$= \int_{t_1 \in (x_1 - \frac{\varepsilon}{2}, x_1 + \frac{\varepsilon}{2})} dt_1 \int_{t_2 \in (x_2 - \frac{\varepsilon}{2}, x_2 + \frac{\varepsilon}{2})} dt_2 \int \cdots \int_{t_n \in (t_n - \frac{\varepsilon}{2}, t_n + \frac{\varepsilon}{2})} dt_n \underbrace{f(t_1, \cdots, t_n)}_{\approx f(x_1, \cdots, x_n)}$$

$$\approx f(x_1, \cdots, x_n) \cdot \varepsilon^n.$$

On the other hand,

$$LHS = n!\ P\left(X_1 \in (x_1 - \tfrac{\varepsilon}{2}, x_1 + \tfrac{\varepsilon}{2}), \cdots, X_n \in (x_n - \tfrac{\varepsilon}{2}, x_n + \tfrac{\varepsilon}{2}) \right)$$

$$= n! \prod_{k=1}^{n} P\left(X_k \in (x_k - \tfrac{\varepsilon}{2}, x_k + \tfrac{\varepsilon}{2}) \right)$$

$$\approx n! \prod_{k=1}^{n} \left[f(x_k) \cdot \varepsilon \right].$$

$$f(x_1, \cdots, x_n)\, \varepsilon^n = n! \left(\prod_{k=1}^{n} f(x_k) \right) \cdot \varepsilon^n$$

$$\Rightarrow f(x_1, \cdots, x_n) = n! \prod_{k=1}^{n} f(x_k).$$

[Remark : $P(X \in (x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2})) \approx f(x) \cdot \varepsilon$.

$P(X = x) \approx f(x)$].

$P(X \in (x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2}))$

$= \int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} f(x)\,dx = \varepsilon \cdot f(\eta)$, $\eta \in (x - \frac{\varepsilon}{2}, x + \frac{\varepsilon}{2})$.

$\varepsilon \to 0$, $P(X = x) = f(x)$.

3. If $X_1, X_2, \cdots, X_n$ are i.i.d $Unif(0,1)$.

then $f_{X_{(1)}, X_{(2)}, \cdots, X_{(n)}}(x_1, x_2, \cdots, x_n) = n!$, $0 < x_1 < \cdots < x_n < 1$.

$$1 = \underbrace{\int \cdots \int}_{n} f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n$$

$$= n! \underbrace{\int \cdots \int}_{n} 1_{0 < x_1 < \cdots < x_n < 1}\, dx_1 \cdots dx_n.$$

$$= n! \int_0^1 dx_n \int_0^{x_n} dx_{n-1} \int_0^{x_{n-1}} dx_{n-2} \cdots \int_0^{x_2} dx_1$$

$$= n! \int_0^1 dx_n \int_0^{x_n} dx_{n-1} \cdots \int_0^{x_3} x_2\, dx_2$$

$$= n! \int_0^1 dx_n \int_0^{x_n} dx_{n-1} \cdots \int_0^{x_4} dx_3 \cdot \frac{x_3^2}{2}$$

$$= \cdots$$

$$= n! \int_0^1 dx_n \frac{x_n^{n-1}}{(n-1)!}$$

$$= n! \cdot \frac{1}{n!} = 1.$$

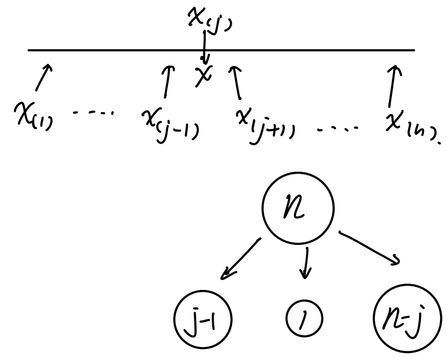
[$X_1, X_2, \cdots, X_n$ are i.i.d $Unif(0,1)$.]

4. For any $1 \leq j \leq n$, the marginal p.d.f of $X_{(j)}$ is

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!\,(j-1)!} x^{j-1} (1-x)^{n-j}, \quad 0 < x < 1.$$

More generally, $f_{X_{(j)}}(x) = \frac{n!}{(n-j)!\,(j-1)!} [F(x)]^{j-1} [1 - F(x)]^{n-j} f(x)$. $F(x)$: c.d.f of X.

$f_{X_{(j)}}(x) \approx P(X_{(j)} = x)$



$= P(\exists\ j-1 \text{ R.V.s } < x$

$\quad \exists\ n-j \text{ R.V.s } > x$

$\quad \exists\ 1 \text{ R.V. } = x).$

$$= \frac{n!}{1!\,(n-j)!\,(j-1)!} P(X_1, \cdots, X_{j-1} < x,\ X_j = x,\ X_{j+1}, \cdots, X_n > x).$$

$$\qquad \| \qquad x^{j-1}(1-x)^{n-j}.$$

$P(X_1 < x) = x$

$P(X_1 > x) = 1 - x.$

$P(X_1 = x) \approx 1.$

5. $EX = \int_{-\infty}^{+\infty} x f(x)\, dx$ given $\int_{-\infty}^{+\infty} |x| f(x)\, dx < +\infty$

$E[g(x)] = \int_{-\infty}^{+\infty} g(x) f(x)\, dx.$

Eg3. If $X \sim N(0,1)$, find $E(|X|^\alpha)$, $\forall \alpha \in \mathbb{R}$.

Proof: $E|X|^\alpha = \int_{-\infty}^{+\infty} |x|^\alpha \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$

$$= 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} x^\alpha e^{-\frac{x^2}{2}} dx.$$

If $\alpha \le -1$, then $\int_0^{+\infty} x^\alpha e^{-\frac{x^2}{2}} dx = \infty$

$$\ge \int_0^1 x^\alpha e^{-\frac{x^2}{2}} dx \ge e^{-\frac{1}{2}} \int_0^1 x^\alpha dx = \infty.$$

If $\alpha > -1$, then $E|X|^\alpha = 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^\infty x^\alpha e^{-\frac{x^2}{2}} dx$

$$\overset{x^2 = t}{\underset{x = \sqrt{t}}{=\!=}} 2 \cdot \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} t^{\frac{\alpha}{2}} \cdot e^{-\frac{t}{2}} \frac{1}{2\sqrt{t}} dt.$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^\infty t^{\frac{\alpha-1}{2}} e^{-\frac{t}{2}} dt.$$

$$\overset{s = \frac{t}{2}}{=\!=} \frac{1}{\sqrt{2\pi}} \int_0^\infty (2s)^{\frac{\alpha-1}{2}} e^{-s} \cdot 2\, ds$$

$$= \frac{1}{\sqrt{2\pi}} 2^{\frac{\alpha+1}{2}} \underbrace{\int_0^{\infty} s^{\frac{\alpha-1}{2}} e^{-s} \, ds}_{\Gamma(\frac{\alpha+1}{2})}.$$

$$\Gamma(\beta) = \int_0^{\infty} s^{\beta-1} e^{-s} ds$$

$$(\beta > 0).$$

Week 11. Lecture 13

Eg 1. $X \sim N(0,1)$ $E(|X|^{\alpha})$.

Eg 2. If $X, Y \sim N(0,1)$ indep. find $E(|X^2+Y^2|^{\alpha})$ for $\alpha \in \mathbb{R}$.

Solution: $E(|X^2+Y^2|^{\alpha}) = \iint (x^2+y^2)^{\alpha} \cdot f_{XY}(x,y) \, dx dy$

$$= \iint (x^2+y^2)^{\alpha} \cdot \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-\frac{1}{2}(x^2+y^2)} dx dy$$

$$= \int_0^{+\infty} r \, dr \int_0^{2\pi} d\theta \, (r^2)^{\alpha} \cdot \frac{1}{2\pi} \cdot e^{-\frac{1}{2}r^2}$$

$$= \int_0^{+\infty} r^{2\alpha+1} e^{-\frac{1}{2}r^2} dr \overset{t=\frac{1}{2}r^2}{\underset{r=\sqrt{2t}}{=}} \int_0^{+\infty} e^{-t} (\sqrt{2t})^{2\alpha+1} \frac{\sqrt{2}}{2\sqrt{t}} dt$$

$$= 2^{\frac{2\alpha+1}{2}} \cdot 2^{-\frac{1}{2}} \int_0^{+\infty} t^{\alpha} e^{-t} dt = \begin{cases} 2^{\alpha} \Gamma(\alpha+1), & \forall \alpha > -1. \\ \infty, & \forall \alpha \le -1. \end{cases}$$

Properties of Expectation:

Let $X_1, \cdots, X_n$ be R.V.s s.t $E|X_i| < +\infty$, $\forall i$.

(a) If $c_0, c_1, \cdots, c_n \in \mathbb{R}$, then

$$E[c_0 + c_1 X_1 + c_2 X_2 + \cdots + c_n X_n] = c_0 + c_1 EX_1 + \cdots + c_n EX_n.$$ (Not requiring independence!)

Proof for binary case:

$$E(X+Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x+y) f(x,y) dx dy$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f(x,y) dx dy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f(x,y) dx dy$$

$$= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx + \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy$$

$$= EX + EY.$$ (Not requiring X, Y indep!)

(b) If $X_1, \cdots, X_n$ are indep. then

$\forall$ functions $g_1, \cdots, g_n$, we have

$$E\left[\prod_{k=1}^{n} g_k(X_k)\right] = \prod_{k=1}^{n} E[g_k(X_k)].$$

Rmk: $g_1(X_1), \cdots, g_n(X_n)$ are also indep.

Eg 3. Ducks: $D_1, D_2, \cdots, D_{10}$

Hunters: $H_1, H_2, \cdots, H_{10}$. each hunter shot at one duck (which duck? random).

each shot: kill rate $= p$

Let X = the total # of ducks that escape.

Find EX.

Solution: Let $X_i = \begin{cases} 1, & \text{if duck } i \text{ escapes.} \\ 0, & \text{otherwise} \end{cases}$

$X = \sum_{i=1}^{10} X_i$.

Then $EX = E\left(\sum_{i=1}^{10} X_i\right) = \sum_{i=1}^{10} EX_i = 10\, EX_1$ by $X_1 \overset{d}{=} X_2 \overset{d}{=} \cdots \overset{d}{=} X_{10}$ (equal in distribution)

$EX_1 = P(X_1=1)\cdot 1 + P(X_1=0)\cdot 0$

$= P(X_1=1) = \left(1-\frac{1}{10}\cdot p\right)^{10}$.

So $EX = 10\, EX_1 = 10\left(1-\frac{p}{10}\right)^{10}$.

Eg 4. A group of n men and n women is lined up at random.

(a) Find the expected # of men who have a woman next to them.

Solution: Let $X_i = \begin{cases} 1, & \text{if man } i \text{ has a woman next to him.} \\ 0, & \text{otherwise.} \end{cases}$

$$EX_{total} = E\left[X_1 + \cdots + X_n\right] = \sum_{k=1}^{n} EX_k = nEX_1.$$

$$EX_1 = P(X_1=1) = \frac{1}{2n}\cdot\frac{n}{2n-1} + \frac{1}{2n}\cdot\frac{n}{2n-1} + \frac{2n-2}{2n}\left(1-\frac{(n-1)(n-2)}{(2n-1)(2n-2)}\right)$$

$$= \frac{1}{2n-1} + \frac{n-1}{n}\cdot\left(1-\frac{n-2}{(2n-1)\cdot 2}\right)$$

$$= \frac{1}{2n-1} + \frac{n-1}{n}\cdot\frac{3n}{4n-2} = \frac{3n-1}{4n-2}. \Rightarrow EX_{total} = \frac{n(3n-1)}{4n-2}.$$

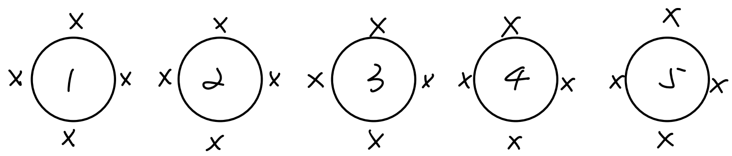
(b) Repeat part (a), but assuming that the group is randomly seated at a round table.

$$EX_{total} = n \cdot 2n \cdot \frac{1}{2n}\left(1 - \frac{(n-1)(n-2)}{(2n-1)(2n-2)}\right)$$

$$= n \cdot \frac{3n}{4n-2} = \frac{3n^2}{4n-2}.$$

Eg5. 20 individuals, 10 married couples.

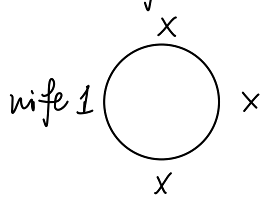
5 tables, each with 4 seats.



(a) If the seating is at random, find the expected # of couples that are seated at the same table.

Solution: $EX_{total} = E(\sum_{i=1}^{10} X_i) = 10\, EX_1.$

Let the wife 1 take seat first.



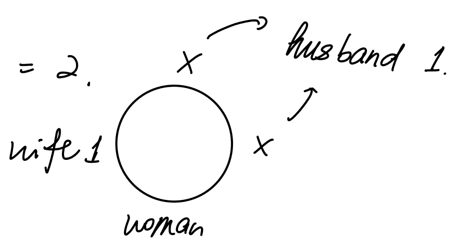
husband 1.

$$\frac{\binom{18}{2}}{\binom{19}{3}} = \frac{3}{19}.$$

(b) If 2 men & 2 women are randomly chosen to be seated at each table, repeat (a).

Solution: $EX_{total} = 10\, EX_1 = 10\, P(X_1 = 1) = 2.$

$$P(X_1 = 1) = \frac{\binom{9}{1}}{\binom{10}{2}} = \frac{2}{10}$$



Eg6 (Coupon - collecting problem)

Suppose there are N different types of coupons, and each time it is equally likely to be any of the N types.

Find expected # of coupons needed before obtaining a complete set of all the N types.

<u>Solution:</u> Def X_i, $0 \le i \le N-1$, to be the number of additional coupons that needed to obtain after i distinct types have been collected.

$$\#\text{ of types: } 0 \xrightarrow{X_0=1} 1 \xrightarrow{X_1} 2 \xrightarrow{X_2} 3 \longrightarrow \cdots \longrightarrow N-1 \xrightarrow{X_{N-1}} N.$$

$$X_{\text{total number}} = X_0 + X_1 + \cdots + X_{N-1}.$$

$$P(X_1 = k) = \underbrace{\left(\tfrac{1}{N}\right)^{k-1}}_{k-1 \text{ same type}} \left(\tfrac{N-1}{N}\right)^1 \longrightarrow \text{remaining } N-1 \text{ types.} \quad \forall k \ge 1.$$

$$X_1 \sim \text{Geometric}\left(\tfrac{N-1}{N}\right).$$

$$P(X_2 = k) = \left(\tfrac{2}{N}\right)^{k-1}\left(\tfrac{N-2}{N}\right)^1. \qquad X_2 \sim \text{Geometric}\left(\tfrac{N-2}{N}\right).$$

$\forall i \in [\![1, N-1]\!]$,

$$P(X_i = k) = \left(\tfrac{i}{N}\right)^{k-1}\left(\tfrac{N-i}{N}\right)^1. \qquad X_i \sim \text{Geometric}\left(\tfrac{N-i}{N}\right).$$

$$EX_{\text{total}} = \sum_{i=0}^{N-1} EX_i$$

also works for $i=0$.

$$= \sum_{i=0}^{N-1} E\left(\text{Geometric}\left(\tfrac{N-i}{N}\right)\right)$$

$$= \sum_{i=1}^{N-1} \frac{N}{N-i} = N \sum_{k=1}^{N} \frac{1}{k} \approx N \ln N.$$

<u>Recall:</u> $E(\text{Geometric}(p)) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = \frac{1}{p}.$

#.

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P\left(\bigcap_{j=1}^{k} E_{i_j}\right).$$

Week 12, Lecture 14

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} P\left(\bigcap_{j=1}^{k} E_{i_j}\right).$$

Let $X_i = \begin{cases} 1, & \text{if } E_i \text{ occurs} \\ 0, & \text{otherwise.} \end{cases} = 1_{E_i}. \quad \forall\, 1 \le i \le n.$

$LHS = E\left(\mathbb{1}_{\bigcup_{i=1}^{n} E_i}\right) = E\left(1 - \prod_{i=1}^{n}(1-X_i)\right)$

If $\bigcup_{i=1}^{n} E_i$ occurs, then $\exists\, i_0$ s.t. E_{i_0} occurs. $\Rightarrow X_{i_0} = 1 \Rightarrow \prod_{i=1}^{n}(1-X_i) = 0 \Rightarrow 1 - \prod_{i=1}^{n}(1-X_i) = 1.$

If $\bigcup_{i=1}^{n} E_i$ doesn't occur, then $\bigcap_{i=1}^{n} E_i^c$ occurs $\Rightarrow \forall\, 1 \le i \le n,\ X_i = 0. \Rightarrow 1 - \prod_{i=1}^{n}(1-X_i) = 0.$

Notice that
$$1 - \prod_{i=1}^{n}(1-X_i) = 1 - \sum_{k=0}^{n} \sum_{1 \le i_1 < \cdots < i_k \le n} \prod_{j=1}^{k}(-X_{i_j}).$$

$$= -\sum_{k=1}^{n} \sum_{1 \le i_1 < \cdots < i_k \le n} (-1)^k \prod_{j=1}^{k} X_{i_j}.$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \prod_{j=1}^{k} X_{i_j}.$$

By using $E\left(\prod_{j=1}^{n} X_{i_j}\right) = 1 \cdot P\left(\prod_{j=1}^{n} X_{i_j} = 1\right)$

$$= P\left(\bigcap_{j=1}^{n} \{X_{i_j} = 1\}\right) = P\left(\bigcap_{j=1}^{n} E_{i_j}\right).$$

Therefore. $P\left(\bigcup_{i=1}^{n} E_i\right) = E\left(1 - \prod_{i=1}^{n}(1-X_i)\right)$

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1 \le i_1 < \cdots < i_k \le n} \underbrace{E\left(\prod_{j=1}^{k} X_{i_j}\right)}_{= P\left(\bigcap_{j=1}^{k} E_{i_j}\right)}.$$

2. The variance of X is given by

$$Var(X) = E\left[(X - \underbrace{EX}_{\mu = EX})^2\right] = \begin{cases} \sum_{m=-\infty}^{\infty} (m-\mu)^2 P(X=m). \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx. \end{cases}$$

3. Properties of variance

(1) If $EX = \mu$, $Var < +\infty$, then

(a) $Var(a+bX) = b^2\, Var(X)$, $\forall\, a, b \in \mathbb{R}$.

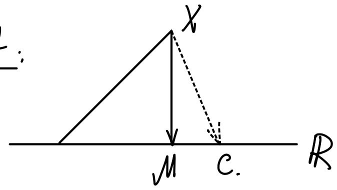
Proof: $\mathrm{Var}(a+bX) = E\left[\left(a+bX - E(a+bX)\right)^2\right]$

$$= E\left[\left(\cancel{a}+bX-\cancel{a}-bEX\right)^2\right]$$

$$= E\left[b^2(X-EX)^2\right] = b^2\, E\left[(X-EX)^2\right]. \qquad \square.$$

(b) $\mathrm{Var}(X) = E\left[(X-\mu)^2\right] < E\left[(X-c)^2\right],\ \forall c \neq \mu.$

Proof:



$$E\left[(X-c)^2\right] = E\left[(X-\mu+\mu-c)^2\right] = E\left[(X-\mu)^2\right] + 2(\mu-c)\,E\left[X-\mu\right] + (\mu-c)^2.$$

Notice that $\quad E[X-\mu] = EX-\mu = 0.$

$$E\left[(X-c)^2\right] = E\left[(X-\mu)^2\right] + (\mu-c)^2 > E\left[(X-\mu)^2\right]. \qquad \square.$$

(c). If $\mathrm{Var}(X)=0$, then $\quad X=\mu \quad$ almost every where. i.e. $P(X=\mu)=1.$

(2) Let $X_1, \cdots, X_n$ be R.V.s with $\mathrm{Var}(X_i) < \infty,\ \forall i.$

then (a) $\quad \mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sum_{j=1}^n \left[E(X_iX_j) - \mu_i\mu_j\right] \qquad (\mu_i = EX_i)$

$$= \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n} \mathrm{Cor}(X_i, X_j)$$

Proof: $\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = E\left[\left(\sum_{i=1}^n X_i - E\left(\sum_{i=1}^n X_i\right)\right)^2\right].$

$$= E\left[\left(\sum_{i=1}^n (X_i-\mu_i)\right)^2\right] = E\left[\left(\sum_{i=1}^n (X_i-\mu_i)\right)\cdot\left(\sum_{j=1}^n (X_j-\mu_j)\right)\right].$$

$$= \sum_{i=1}^n \sum_{j=1}^n E\left[(X_i-\mu_i)(X_j-\mu_j)\right].$$

$$= \sum_{i=1}^n \sum_{j=1}^n E\left[X_iX_j - \mu_iX_j - \mu_jX_i + \mu_i\mu_j\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[E(X_iX_j) - \mu_i \underbrace{E(X_j)}_{\mu_j} - \mu_j \underbrace{E(X_i)}_{\mu_i} + \mu_i\mu_j\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\left[E(X_iX_j) - \mu_i\mu_j\right].$$

Recall $\mathrm{Cov}(X_i, X_j) = E[(X_i - EX_i)(X_j - EX_j)]$ $\quad$ [If $i=j$, $\mathrm{Cov}(X_i, X_i) = \mathrm{Var}(X_i)$.]

$$= E(X_iX_j) - \mu_i\mu_j.$$

Thus $\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathrm{Var}(X_i) + 2\sum_{1\le i<j\le n}\mathrm{Cov}(X_i, X_j)$. $\square$

(b) If $X_1, \dots, X_n$ are independent, then

$$\mathrm{Var}\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n}\mathrm{Var}(X_j).$$

Proof: If X_i, X_j indep. $\mathrm{Cov}(X_i, X_j) = 0$. $\forall i \ne j$. $\square$

Independence: $P(X_1 \in A_1, \dots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$. $\forall A_i \subseteq \mathbb{R}$.

$\iff$ $\forall$ function g_i, $E\left(\prod_{i=1}^{n} g_i(X_i)\right) = \prod_{i=1}^{n} E(g_i(X_i))$.

$\forall i \ne j$, let $\begin{cases} g(X_i) = X_i - \mu_i \\ g(X_j) = X_j - \mu_j. \end{cases}$

$$\Rightarrow E(g(X_i)\,g(X_j)) = E[(X_i - \mu_i)(X_j - \mu_j)]$$

$$\parallel$$

$$E(g(X_i))\,E(g(X_j)) = E(X_i - \mu_i)\cdot E(X_j - \mu_j) = 0.$$

4. Covariance: $\sigma_{XY} = \mathrm{Cov}(X, Y) = E[(X - EX)(Y - EY)]$

Correlation: $\rho = \dfrac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1, 1]$.

5. (Cauchy-Schuwatz Inequality) $\forall$ R.V.s X, Y.

$$|E(XY)| \le E|XY| \le \sqrt{EX^2 \cdot EY^2}.$$

<u>Proof</u>: Consider $E[(aX+Y)^2]$.

$$= E[(aX)^2 + 2aXY + Y^2]$$

$$= a^2 EX^2 + 2a E(XY) + EY^2 \ge 0. \quad \forall a \in \mathbb{R}.$$

So $\Delta = (2EXY)^2 - 4 EX^2 \cdot EY^2 \le 0.$

$$(E(XY))^2 \le EX^2 \cdot EY^2. \qquad \Rightarrow (E|XY|)^2 \le EX^2 \cdot EY^2.$$

$$-|XY| \le XY \le |XY| \Rightarrow -E|XY| \le E(XY) \le E|XY|. \Rightarrow |E(XY)| \le E|XY|.$$

6. $$|\rho| = \frac{|Cov(X,Y)|}{\sqrt{Var(X) \cdot Var(Y)}} \le 1.$$

$$|Cov(X,Y)| = |E[(X-EX)(Y-EY)]| \le \sqrt{E(X-EX)^2 \cdot E(Y-EY)^2}.$$

$$= \sqrt{Var X \cdot Var Y}.$$

If $\rho = \pm 1$, then $Y = a + bX$ almost every where for some $a, b \in \mathbb{R}$.

7. If $\rho = 0$, or $Cov(X,Y) = 0$. we say X and Y uncorrelated.

$$\text{Independent} \Rightarrow \text{Uncorrelated.} \quad \checkmark.$$

$$\not\Leftarrow.$$

"$\not\Leftarrow$": <u>Eg.</u> $\begin{cases} X \sim N(0,1) \\ Y = |X| \end{cases}$ Not independent.

$$Cov(X,Y) = E[(X-EX)(|X| - E|X|)]$$

$$= \underbrace{E(X \cdot |X|)}_{=0.} - \underbrace{EX}_{=0} \cdot E|X|.$$

$$= \int_{-\infty}^{\infty} x|x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0.$$

8. Recall the conditional p.d.f of X given $Y=y$ is

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

The conditional expectation is

$$E(X|Y=y) = \int x \cdot f_{X|Y}(x|y)\,dx$$

$$= \int x \,\frac{f(x,y)}{f_Y(y)}\,dx.$$

<u>Thm</u>. If X and Y are continuous R.V.s then

$$E(X) = \int E(X|Y=y)\, f_Y(y)\,dy$$

$$\iint x\, f(x,y)\,dx\,dy = \int \int x\, \frac{f(x,y)}{\cancel{f_Y(y)}}\,dx\, \cancel{f_Y(y)}\,dy.$$

Week 12. Lecture 15

<u>Thm</u>. (Total Probability Formula).

For any R.V. X, if Y is a continuous R.V. then

$$E(X) = \int E(X|Y=y)\, f_Y(y)\,dy.$$

If Y is a discrete R.V. $E(X) = \sum_m E(X|Y=m)\,P(Y=m).$

Example 1: A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that will take him to safety after 3 hours of travel. The second door leads to a tunnel that will return him to the mine after 5 hours of travel. The third door leads to a tunnel that will return him to the mine after 7 hours. If we assume that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until he reaches safety?

X = total time until safety.

$$= \begin{cases} 3 & \frac{1}{3} \leftarrow \text{first door.} \\ 5+\tilde{X} & \frac{1}{3} \leftarrow \text{second door.} \\ 7+\tilde{X} & \frac{1}{3} \leftarrow \text{third door.} \end{cases}$$

Y = first choice of door.

$$EX = \underset{\overset{\shortparallel}{3}}{E(X|Y=1)} \cdot \frac{1}{3} + \underset{\overset{\shortparallel}{5+EX}}{E(X|Y=2)} \cdot \frac{1}{3} + \underset{\overset{\shortparallel}{7+EX}}{E(X|Y=3)} \cdot \frac{1}{3}$$

$$\Rightarrow\ EX = \frac{1}{3}(3+5+7) + \frac{2}{3}EX \Rightarrow EX = 15.$$

Example 2: Suppose that the number of people entering a department store on a given day is a random variable with mean 50. Suppose further that the amounts of money spent by these customers are independent random variables having a common mean of 8. Finally, suppose also that the amount of money spent by a customer is also independent of the total number of customers who enter the store. What is the expected amount of money spent in the store on a given day?

Let N = number of people entering.

Let $X_1, X_2, \cdots, X_N$ be the amount of money spent by the customer i. Then $\text{Total} = \sum_{k=1}^{N} X_k$.

$$E(\text{Total}) = E\left(\sum_{k=1}^{N} X_k\right) = \sum_{n=0}^{\infty} E\left(\sum_{k=1}^{N} X_k \,\middle|\, N=n\right) P(N=n)$$

$$\overset{\text{indep.}}{=} E\left(\sum_{k=1}^{n} X_k \,\middle|\, \cancel{N=n}\right)$$

$$= \sum_{n=0}^{\infty} E\left(\sum_{k=1}^{n} X_k\right) P(N=n) = \sum_{n=0}^{\infty} 8n\, P(N=n) = 8 \sum_{n=0}^{\infty} n\, P(N=n) = 8\, EN = 8 \cdot 50 = 400.$$

/. Probability generating function (for discrete R.V.)

For a non-negative discrete R.V. X, define:

$$\forall s \in \mathbb{R}, \quad G_X(s) = E(s^X) = \sum_{n=0}^{\infty} s^n P(X=n).$$

$$= P(X=0) + s \cdot P(X=1) + s^2 \cdot P(X=2) + \cdots.$$

The series is absolutely convergent at least for $|s| \le 1$.

Eg. 1 If $X \sim \text{Bernoulli}(p)$, then

$$G_X(s) = 1-p + s \cdot p.$$

Eg. 2 If $X \sim \text{Bin}(n, p)$, then

$$G_X(s) = \sum_{k=0}^{n} s^k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (sp)^k (1-p)^{n-k}$$

$$= (sp + 1-p)^n.$$

Eg. 3 If $X \sim \text{Poisson}(\lambda)$, then

$$G_X(s) = \sum_{n=0}^{\infty} s^n \cdot e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(s\lambda)^n}{n!} = e^{-\lambda} e^{s\lambda} = e^{(s-1)\lambda}, \quad \forall s \in \mathbb{R}.$$

2. Properties of prob. generating function.

(a) $\forall k \geq 0$, $P(X=k) = \frac{1}{k!} G^{(k)}(0)$ ← k-th derivative.

Proof: $G(s) = P(X=0) + sP(X=1) + s^2 P(X=2) + \cdots$

$$G^{(k)}(s) = \left(\sum_{n=k}^{\infty} s^n P(X=n)\right)^{(k)} = \sum_{n=k}^{\infty} \left[n(n-1)\cdots(n-k+1)\, s^{n-k} P(X=n)\right].$$

$\Rightarrow G^{(k)}(0) = k!\, P(X=k)$. □.

(b) $G(1) = 1$ $\quad \left(G(1) = \sum_{n=0}^{\infty} P(X=n) = 1 \right).$

$G'(1) = EX$ $\quad \left(G'(s) = \sum_{n=1}^{\infty} n s^{n-1} P(X=n) \right).$ $(n=0 \quad 0 \cdot P(X=0) = 0)$

$G''(1) = E[X(X-1)]$. $\quad \left(G''(s) = \sum_{n=2}^{\infty} n(n-1) s^{n-2} P(X=n) \right)$ $\quad (n=0, 1, \quad n \cdot (n-1) \cdot P(X=n) = 0)$

$Var(X) = EX^2 - (EX)^2 = G''(1) + G'(1) - (G'(1))^2$.

(c). If $X_1, X_2, \cdots, X_n$ indep. then

for $Y = X_1 + \cdots + X_n$, we have

$$G_Y(s) = E(s^Y) = E(s^{X_1 + \cdots + X_n}) = E(s^{X_1} \cdots s^{X_n}) = E(s^{X_1}) \cdot E(s^{X_2}) \cdots E(s^{X_n}).$$

Eg. Recall: $T_{total} = 1 + X_1 + \cdots + X_{N-1}$. $\quad X_k \sim Geometric\left(\frac{N-k}{N}\right)$.

$X \sim Geometric(p)$. $\quad G_X(s) = \sum_{k=1}^{\infty} s^k (1-p)^{k-1} p = sp \frac{1}{1 - s(1-p)}.$

$$G_T(s) = \prod_{k=0}^{N-1} \frac{s \cdot \frac{N-k}{N}}{1 - s \cdot \frac{k}{N}} = \prod_{k=0}^{N-1} \frac{sN - sk}{N - sk} = s^N \prod_{k=0}^{N-1} \frac{N-k}{N-sk}.$$

3. Moment generating function.

Define $\quad M_X(t) = E(e^{tX}) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$. $\quad \forall t \in \mathbb{R}$.

Remark: $e^{tX} = \sum_{n=0}^{\infty} \frac{(tX)^n}{n!}$. $\Rightarrow E(e^{tX}) = E\left(\sum_{n=0}^{\infty} \frac{(tX)^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$.

Hence, $\forall k \geq 0$, $E(X^k) = M^{(k)}(0)$.

It's possible that $E(X^n) < \infty$, $\forall n \geq 0$, but $M_X(t) = \infty$.

eg. $P(X \leq x) = \begin{cases} 1 - e^{-\sqrt{x}}, & x \geq 0. \\ 0, & x < 0. \end{cases}$

$\Rightarrow$ p.d.f $= e^{-\sqrt{x}} \cdot \frac{1}{2\sqrt{x}}$, $x \geq 0$.

$\forall n \geq 0$, $EX^n = \int_0^\infty x^n \cdot e^{-\sqrt{x}} \frac{1}{2\sqrt{x}} \, dx < \infty$.

but $\forall t > 0$, $M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \cdot e^{-\sqrt{x}} \frac{1}{2\sqrt{x}} \, dx = \infty$

4. The characteristic function of X is defined by

$$\phi_X(t) = E(e^{itX}) = E(\cos tX + i \sin tX) \quad . \quad \forall t \in \mathbb{R}.$$

The c.d.f of X can be uniquely determined by $\phi(t)$.

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} d F(x).$$

Inversion formula: $\forall a < b \in \mathbb{R}$.

$$\frac{1}{2\pi} \lim_{T \to \infty} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi(t) \, dt = \frac{1}{2}(F(b) + F(b^-)) - \frac{1}{2}(F(a) + F(a^-))$$

$(= F(b) - F(a)$ if continuous $)$.

Eg 1. $X \sim$ Bernoulli (p) $\quad \phi_X(t) = E(e^{itX})$

$$= 1 \cdot P(X=0) + e^{it} \cdot P(X=1) = 1 - p + p \cdot e^{it}.$$

Eg 2. $X \sim Bin(n, p)$ $\quad \phi_X(t) = E(e^{itX})$

$$= \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} (p e^{it})^k (1-p)^{n-k}$$

$$= (p \cdot e^{it} + 1 - p)^n.$$

Eg 3. $X \sim \text{Poisson}(\lambda)$ $\phi_X(t) = \sum_{k=0}^{\infty} e^{itk} e^{-\lambda} \frac{\lambda^k}{k!}$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} (\lambda e^{it})^k = e^{-\lambda} e^{\lambda e^{it}}.$$

Eg 4. $X \sim N(\mu, \sigma^2)$. $\phi_X(t) = E(e^{itX})$

$$= \int_{-\infty}^{\infty} e^{itx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x^2 - 2\mu x + \mu^2 - 2\sigma^2 itx]} dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[x^2 - 2(\mu+\sigma^2 it)x + (\mu+\sigma^2 it)^2 - (\mu+\sigma^2 it)^2 + \mu^2]} dx.$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}[(x-\mu-\sigma^2 it)^2 - (\cancel{\mu^2} + 2\mu\sigma^2 it - \sigma^4 t^2) + \cancel{\mu^2}]} dx$$

$$= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x-\mu-\sigma^2 it)^2} dx}_{=1} \cdot e^{-\frac{1}{2\sigma^2}(-2\mu\sigma^2 it + \sigma^4 t^2)}$$

$$= e^{i\mu t - \frac{1}{2}\sigma^2 t^2}.$$

Week 13. Lecture 16.

Characteristic function. $\phi(t) = E(e^{itX}) = E\cos tX + i\,E\sin tX$, $\forall t \in \mathbb{R}$.

1. Properties of characteristic function.

(1) $\phi(0) = 1$. $|\phi(t)| \le 1$. $\phi(-t) = \overline{\phi(t)}$.

Proof: $|E(e^{itX})| \le E|e^{itX}| = 1$. $(|EX| \le E|X|)$.

$$\phi(-t) = E\cos(-tX) + i\,E\sin(-tX)$$

$$= E\cos tX - i\,E\sin tX = \overline{E\cos tX + i\,E\sin tX}.$$

(2) For any $k \ge 1$, if $E|X|^k < \infty$, then

$$\phi^{(k)}(0) = i^k EX^k.$$

Proof: $\frac{d}{dt}\phi(t) = \frac{d}{dt} E e^{itX} = \lim_{h\to 0} \frac{E e^{i(t+h)X} - E e^{itX}}{h}.$

$$= \lim_{h\to 0} E\left[\frac{e^{i(t+h)X} - e^{itX}}{h}\right].$$

Note $\left|\frac{e^{i(t+h)X} - e^{itX}}{h}\right| = \underbrace{|e^{itX}|}_{\le 1} \cdot \underbrace{\left|\frac{e^{ihX}-1}{h}\right|}_{=|iXe^{i\delta X}|,\ \delta\in(0,h).} \le |X|.$

and $E|X| < \infty$.

So $\frac{d}{dt}\phi(t) = E\left[\lim_{h\to 0} \frac{e^{i(t+h)X} - e^{itX}}{h}\right].$ Dominated Convergence Theorem.

(Similar process to using Weierstrass's Law).

$$= E\left[e^{itX}\cdot iX\right].$$

□.

Let $t=0$ to see $\phi'(0) = iEX$ $[k=1]$.

Remark $\frac{d}{dt} E e^{itX} = E(iXe^{itX})$. $\cdots$ $\frac{d^k}{dt^k} E e^{itX} = E\left[(iX)^k e^{itX}\right].$

$$t=0 \Rightarrow \phi^{(k)}(0) = E\left[(iX)^k e^{itX}\right].$$

(3). For any $t\in\mathbb{R}$. $|\phi(t+h) - \phi(t)| \le E|e^{itX} - 1|$.

So $\phi(t)$ is uniformly continuous on $\mathbb{R}$.

i.e. $\forall \varepsilon>0, \exists \delta>0, \forall t\in\mathbb{R}.$ $|f(t+h) - f(t)| < \varepsilon. \ \forall |h| < \varepsilon.$

Proof: $|\phi(t+h) - \phi(t)| = |E e^{i(t+h)X} - E e^{itX}|$

$$= E\left[e^{itX}\cdot(e^{ihX}-1)\right] \le E|e^{itX}\cdot(e^{ihX}-1)| = E|e^{itX}|\cdot|e^{ihX}-1| = E|e^{ihX}-1|.$$

Since $\lim_{h\to 0} E\left|e^{ihX}-1\right| = E\left[\lim_{h\to 0}\left|e^{ihX}-1\right|\right] = 0$ by $\left|e^{ihX}-1\right| \le \left|e^{ihX}\right|+1 = 2 < +\infty$.

$\forall \varepsilon>0, \exists \delta>0, \forall |h|<\delta, E\left|e^{ihX}-1\right| < \varepsilon.$ $\square$.

(4). If $X_1, \cdots, X_n$ are indep. then if we let $Y = X_1 + \cdots + X_n$, we have

$$\phi_Y(t) = E e^{itY} = \prod_{k=1}^{n} \phi_{X_k}(t).$$

<u>Proof</u>: $E e^{itY} = E\left[e^{itX_1} e^{itX_2} \cdots e^{itX_n}\right]$

indep.

$$= E e^{itX_1} \cdot E e^{itX_2} \cdots\cdots E e^{itX_n}$$

$$= \prod_{k=1}^{n} \phi_{X_k}(t). \qquad \square.$$

2. Multivariate Normal Distribution.

Let $Z_1, \cdots, Z_n \sim N(0,1)$ be independent.

Then $(Z_1, \cdots, Z_n)$ is a standard n-dimentional normal distribution.

Define
$$\underbrace{\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}}_{\vec{X}} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}}_{A.} \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix} + \underbrace{\begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}}_{\vec{\mu}}$$
to be a multivariate normal random vector.

We let $\vec{\mu} = (\mu_1, \cdots, \mu_n)$ and $\Sigma = AA^T$.

Then $E\, e^{i\vec{t}\cdot\vec{X}} = e^{i\vec{t}\cdot\vec{\mu} - \frac{1}{2}\vec{t}\,\Sigma\,\vec{t}^T}$.

In particular, if Σ is positive definite, then

$$f(\vec{X}) = \frac{1}{(\sqrt{2\pi})^n \sqrt{\det(\Sigma)}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})\Sigma^{-1}(\vec{X}-\vec{\mu})^T}, \quad \vec{X} = (X_1, X_2, \cdots, X_n)^T.$$

<u>Remark</u>. $\Sigma = \left(Cov(X_i, X_j)\right)_{1\le i,j\le n}.$

Proof: $E e^{i\vec{t}\cdot\vec{z}} = E e^{i(t_1 z_1 + t_2 z_2 + \cdots + t_n z_n)}$

$$= \prod_{k=1}^{n} E e^{i t_k z_k} = \prod_{k=1}^{n} \Big(e^{i t_k \cdot 0 - \frac{1}{2} t_k^2 \cdot 1} \Big) = e^{-\frac{1}{2}\sum_{k=1}^{n} t_k^2} = e^{-\frac{1}{2}\vec{t}\cdot\vec{t}^T}.$$

$N(0,1)$

Then, $E e^{i\vec{t}\cdot\vec{X}} = E e^{i\vec{t}(A\vec{z}+\vec{\mu})}$

$$= e^{i\vec{t}\vec{\mu}} \cdot E e^{i\underline{\vec{t}A}\vec{z}}$$

$$= e^{i\vec{t}\vec{\mu}} \cdot E e^{-\frac{1}{2}(\vec{t}A)(\vec{t}A)^T} = e^{i\vec{t}\vec{\mu}} \cdot e^{-\frac{1}{2}\vec{t}AA^T\vec{t}^T}.$$

$$\left.\begin{aligned} X_i &= \sum_{k=1}^{n} a_{ik} Z_k + \mu_i \\ X_j &= \sum_{l=1}^{n} a_{jl} Z_l + \mu_j \end{aligned}\right\} \quad \begin{aligned} Cov(X_i, X_j) &= E\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\ &= E\left[\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ik} a_{jl} Z_k Z_l\right]. \end{aligned}$$

$$E(Z_k Z_l) = \begin{cases} 1, & k = l \\ 0, & k \neq l. \end{cases} \Rightarrow Cov(X_i, X_j) = \sum_{k=1}^{n} a_{ik} a_{jk} = \sum_{k=1}^{n} A_{ik} A^T_{kj} = (AA^T)_{ij}. \quad \square$$

If $Cov(X_i, X_j) = 0 \quad \forall i \neq j.$

$\Leftrightarrow X_i, X_j$ indep.

then $\Sigma = \begin{pmatrix} \Sigma_{11} & & 0 \\ & \Sigma_{22} & \\ 0 & & \ddots \ \Sigma_{nn} \end{pmatrix}.$

Then one can check

$$E e^{i\vec{t}\cdot\vec{X}} = \prod_{k=1}^{n} E e^{i t_k X_k}$$

$$\| \qquad\qquad\qquad \|$$

$$e^{i\vec{t}\vec{\mu} - \frac{1}{2}\vec{t}\Sigma\vec{t}^T} = \prod_{k=1}^{n} e^{i t_k \mu_k - \frac{1}{2} t_k^2 \Sigma_{kk}}.$$

$$\Rightarrow E e^{i\vec{t}\cdot\vec{X}} = \prod_{k=1}^{n} E e^{i t_k X_k}$$

$$\Updownarrow$$

$X_1, \dots, X_n$ independent.

If $X \sim N(\mu, \sigma^2)$, then $E\, e^{itX} = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$

$$E\, e^{it_k X_k} = e^{it_k \mu_k - \frac{1}{2} t_k^2 \sigma_k^2}. \quad \text{where } \sigma_k^2 = E\left[(X_k - \mu_k)^2\right] = Cov(X_k, X_k) = \Sigma_{kk}.$$

3. If $X_1, \cdots, X_n$ are iid $N(\mu, \sigma^2)$, then

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_n. \quad \text{sample mean.}$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2. \quad \text{sample variance.}$$

$$E(\hat{\mu}) = \mu, \quad E(\hat{\sigma}^2) = \sigma^2.$$

$$E(\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu.$$

$$(n-1)\hat{\sigma}^2 = \sum_{i=1}^{n} (X_i - \hat{\mu})^2 = \sum_{i=1}^{n} (X_i^2 - 2\hat{\mu} X_i + \hat{\mu}^2) = \sum_{i=1}^{n} X_i^2 - 2\hat{\mu} \underbrace{\sum_{i=1}^{n} X_i}_{= n\hat{\mu}} + n\hat{\mu}^2.$$

$$= \sum_{i=1}^{n} X_i^2 - n\hat{\mu}^2.$$

$$n\hat{\mu}^2 = \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 + \frac{1}{n} \sum_{i \neq j} X_i X_j.$$

So $(n-1)\hat{\sigma}^2 = (1 - \frac{1}{n}) \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \sum_{i \neq j} X_i X_j.$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \frac{1}{n-1} \sum_{i \neq j} X_i X_j.$$

$$E(\hat{\sigma}^2) = \frac{1}{n} \sum_{i=1}^{n} E X_i^2 - \frac{1}{n} \frac{1}{n-1} \sum_{i \neq j} E X_i X_j.$$

$$= \frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \mu^2) - \frac{1}{n} \frac{1}{n-1} n(n-1) \cdot \mu^2$$

$$= \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

4. $X_1, \cdots, X_n \sim N(0,1)$. i.i.d.

Define $Y = X_1^2 + \cdots + X_n^2 \sim \chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$

It suffices to prove that $X_1^2 \sim \Gamma(\frac{1}{2}, \frac{1}{2})$.

Then $X_k^2 \sim \Gamma(\frac{1}{2},\frac{1}{2})$, $\forall k \le n$.

$$\Rightarrow \sum_{k=1}^{n} X_k^2 \sim \Gamma(\tfrac{n}{2},\tfrac{1}{2}).$$

$$\Gamma(\alpha,\lambda) \sim \lambda e^{-\lambda x} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}.$$

$$\Gamma(\tfrac{1}{2},\tfrac{1}{2}) \sim \tfrac{1}{2} e^{-\frac{1}{2}x} \frac{(\frac{1}{2}x)^{-\frac{1}{2}}}{\underset{\sqrt{\pi}}{\overset{\shortparallel}{\Gamma(\frac{1}{2})}}} = x^{-\frac{1}{2}} \frac{\frac{1}{\sqrt{2}}}{\sqrt{\pi}} e^{-\frac{1}{2}x} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}}.$$

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty x^{\frac{1}{2}-1} e^{-x} dx \overset{t=\sqrt{x}}{\underset{x=t^2}{=\!=}} \int_0^\infty \frac{1}{t} e^{-t^2} 2t\, dt = 2\int_0^\infty e^{-t^2} dt$$

$$= \int_{-\infty}^{\infty} e^{-t^2} dt. = \sqrt{\pi}$$

$$\uparrow \sigma^2 = \tfrac{1}{2}.$$

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx.$$

$$P(X_1^2 \le x) = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2\int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \ \forall x > 0.$$

Take derivative to get the p.d.f of X_1^2.

$$2 \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{x})^2}{2}} \cdot \frac{1}{2\sqrt{x}} = \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \sim \Gamma(\tfrac{1}{2},\tfrac{1}{2})$$

Week 14 Lecture 17.

1. (Markov's Inequality)

For any R.V. X, we have

$$\boxed{P(|X| \ge t) \le \frac{E|X|}{t}, \ \forall t > 0.}$$

Proof: $1_{|X|\ge t} \le \frac{|X|}{t} 1_{|X| \ge t}$.

$$\begin{cases} |X| < t: & 0 \le 0. \ \checkmark. \\ |X| \le t: & 1 \le \frac{|X|}{t} \ \checkmark. \end{cases}$$

In fact, $1_{|X|\ge t} \le \frac{|X|}{t}$

$$\Rightarrow E(1_{|X|\ge t}) \le \frac{E|X|}{t}.$$

$$P(|X| \ge t) = E(1_{|X|\ge t}) \le E\left(\frac{|X|}{t} 1_{|X|\ge t}\right)$$

$$\le E\left(\frac{|X|}{t}\right) = \frac{E|X|}{t}. \quad \square.$$

2. (Chebyshev's Inequality).

For any R.V. X with mean μ, variance σ^2.

we have $\boxed{P(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}. \ \forall t>0.}$ $\qquad P(|X-EX| \geq t) \leq \frac{\mathrm{Var} X}{t^2}.$

Proof: $P(|X-\mu| \geq t)$ $t>0.$

$= P(|X-\mu|^2 \geq t^2)$

(Markov's Inequality). $\leq \frac{E|X-\mu|^2}{t^2} = \frac{\sigma^2}{t^2}.$ $\square$

Eg 1. If $X \sim \mathrm{Unif}(0,10)$, then

$\mu = EX = 5$, $\sigma^2 = E(X-\mu)^2 = \frac{25}{3}$.

Estimate $P(X>9) = \int_9^{10} \frac{1}{10} dx = \frac{1}{10}.$

(1) Using Markov's Inequality.

$$P(X>9) = P(|X| \geq 9) \leq \frac{E|X|}{9} = \frac{5}{9} \approx 0.56.$$

(2) Using Chebyshev's Inequality.

$$P(X>9) = P(X-5 \geq 4) \leq P(|X-5| \geq 4) \leq \frac{\sigma^2}{4^2} = \frac{25}{48} \approx 0.52.$$

3. If $\mathrm{Var}(X)=0$, the $P(X=\mu)=1$.

In fact, if $E|X-\mu|^p = 0$ for some $p>0$, then $X=\mu$ a.s.

Proof: $\forall n \geq 1$, $P(|X-\mu| \geq \frac{1}{n}) = P(|X-\mu|^p \geq \frac{1}{n^p}) \leq \frac{E|X-\mu|^p}{\frac{1}{n^p}} = 0.$

Markov Inequality.

$$P\left(\bigcup_{n=1}^{\infty}\{|X-\mu|\geq \tfrac{1}{n}\}\right) \leq \sum_{n=1}^{\infty} P(|X-\mu|\geq \tfrac{1}{n}) = 0. \Rightarrow P\left(\bigcap_{n=1}^{\infty}\{|X-\mu|<\tfrac{1}{n}\}\right)=1.$$

$$\{X=\mu\} = \bigcap_{n=1}^{\infty}\{|X-\mu|<\tfrac{1}{n}\} \quad \Rightarrow P(X=\mu)=1. \quad \square.$$

$\subseteq$ ✓

$\supseteq$ $|X-\mu|<\frac{1}{n}\to 0$ $(n\to\infty)$.

4. Weak Law of Large Numbers.

Let $X_1, X_2, \cdots$ i.i.d. R.V. s. with mean μ and varience σ^2, then

$$\forall \varepsilon>0, \lim_{n\to\infty} P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|>\varepsilon\right)=0.$$

<u>Proof:</u> By Chebyshev's Inequality.

$$P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|>\varepsilon\right) = \leq \frac{Var\left(\frac{X_1+\cdots+X_n}{n}\right)}{\varepsilon^2}$$

Recall $Var(a+bX)=b^2\,Var(X)$. $\forall a,b\in\mathbb{R}$.

So $Var\left(\frac{X_1+\cdots+X_n}{n}\right)=\frac{1}{n^2}Var(X_1+\cdots+X_n)$.

By independence of $X_1,\cdots,X_n$, $Var(X_1+\cdots+X_n)=Var(X_1)+\cdots+Var(X_n)=n\sigma^2$.

So $P\left(\left|\frac{X_1+X_2+\cdots+X_n}{n}-\mu\right|>\varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \to 0$ $(n\to\infty)$.

5. Four Convergence Modes.

(1) Convergence in probability.

Let $X_1,\cdots,X_n,\cdots$ be a sequence of R.V.s.

We say X_n converges to X in probability if

$$\forall \varepsilon>0, \ P(|X_n-X|>\varepsilon)\to 0 \ (n\to\infty).$$

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

$X_n \xrightarrow{P} X$.

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

Eg. $\frac{X_1+\cdots+X_n}{n}\xrightarrow{P}\mu$.

(2). a.s. convergence.

$$P(\lim_{n\to\infty} X_n = X) = 1.$$

w.p.1. (strictly equals 1).

$X_n(w) \longrightarrow X(w)$. (Sample Space: $S = (w_1, w_2, \cdots)$, w_i: experiment. $X_n(w)$: when w happens. X_n represents what number).

$X_n: S \to \mathbb{R}$

$w \mapsto X_n(w)$.

$X_n(w) \longrightarrow X(w)$ $(n\to\infty)$ as real numbers.

$$\{\lim_{n\to\infty} X_n = X\} = \bigcap_{\varepsilon>0} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|X_n - X| < \varepsilon\}.$$

$$\Updownarrow$$

$$\forall \varepsilon>0,\ \exists N\geq 1,\ \forall n\geq N,\ |X_n - X| < \varepsilon.$$

<u>Remark</u>: $\forall x \in \bigcap_{\varepsilon>0} A_\varepsilon$, we have $\forall \varepsilon>0$, $x\in A_\varepsilon$.

So a.s. convergence implies

$$P\left(\bigcap_{\varepsilon>0} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \{|X_n - X| < \varepsilon\}\right) = 1.$$

$$\Leftrightarrow P\left(\bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} |X_n - X| < \tfrac{1}{k}\right) = 1.$$

$$\Leftrightarrow P\left(\bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} |X_n - X| \geq \tfrac{1}{k}\right) = 0.$$

<u>Recall</u>: $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_k \subseteq \cdots$

then $P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} P(A_k)$.

Let $A_k = \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} \{|X_n - X| > \tfrac{1}{k}\}$.

then $A_k \subseteq A_{k+1}$. So $P\left(\bigcup_{k=1}^{\infty} A_k\right) = 0 \Rightarrow P(A_k) = 0$. $\forall k\geq 1$.

Thus, a.s. convergence $\Rightarrow$ $\forall k\geq 1$, $P\left(\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} |X_n - X| > \tfrac{1}{k}\right) = 0$.

Let $B_N = \bigcup_{n=N}^{\infty} |X_n - X| > \varepsilon$.

then $B_N \supseteq B_{N+1} \supseteq B_{N+2} \supseteq \cdots$. So $P\left(\bigcap_{N=1}^{\infty} B_N\right) = \lim_{N\to\infty} P(B_N)$.

$$\forall \varepsilon>0, \quad \lim_{N\to\infty} P\Big(\bigcup_{n=N}^{\infty} |X_n-X|>\varepsilon\Big)=0. \Rightarrow \lim_{N\to\infty} P(|X_N-X|>\varepsilon)=0.$$

So a.s convergence implies convergence in probability.

<u>Upshot</u>: $X=\mu$ a.s. $\iff$ $P(X=\mu)=1$.

$X_n \to X$ a.s. $\iff$ $P(X_n\to X)=1$.

$\uparrow$ converge as real numbers.

(3) Convergence in distribution. $(X_n \xrightarrow{d} X.)$

<u>Def</u>: For any $x\in\mathbb{R}$ s.t. x is a continuity point of $F_X(b)=P(X\le b)$. we have

$$\lim_{n\to\infty} P(X_n\le x) = P(X\le x) \qquad \Big(\lim_{n\to\infty} E(1_{X_n\le x}) = E\,1_{X\le x}\Big).$$

(a) Why only continuity point: $X_n=\frac{1}{n}$. $X=0$.

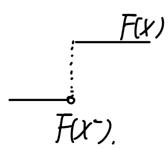
$$P(X_n\le 0)=0. \;\not\to\; P(X\le 0)=1. \qquad F_X(b)=\begin{cases}1, & b\ge 0\\ 0, & b<0.\end{cases} = P(X\le b).$$

(b) Discontinuity point is at most countable.

($\forall x\in\mathbb{R}$, $F_n(x)\to F(x)$ is too strong.).

Let $D=\{x\in\mathbb{R}: F(x)$ is discontinuous at $x\}$.

So $\forall x\in D$, $F(x)>F(x^-)$



$\Rightarrow \exists N\ge 1$, $F(x)-F(x^-)>\frac{1}{n}$.

$D=\bigcup_{n=1}^{\infty} D_n$, with $D_n=\{x\in\mathbb{R}: F(x)-F(x^-)>\frac{1}{n}\}$.

By $\lim_{x\to\infty}F(x)=1$, $\lim_{x\to-\infty}F(x)=0$, we get $|D_n|<\infty$.

So $D=\bigcup_{n=1}^{\infty} D_n$ is countable.

Equivalent Definition: (Weak Convergence).

$\forall$ bounded continuous function $f: \mathbb{R} \to \mathbb{R}$.

$$\lim_{n\to\infty} E[f(X_n)] = E[f(X)].$$

(4). L^p convergence for $p>0$:

$$\lim_{n\to\infty} E|X_n - X|^p = 0.$$

Upshot:

$$\begin{array}{l} a.s \xrightarrow{\checkmark} p \longrightarrow d. \\ L^p \nearrow \end{array}$$

Strong $\longrightarrow$ Weak.

6. L^p convergence $\Rightarrow$ Convergence in probability.

If $E|X_n - X|^p \to 0$ $(n\to\infty)$, then $\forall \varepsilon > 0$.

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^p > \varepsilon^p) \underset{\uparrow}{\leq} \frac{E|X_n - X|^p \to 0}{\varepsilon^p} \to 0 \ (n\to\infty).$$

(Markov Inequality).

Eg. If $X_n \xrightarrow{d} c$, then $X_n \xrightarrow{P} c$, where c is a constant.

$$F_c(X) = \begin{cases} 1, & (x \geq c) \\ 0, & (x < c) \end{cases} = P(c \leq X).$$

So $\forall x \neq c$, $P(X_n \leq x) \to P(X \leq x)$

$$\begin{cases} x \geq c. & P(X_n \leq x) \to 1. \\ x < c, & P(X_n \leq x) \to 0. \end{cases}$$

$\forall \varepsilon > 0$, $P(|X_n - c| > \varepsilon)$

$= P(X_n - c > \varepsilon \text{ or } X_n - c < -\varepsilon)$

$\leq P(X_n - c > \varepsilon) + P(X_n - c < -\varepsilon)$

$= 1 - P(X_n \leq c + \varepsilon) + P(X_n < c - \varepsilon) \to 1 - 1 + 0 = 0.$

Week 14. Lecture 18.

$$\begin{array}{l} a.s \searrow \\ \qquad\quad p \xrightarrow{\not\Leftarrow} d. \\ L^p \nearrow \end{array}$$

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

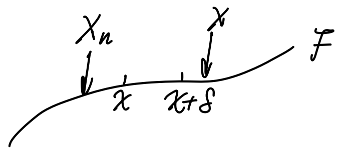
$X_n \xrightarrow{d} X$ if: $\forall$ continuity point x of the c.d.f of X, $\lim_{n\to\infty} P(X_n \le x) = P(X \le x)$.

1. $X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$

Proof: $\forall x$ is a continuity point of the c.d.f $F(x)$ of X.

we will prove $F_n(x) = P(X_n \le x) \to F(x)$. $(n \to \infty)$.

$$F_n(x) - F(x) = P(X_n \le x) - P(X \le x) = E(1_{X_n \le x} - 1_{X \le x}).$$



$$P(X_n \le x) = P(X_n \le x, X > x+\delta) + P(X_n \le x, X \le x+\delta)$$

$$\le P(|X_n - X| > \delta) + P(X \le x+\delta)$$

$$\Rightarrow P(X_n \le x) - P(X \le x) \le P(|X_n - X| > \delta) + P(X \le x+\delta) - P(X \le x)$$

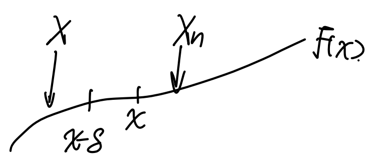
$$= P(|X_n - X| > \delta) + F(x+\delta) - F(x)$$

$$\Rightarrow \overline{\lim_{n\to\infty}} \big(P(X_n \le x) - P(X \le x)\big) \le F(x+\delta) - F(x)$$

$$F(x) - F_n(x) = P(X \le x) - P(X_n \le x)$$

$$= 1 - P(X > x) - (1 - P(X_n > x))$$

$$= P(X_n > x) - P(X > x).$$



$$P(X_n > x) = P(X_n > x, X \le x-\delta) + P(X_n > x, X > x-\delta)$$

$$\le P(|X_n - X| > \delta) + P(X > x-\delta)$$

$$F(x) - F_n(x) \le P(|X_n - X| > \delta) + 1 - F(x-\delta) - (1 - F(x))$$

$$= P(|X_n - X| > \delta) + F(x) - F(x-\delta)$$

$$\Rightarrow \overline{\lim_{n\to\infty}} \big(F(x) - F_n(x)\big) \le F(x) - F(x-\delta)$$

So $$F(x-\delta) - F(x) \le \underline{\lim_{n\to\infty}} (F_n(x) - F(x)) \le \overline{\lim_{n\to\infty}} (F_n(x) - F(x)) \le F(x+\delta) - F(x).$$

Let $\delta \to 0^+$ to get

$$0 \le \varliminf_{n\to\infty} (F_n(x) - F(x)) \le \varlimsup_{n\to\infty} (F_n(x) - F(x)) \le 0.$$

by x is a continuity point of $F(x)$. $\square$

<u>Second proof</u>: prove by using $X_n \xrightarrow{d} X \overset{\text{equivalent def.}}{\Longleftrightarrow} \forall$ bounded continuous function $f: \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} E[f(X_n)] = E[f(X)].$$

Use $P(|X_n - X| > \delta) \to 0 \quad \forall \delta > 0, \ (n \to \infty)$.

$$E|f(X_n) - f(X)| = \underbrace{E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| > \delta}\right]}_{\to 0 \ (n\to\infty)} + E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta}\right]$$

Pick $M > 0$ large s.t.

$$E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta}\right] = E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta} 1_{|X| \ge M}\right] + E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta} 1_{|X| < M}\right].$$

$\forall \varepsilon > 0, \exists M = M(\varepsilon) > 0$, s.t. $P(|X| > M) = \varepsilon$.

For f on $[-M-1, M+1]$, f continuous $\Rightarrow f$ uniformly continuous.

$\exists \delta = \delta(\varepsilon) \in (0,1)$, s.t. $\forall |x-y| < \delta$, $|x|, |y| \le M+1$, $|f(x) - f(y)| < \varepsilon$.

$$\forall n \ge 1, \quad E|f(X_n) - f(X)| = \underbrace{E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| > \delta}\right]}_{I_1} + \underbrace{E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta} 1_{|X| \le M}\right]}_{I_2}$$
$$+ \underbrace{E\left[|f(X_n) - f(X)| \, 1_{|X_n - X| \le \delta} 1_{|X| > M}\right]}_{I_3}.$$

$$I_1 \le 2\|f\|_\infty \cdot \underbrace{E\left[1_{|X_n - X| > \delta}\right]}_{= P(|X_n - X| \ge \delta)} \qquad \|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)| < \infty.$$

$\exists N = N(\delta) = N(\delta, M) = N(\varepsilon) > 0$, $\forall n \ge N$, $P(|X_n - X| \ge \delta) < \varepsilon$.

$$I_2 \le E\left[\underbrace{|f(X_n)-f(X)|}_{<\varepsilon}\ 1_{|X_n|\le M+1}\ 1_{|X|\le M+1}\right] \le \varepsilon.$$

$$I_3 \le E\left[2\|f\|_\infty\ 1_{|X|>M}\right] = 2\|f\|_\infty\ P(|X|>M) \le 2\|f\|_\infty\ \varepsilon$$

So $I_1+I_2+I_3 \le [4\|f\|_\infty+1]\,\varepsilon. \quad \forall n>N.$

i.e. $\forall \varepsilon>0,\ \exists N=N(\varepsilon),$ s.t. $\forall n\ge N,\quad E|f(X_n)-f(X)|<\varepsilon.$

$\Leftrightarrow \lim_{n\to\infty} E\left[\,|f(X_n)-f(X)|\,\right]=0.$

$\Rightarrow \left|E[f(X_n)]-E[f(X)]\right| = \left|E[f(X_n)-f(X)]\right| \le E\left[|f(X_n)-f(X)|\right] \longrightarrow 0 \quad (n\to+\infty)$ $\square$.

Upshot:

$$\begin{tikzcd} a.s \arrow[dr, "\checkmark"] & & \\ & p \arrow[r, "\checkmark"] & d \arrow[l, bend left] \\ \mathcal{L}^p \arrow[ur, "\checkmark"] & & \end{tikzcd}$$

If the limiting R.V. is a constant. i.e. $X_n \xrightarrow{d} C \Leftrightarrow X_n \xrightarrow{P} C$.

<u>Counter Examples:</u>

<u>Eg 1.</u> $X_n \xrightarrow{d} X \not\Rightarrow X_n \xrightarrow{P} X$:

$X_n = Z \sim N(0,1), \quad \forall n\ge 1.$

$X = -Z \sim N(0,1).$

$P(X_n\le x) = P(Z\le x) = P(-Z\le x) = P(X\le x).$ So $X_n \xrightarrow{d} X.$

but $\forall \varepsilon>0,\ P(|X_n-X|>\varepsilon) = P(2|Z|>\varepsilon) = P(|Z|>\frac{\varepsilon}{2})$

$$=\left(\int_{\frac{\varepsilon}{2}}^{+\infty}+\int_{-\infty}^{-\frac{\varepsilon}{2}}\right)\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \not\to 0.$$

<u>Remark</u>: We compare X_n and X directly in a.s, $\mathcal{L}^p$, p convergence.

But we don't require this in d. convergence

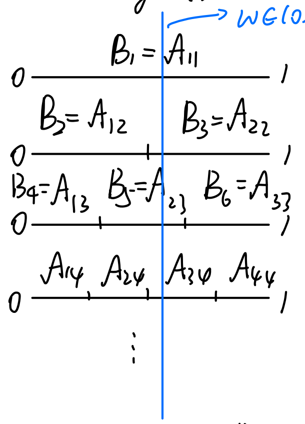
$\Rightarrow$ in a.s, $\mathcal{L}^p$, p: X_n, X are in the same sample space but in d they may not!

Eg 2: $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{a.s} X$.

$S = (0,1)$. P = Lebesgue measure.

$P((0,\frac{1}{2})) = \frac{1}{2}$ = length of the integral.

$\rightarrow w \in (0,1)$ take this w as an example.



$X_n : (0,1) \to \mathbb{R}$.

$X_n = 1_{B_n}$, $\forall n \geq 1$ where

$(B_1, B_2, B_3, \cdots) = (A_{11}, A_{12}, A_{22}, A_{13}, \cdots)$

$A_{jk} = \left[\frac{j-1}{k}, \frac{j}{k}\right]$

$\forall k \geq 1$, $1 \leq j \leq k$.

$\forall w \in (0,1) \in S$, $X_n(w) = 1_{B_n}(w) = 1_{w \in B_n} = \begin{cases} 1, & w \in B_n \\ 0, & w \in B_n^c. \end{cases}$

$X : S \to \mathbb{R}$, $X(w) = 0$, $\forall w \in S$.

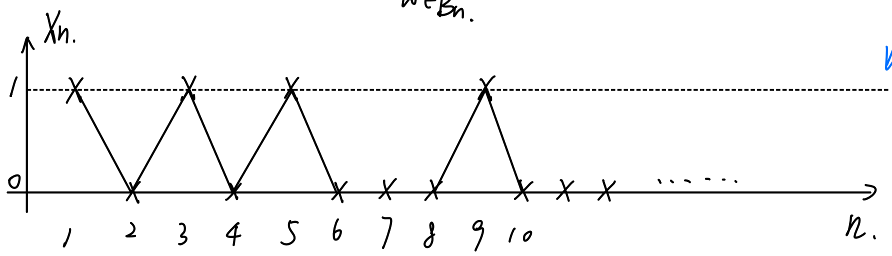
We will prove $X_n \xrightarrow{P} X$ but $X_n \not\xrightarrow{a.s} X$.

$\forall \varepsilon > 0$, $P(|X_n - X| > \varepsilon) = P(1_{B_n} > \varepsilon)$ $(= 0$ if $\varepsilon \geq 1)$.

$\forall \varepsilon \in (0,1)$, $P(|X_n - X| > \varepsilon) = P(1_{B_n} > \varepsilon)$

$\leq P(B_n) = |A_{jk}| = \frac{1}{k} \to 0$.

Next, $\forall w \in (0,1)$, $X_n(w) = 1_{w \in B_n}$.



never convergent to 0 since there will always have large n, $X_n(w) = 1$.

Thus, $X_n(w) = 1_{w \in B_n} \not\to 0$.

Eg 3. $X_n \xrightarrow{L^P} X \not\Rightarrow X_n \xrightarrow{a.s} X$.

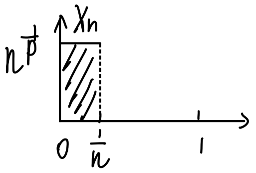
Same as $X_n \xrightarrow{P} X \not\Rightarrow X_n \xrightarrow{a.s} X$.

$\forall p > 0$, $E|X_n - X|^P = E|1_{B_n}|^P = E\, 1_{B_n} = P(B_n) \to 0$.

Eg 4. $X_n \xrightarrow{a.s} X \not\Rightarrow X_n \xrightarrow{L^p} X$.

$S = (0,1)$. P = Lebesgue measure

$X_n : (0,1) \to \mathbb{R}$. $X_n(w) = n^{\frac{1}{p}} \mathbb{1}_{w \in (0, \frac{1}{n})}$.



$X(w) = 0$.

$X_n \xrightarrow{a.s} X \not\Rightarrow X_n \xrightarrow{L^p} X$.

$\forall w \in (0,1)$, $\exists N \geq 1$, s.t $\forall n \geq N$, $w > \frac{1}{n}$. So $X_n(w) = n^{\frac{1}{p}} \mathbb{1}_{w \in (0, \frac{1}{n})} = 0$. $\forall n \geq N$.

$\Rightarrow X_n(w) \to 0 = X(w)$. so $X_n \xrightarrow{a.s.} X$.

$$E|X_n - X|^p = E|X_n|^p = P(X_n = n^{\frac{1}{p}})(n^{\frac{1}{p}})^p = n \times \frac{1}{n} = 1.$$

$$= P(w \in (0, \tfrac{1}{n})) \times n$$

$$= \tfrac{1}{n} \times n = 1. \quad \text{So } X_n \not\xrightarrow{L^p} X.$$

Week 15. Lecture 19.

Table 2: cdf of normal distribution

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

$\sim\!\!\to$ $P(Z \leq x)$, $Z \sim N(0,1)$. $x > 0$.

$P(Z \geq 2.1) = 1 - P(Z \leq 2.1) = 1 - 0.9821 = 0.0179$.

$P(-1.4 < Z < 3.05) = P(Z \leq 3.05) - P(Z \leq -1.4)$

$= 0.9989 - (1 - P(Z \leq 1.4))$

$= 0.9989 - (1 - 0.9192)$

$= 0.9181$.

1. (Central Limit Theorem)

Let $X_1, X_2, \cdots$ be a sequence of i.i.d R.V.s, with mean μ and variance σ^2.

Then
$$\lim_{n \to \infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} \leq x \right) = P(Z \leq x), \quad \forall x \in \mathbb{R},$$

where $Z \sim N(0,1)$. That is

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} Z.$$

Idea:

Weak Law of Large Numbers. $\frac{\sum_{k=1}^{n} X_k}{n} \xrightarrow{P} \mu.$

Strong Law of Large Numbers. $\frac{\sum_{k=1}^{n} X_k}{n} \xrightarrow{a.s.} \mu.$

$\Rightarrow \sum_{k=1}^{n} X_k \approx n\mu$ if n is large.

C.L.T: $\sum_{k=1}^{n} X_k - n\mu = O(\sqrt{n})$.

$\Rightarrow \frac{\sum_{k=1}^{n} X_k}{n} - \mu = O(\frac{1}{\sqrt{n}})$

$\Rightarrow \sum_{k=1}^{n} X_k \approx n\mu + C\sqrt{n}.$

2. (Continuity Theorem)

Let $\xi_1, \xi_2, \cdots$ be a sequence of R.V.s with characteristic function $\phi_n(t) = E e^{it\xi_n}$.

If $\phi_n(t) \to \phi(t)$, $\forall t \in \mathbb{R}$ and $\phi(t) = Ee^{itX}$ for some R.V. X and $\phi(t)$ is continuous at 0.

Then $\xi_n \xrightarrow{d} X$.

3. Proof of C.L.T.:

We first assume that $\mu = 0$, $\sigma^2 = 1$. Let $\xi_n = \frac{\sum_{k=1}^{n} X_k}{\sqrt{n}}$ for each $n \geq 1$.

We'll prove $\forall t \in \mathbb{R}$.

$$\phi_n(t) = E e^{it\xi_n} \longrightarrow \phi_Z(t) = e^{-\frac{1}{2}t^2}.$$

[Recall: $X \sim N(\mu, \sigma^2)$. then $E e^{itX} = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$]

$\forall n \geq 1$, $\phi_n(t) = E e^{it\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k} = E\left(\prod_{k=1}^{n} e^{it\frac{X_k}{\sqrt{n}}}\right) \overset{\text{indep.}}{=} \prod_{k=1}^{n} E e^{it\frac{1}{\sqrt{n}}X_k}$.

Let $\phi(t) = E(e^{it X_1})$, then $\phi_n(t) = (\phi(\frac{t}{\sqrt{n}}))^n$

$\Rightarrow \phi(\frac{t}{\sqrt{n}}) = E(e^{i\frac{t}{\sqrt{n}}X_1})$ Note $n \to \infty$, $\frac{t}{\sqrt{n}} = o(1)$.

By Taylor expansion of $\phi(t)$ at 0, we get

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + o(t^2).$$

Here $o(t^2)$ satisfies $\lim_{t\to 0} \frac{o(t^2)}{t^2} = 0$.

$$\phi(0) = E e^{i0X_1} = 1.$$

$$\phi'(t) = \frac{d}{dt}\left[E(e^{itX_1})\right] = E\left(\frac{d}{dt}e^{itX_1}\right) = E(iX_1 e^{itX_1})$$

$$\Rightarrow \phi'(0) = E(iX_1) = iEX_1 = i\mu = 0$$

$$\phi''(t) = \frac{d}{dt}\left[E(iX_1 e^{itX_1})\right] = E\left(\frac{d}{dt} iX_1 e^{itX_1}\right) = E(-X_1^2 e^{itX_1}) = -E(X_1^2 e^{itX_1})$$

$$\Rightarrow \phi''(0) = -EX_1^2 = -\sigma^2 = -1.$$

So $\phi(t) = 1 - \frac{1}{2}t^2 + o(t^2)$, $(t\to 0)$.

$$\Rightarrow \forall t\in\mathbb{R},\ \phi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right),\ (n\to\infty).$$

$$\phi_n(t) = \left(\phi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left[1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right]^n \xrightarrow{n\to\infty} e^{-\frac{1}{2}t^2}.$$

$$= e^{\ln\left[1 - \frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right]\cdot n}$$

$$= e^{\left[-\frac{1}{2}\frac{t^2}{n} + o\left(\frac{t^2}{n}\right) + o\left(\frac{t^2}{n}\right)\right]\cdot n}$$

$$= e^{-\frac{1}{2}t^2 + o(t^2)} \longrightarrow e^{-\frac{1}{2}t^2}.$$

$$\left(1+\frac{x}{n}\right)^n \longrightarrow e^x,\ \forall x\in\mathbb{R}.$$

$$\overset{\parallel}{e}^{\ln\left(1+\frac{x}{n}\right)\cdot n} \qquad \ln(1+t) = t - \frac{1}{2}t^2 + \frac{1}{3}t^3 - \frac{1}{4}t^4 + \cdots = t + o(t)\ (t\to 0).$$

$$= e^{\left[\frac{x}{n} + o\left(\frac{x}{n}\right)\right]\cdot n}$$

$$= e^{x + o(x)} \to e^x.$$

Thus, if $\mu=0$, $\sigma^2=1$, then $\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \xrightarrow{d} Z$.

For general μ and σ^2, we let $Y_k = \frac{X_k-\mu}{\sigma}$, $\forall k\geq 1$.

Then $EY_k = 0$, $EY_k^2 = \frac{E(X_k-\mu)^2}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1$.

So $\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k \xrightarrow{d} Z$ $\square$.

$$\frac{\sum_{k=1}^n (X_k-\mu)}{\sqrt{n\sigma^2}} = \frac{1}{\sqrt{n\sigma^2}}\left[\sum_{k=1}^n (X_k-\mu)\right] \xrightarrow{d} Z.$$

$$n\sigma^2 = Var\left(\sum_{k=1}^n (X_k-\mu)\right) = \sum_{k=1}^n Var(X_k-\mu)$$

Eg1. Let $X \sim \text{Poisson}(100)$.

Find $P(X \geq 120) = \sum_{k=120}^{\infty} P(X=k)$

$$= \sum_{k=120}^{\infty} e^{-100} \frac{100^k}{k!}.$$

Recall: $X_1 \sim \text{Poisson}(1)$

$X_2 \sim \text{Poisson}(1)$ indep. $\Rightarrow X_1 + X_2 \sim \text{Poisson}(2)$.

$\vdots$

$X_{100} \sim \text{Poisson}(1)$. $\Rightarrow \sum_{k=1}^{100} X_k \sim \text{Poisson}(100)$.

So $P(X \geq 100) = P\left(\sum_{k=1}^{100} X_k \geq 120\right)$.

Continuity correction. $= P\left(\sum_{k=1}^{100} X_k > 119.5\right)$.

$$= P\left(\frac{\sum_{k=1}^{100} X_k - 100 \cdot 1}{\sqrt{100 \cdot 1}} > \frac{119.5 - 100}{\sqrt{100}} \right).$$

$$\frac{\sum_{k=1}^{100} X_k - n\mu}{\sqrt{n\alpha^2}} \xrightarrow{d} Z.$$

$\approx P(Z > 1.95)$

$= 1 - P(Z \leq 1.95)$

$= 1 - 0.9744 = 0.0256.$

Why continuity correction?

$$P\left(\sum_{k=1}^{100} X_k \geq 120\right) + P\left(\sum_{k=1}^{100} X_k \leq 119\right) = 1.$$

$$\approx P\left(Z \geq \underbrace{\frac{120-100}{\sqrt{100}}}_{=2}\right) + P\left(Z \geq \underbrace{\frac{119-100}{\sqrt{100}}}_{=1.9}\right) < 1$$

$$P\left(\sum_{k=1}^{100} X_k > 119.5\right) + P\left(\sum_{k=1}^{100} X_k < 119.5\right) = 1$$

$\approx P(Z > 1.95) + P(Z < 1.95) = 1.$

5. Chernoff bound.

$\forall a > 0, \quad P(X \geq a) \leq E\left(\frac{e^{tX}}{e^{ta}}\right), \quad \forall t > 0.$

$P(X \leq a) \leq E\left(\frac{e^{tX}}{e^{ta}}\right), \quad \forall t < 0.$

Proof: $\forall t > 0, \quad P(X \geq a) = P(tX \geq ta)$

$$= P(e^{tX} \geq e^{ta}) \leq \frac{E e^{tX}}{e^{ta}}.$$

$\forall t < 0, \quad P(X \leq a) = P(tX \geq ta)$

$$= P(e^{tX} \geq e^{ta}) \leq \frac{E e^{tX}}{e^{ta}}.$$

Application: For $Z \sim N(0,1)$.

$$P(Z \geq a) \leq \frac{E e^{tX}}{e^{ta}} = \frac{e^{\frac{1}{2}t^2}}{e^{ta}}.$$

$$t = -it \begin{cases} E e^{itZ} = e^{-\frac{1}{2}t^2} = e^{\frac{1}{2}(it)^2} \\ E e^{tZ} = e^{\frac{1}{2}t^2}. \end{cases}$$

$$= \int_{-\infty}^{+\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

i.e. $\forall t > 0, \quad P(Z \geq a) \leq e^{\frac{1}{2}t^2 - ta}.$

So $P(Z \geq a) \leq \inf_{t > 0} e^{\frac{1}{2}t^2 - ta} = e^{-\frac{1}{2}a^2}$

$\Rightarrow P(Z \geq a) \leq e^{-\frac{1}{2}a^2}, \quad \forall a > 0.$

$$\| $$

$$\int_a^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

6. Jesen's Inequality: $f(EX) \le E[f(x)]$,

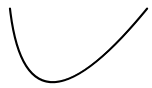
if $f(x)$ is a convex function.

$\downarrow$

convex

or

concave up.



$f''(x) \ge 0, \ \forall x \in \mathbb{R}$.

or

$f(\theta a + (1-\theta) b) \le \theta f(a) + (1-\theta) f(b), \ \forall a, b \in \mathbb{R}, \ \theta \in [0,1]$.

Week 16. Lecture 20.

1. Jensen's Inequality.

If $f(x)$ is concave up, then

$$f(EX) \le E f(x).$$

Eg. $f(x) = |x| \Rightarrow |EX| \le E|X|$.

(Concave up $=$ $f''(x) \ge 0$.)

Proof: Let $\mu = EX \in \mathbb{R}$, then taylor expansion gives

$$f(x) = f(\mu) + f'(\mu)(x-\mu) + \frac{1}{2} \underbrace{f''(\xi)}_{\ge 0} \underbrace{(x-\mu)^2}_{\ge 0}. \qquad \xi \in (\mu, x) \text{ or } (x, \mu)$$

$$\ge f(\mu) + f'(\mu)(x-\mu)$$

So $f(X) \ge f(\mu) + f'(\mu)(X-\mu)$.

$$\Rightarrow E[f(X)] \ge E[f(\mu) + f'(\mu)(X-\mu)].$$

$$= f(\mu) + f'(\mu) E[X-\mu].$$

Notice that $E[X-\mu] = EX - \mu = 0$

Hence $E[f(X)] \ge f(\mu) = f(EX)$. $\square$.

$\forall n \geq 1,\ (E|X|^n)^{\frac{1}{n}} \leq (E|X|^{n+1})^{\frac{1}{n+1}}.$

$\Rightarrow \forall n \geq m,$ if $E|X|^n < \infty$, then $E|X|^m < \infty.$

$$E|X|^n \leq (E|X|^{n+1})^{\frac{n}{n+1}}.$$

To show $f(EX) \leq Ef(X).$

$$(E|X|^n)^{\frac{n+1}{n}} \leq E|X|^{n+1}.$$

$$\begin{cases} f(x) = x^{\frac{n+1}{n}}. \\ Y = |X|^n. \end{cases} \Rightarrow f(Y) = |X|^{n+1}.$$

$$f(EY) \leq E f(Y) = E|X|^{n+1}.$$

$$(E|X|^n)^{\frac{n+1}{n}} \leq E|X|^{n+1}.$$

Here, $f(x) = x^{1+\frac{1}{n}}.$

$f'(x) = (1+\frac{1}{n})\, x^{\frac{1}{n}}$

$f''(x) = (1+\frac{1}{n})\frac{1}{n} x^{\frac{1}{n}-1} > 0, \quad \forall x > 0.$

2. (Strong Law of Large Numbers.)

Let $X_1, X_2, \cdots$ be a sequence of i.i.d. R.V.s,

each having finite expectation μ. ($E|X_k| < \infty$, $EX_k = \mu$, $\forall k \geq 1$)

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

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

Theorem 16.1 $\quad X_n \xrightarrow{P} X \iff \forall$ sub sequence X_{n_k}

$\exists$ further subsequence $X_{n_{k(m)}}$ s.t. $X_{n_{k(m)}} \xrightarrow{a.s} X$.

Eg. $X_{1^2}, X_{2^2}, X_{3^2}, \cdots, X_{n^2}, \cdots$

$\Rightarrow X_{s_1^2}, X_{s_2^2}, \cdots, X_{s_n^2}$ s.t. $X_{s_n^2} \xrightarrow{a.s} X$.

$(s_1 < s_2 < \cdots < s_n \in \mathbb{Z}_+)$.

Theorem 16.2 (Borel — Cantelli Lemma)

For a sequence of events $\{A_n\}_{n=1}^{\infty}$,

if $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n \text{ infinitely often}) = 0$.

Rmk: A_n infinitely often $= \{A_n\}$ occurs infinitely many times

$$= \bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} A_n.$$

$N=1 \Rightarrow A_{n_1}$ occurs for $n_1 \geq 1$.

$N = n_1+1 \Rightarrow A_{n_2}$ occurs for $n_2 \geq n_1+1$.

$\cdots \Rightarrow A_{n_k} \quad \cdots$

$\Rightarrow A_{n_1}, A_{n_2}, \cdots, A_{n_k}, \cdots$ occur.

$\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} A_n^c = \exists N \geq 1$, st $\forall n \geq N$, A_n^c occurs

(A_n doesn't occur).

Proof of Borel - Cantelli:

$$E\left(\sum_{n=1}^{\infty} 1_{A_n}\right) = \sum_{n=1}^{\infty} E(1_{A_n}) = \sum_{n=1}^{\infty} P(A_n) < \infty.$$

So $P(\sum_{n=1}^{\infty} 1_{A_n} < \infty) = 1$.

$[E|X| < \infty \Rightarrow P(|X| < \infty) = 1]$.

$\sum_{n=1}^{\infty} 1_{A_n} < \infty = \{A_n\}$ occurs for finite times.

$\sum_{n=1}^{\infty} 1_{A_n} = \infty = \{A_n\}$ infinitely often.

So $P(\sum_{n=1}^{\infty} 1_{A_n} = \infty) = P(A_n \text{ i.o.}) = 0$. $\square$.

<u>Proof of thm 16.1:</u>

$X_n \xrightarrow{P} X \Leftrightarrow \forall$ sub sequence X_{n_k}

$\exists$ further sub sequence $X_{n_{k(m)}}$ s.t. $X_{n_{k(m)}} \xrightarrow{a.s} X$.

"$\Leftarrow$": If $\forall X_{n_k}$, $\exists X_{n_{k(m)}}$ s.t $X_{n_{k(m)}} \xrightarrow{a.s} X$.

$\Rightarrow X_{n_{k(m)}} \xrightarrow{P} X$.

$\forall \varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$. let $y_n = P(|X_n - X| > \varepsilon)$

$\lim_{n\to\infty} y_n = 0$. $\Leftrightarrow \forall y_{n_k}$, $\exists y_{n_{k(m)}}$ s.t $y_{n_{k(m)}} \to 0$.

"$\Rightarrow$" ✓. "$\Leftarrow$": By contradiction, if not,

$\exists \varepsilon > 0$, $\forall N \geq 1$, $\exists n > N$, $|y_n| > \varepsilon$.

So $\exists$ subsequence y_{n_k} s.t $|y_{n_k}| > \varepsilon$, $\forall k \geq 1$.

$\updownarrow$

By assumption, $\exists y_{n_{k(m)}}$ s.t $y_{n_{k(m)}} \to 0$ ↯.

"$\Rightarrow$": If $X_n \xrightarrow{P} X$.

then $\forall k \geq 1$, $P(|X_n - X| \geq \frac{1}{k}) \to 0$.

So $\exists n_k \geq 1$. s.t $P(|X_{n_k} - X| > \frac{1}{k}) < \frac{1}{2^k}$.

$\Rightarrow \sum_{k=1}^{\infty} P(|X_{n_k} - X| > \frac{1}{k}) \leq \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty$.

$\Rightarrow P(|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}) = 0$ by Borel - Cantelli.

That is $P(\bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} |X_{n_k} - X| \leq \frac{1}{k}) = 1$.

$\Downarrow$

$P(\lim_{k\to\infty} X_{n_k} = X) = 1$. $\Rightarrow X_{n_k} \xrightarrow{a.s} X$.

$\forall X_{n_k}$ $\exists X_{n_{k(m)}}$ s.t $X_{n_{k(m)}} \xrightarrow{a.s} X$. $\square$

$X_{n_k} \xrightarrow{P} X$

<u>Eg 1.</u> If $X_1 \geq X_2 \geq X_3 \geq \cdots$ are positive R.V.s.

and $X_n \xrightarrow{P} 0$, then $X_n \xrightarrow{a.s} 0$.

<u>Proof:</u> By $X_n \xrightarrow{P} 0$, we get $\exists X_{n_k} \xrightarrow{a.s} 0$.

$P(\lim_{k\to\infty} X_{n_k} = 0) = 1$.

$\downarrow$

$\forall \varepsilon > 0$, $\exists K_0 > 0$, s.t $\forall k \geq K_0$, $|\overset{X_{n_k}}{\overset{\shortparallel}{X_{n_k}}}| < \varepsilon$.

$\Rightarrow X_{n_{K_0}} < \varepsilon$, $\forall n \geq n_{k_0}$. $X_n \leq X_{n_{k_0}} < \varepsilon$.

$\Rightarrow \forall \varepsilon > 0$, $\exists n_{K_0} > 0$ s.t $\forall n \geq n_{K_0}$, $|X_n| \leq |X_{n_{K_0}}| < \varepsilon$. $\square$

Proof of Strong Law of Large Numbers.

By assuming $E|X_1|^4 < \infty$.

$$\frac{\sum_{k=1}^n X_k}{n} \xrightarrow{a.s} \mu \iff \frac{\sum_{k=1}^n (X_k-\mu)}{n} \xrightarrow{a.s} 0.$$

We assume $\mu=0$.

Let $S_n=\sum_{k=1}^n X_k$, $\forall n\geq 1$. We will calculate $E(S_n^4)$

$$S_n^4=(S_n^2)^2=\left(\sum_{k=1}^n X_k^2+2\sum_{1\le i<j\le n} X_iX_j\right)^2.$$

$$=\left(\sum_{k=1}^n X_k^2\right)^2+4\sum_{k=1}^n X_k^2\sum_{1\le i<j\le n}X_iX_j+4\left(\sum_{1\le i<j\le n}X_iX_j\right)^2.$$

$$=\sum_{k=1}^n X_k^4+2\sum_{1\le i<j\le n}X_i^2X_j^2+4\sum_{k=1}^n X_k^2\sum_{1\le i<j\le n}X_iX_j+4\sum_{1\le i<j\le n}X_i^2X_j^2+4\sum_{\substack{1\le i<j\le n\\ 1\le \tilde{i}<\tilde{j}\le n\\ (i,j)\neq(\tilde{i},\tilde{j})}}X_iX_jX_{\tilde{i}}X_{\tilde{j}}.$$

$$=\sum_{k=1}^n X_k^4+6\sum_{1\le i<j\le n}X_i^2X_j^2+4\sum_{k=1}^n X_k^2\sum_{1\le i<j\le n}X_iX_j+4\sum_{\substack{1\le i<j\le n\\ 1\le \tilde{i}<\tilde{j}\le n\\ (i,j)\neq(\tilde{i},\tilde{j})}}X_iX_jX_{\tilde{i}}X_{\tilde{j}}.$$

$E X_k^4 = EX_1^4 = C_0.$

$E\left[X_i^2X_j^2\right] = EX_i^2\,EX_j^2=(EX_1^2)^2=C_1.$

$$E\left[X_k^2X_iX_j\right]=0,\ \forall i<j\ \begin{cases} k\notin\{i,j\}. & EX_k^2\,EX_i\,EX_j=0\\ k=i. & EX_i^3\,EX_j=0\\ k=j. & EX_j^3\,EX_i=0.\end{cases}$$

$$E\left[X_iX_jX_{\tilde{i}}X_{\tilde{j}}\right]=\begin{cases}\tilde{i}\neq i,j. & EX_{\tilde{i}}=0.\\ \tilde{i}=i\quad \tilde{j}\neq j & \begin{cases}\tilde{j}=i & EX_j=0.\\ \tilde{j}\neq i & EX_j=EX_{\tilde{j}}=0.\end{cases}\end{cases} =0.$$

$$\Rightarrow E[S_n^4] = E\left[\left(\sum_{k=1}^{n} X_k\right)^4\right] = E\left[\sum_{k=1}^{n} X_k^4\right] + 6\, E\left(\sum_{i<j} X_i^2 X_j^2\right)$$

$$= n C_0 + 6 \binom{n}{2} C_1$$

$$\leq C_2 \cdot n^2.$$

$$P\left(\left|\frac{1}{n} S_n\right| > n^{-\frac{1}{8}}\right) = P\left(|S_n| > n^{\frac{7}{8}}\right) = P\left(|S_n|^4 > n^{\frac{7}{2}}\right)$$

$$\leq \frac{E S_n^4}{n^{\frac{7}{2}}} \leq \frac{C_2 \cdot n^2}{n^{\frac{7}{2}}} \leq C_2 \cdot n^{-\frac{3}{2}}.$$

$$\Rightarrow \sum_{n=1}^{\infty} P\left(\left|\frac{1}{n} S_n\right| > n^{-\frac{1}{8}}\right) \leq \sum_{n=1}^{\infty} C_2 \cdot n^{-\frac{3}{2}} < \infty.$$

By Borel-Cantelli

$$P\left(\left|\frac{1}{n} S_n\right| > n^{-\frac{1}{8}} \text{ i.o}\right) = 0. \Rightarrow P\left(\exists N \geq 1, \forall n \geq N, \left|\frac{1}{n} S_n\right| \leq n^{-\frac{1}{8}}\right) = 1.$$

$$\Rightarrow P\left(\lim_{n\to\infty} \frac{1}{n} S_n = 0\right) = 1. \qquad \text{i.e.} \quad \frac{1}{n} S_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{a.s} 0. \qquad \square.$$

<u>Theorem (Dominated Convergence)</u>

If $X_n \xrightarrow{a.s.} X$ and $|X_n|^p \leq Y$, $\forall n \geq 1$,

and $EY < \infty$, then $E|X_n - X|^p \to 0$.