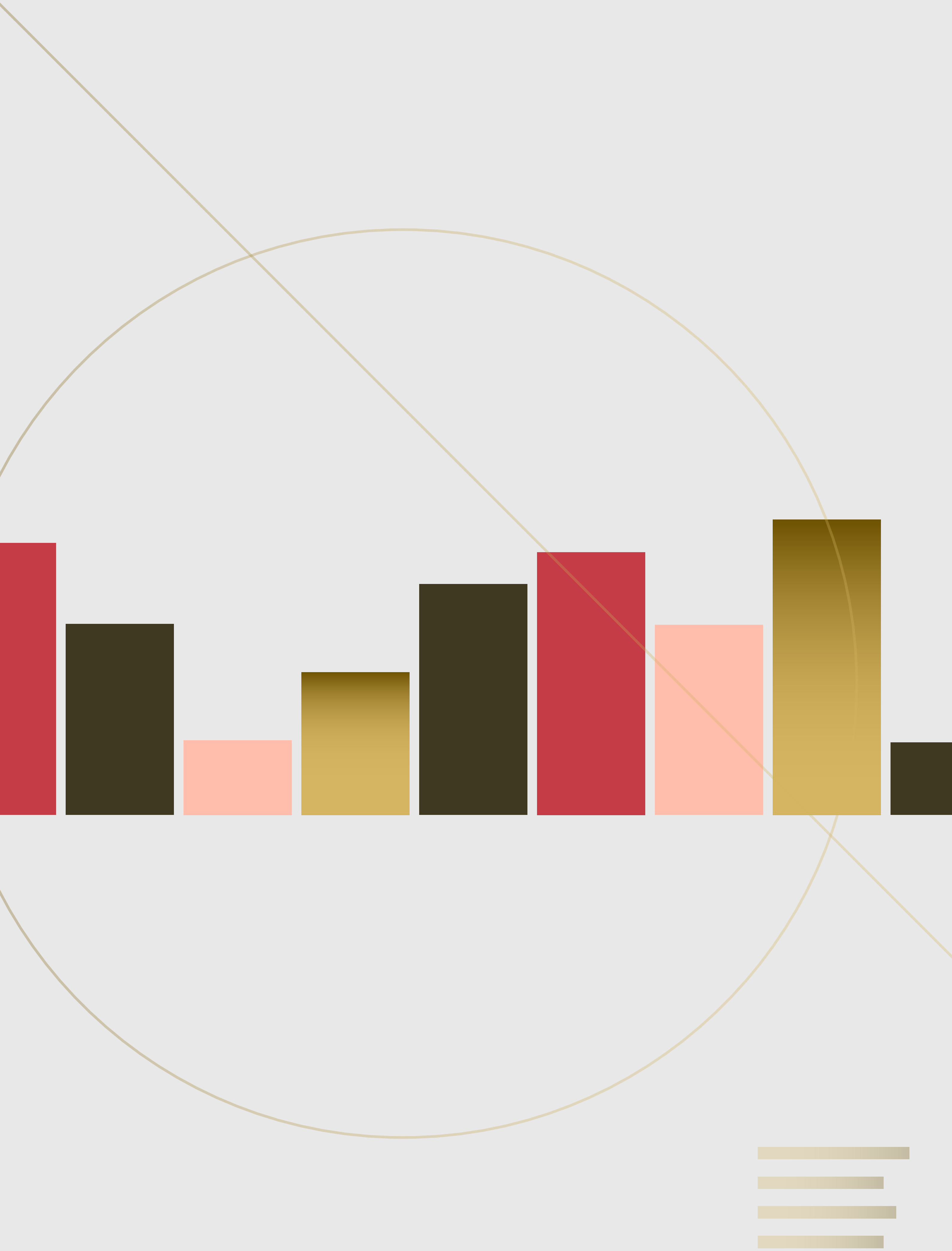


Lec 1

1. Probability theory & Statistics

① Analysis of random phenomena & Analysis of data

② Foundation & Application

2. Fundamental Concepts

① 「Experiment」: a. can be repeated infinitely
(= 「phenomenon」) b. has a well-defined set of possible outcomes

② 「Random Experiment」: An experiment has more than one possible outcomes: a. Assumption 1: S contains m possible outcomes
(Oppositely, 「Determinative Experiment」) b. Assumption 2: The m outcomes are equally likely

③ 「Sample Space」: a. the collection of all possible outcomes
b. denoted by S

④ 「Event」: A set of S

⑤ 「An event A has occurred」: When a random experiment is performed, the outcome is in A

3. Set

① $\emptyset$: the null set | empty set

② $A \subseteq B$: Subset
$A \cup B$: Union
$A \cap B$: Intersection
A': Complement

③ mutually exclusive
exhaustive
mutually exclusive and exhaustive

④ Laws

a. Commutative Law

b. Associative Law

c. Distributive Law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$; $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

d. De Morgan's Law: $(A \cup B)' = A' \cap B'$; $(A \cap B)' = A' \cup B'$

4. Definition of probability

① Intuitive one: $P(A) = \lim_{n\to\infty} \boxed{\frac{N(A)}{n}}$ → 'Relative frequency'

② A real-valued, set function P with three properties: (First order property)

a. $P(A) \geq 0$

b. $P(S) = 1$

c. if $A_1, A_2 \cdots$ are countable and mutually exclusive events, then $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$

5. Properties of probability and their proofs (Second order property)

① $P(A) = 1 - P(A')$

Proof: $P(A) = P(A \cup A') - P(A') = P(A) + P(A') - P(A') = P(A)$

② $P(\phi) = 0$

Proof: $P(\phi) = 1 - P(\phi') = 1 - P(S) = 0$

③ If $A \subseteq B$, $P(A) \leq P(B)$

Proof: $B = B \cup A = \underbrace{(B \cup A) \cap S = (B \cup A) \cap (A' \cup A) = (B \cap A') \cup A}_{\text{Make up mutually exclusive events}}$

$\Rightarrow P(B) = P((B \cap A') \cup A) = P(B \cap A') + P(A) \geq P(A)$

$\cap S \to \cap(A \cup A') \to$ Distributive Law

④ $P(A) \leq 1$

Proof: $P(A) = P(A \cup A') - P(A') = 1 - P(A') \leq 1$

⑤ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof: $A \cup B = (A \cup B) \cap (A \cup A') = A \cup (A' \cap B) \Rightarrow P(A \cup B) = P(A) + P(A' \cap B)$ (1)

$B = B \cap (A \cup A') = (A \cap B) \cup (A' \cap B) \Rightarrow P(B) = P(A \cap B) + P(A' \cap B)$ (2)

(1) + (2) $\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

* Glossary:
① Preamble 序言
② Roll a die 掷骰子 (die → dice (pl.))
③ Flip a coin 抛硬币
④ Distinct 确定的
⑤ Countable 可列举的
⑥ Variate = Variable 变量

Lec 2

1. Enumeration

① $P(A) = \frac{N(A)}{N(S)}$

Two ways to define P: $P = \lim_{n\to\infty} \frac{N(A)}{n}$ (Taking objects) ☆ ; $P = \frac{N(A)}{N(S)}$ (Enumeration)

② Counting Technique ⇒ Multiplication Principle

2. Permutation

① Permutation of n objects : $n!$

② Permutation of n objects taken r at a time : ${}_nP_r = n\cdot(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$

a. Ordered sample of size r : the set of r objects whose order is noted

b. Sampling with replacement : n^r

Sampling without replacement : ${}_nP_r$

3. Combination

① Inference in an indirect way:

$${}_nP_r = X\cdot r! \Rightarrow X = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} \triangleq {}_nC_r / {}_nC_{n-r} / \binom{n}{r} / \binom{n}{n-r}$$

② Can be denoted by " a combination of n objects taken r at a time "

③ $(a+b)^n = \sum_{r=0}^{n} {}_nC_r a^{n-r} b^r \Rightarrow$ ${}_nC_r$ is called 「Binomial Coefficient」

4. Distinguishable Permutation

① Distinguishable Permutation of n objects of two types

$n! = X\cdot r!\cdot(n-r)! \Rightarrow X = {}_nC_r$

S { A $(a_1, a_2 \cdots)$; B $(b_1, b_2 \cdots)$ } ⇒ the elements of S have two types, we permute them only noting their types, instead of identities.

② Distinguishable Permutation of n objects of m types

$n! = X\cdot n_1!\cdot n_2!\cdots n_m! \Rightarrow X = \frac{n!}{n_1!\cdots n_m!}$ $(n_1+n_2+\cdots n_m = n)$

* Glossary: ① Enumeration 计数,列举

② Equally likely 等概的

③ Extension 延长部分

④ Sequential implementation 有顺序的实施

⑤ Placebo 安慰剂

Lec 3

1. Conditional probability

① Definition: (it is a new function with regard to "probability")

a. Direct: Define the probability function for the reduced sample space

b. Indirect: By linking to the probability function for the original sample space

↳ $P(R|E) = \underbrace{\frac{N(R\cap E)}{N(E)}}_{(a)} = \underbrace{\frac{N(R\cap E)/N(S)}{N(E)/N(S)} = \frac{P(R\cap E)}{P(E)}}_{(b)}$

c. Totally: The conditional probability of an event A, given that event B has occurred (let $P(B)>0$), is defined by $P(A|B) = \frac{P(A\cap B)}{P(B)}$

↳ (pronounced as "given")

② Some properties:

a. $P(A'|B) = 1 - P(A|B)$

proof: $1 = P[(A'\cup A)|B] = \frac{P[(A'\cup A)\cap B]}{P(B)} = \frac{P[(A'\cap B)\cup(A\cap B)]}{P(B)} = \frac{P(A'\cap B)+P(A\cap B)}{P(B)} = \frac{P(A'\cap B)}{P(B)} + \frac{P(A\cap B)}{P(B)} = P(A'|B) + P(A|B)$

b. Multiplication Rule:

$$P(A\cap B) = \begin{cases} P(A|B)\cdot P(B), & \text{provided } P(B)>0 \\ P(B|A)\cdot P(A), & \text{provided } P(A)>0 \end{cases}$$

c. $P(A\cap B\cap C) = P[A\cap(B\cap C)] = P[A|(B\cap C)]\cdot P(B\cap C) = P[A|(B\cap C)]\cdot P(B|C)\cdot P(C)$

③ $P[(A\cup B)|C] = P(A|C) + P(B|C)$

④ Summary of notations

- $\cup$ "or"
- $\cap$ "and"
- $|$ "when/after" (change "before" into "after") ⟶ cont'd

⑤ A relatively tricky exercise

Question

Roll a pair of 4-sided dice and observe the sum of the dice

A = {a sum of 3 is rolled}

B = {a sum of 3 or a sum of 5 is rolled}

C = {a sum of 3 is rolled before a sum of 5 is rolled}

What are $P(A)$, $P(B)$, $P(C)$?

⟶ It applies the P enumeration definition, $P = \frac{N(3)}{N(3 \text{ and } 5)}$

a. Method 1: Discuss in a simplified sample space

$P(C) = \frac{N(\{3 \text{ first}\})}{N(\{3 \text{ first or } 5 \text{ first}\})} = \frac{1}{3}$

b. Method 2: (By conditional probability)

$P(C) = P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{\frac{2}{16}}{\frac{6}{16}} = \frac{1}{3}$

↳ How to select the events of a conditional probability?

⚠ First & Second

⚠ Individual & Total (with definite whole-partial relationship: $P(p|w)$)

2. Independent events

① Definition: Events A and B are independent (= mutually independent) if and only if (necessary and sufficient condition) $P(A\cap B)=P(A)\cdot P(B) = P(A)\cdot P(B|A)$; $P(B)=P(B|A)$ (not relevant to the given event)

Otherwise, event A and B are called dependent events.

② Theorem

If A and B are independent, then $A(A')$ and $B(B')$ are independent.

For instance:

△ A and B'

proof: $P[A\cap(B\cup B')] = P[(A\cap B)\cup(A\cap B')] = P(A\cap B)+P(A\cap B') = P(A)\cdot P(B)+P(A\cap B') \Rightarrow P(A\cap B') = P(A)[1-P(B)] = P(A)\cdot P(B')$

△ A' and B'

proof: $P(A'\cap B') = 1-P(A\cup B) = 1-P(A)-P(B)+P(A\cap B) = [1-P(A)][1-P(B)] = P(A')P(B')$

③ Mutually independent

a. Definition: Events A, B and C are mutually independent if and only if: (both)

△ A, B, C are pairwise independent

△ $P(A\cap B\cap C) = P(A)P(B)P(C)$

b. Extension: △ Each pair, triple, quartet of the events are independent, and

△ $P(A_1\cap A_2 \cdots A_n) = P(A_1)\cdots P(A_n)$, then $A_1, A_2 \cdots A_n$ are mutually independent

c. Theorem

If A, B and C are independent, then $A(A')$ and $(B\cup C)/(B\cap C)$ are independent

n events are independent: $n>2 \Rightarrow$; (mutually exclusive, prove) ; $n>2 \Rightarrow$; & $\cup/\cap$ (prove) — Transitivity

For instance: 证3项的结论前先证两项的

△ A' and $(B\cap C')$

proof: by theorem of independent events, we just need to prove that A and $(B\cap C')$ are independent.

$P(A\cap B\cap C') = P(A\cap B\cap S)-P(A\cap B\cap C) = P(A\cap B)-P(A\cap B\cap C) = P(A\cap B)P(C') = P(A)P(B)P(C') = P(A)P(B\cap C')$

△ A and $(B\cup C)$

proof: $P[A\cap(B\cup C)] = P[(A\cap B)\cup(A\cap C)] = P(A\cap B)+P(A\cap C)-P(A\cap B\cap C) = P(A)P(B)+P(A)P(C)-P(A)P(B)P(C) = P(A)[P(B)+P(C)-P(B)P(C)] = P(A)P(B\cup C)$

△ A', B' and C'

proof: from △, we know that A' and $(B'\cap C')$ are independent, then

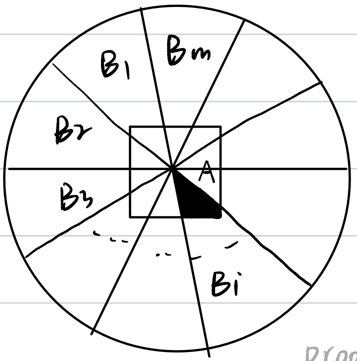
$P[A'\cap B'\cap C'] = P[A'\cap(B'\cap C')] = P(A')\cdot P(B'\cap C') = P(A')P(B')P(C')$ (De Morgen's Law)

Techniques: △ $A'\to A$ △ 3 terms → 2 terms

* Glossary: ① Successive 相继的 ② Die → Dice (pl)

Lec 4

1. Bayes' Theorem



Assume that

① $B_1, B_2 \cdots B_m$ are mutually exclusive and exhausive

② $P(B_i) > 0$.

Then we have:

① $P(A) = \sum_{i=1}^{m} P(A \cap B_i) = \sum_{i=1}^{m} P(B_i)P(A|B_i)$

proof: $P(A) = P(A \cap S) = P[A \cap (B_1 \cup B_2 \cdots \cup B_m)] = P[(A \cap B_1) \cup (A \cap B_2) \cup \cdots (A \cap B_m)] = P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_m)$
$= \sum_{i=1}^{m} P(A \cap B_i) = \sum_{i=1}^{m} P(B_i)P(A|B_i)$

prior probability (B)

② If $P(A) > 0$, then

posterior probability — $P(B_i|A) = \dfrac{P(B_i \cap A)}{P(A)} = \boxed{\dfrac{P(B_i)P(A|B_i)}{\sum_{i=1}^{m} P(B_i)P(A|B_i)}}$ → Bayes' Theorem

total probability $P(A)$

↳ used to compute the probability of big event given small event

↓ essence

convert $P(B|S)$ to $P(S|B)$ with conditional probability

principle: $P(B|S) = \dfrac{(\qquad\qquad)}{(\qquad\qquad)}$ → By definition of conditional probability
→ By definition of conditional probability & By events conversion

Lec 5

1. Random Variable

① Definition: Given a random experiment with sample space S, a function $X: S \to \bar{S} \subseteq \mathbb{R}$ that assign one real number $X(s)=x$ to each $s \in S$ is called Random Variable (RV)

② $\begin{cases} S: \text{original sample space} \\ \bar{S}: \text{numeric sample space} \end{cases}$

RV $\begin{cases} \text{A function} \\ \text{A dependent variable of itself function} \\ \text{An independent variable of the function about } X \end{cases}$

③ a. one to one

b. can be many to one

④ Conventions

a. uppercase & lowercase

b. $\begin{cases} Pr & \text{probability function associated with } S \\ P & \text{probability function associated with } \bar{S} \end{cases}$

c. $\begin{cases} P(X=x) \triangleq P(\{X=x\}) = Pr(\{s \mid X(s)=x,\ s \in S\}) \\ P(X \in A) \triangleq P(\{X \in A\}) = Pr(\{s \mid X(s) \in A,\ s \in S\}) \end{cases}$

event → set → associated with S

2. Discrete Random Variable

① Definition: A RV is called discrete if its range $\bar{S}$ is finite or countably infinite.

↳ The range of X is the set $X(s) = \bar{S} = \{x \mid X(s)=x,\ s \in S\}$.

3. Probability Mass Function (pmf)

① Definition: Suppose that X is a RV with range $\bar{S}$. Then a function $f(x): \bar{S} \to (0,1]$ is called pmf, if:

associated with the definition of probability function:

a. $f(x) > 0,\ x \in \bar{S}$ — a. $P(A_i) \geq 0,\ A_i \in S$

b. $\sum_{x \in \bar{S}} f(x) = 1$ — b. $P(S) = 1$

c. $P(X \in A) = \sum_{x \in A} f(x),\ A \subseteq \bar{S}$ — c. $P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$ ($A_1, A_2, A_3 \cdots$ are mutually exclusive)

② The domain of $f(x)$ is $\mathbb{R}$, then $f(x)=0$ when $x \notin \bar{S}$.

In this case, $f(x): \mathbb{R} \to [0,1]$ and $\bar{S}$ is called the support of X (meaningful domain)

4. Cumulative Distribution Function (cdf)

① Definition: The function $F(x) = P(X \leq x)$ is called the cumulative distribution function (cdf)

② $F(a < x \leq b) = F(b) - F(a)$

5. Uniform Distribution

① Definition: A RV X is said to have a uniform distribution if $f(x) = C$ for $x \in \bar{S}$

* Glossary: ① Uppercase 大号的

② Lowercase 小号的

Lec 6

1. The key characteristics of the probability distribution

① Mean

② Variance

③ Moment

④ Moment generating function

2. Mathematical Expectation (Mean)

① Definition: $E[g(X)] = \sum_{x \in S} g(x) f(x)$ ($E \to$ pmf. RV)

a. X should be a discrete RV

② Mathematical expectation is a weighted sum. the weight is it's pmf, $f(x)$.

a. Arithmetic average is one of the Mathematical expectation, and its pmf is $\frac{1}{N}$ $\Rightarrow$ uniform distribution

③ Mathematical expectation is a constant

④ Theorem ($E_1 \to E_2$)

a. $E(c) = c$

proof: $E(c) = \sum_{x \in S} c \cdot f(x) = c \cdot \sum_{x \in S} f(x) = c$

b. $E[cg(x)] = c\, E[g(x)]$

proof: $E[cg(x)] = \sum_{x \in S} c\, g(x) \cdot f(x) = c \sum_{x \in S} g(x) \cdot f(x) = c\, E[g(x)]$

c. $E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(x)]$

proof: $E[c_1 g_1(x) + c_2 g_2(x)] = \sum_{x \in S} [c_1 g_1(x) f(x) + c_2 g_2(x) f(x)] = \sum_{x \in S} [c_1 g_1(x) f(x) + c_2 g_2(x) f(x)] = c_1 E[g_1(x)] + c_2 E[g_2(x)]$

(Technique: △ Constant C can be extracted out of the parenthesis)

△ Summation parts can be divided.

3. Variance

① Definition: $Var(X) = E(X - EX)^2 = E(X^2 + E^2X - 2EX) = EX^2 + (EX)^2 - 2(EX)^2 = EX^2 - (EX)^2$

a. $EX^2 = E(X)^2 = E(X^2) \neq (EX)^2$ (omit the parenthesis of the RV)

b. Terminal result can be remembered as "internal² − external²"

c. It's called variance of X

② Standard deviation: $\sigma(X) = \sqrt{Var(X)}$

$SD(x)$

③ Properties:

a. $Var(c)=0$

proof: $Var(c)=E[c-E(c)]^2=E(0)=0$

b. $Var(cX)=c^2Var(X)$

proof: $Var(cX)=E(cX)^2-(E(cX))^2=c^2[E(X)^2-(EX)^2]=c^2Var(X)$

④ Variance or standard deviation is a measure of the dispersion or spread of the value of X with respect to its mean.

4. Moment

① Definition

a. r^{th} moment of the distribution about the origin: $E(X^r)=\sum_{x\in S}x^r f(x)$

b. r^{th} moment of the distribution about b: $E(X-b)^r=\sum_{x\in S}(x-b)^r f(x)$

c. r^{th} factorial moment: $E[(X)_r]=E[X(X-1)\cdots(X-r+1)]=\sum_{x\in S}(X)_r f(x)$

$E(X) \longrightarrow$ Moment

$X \longrightarrow X^r$ (r^{th} power function)

$X \longrightarrow (X)_r$

5. Moment Generating Function (mgf)

① Definition: $M(t)=E(e^{tX})=\sum_{x\in S}e^{tx}f(x)$ (the variable is t) exponential

② Properties:

a. $M(0)=1$

proof: $M(0)=E(1)=1$

b. The same mgf $\Rightarrow$ the same probability distribution. (RV. pmf…) $\longrightarrow$ it can deduce the distribution

c. Derivative of t

$M'(t)=\sum_{x\in S}xe^{tx}f(x)$

$M''(t)=\sum_{x\in S}x^2e^{tx}f(x)$

$M^{(r)}(t)=\sum_{x\in S}x^re^{tx}f(x)$

(Summation and differentiation can exchange the place here)

(t has restrictions)

d. Derivative of 0

$M'(0)=EX$

$M''(0)=EX^2$

$M^{(r)}(0)=EX^r$

$\Longrightarrow M^r(0)=EX^r=r^{th}$ moment about the origin

③ Geometric distribution

a. Suppose X has the geometric distribution, that is its pmf is $f(x)=q^{x-1}p$. ($\triangle p=1-q$; $\triangle q\in(0,1)$; $\triangle x=1,2,3\cdots\infty$)

* Glossary: ① estimator 估计量

② dispersion 离散

③ with respect to 对于…

④ intercharge 相互作用

⑤ Origin 原点

Lec 7

1. Bernoulli distribution p

① RV: $X: S \to \bar{S}$, $S = \{\text{success}, \text{failure}\}$, $\bar{S} = \{1, 0\}$, $X\{\text{success}\} = 1$, $X\{\text{failure}\} = 0$

② pmf: $f(x) = p^x (1-p)^{1-x}$, $x \in \bar{S} = \{0, 1\}$

a. it's a general formula

b. we say X has a Bernoulli distribution with probability of success p

③ characteristics

a. $E(X) = p$

b. $Var(X) = p(1-p)$

c. $M(t) = e^t p - p + 1$

2. Binomial distribution n, p

① Bernoulli experiment $\xrightarrow{n \text{ repetitions}}$ Bernoulli trials

$\downarrow$ Bernoulli distribution $\qquad$ $\downarrow$ Binomial distribution

② RV: $X: S \to \bar{S}$, $S = \{\text{the successful times}\}$, $\bar{S} = \{0, 1, 2, \cdots n\}$

(actually, it is the same as RV in Bernoulli distribution, $S = \{s, f\} = \{\text{the successful times}\}$)

a. let X_i denote the Bernoulli RV associated with i^{th} trial

③ pmf: $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x \in \bar{S} = \{0, 1, 2, \cdots n\}$

a. it can be denoted by $X \sim b(n, p)$ (n, p are called parameter of the distribution)

b. the order is not concerned to P of Bernoulli trials, but it is concerned to pmf of Binomial distribution

④ cdf: $F(x) = P(X \le x) = \sum_{X_i,\, X_i \le x} f(X_i) = \sum_{X_i=0}^{\lfloor x \rfloor} \binom{n}{X_i} p^{X_i} (1-p)^{n-X_i}$ ($\lfloor x \rfloor$ is the largest integer that $\le x$)

⑤ characteristics

a. $E(X) = np$

proof: $M'(t) = n(pe^t - p + 1)^{n-1} pe^t \Rightarrow M'(0) = EX = np$ (we can't prove it by definition)

b. $Var(X) = np(1-p)$

proof: $M''(t) = n(n-1)(pe^t - p + 1)^{n-2} (pe^t)^2 + n(pe^t - p + 1)^{n-1} \cdot pe^t \Rightarrow M''(0) = EX^2 = n(n-1)p^2 + np$

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

c. $M(t) = (e^t p - p + 1)^n = [e^t \cdot p + e^0 (1-p)]^n \Rightarrow$ it can conduct the p

proof: $M(t) = \sum_{x \in \bar{S}} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x \in \bar{S}} \binom{n}{x} (pe^t)^x (1-p)^{n-x} = \underbrace{(pe^t - p + 1)^n}_{[e^{1 \cdot t} p + e^{0 \cdot t}(1-p)]^n}$, $t \in R$

二项展开式逆运算

* Glossary: ① tuple: 元组 (= element) ② hack: 劈砍

Lec 8

1. Negative Binomial Distribution r, p

① RV: the trial number at which the r^{th} success is observed

a. $S = \{r, r+1, \cdots\}$ (countably infinite)

② pmf: $f(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, \cdots$ (pay attention to the position of x)

a. Reasoning:

$f(x) = P(\{\text{At the } x^{th} \text{ trial, } r^{th} \text{ success is observed}\})$

$\overset{\S_1}{=} P(A \cap B)$

$\S_1$: $A = \{$For the first $(x-1)$ trials, $(r-1)$ successes have been observed$\}$ (Binomial distribution)

$B = \{$At the x^{th} trial, the outcome is success$\}$

$= P(A)P(B)$

$= \binom{x-1}{r-1} p^{r-1} (1-p)^{x-r} \cdot p$

$= \binom{x-1}{r-1} p^r (1-p)^{x-r}$, $x = r, r+1, \cdots$

b. A supplementary formula:

$\sum_{x=r}^{\infty} \binom{x-1}{r-1} w^{x-r} = (1-w)^{-r}$ (When: total binomial element is different)

proof: Let $y = x - r$, then left $= \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} w^y = \sum_{y=0}^{\infty} \frac{(y+r-1)!}{(r-1)!\,y!} w^y = \sum_{y=0}^{\infty} \frac{(y+r-1)(y+r-2)\cdots(r)}{y!} w^y$

change the variable — expand

$= \sum_{y=0}^{\infty} \frac{(-y-r+1)(-y-r+2)\cdots(-r)}{y!} (-1)^y w^y = \sum_{y=0}^{\infty} \frac{(-r)!}{y!\,(-r-y)!} (-w)^y$

take the $(-1)^y$ out — merge $(-1)^y$ & $(w)^y$

$= \sum_{y=0}^{\infty} \binom{-r}{y} (1)^{-r-y} (-w)^y = (1-w)^{-r}$

merge the binomial expansion

③ characteristics

a. $E(X) = \frac{r}{p}$ $\quad np \to \frac{r}{p}$

proof: $M'(0) = \frac{r}{p}$

b. $Var(X) = \frac{r(1-p)}{p^2}$ $\quad np(1-p) \to \frac{r(1-p)}{p^2}$

proof: $M''(0) - [M'(0)]^2 = \frac{r(1-p)}{p^2}$

③ $M(t)=\dfrac{(pe^t)^r}{[1-(1-p)e^t]^r}$

proof: $M(t)=\sum_{x\in S} e^{tx}\binom{x-1}{r-1}p^r(1-p)^{x-r}$

$$=\sum_{x=r}^{\infty} e^{tx}\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

$$=\sum_{x=r}^{\infty} e^{(r+x-r)t}\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

$$=(pe^t)^r\sum_{x=r}^{\infty}\binom{x-1}{r-1}[(1-p)e^t]^{x-r}$$ (Take the e^{tr} out then there is only one term left)

$$=(pe^t)^r\frac{1}{[1-(1-p)e^t]^r}=\frac{(pe^t)^r}{[1-(1-p)e^t]^r}$$

④ $r=1 \Rightarrow$ Geometric distribution

a. pmf: $f(x)=(1-p)^{x-1}p$, $x=1,2,\cdots$ (its pmf is a geometric series)

2. Poisson distribution λ (RV is discrete, process is continuous)

① APP (Approximate Poisson Process):

a. definition: the experiment that counts the successful times in a continuous interval with a parameter λ ($\lambda>0$)

b. conditions:

- C_1 subexperiment (Bernoulli experiment) are independent
- C_2 The probability of one occurence in a sufficient short subinterval of length h is λh
- C_3 The probability of two or more occurences in a sufficient short subinterval is 0

As for subintervals ($h\le 1$)
As for one subinterval

② Poisson distribution

a. description: the distribution of APP with an interval of length 1 (unit interval) define artificially

△ partition the unit interval into n equally spaced subintervals

△ each subinterval has one occurrence at most $\Rightarrow$ n is sufficiently large ($n>>x$) (= infinitely)

△ the difference between Binomial distribution & Poisson distribution

Binomial distribution: $n=C$

Poisson distribution: $n=\infty$, P is small

b. RV: the number of successes in occurrence of the whole unit interval

c. pmf. $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, 2, \cdots$

△ proof: $f(x) = P(X=x)$, $X \sim b(n, \frac{\lambda}{n})$ lim pmf(Binomial)

$$= \lim_{n\to\infty} \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \lim_{n\to\infty} \underbrace{\frac{n!}{(n-x)!\,n^x}}_{(A)} \frac{\lambda^x}{x!} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-x}}_{(B)} \underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{(C)}$$

$$= \frac{\lambda^x}{x!} A \cdot B \cdot C$$

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

$$= \frac{\lambda^x}{x!} \cdot 1 \cdot 1 \cdot e^{-\lambda} = \frac{\lambda^x e^{-\lambda}}{x!}$$ $\frac{CE}{F}$

△ We denote it by $X \sim \text{Poisson}(\lambda)$

d. characteristics (with the interpretation of λ)

① $E(X) = \lambda$

proof: $M'(0) = \lambda$

② $Var(X) = \lambda$!!!

proof: $M''(0) - [M'(0)]^2 = \lambda$

③ $M(t) = e^{\lambda(e^t - 1)}$

Test result

proof: $M(t) = \sum_{x=0}^{\infty} e^{tx} \cdot \frac{\lambda^x e^{-\lambda}}{x!} = e^{-\lambda} \boxed{\sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}} = e^{-\lambda} \boxed{e^{\lambda e^t}} = e^{\lambda(e^t - 1)}$

Taylor: $e^x = 1 + \frac{x^1}{1!} + \frac{x^2}{2!} + \cdots \frac{x^n}{n!}$

(expand e^x at $x_0 = 0$)

Lec 9

1. Random Variable of Continuous Type

① Definition: A RV with a continuous range is said to be continuous RV

2. Probability density function (pdf), $f(x): \bar{S} \to [0,\infty)$, such that

① Three properties:
- $f(x) > 0,\ x \in \bar{S}$
- $\int_{\bar{S}} f(x)dx = 1$
- $P(a \le X \le b) \hat{=} \int_a^b f(x)dx$

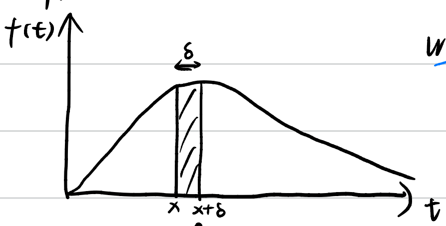
Comparison with discrete RV & pmf
i) property of range $\bar{S}$
ii) the range of pmf & pdf
iii) $\sum \Rightarrow \int$

Pay attention:
which variable is to $\sum$ | $\int$,
others are seen as constants

② If we extend $\bar{S} \Rightarrow R$, then we call $\bar{S}$ "the support of X"
$\Rightarrow$ significative range

③ Some remarks

a. An approximation:



When $\delta \to 0$, then $P(X \le t \le x+\delta) = \int_x^{x+\delta} f(t)dt \approx f(x)\delta$

b. $P(X=a) = \int_a^a f(x)dx = 0$

c. $P(a \le X \le b) = P(a < X \le b) = P(a \le X < b)$

d. pdf doesn't need to be continuous nor bounded

3. Cumulative distribution function $\to$ Be clear: ① which is "upper limit"; ② which is variable of integration

① Definition: $R \to [0,1]$, $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$

② $P(a \le X \le b) = F(b) - F(a)$ — Pay attention to piecewise!

③ $F'(x) = f(x)$ $\to$ Often, we don't actually compute the (integral of pdf / cdf) before we take derivative

- Discrete RV: $P = $ pmf
- Continuous RV: $P = $ cdf

4. Uniform distribution a, b

① It's denoted by $X \sim U(a,b)$

② pdf: $f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$

③ cdf: $F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$

④ Characteristics

a. $E(X) = \frac{1}{b-a} \cdot \left[(b-a) \cdot \frac{a+b}{2}\right] = \frac{a+b}{2}$

b. $Var(X) = \frac{(b-a)^2}{12}$

c. $M(t) = \begin{cases} \frac{e^{tb} - e^{ta}}{t(b-a)}, & t \neq 0 \\ 1, & t = 0 \end{cases}$

Lec 10

1. Key characteristics

① Mathematical Expectation

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

b. Linear operator $\Rightarrow E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$

② Variance: $Var(X) = \int_S (X - EX)^2 f(x)dx$

③ Moment: $E(X^r) = \int_S x^r f(x)dx$

④ mgf: $M(t) = E(e^{tX}) = \int_S e^{tx} f(x)dx$

How to determine "t" ? $\Rightarrow$ ① Let $M^{n}(t)$ exist ② Let $M(t)$ finite

2. $(100p)^{th}$ percentile

$p = F(\pi_p) = \int_{-\infty}^{\pi_p} f(x)dx$, $p \in [0,1]$ (π_p is a parameter)

$\begin{cases} \pi_p \\ p \end{cases}$

① 25^{th}, 75^{th}: first, third quartile ; 50^{th}: median / second quartile

3. Exponential distribution θ

① Description: $\boxed{\text{Poisson distribution}}$

$\Downarrow$ $\lambda^* = \lambda T$

$\boxed{\text{Extended Poisson distribution}}$: pmf: $\frac{(\lambda T)^x e^{-\lambda T}}{x!} \Rightarrow P(X=0) = e^{-\lambda T}$

$\Downarrow$ We focus on the waiting time W until the first occurrence for the App

$\boxed{\text{Elementary Exponential distribution}}$: $F(w) = P(W \le w) = 1 - P(W > w) = 1 - P(X=0) = 1 - e^{-\lambda w}$

$\Rightarrow f(w) = F'(w) = \lambda e^{-\lambda w}$

$\Downarrow$ $\theta = \frac{1}{\lambda}$

$\boxed{\text{Exponential distribution}}$

$\begin{cases} \lambda: 次/单时 \\ \theta: 时/单次 \end{cases} \longrightarrow \theta = \frac{1}{\lambda}$

② pdf: $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$, $x \geq 0$, $\theta > 0$

θ: the average waiting time until the first occurrence

③ characteristics:

a. $E(X) = \theta$

proof: $M'(t) = -\frac{1}{(1-\theta t)^2} \cdot (-\theta) = \frac{\theta}{(1-\theta t)^2} \Rightarrow M'(0) = \theta$

b. $Var(X) = \theta^2$

proof: $M''(t) = (-2)\frac{\theta}{(1-\theta t)^3} \cdot (-\theta) = \frac{2\theta^2}{(1-\theta t)^3} \Rightarrow M''(0) = 2\theta^2 \Rightarrow M''(0) - [M'(0)]^2 = 2\theta^2 - \theta^2 = \theta^2$

c. $M(t) = \frac{1}{1 - t\theta}$

proof: $M(t) = E(e^{tX}) = \int_0^\infty e^{tx} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \frac{1}{\theta} \int_0^\infty e^{x(t-\frac{1}{\theta})} dx = \frac{1}{\theta} \left[\frac{e^{x(t-\frac{1}{\theta})}}{t - \frac{1}{\theta}} \right]_0^\infty = \frac{1}{\theta} \left[0 - \frac{1}{t - \frac{1}{\theta}} \right] = \frac{1}{1 - \theta t}$

When we compute the limit,
we recall that M(t) should be finite

Reflection (How to deal with DISTRIBUTION)

① Confirm the distribution type

$\Downarrow$

② Confirm RV

$\Downarrow$

③ Confirm the parameter

$\Downarrow$

④ Substitude into the formula

Some matters:

① the domain of variable

② the function is piecewise or not

$$F'(w) = \lambda e^{-\lambda w} - e^{-\lambda w}\sum_{k=1}^{\alpha-1}\left[\frac{k(\lambda w)^{k-1}\lambda}{k!} - \frac{(\lambda w)^k\lambda}{k!}\right]$$
$$= \lambda e^{-\lambda w} - e^{-\lambda w}\left[\lambda - \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!}\right]$$
$$= \frac{\lambda(\lambda w)^{\alpha-1}}{(\alpha-1)!}e^{-\lambda w}.$$

有重复项

Lec 11

1. Gamma distribution θ、α

① Discription: the waiting time W until the α^{th} occurrence $(\alpha=1,2,3,\cdots)$

② Derivative process of pdf

定义W

Poisson Distribution for length T: $f(y)=\frac{(\lambda T)^y e^{-\lambda T}}{y!}$ $\Rightarrow$ $P(W>w)=P(\text{number of occurrences in } [0,w] \text{ smaller than } \alpha)$ $=\sum_{k=0}^{\alpha-1}\frac{(\lambda w)^k e^{-\lambda w}}{k!}$ Σ $\Rightarrow$ $f(w)=F'(w)=\frac{\lambda^\alpha w^{\alpha-1}}{(\alpha-1)!}e^{-\lambda w}$

$\Downarrow$

Gamma Function:
a. $\Gamma(t)=\int_0^\infty y^{t-1}e^{-y}dy,\ t>0$

b. $\Gamma(t)=-y^{t-1}e^{-y}\Big|_0^\infty+\int_0^\infty (t-1)y^{t-2}e^{-y}dy=(t-1)\Gamma(t-1)$ $\Rightarrow$ pdf: $f(x)=\frac{1}{\Gamma(\alpha)\theta^\alpha}x^{\alpha-1}e^{-\frac{x}{\theta}}$

c. $\Gamma(n)=(n-1)!$ $\longrightarrow$ 所有整数$\Gamma(n)$

d. $\Gamma(1)=1$, $\Gamma(\frac{1}{2})=\sqrt{\pi}$

③ characteristics

a. $E(X)=\alpha\theta$

b. $Var(X)=\alpha\theta^2$

c. $M(t)=\frac{1}{(1-\theta t)^\alpha}$ proof

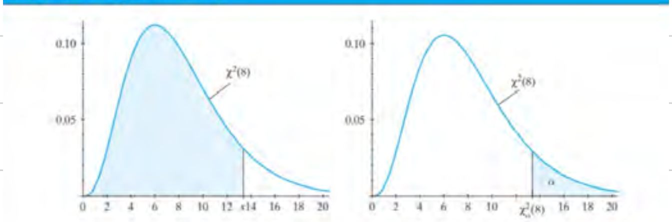
2. Chi-square distribution (卡方分布) r

① Discription: Let $\theta=2$, $\alpha=\frac{r}{2}$ (r is an integer) $\Rightarrow$ pdf: $f(x)=\frac{1}{\Gamma(\frac{r}{2})2^{\frac{r}{2}}}x^{\frac{r}{2}-1}e^{-\frac{x}{2}}$

② It is denoted by $X\sim\chi^2(r)$, r is "degree of freedom"

③ cdf

Table IV The Chi-Square Distribution



Confirm $\left\langle \begin{matrix} r \\ p \end{matrix}\right.$ $\Longleftrightarrow$ find x

$$P(X \le x)=\int_0^x \frac{1}{\Gamma(r/2)2^{r/2}}w^{r/2-1}e^{-w/2}dw$$

	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09

$\chi^2_\alpha(r)$

④ characteristics

a. $E(X)=\alpha\theta=r$

b. $Var(X)=\alpha\theta^2=2r$

c. $M(t)=(1-2t)^{-\frac{r}{2}}$, $t<\frac{1}{2}$

Normal distribution (Gaussian distribution) μ, σ

① "Bell-shaped"
It is denoted by $X \sim N(\mu, \sigma^2)$

② pdf: $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \cdot \frac{(x-\mu)^2}{\sigma^2}\right)$

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

Let $I = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) dx$

let $z = \frac{x-\mu}{\sigma}$, then $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$

$I^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \int_{-\infty}^{\infty} e^{\frac{-y^2}{2}} dy = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{y^2+z^2}{2}} dy\,dz$

let $\begin{cases} y = r\cos\theta \\ z = r\sin\theta \end{cases}$, $I^2 = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-\frac{r^2}{2}} r\,dr\,d\theta = \frac{1}{2\pi} \int_0^{2\pi} d\theta \int_0^{\infty} e^{-\frac{r^2}{2}} d\frac{r^2}{2} = \frac{1}{2\pi} \cdot 2\pi \cdot (-1) \cdot e^{-\frac{r^2}{2}} \Big|_0^{\infty} = 1$

③ characteristics

a. $E(X) = \mu$

b. $Var(X) = \sigma^2$

c. $M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

proof:

$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} dx$

since $e^{tx} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} = \exp\left\{-\frac{1}{2\sigma^2}[x^2 - 2(\mu+\sigma^2 t)x + \mu^2]\right\} = \exp\left(-\frac{1}{2\sigma^2}\left\{[x-(\mu+\sigma^2 t)]^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right\}\right)$

$\Rightarrow M(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}[x-(\mu+\sigma^2 t)]^2\right) dx \cdot \exp\left(\frac{-2\mu\sigma^2 t - \sigma^4 t^2}{-2\sigma^2}\right)$

Recall that $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1 \Rightarrow \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}[x-(\mu+\sigma t)]^2\right) dx = 1$

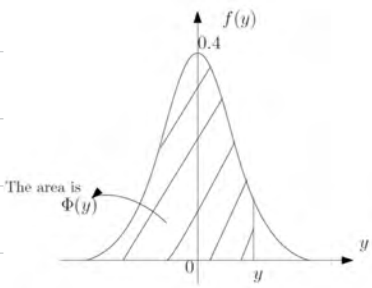
$\Rightarrow M(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$

Lec 12

1. Standard Normal distribution

① $Y \sim N(0,1)$

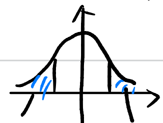
② Graph



a. symmetric

b. max ≈ 0.4

c.



left tail   right tail

③ Some statements

a. How to prove the pmf/pdf? → $\frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2})$

i) $f(x) \geq 0$, $x \in R$

ii) Σ pmf / $\int$ pdf $dx = 1$ (irrespective of the parameters)

b. we should focus on the properties of previous D ⟶ they can be used to simplify the computation

④ cdf

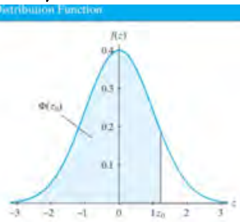
$\Phi(y) = F(y) = P(Y \leq y) = \int_{-\infty}^{y} f(z)\,dz = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\,dz$

It's the unique denotation for cdf of Standard Normal D

a. Symmetry property: $\Phi(-y) = 1 - \Phi(y)$

b. Two appendixes

Table Va The Standard Normal Distribution Function



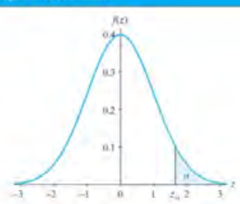
$P(Z \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\,dw$

$\Phi(-z) = 1 - \Phi(z)$

z	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	0.7703	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

A ⟷ B

Table Vb The Standard Normal Right-Tail Probabilities



$P(Z > z_\alpha) = \alpha$

$P(Z > z) = 1 - \Phi(z) = \Phi(-z)$

z_α	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823

i) A is for left tail (cdf / π_p), B is for right tail (z_α)

ii) Confirm $\begin{cases} x / \pi_p \\ z_\alpha \end{cases}$ ⟺ find p

"|" + "—"

$=100(1-\alpha)$ percent

2. The upper 100α percent point

① $P(Z \geq z_\alpha) = \alpha$

② How to transform Π_p & z_α ?

$p = 1-\alpha$ / $\alpha = 1-p$ & combine the graph (symmetric) $\Rightarrow$ Total parameter: α / p, $1-\alpha$ / $1-p$, $-\alpha$ / $-p$, $-(1-\alpha)$ / $-(1-p)$

3. Two theorems about normal D

① $X \sim N(\mu, \sigma^2) \Rightarrow Y = \frac{X-\mu}{\sigma} \sim N(0,1)$

a. equivalent set $\longleftrightarrow$ with same probability

② $X \sim N(\mu, \sigma^2) \Rightarrow Y = \frac{(X-\mu)^2}{\sigma^2} \sim \chi^2(1)$

* Glossary: ① exponent: 指数 ② irrespective / regardless: 不顾

Lec 14

1. Bivariate RV

① Definition: (X, Y) , $\bar{S} \subseteq R^2$

② Details

a. $X \to \bar{S}_X$, $Y \to \bar{S}_Y$ $\Rightarrow$ $\bar{S} \subseteq \bar{S}_X \times \bar{S}_Y = \{(x,y) \mid x \in \bar{S}_X, y \in \bar{S}_Y\}$ (pay attention to the weight in $\bar{S}$)

b. Two interpretations of bivariate RV
- by definition, a pair of RVs
- two univariate RVs are studied jointly

2. Joint pmf

① Definition: $f(x,y): \bar{S} \to [0,1]$, $f(x,y) = P(X=x, Y=y)$

② Properties:

a. $f(x,y) \in [0,1]$

b. $\sum_{(x,y) \in \bar{S}} f(x,y) = 1$

c. For $A \subseteq \bar{S}$, $P[(X,Y) \in A] = \sum_{(x,y) \in A} f(x,y)$

3. Marginal pmf

① Definition:

Knowing joint pmf $\Rightarrow$
- $f_X(x) = P_X(X=x) = P(X=x, Y \in \bar{S}_Y(x)) = \sum_{y \in \bar{S}_Y(x)} f(x,y)$
- $f_Y(y) = P_Y(Y=y) = P(Y=y, X \in \bar{S}_X(y)) = \sum_{x \in \bar{S}_X(y)} f(x,y)$

② Remarks on sample space

a. $\bar{S} = \{$all possible values of $(X, Y)\}$

b. $\bar{S}_X$, $\bar{S}_Y$ = {all possible values of X} $= \{x \mid (x,y) \in \bar{S}, y \in \bar{S}_Y\}$ / ~

c. $\bar{S}_X(y)$, $\bar{S}_Y(x) = \{x \mid (x,y) \in \bar{S}, y = y\}$ / ~

$y \in \bar{S}_Y \Rightarrow y = y$

$\bar{S}_X(y) \subseteq \bar{S}_X$

4. Independent RV

① Definition: $P(A \cap B) = P(A) \cdot P(B)$

$\Downarrow$ let $A = \{X=x, Y \in \bar{S}_Y(x)\}$, $B = \{Y=y, X \in \bar{S}_X(y)\}$

$P(X=x, Y=y) = P(X=x) P(Y=y) \Leftrightarrow f(x,y) = f_X(x) \cdot f_Y(y)$

② Collary: $\bar{S} = \bar{S}_X \cdot \bar{S}_Y$ (Equivalent function $\Rightarrow$ Equivalent domain)

$\bar{S}$ now should be rectangular

- If not, they are dependent
- If so, they are not always independent

Lec 15

八、Mathematical expectation

① Definition: $E[g(X,Y)] = \sum_{(x,y)\in S} g(x,y)\,f(x,y)$

a. $\sum_{(x,y)\in S} = \begin{cases} \text{i) } \sum_{x\in S_X}\sum_{y\in S_Y(x)} \\ \text{ii) } \sum_{(x,y)\in S_1} \sim + \sum_{(x,y)\in S_2} \sim \quad (S = S_1 + S_2) \end{cases}$ classify according to piecewise function

2. Covariance

① Definition: $Cov(X,Y) \triangleq E[(X-EX)(Y-EY)] = \sum_{(x,y)\in S} (x-E(X))(y-E(Y))f(x,y)$

$= E(XY) - E(X)E(Y)$

proof: $E[(X-EX)(Y-EY)] = E(XY - XEY - YEX + EXEY) = E(XY) - EY\,EX - EX\,EY + EX\,EY = E(XY) - E(X)E(Y)$

② Sign of covariance $\begin{cases} = 0 & \text{uncorrelated} \\ > 0 & \text{positively correlated} \\ < 0 & \text{negatively correlated} \end{cases}$

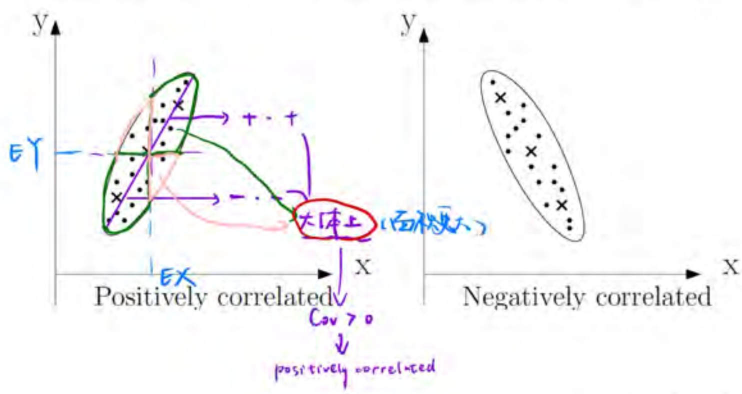
interpretation of Cov

indicate that the values of $X - E(X)$ and $Y - E(Y)$ tend to have the same or opposite sign respectively

study the sign property of these two value

a. positively & negatively correlated

Assume that X and Y are uniformly distributed over the ellipses.



$|\rho|\uparrow$, closer $\Rightarrow |\rho| = 1$, linear

b. Independent $\underset{\times}{\overset{\checkmark}{\rightleftharpoons}}$ Uncorrelated

I ⊆ U

i) Independent: $f(x,y) = f_X(x)\cdot f_Y(y)$

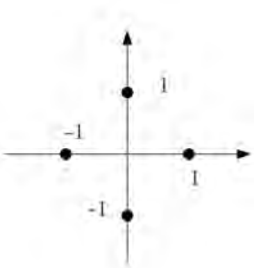
Uncorrelated: $E(XY) = E(X)E(Y)$ $(Cov = 0)$

ii) $I \xrightarrow{\checkmark} U$

proof: $E(XY) = \sum_{(x,y)\in S}(xy)f(x,y) = \sum_{(x,y)\in S}(xy)f_X(x)f_Y(y) = \sum_{x\in S_X} x f_X(x) \sum_{y\in S_Y(x)} y f_Y(y) = \sum_{x\in S_X} x f_X(x) \sum_{y\in S_Y} y f_Y(y) = E(X)E(Y)$

iii> $I \xleftarrow{\times} U$

Counter instance



$(X,Y): (1,0), (0,1), (-1,0), (0,-1)$

$\triangle$ $Cov = E(XY) - E(X)E(Y) = 0$

$\triangle$ $f_X(x) = \begin{cases} \frac{1}{4}, & x=1 \\ \frac{1}{2}, & x=0 \\ \frac{1}{4}, & x=-1 \end{cases}$, $f_Y(y) = \begin{cases} \frac{1}{4}, & y=1 \\ \frac{1}{2}, & y=0 \\ \frac{1}{4}, & y=-1 \end{cases}$ $\Rightarrow f_X(0) f_Y(1) = \frac{1}{8} \neq f(0,1) = \frac{1}{4}$

③ Correlation Coefficient

a. $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

pearson coefficient

i> interpretation: the same as the one of covariance $(SD(x)/SD(y) \geq 0)$

ii> Two properties

$\triangle$ $\rho \in [-1, 1]$

So it is seen as the normalized version of $Cov(X,Y)$

$\triangle$ $\rho = +1/-1 \iff$ There exists a positive/negative constant C s.t. $Y - E(Y) = c[X - E(X)]$, $c = \begin{cases} \frac{\sigma_X}{\sigma_Y}, & \rho = +1 \\ -\frac{\sigma_X}{\sigma_Y}, & \rho = -1 \end{cases}$

if and only if

proofs:

$\triangle$ Let $V = X - \mu_X$, $W = Y - \mu_Y$, t (a constant)

Then $E[(V+tW)^2] = E(V^2) + 2tEVW + t^2E(W^2) = Var(X) + 2t\,Cov(X,Y) + t^2 Var(Y) \geq 0$ (Eq)

Since this (Eq) is true for any t, $4Cov(X,Y)^2 - 4\sigma_X^2\sigma_Y^2 \leq 0 \Rightarrow |\rho(X,Y)| \leq 1 \Rightarrow \rho \in [-1,1]$

$\triangle$ $\rho = \pm 1 \iff Cov(X,Y) = \pm\sigma_X\sigma_Y$ ($\triangle$ now is equal to 0)

$t = -\dfrac{Cov(X,Y)}{\sigma^2_Y} = \pm\dfrac{\sigma_X}{\sigma_Y}$

$V + tW$ $= 0$.

$\Rightarrow X - E(X) = -t[Y - E(Y)] = \pm\dfrac{\sigma_X}{\sigma_Y}[Y - E(Y)]$

$\rho = +1 \Rightarrow \dfrac{\sigma_X}{\sigma_Y}$

$\rho = -1 \Rightarrow -\dfrac{\sigma_X}{\sigma_Y}$

$$\boxed{y - \mu_Y = \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X) \qquad y = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X).}$$

$\rho \to \pm 1 \Rightarrow$ more linear relationship between $(X - \mu_X)$ & $(Y - \mu_Y)$

Question

Consider n independent tosses of a coin with probability of a head equal to p. Let X and Y be the number of heads and of tails in the n tosses, respectively. Calculate the correlation coefficient of X and Y.

☆

$X + Y = n \quad \Rightarrow E[X] + E[Y] = n \Rightarrow X - E(X) = -[Y - E(Y)]$

$Cov(X,Y) = E[(X - EX)(Y - EX)] = -E(Y - EY)^2 = -Var(Y)$

$Var(X) = E(X - EX)^2 = E(Y - EY)^2 = Var(Y)$

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{-Var(Y)}{\sqrt{Var(Y)}\sqrt{Var(Y)}} = -1$$

Lec 16

1. Conditional distribution

① Conditional pmf

a. Definition: Conditional pmf of X given $Y=y$: $g(x|y) = \dfrac{f(x,y)}{\underbrace{f_Y(y)}_{>0}}$, $x \in \overline{S_X(y)}$

Conditional pmf of Y given $X=x$: $h(y|x) = \dfrac{f(x,y)}{\underbrace{f_X(x)}_{>0}}$, $y \in \overline{S_Y(x)}$

b. if X & Y are independent $\Rightarrow$ $g(x|y) = f_X(x)$

② Conditional Mathematical Expectation

a. Definition: $\underbrace{E[g(Y) \mid X=x]}_{\text{Conditional RV}} = \sum_{y \in S_Y(x)} g(y)h(y|x)$ it is a function of x

b. Special types:

△ Let $g(Y) = Y \Rightarrow E[Y|X=x] \Rightarrow$ conditional mean

△ Let $g(Y) = [Y - \underbrace{E(Y|X=x)}_{\text{conditional mean}}]^2 \Rightarrow Var(Y|X=x) = E[Y^2|X=x] - (E[Y|X=x])^2 \Rightarrow$ conditional variance

* Glossary: ① scalar: 标量 ("无向量")

Lec 17

1. Joint pdf

① Definition: $f(x,y): \overline{S} \longrightarrow (0,\infty)$

② Properties:

a. $f(x,y)>0$

b. $\iint_{\overline{S}} f(x,y)\,dx\,dy=1$

c. $P((X,Y)\in A)=\iint_A f(x,y)\,dx\,dy$, $A\subseteq\overline{S}$

2. Marginal pdf

$$\begin{cases} f_X(x)=\int_{\overline{S_Y(x)}} f(x,y)\,dy: \overline{S_X}\longrightarrow(0,\infty) \\ f_Y(y)=\int_{\overline{S_X(y)}} f(x,y)\,dx: \overline{S_Y}\longrightarrow(0,\infty) \end{cases}$$

3. Mathematical expectation:

$$E[g(X,Y)]=\iint_{\overline{S}} g(x,y)f(x,y)\,dx\,dy$$

4. Covariance

① $Cov(X,Y)=E[(X-EX)(Y-EY)]=E(XY)-E(X)E(Y)$

② Correlation coefficient: $\rho(X,Y)=\dfrac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$

5. Independent Continuous RVs

$f(x,y)=f_X(x)\cdot f_Y(y)$ ($\overline{S}$ is rectangular)

6. Conditional pdf

① $h(y|x)=\dfrac{f(x,y)}{f_X(x)}$, $y\in\overline{S_Y(x)}$

② Conditional mean: $E(Y|X=x)=\int_{\overline{S_Y(x)}} y\,h(y|x)\,dy$

③ Conditional variance: $Var(Y|X=x)=\int_{\overline{S_Y(x)}}[y-E(Y|X=x)]^2h(y|x)\,dy=E(Y^2|X=x)-[E(Y|X=x)]^2$

Double Integral

1、Concept

	Univariate	Bivariate
①	$(x,f(x))\Rightarrow$ 曲线	$(x,y,f(x,y))\Rightarrow$ 曲面
②	投影为直线 $(\overline{S})$	投影为平面 $(\overline{S})$
③	积分为f在$\overline{S}$上的覆盖	
	S	V
④	微分为线	微分为切片

2、Computation

$\iint dx\,dy$

对于每个x，x在$(\ ,\ )$上积 (about y) ⇓ y在定义域上积

$\iint dy\,dx$

对于每个y，y在$(\ ,\ )$上积 (about x) ⇓ x在定义域上积

Lec 18

1、Bivariate normal distribution

① joint pdf

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}q(x,y)\right]\ ;\ q(x,y)=\frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2-2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]$$

a. $x\in R,\ y\in R$ $(\in(-\infty,+\infty))$

b. 5 parameters: $\mu_X, \mu_Y\in R$, $\sigma_X, \sigma_Y>0$, $\rho\in(-1,1)$ $(\sqrt{1-\rho^2}\neq 0)$

c. Properties: ⚠⚠ for CP

⚠ Marginal pdfs are normal with $X\sim N(\mu_X,\sigma_X^2)$, $Y\sim N(\mu_Y,\sigma_Y^2)$

proof: $f_X(x)=\int_{S_Y(x)} f(x,y)\,dy=\frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}\exp\left\{-\frac{1}{2}\left[\left(\underbrace{\frac{\left(\frac{y-\mu_Y}{\sigma_Y}\right)-\left(\rho\frac{x-\mu_X}{\sigma_X}\right)}{\sqrt{1-\rho^2}}}_{(t)}\right)^2+\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]\right\}d\frac{y-\mu_Y}{\sigma_Y}$

$=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2}\cdot\underbrace{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-\frac{(t)^2}{2}}d(t)}_{=1}=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{\left(\frac{x-\mu_X}{\sigma_X}\right)^2}{2}}=\frac{1}{\sqrt{2\pi}\sigma_X}e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}}$

What's more, we can prove $\iint_S f(x,y)\,dxdy=1$ by this property $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dxdy=\int_{-\infty}^{+\infty}f(y)\,dy=1$

⚠ ρ is the correlation coefficient

$E(XY)=\iint_S xy\,f(x,y)\,dxdy=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}x\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2}\right)\right\}dx\int_{-\infty}^{+\infty}y\exp\left\{-\frac{1}{2(1-\rho^2)}\left(-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}+\frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)\right\}dy$

$=\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{+\infty}(\sigma_X x+\mu_X)\exp\left\{-\frac{1}{2(1-\rho^2)}x^2\right\}dx\int_{-\infty}^{+\infty}(\sigma_Y y+\mu_Y)\exp\left\{-\frac{1}{2(1-\rho^2)}(-2\rho xy+y^2)\right\}dy$

$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}(\sigma_X x+\mu_X)(\sigma_Y\rho x+\mu_Y)\exp\left\{-\frac{x^2}{2}\right\}dx=\mu_X\mu_Y+\rho\sigma_X\sigma_Y$

⚠ Conditional pdfs are normal with

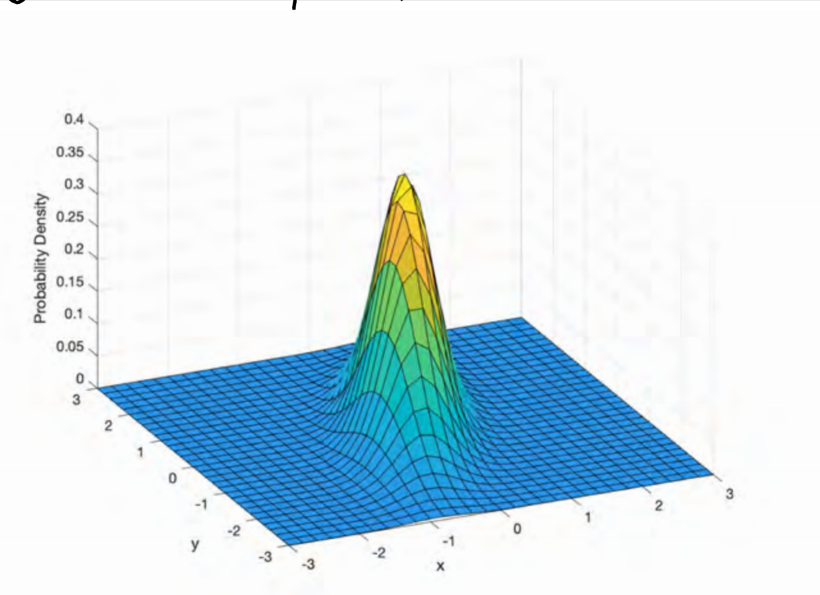
$$\begin{cases}X\mid Y=y\sim N\left(\mu_X+\frac{\sigma_X}{\sigma_Y}\rho(y-\mu_Y),\ (1-\rho^2)\sigma_X^2\right)\\ Y\mid X=x\sim N\left(\mu_Y+\frac{\sigma_Y}{\sigma_X}\rho(x-\mu_X),\ (1-\rho^2)\sigma_Y^2\right)\end{cases}$$

⇒ conditional mean & conditional variance are given

⇒ X & Y can be $\begin{cases}\text{independent}\\ \text{dependent}\end{cases}$

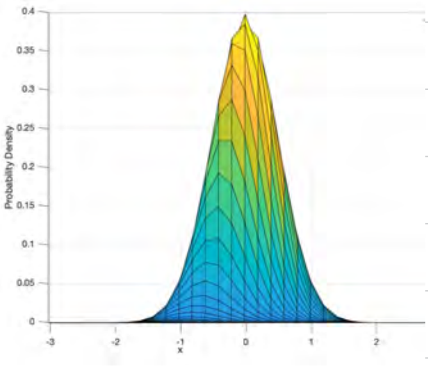
⚠ Independent $\triangleq$ Uncorrelated

② Geometrical interpretation



a. let $x = x_0$, consider $z = f(x_0, y)$ parallel to the yz-plane
let $y = y_0$, consider $z = f(x, y_0)$ parallel to the xz-plane

$z = f(x_0, y) = \underbrace{f_X(x_0)}_{\text{Constant}}\ \underbrace{h(y \mid x_0)}_{\text{Normal}}$



(bell-shaped)

b. let $z = z_0$, $z_0 \in \left(0, \boxed{\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}}\right)$

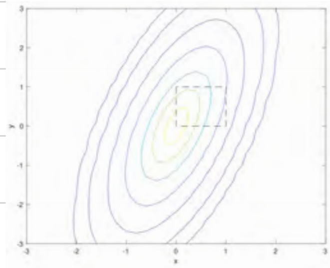
$\rho \neq \pm 1$
$q(x,y) > 0$

joint pdf has its maximum

parallel to the xy-plane

$\exp\left[-\frac{1}{2}q(x,y)\right] = z_0 \cdot 2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}$

logarithm yields $\rightarrow$ function of ellipse



(ellipse)

called level curve / contour

2. Distribution of g(RV)

① Assumption:

a. $u(X)$ is one to one (strictly monotonous) $\Rightarrow$ $v(Y)$ exists

b. defined discrete / continuous

② Core thought:

$\boxed{\begin{array}{l} u(X) \text{ is strictly monotonous} \\ u(X) \text{ is discrete / continuous} \end{array}} \longrightarrow \boxed{X,\ \overline{S_X},\ f(x)} \longrightarrow \boxed{\overline{S_Y}} \longrightarrow \boxed{g(y)}$

probability function

$Y = u(X) \Rightarrow \overline{S_Y} = \{y \mid y = u(x), x \in \overline{S_X}\}$

(compound function)

③ Two cases

a. Discrete case

$$\begin{cases} \overline{S_Y} = \{y \mid y = u(X),\ x \in \overline{S_X}\} \\ g(y) = P(Y = y) = P(u(X) = y) = P(X = v(y)) \end{cases} \Rightarrow \boxed{g(y) = f(v(y))}$$

b. Continuous case

△反函数与正函数拥有相同的单调性 △ $v(u(x)) = v(y) = x$

$$\begin{cases} \overline{S_Y} = \{y \mid y = u(X),\ x \in \overline{S_X}\} \\ G(y) = \begin{cases} \text{strictly} \uparrow \quad P(Y \le y) = P(u(X) \le y) = P(X \le v(y)) = \int_{c_1}^{v(y)} f(x)\,dx \\ \text{strictly} \downarrow \quad P(Y \le y) = P(u(X) \le y) = P(X \ge v(y)) = 1 - P(X \le v(y)) = 1 - \int_{c_1}^{v(y)} f(x)\,dx \end{cases} \end{cases}$$

$\Downarrow$

$$g(y) = G'(y) = \begin{cases} f(v(y))\, \underbrace{v'(y)}_{(+)} \\ -f(v(y))\, v'(y) = -f(v(y))\, \underbrace{v'(y)}_{(-)} \end{cases} = \boxed{f(v(y)) \left| \frac{dv(y)}{dy} \right|}$$

the difference is $v'(y)$

Lec 19

1. Theorem [Random Number Generator]

① Definition:

a. $Y \sim U(0,1)$

b. $F(x)$ has the properties of cdf of $U(0,1)$ with $F(a)=0$, $F(b)=1$. a could be $-\infty$, b could be $+\infty$ s.t. $F(x): (a,b) \longrightarrow [0,1]$

$F(x)$ is strictly increasing (unique assumption for this theorem) → continuous type

c. Then $X = F^{-1}(Y)$ is a continuous RV with cdf $F(x)$

Proof: (Idea: we need to show $P(X \le x) = F(x)$)

$P(X \le x) = P(F^{-1}(Y) \le x) = P(Y \le F(x))$

Since $Y \sim (0,1)$, $F(y) = \int_0^y 1\,dy = y$

Then $P(Y \le F(x)) = F(x) \Rightarrow P(X \le x) = F(x)$, prove right

② Expansion:

a. Generate a random number y from $U(0,1)$

b. Take $x = F^{-1}(y)$

Then x is a random number generated from the distribution with cdf $F(x)$

③ Description:

a y from $U(0,1)$

a distribution with cdf $F(x)$

$\xrightarrow{\text{Generate}}$ $x = F^{-1}(y)$ [Application] (simulate the observation)

$X = F^{-1}(Y)$

↑ Generate

$Y \sim U(0,1)$

uniform distribution with cdf $F(X)$

④ Verification (by Histogram)

a. Plotting process

△ x: sample space $\xrightarrow{\text{divided into}}$ adjecent, non-overlapping and equally spaced subintervals

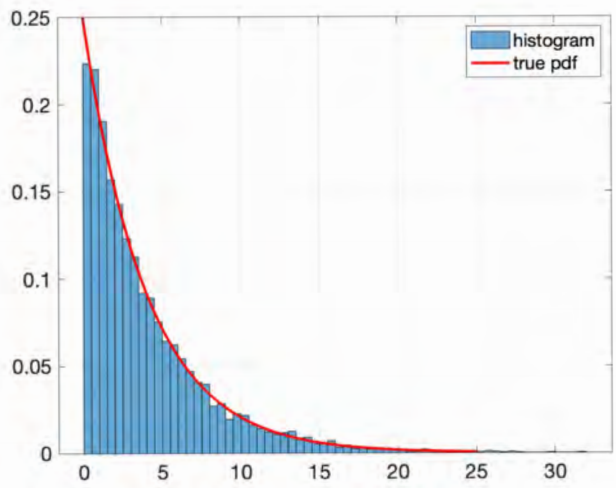
△ y: count # of times in each interval ⟶ compute the relative frequency ⟶ let $y = \dfrac{\text{rel. f}}{\text{length of each interval}}$

△ x-y rectangular graph

b. Property

Area ⇒ Rel. f

Total area ⇒ 1



fit the histogram & the true pdf graph ⇒ prove right

Rel.f ≈ Prob.

2. Theorem [Counter RNG]

X is a continuous RV and $\overline{S_X} = (a,b)$, $F(x)$ is strictly increasing

$\Downarrow$

$Y = F(X) \sim U(0,1)$

Proof:

Since $F(a)=0$, $F(b)=1$ and $F(x)$ is strictly increasing,

then $Y = F(X)$, $\overline{S_Y} = (0,1)$

$P(Y \le y) = P(F(X) \le y) = P(X \le F^{-1}(y)) = F(F^{-1}(y)) = y$, $y \in (0,1) \longrightarrow$ cdf of $U(0,1)$

Lec 20

1. Multivariate RV (we assume that RVs are independent)

① Joint pmf / pdf for independent RV

$$f(x_1, \cdots x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

↳ $X_1, \cdots X_n$ are independent $\Leftrightarrow$ any pair, any triple, …, any $(n-1)$ of them are also independent

proof method: take Σ / $\int$ to others

② Random sample of size n from a common distribution / i.i.d.

a. Definition: the RVs are independently and identically distributed

$f(x_1, \cdots x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$ $\qquad$ $f_{X_1}(x) = f_{X_2}(x) = \cdots = f_X(x)$

b. property: $f(x_1, x_2, \cdots x_n) = f_X(x_1) \cdots f_X(x_n)$

③ Theorem 5.3-1

X_i's are independent, $Y = u_1(X_1) \cdot u_2(X_2) \cdot u_3(X_3) \cdots u_n(X_n)$, $E[u_i(X_i)]$ all exist

$\Rightarrow E(Y) = E[u_1(X_1) u_2(X_2) \cdots u_n(X_n)] = E[u_1(X_1)] \cdot E[u_2(X_2)] \cdot E[u_3(X_3)] \cdots E[u_n(X_n)]$

from RV to F $\rightarrow$ the key is to determine the $u_i(X_i)$

proof:

a. for discrete type

n 维求和

$$E[u_1(X_1) \cdots u_n(X_n)] = \sum_S [\ ]\, f(x_1, x_2, \cdots, x_n) = \sum_{S_{X_1}} u_1(x_1) f_{X_1}(x_1) \cdots \sum_{S_{X_n}} u_n(x_n) f_{X_n}(x_n) = E[u_1(X_1)] \cdot E[u_2(X_2)] \times \cdots \times E[u_n(X_n)]$$

b. for continuous type

$$E[\ ] = \int_S [\ ]\, f(x_1, x_2, \cdots x_n) = \int_{S_{X_1}} u_1(x_1) f_{X_1}(x_1) \cdot \int_{S_{X_2}} u_2(x_2) f_{X_2}(x_2) \times \cdots \times \int_{S_{X_n}} u_n(x_n) f_{X_n}(x_n) = E[u_1(X_1)] \cdot E[u_2(X_2)] \times \cdots \times E[u_n(X_n)]$$

④ Theorem 5.3-2

a. X_i's are independent, $Y = \sum_{i=1}^{n} a_i X_i$

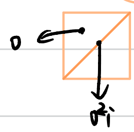
$\Rightarrow E(Y) = \sum_{i=1}^{n} a_i E(X_i)$, $Var(Y) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$

proof:

$$E(Y) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu_i$$

independent

$$Var(Y) = E(Y - \mu_Y)^2 = E\left(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i \mu_i\right)^2 = E\left(\sum_{i=1}^{n} a_i (X_i - \mu_i)\right)^2 = E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (X_i - \mu_i)(X_j - \mu_j)\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[(X_i - \mu_i)(X_j - \mu_j)] = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$



b. Corollary: X_i's are i.i.d., let $Y = \bar{X} = \sum_{i=1}^{n} \frac{1}{n} X_i$ → sample mean / arithmetic expectation

then
- $E(\bar{X}) = E\left(\sum_{i=1}^{n} \frac{1}{n} X_i\right) = n \cdot \frac{1}{n} E(X_i) = \mu$
- $Var(\bar{X}) = Var\left(\sum_{i=1}^{n} \frac{1}{n} X_i\right) = n \left(\frac{1}{n}\right)^2 Var(X_i) = \frac{\sigma^2}{n}$

2. Statistic

① Definition: a function of the random variables $X_1, X_2, \cdots, X_n$ that doesn't have unknown parameters

② Example: $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

→ also an estimator of μ ($n \to \infty$, $Var(\bar{X}) \to 0$, $\bar{X} \to \mu$)

Lec 21

1. Moment generating function technique

change variable < P 1~2, mgf 1~n

① Motivation: mgf $\xrightarrow{\text{if exists}}$ distribution < multivariate RV, pmf / pdf

② Theorem 5.4-1

a. X_is are independent, $Y=\sum_{i=1}^{n} a_i X_i$ (linear combination)

$\Rightarrow M_Y(t)=\prod_{i=1}^{n} M_{X_i}(a_i t)$, $|a_i t|<h_i$ ($M_{X_i}(t)$ is defined for $|t|<h_i$, $M_{X_i}(a_i t)$ is defined for $|a_i t|<h_i$)

key: terms in the parentheses $<h_i$

proof:

$$M_Y(t)=E(e^{tY})=E(e^{t\sum_{i=1}^{n} a_i X_i})=E(e^{ta_1X_1}\times e^{ta_2X_2}\times\cdots e^{ta_nX_n})$$

$$=E(e^{a_1tX_1})\times\cdots\times E(e^{a_ntX_n})$$

$$=\prod_{i=1}^{n} M_{X_i}(a_i t)$$

b. Corollary:

If X_is are i.i.d. with mgf $M(t)$, then

△ The mgf of $Y=\sum_{i=1}^{n} X_i$ is:

$M_Y(t)=\prod_{i=1}^{n} M(t)=[M(t)]^n$, $|t|<h$

→ Bernoulli D ⟹ Binomial D

success for one trial: $M(t)=1-p+pe^t$; success for one experience: $M(t)=(1-p+pe^t)^n$

△ The mgf of $\bar{X}=\sum_{i=1}^{n}\frac{1}{n}X_i$ is:

$M_{\bar{X}}(t)=\prod_{i=1}^{n} M(\frac{1}{n}t)=[M(\frac{t}{n})]^n$, $|\frac{t}{n}|<h$

③ Theorem 5.4-2

a. X_is are independent and each X_i has chi-square D with $r_1, r_2, \cdots, r_n$, that is $X_i\sim\chi^2(r_i)$, $i=1,2,\cdots,n$

then $Y=(X_1+X_2+X_3+\cdots X_n)\sim\chi^2(r_1+r_2+r_3+\cdots r_n)$

$2X_1 \longrightarrow$ ⊗ (X_1 & X_1 are not independent)

RV $\xleftrightarrow[\text{(+)}]{\text{linear relation}}$ r

proof:

$M_Y(t)=\prod_{i=1}^{n} M_{X_i}(t)$, $|t|<h_i$

$M_{X_i}(t)=(1-2t)^{-\frac{r_i}{2}}$

$$M_Y(t)=\prod_{i=1}^{n} M_{X_i}(t)=(1-2t)^{-\frac{r_1}{2}}(1-2t)^{-\frac{r_2}{2}}(1-2t)^{-\frac{r_3}{2}}\cdots(1-2t)^{-\frac{r_n}{2}}=(1-2t)^{-\frac{1}{2}(r_1+r_2+\cdots r_n)}\longrightarrow Y\sim\chi^2(r_1+r_2+r_3+\cdots+r_n)$$

b. Corollary:

△ Z_is are independent and each $Z_i\sim N(0,1)$

then $W=Z_1^2+Z_2^2+\cdots+Z_n^2\sim\chi^2(n)$

SND

proof:

$Z_i^2\sim\chi^2(1)\longrightarrow W\sim\chi^2(n)$

ND

△ X_is are independent and each $X_i\sim N(\mu_i,\sigma_i^2)$

then $W=\sum_{i=1}^{n}(\frac{X_i-\mu_i}{\sigma_i})^2\sim\chi^2(n)$

Lec 22

一、Random function associated with normal distribution

① Theorem 5.5-1

a. X_is are independent normal variables with mean $\mu_1, \mu_2, \cdots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2$, then $Y = \sum_{i=1}^{n} a_i X_i \sim N(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2)$

proof:

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(a_i t) = \prod_{i=1}^{n} \exp(\mu_i a_i t + \tfrac{1}{2}\sigma_i^2 a_i^2 t^2) = \exp\{(\sum_{i=1}^{n} \mu_i a_i) t + \tfrac{1}{2}(\sum_{i=1}^{n} a_i^2 \sigma_i^2) t^2\}$$

b. Corollary:

If X_is are i.i.d., $X_i \sim N(\mu, \sigma^2)$

then $\begin{cases} \triangle\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n}) \\ \triangle\ \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \end{cases}$

② Theorem 5.5-2

X_is are i.i.d, $X_i \sim N(\mu, \sigma^2)$

$E(S^2) = \sigma^2$ (also a statistic)

then $\begin{cases} \text{a. Sample mean } \bar{X} \text{ \& Sample variance } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \text{ are independent} \\ \text{b. } \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1) \end{cases}$

prove

proof of b.

$$\underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}_{\text{let } W = \nearrow} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}+\bar{X}-\mu}{\sigma}\right)^2 = \underbrace{\sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2}_{=\frac{(n-1)S^2}{\sigma^2}} + \underbrace{\frac{2}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})(\bar{X}-\mu)}_{=0} + \underbrace{\sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2}_{\text{let } Z = \nearrow}$$

$$\Rightarrow W = \frac{(n-1)S^2}{\sigma^2} + Z^2 \qquad \sum_{i=1}^{n}\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

pay attention to the conversion

Note that $W \sim \chi^2(n)$, $Z^2 \sim \chi^2(1)$, $\frac{(n-1)S^2}{\sigma^2}$ & Z^2 are independent

$$\Rightarrow E[e^{tW}] = E[e^{t\frac{(n-1)S^2}{\sigma^2}}]E[e^{tZ^2}] \Rightarrow E[e^{t\frac{(n-1)S^2}{\sigma^2}}] = (1-2t)^{-\frac{n-1}{2}},\ t < \tfrac{1}{2}$$

$$\Rightarrow \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

A remark: $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n) \xrightarrow{\text{Lose one degree of freedom}} \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$

A mistake $\begin{cases} b(n, p) \\ N(\mu, \sigma^2) \end{cases}$

③ Theorem 5.5-3 —— Student's t distribution

a. Definition

Let $T=\dfrac{Z}{\sqrt{U/r}}$, $Z\sim N(0,1)$, $U\sim\chi^2(r)$, Z and U are independent

$$f(t)=\frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\,\Gamma(\frac{r}{2})}\,\frac{1}{(1+\frac{t^2}{r})^{\frac{r+1}{2}}}, \quad t\in R, \quad T\sim t(r)$$

just one parameter

$\chi^2(r)$ 我是 r (U)

proof:

△ $F(t)=P(T\le t)=P(\frac{Z}{\sqrt{\frac{U}{r}}}\le t)=P(Z\le\sqrt{\frac{U}{r}}\,t)=\int_0^{\infty}\int_{-\infty}^{\sqrt{\frac{u}{r}}t} g(z,u)\,dz\,du=\int_0^{\infty}\int_{-\infty}^{\sqrt{\frac{u}{r}}t} f_Z(z)\cdot f_U(u)\,dz\,du$

△ $f(t)=F'(t)$

b. Properties

△ "Student's t Distribution is a heavy tailed distribution"

Standard Normal distribution :

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

Students' t distribution with $r=1$:

$$f(x)=\frac{1}{\pi(1+x^2)}, \quad x\in(-\infty,\infty)$$

$\boxed{x\to\infty,\ f_{SDN}(x)\to 0 \text{ faster}}$

to better estimate the sample in the tail

check the table

△ $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

proof: let $Z=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, $U=\dfrac{(n-1)S^2}{\sigma^2}$

$Z\sim N(0,1)$ $U\sim\chi^2(n-1)$

then $T=\dfrac{Z}{\sqrt{U/(n-1)}}=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

Normal distribution

$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

Lec 23

1. The central limit theorem

① Convergence in distribution

a. Concepts: converge in distribution = converge weakly = converge in law to a random variable

b. Definition: $\lim\limits_{n\to\infty} F_n(x) = F(x)$, for $\forall x \in \mathbb{R}$ whose $F(x)$ is continuous $\longrightarrow$ A sequence $X_1, X_2, \cdots$ is convergent

$F_n(x)$: cdf of X_n

对于任意x，函数列都收敛到对应函数

② Central limit theorem (CLT)

a. Definition

Only two assumptions

X_is are i.i.d. from a distribution with $\begin{cases} \text{finite } \mu \\ \text{finite, non-zero } \sigma^2 \end{cases}$,

then as $n \to \infty$, $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ converges to $N(0,1)$

- Normal D: $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- Arbitrary D: $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \to N(0,1)$

b. Practical uses —— Approximate the probability for large n ($n\uparrow$, accuracy$\uparrow$)

△ $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \longrightarrow N(0,1)$

△ $\bar{X} \longrightarrow N(\mu, \frac{\sigma^2}{n})$ $\longrightarrow$ SND

△ $\sum\limits_{i=1}^{n} X_i \longrightarrow N(n\mu, n\sigma^2)$

① Find RV ② compute $\begin{cases} \mu \\ \sigma^2 \end{cases}$ ③ $\dfrac{RV-\mu}{\sigma} \to N(0,1)$

$N(\mu, \sigma^2)$

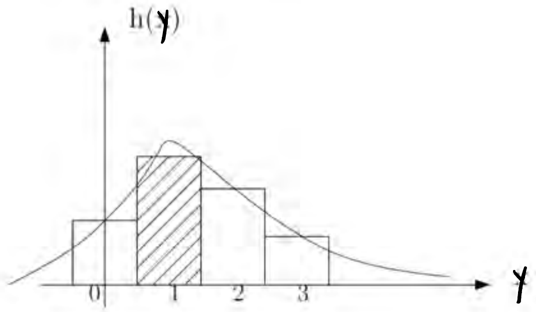
$\to$ Computing pmf / pdf for multivariate RV directly is difficult, so we use CLT to approximate.

2. Approximations for discrete distributions

① Idea: In CLT, X_is' distribution can be a discrete distribution.

Discrete D $\xleftarrow{\text{Approximate}}$ Normal D (Continuous D)

② Proof with histogram



half-unit correction for continuity $\to$ CLT

$P(Y=k) = \underbrace{P(k-\frac{1}{2} \le Y \le k+\frac{1}{2})}_{\text{Histogram}} \approx \underbrace{P(k-\frac{1}{2} \le Y \le k+\frac{1}{2})}_{\text{Curve}}$

确定实点

What about continuous D & normal D?

$\longrightarrow$ Directly draw two graphs.

Lec 24

1. Chebyshev's inequality

① Motivation:

- CLT: $\mu, \sigma^2 \Rightarrow$ approximate the P of $\bar{X} / \sum_{i=1}^{n} X_i$ $\longrightarrow$ $P(\bar{X} \in A) / P(\sum_{i=1}^{n} X_i \in A)$
- Chebyshev's inequality: $\mu, \sigma^2 \Rightarrow$ approximate the P of X (the distribution) $\longrightarrow$ $P(X \in A)$

arbitrary distribution

$\hookrightarrow$ finite μ & finite, nonzero σ^2

② Theorem:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}, \quad \boxed{k \geq 1} \leftarrow \frac{1}{k^2} \leq 1$$

$\longrightarrow$ P(two tails) is upper bounded

$$P\left(\frac{|X-\mu|}{\sigma} \geq k\right) \leq \frac{1}{k^2} \longrightarrow k\uparrow, P\downarrow$$

proof (for discrete case):

$$\sigma^2 = E(X-\mu)^2 = \sum_{x \in S} (x-\mu)^2 f(x)$$

$$= \sum_{x \in A} (x-\mu)^2 f(x) + \sum_{x \in A'} (x-\mu)^2 f(x), \quad \text{where } A = \{x \mid |x-\mu| \geq k\sigma\}$$

$$\Rightarrow \sigma^2 \geq \underline{\sum_{x \in A} (x-\mu)^2 f(x) \geq k^2\sigma^2 \sum_{x \in A}} f(x) = k^2\sigma^2 P(X \in A)$$

$\downarrow$ for $x \in A$

$$\Rightarrow P(X \in A) \leq \frac{1}{k^2}$$

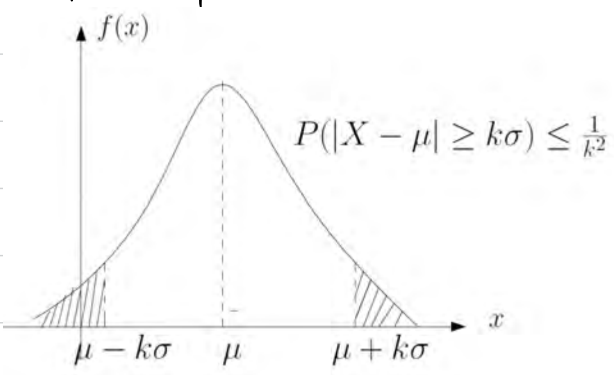
$$\Rightarrow P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

③ Corollary

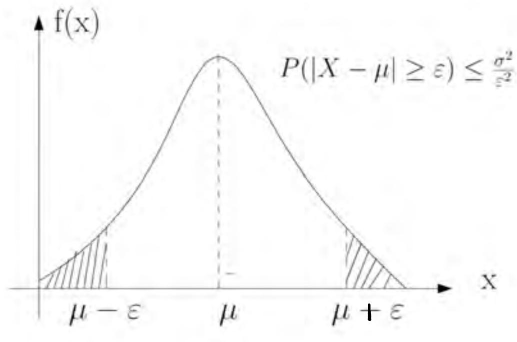
If $\varepsilon = k\sigma$, then $P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$

- k: unit / number of σ
- ε (epsilon)

④ Graphical interpretation



Theorem



Corollary

⑤ Application

Approximate the P of middle / two tails

- P(middle) $\geq C_1$ Lower bounded
- P(two tails) $\leq C_2$ Upper bounded

2. Markov's inequality

① Assumptions

a. X only takes nonnegative values, that is, $\bar{S} \in [0, +\infty)$;

b. Finite mean, μ; Finite, nonnegative σ

c. $a > 0$.

② Theorem:

$$P(X \geq a) \leq \frac{E(X)}{a}$$

proof.

$$E(X) = \sum_{x \in \bar{S}} x f(x) = \sum_{x < a, x \in \bar{S}} x f(x) + \sum_{x \geq a, x \in \bar{S}} x f(x)$$

$$\geq \sum_{x \geq a, x \in \bar{S}} x f(x) \geq a \sum_{x \geq a, x \in \bar{S}} f(x) = a P(X \geq a)$$

Lec 25

1. Theorem: Law of large number

① Convergence in probability

A sequence of RVs $\{Y_n\}_{n=1}^{\infty}$ is said to converge in probability to a RV Y

when $\lim_{n\to\infty} P(|Y_n - Y| \geq \varepsilon) = 0$

② Law of large number

X_is are i.i.d. with finite μ and finite, nonzero σ^2

then $\bar{X}$ converges in probability to μ,

i.e. $\lim_{n\to\infty} P(|\bar{X}-\mu| \geq \varepsilon) = 0$

$\bar{X}$ is an estimator of μ

proof:

$0 \leq P(|\bar{X}-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$

$\Rightarrow 0 \leq \lim_{n\to\infty} P(|\bar{X}-\mu| \geq \varepsilon) \leq \lim \frac{\sigma^2}{n\varepsilon^2} = 0$

$\Rightarrow \lim_{n\to\infty} P(|\bar{X}-\mu| \geq \varepsilon) = 0$

2. Limiting mgf technique.

Application: binomial → poisson
- pmf: take the limit
- mgf: limiting mgf technique

① Theorem

If $\lim_{n\to\infty} M_n(t) = M(t)$, for t in the open interval around $t=0$

then the D associated with $M_n(t)$ converge in D associated with $M(t)$

mgf → D

② Proof for CLT

a. Overview:

let $W = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$, then $\begin{cases} M_n(t): \text{mgf of } W \\ M(t): \text{mgf of SND} \end{cases}$ prove $\boxed{\lim_{n\to\infty} M_n(t) = M(t)}$

b. Detailed process:

$E(e^{tW}) = E\{\exp[\frac{t}{\sqrt{n}\sigma}(\sum_{i=1}^{n} x_i - n\mu)]\}$ $\qquad \bar{x} \to \frac{1}{n}\sum_{i=1}^{n} x_i$

$= E\{\exp\frac{t(X_1-\mu)}{\sqrt{n}\sigma} \cdots \exp\frac{t(X_n-\mu)}{\sqrt{n}\sigma}\}$

$= E\{\exp[\frac{t(X_1-\mu)}{\sqrt{n}\sigma}]\} \cdots E\{\exp[\frac{t(X_n-\mu)}{\sqrt{n}\sigma}]\}$ $\qquad$ Theorem 5.3-1

Let $Z_i = \frac{X_i-\mu}{\sigma}$, $i=1,\cdots,n$ $\qquad g(X_i) \longrightarrow Z_i$

then $Z_1 \cdots Z_n$ are i.i.d. with $\begin{cases} \mu=0 \\ \sigma^2=1 \end{cases}$ $\qquad E(Z_i) = \mu-\mu=0$; $Var(Z_i) = \frac{\sigma^2}{\sigma^2} = 1$

let $M(t) = E\{\exp(tZ_i)\}$, $|t|<h$

then $E[e^{tW}] = [M(\frac{t}{\sqrt{n}})]^n$, $|\frac{t}{\sqrt{n}}| < h$

By Taylor's expansion, there exists $t_1 \in (0, t)$, s.t.

$M(t) = M(0) + M'(0)t + \frac{1}{2}M''(t_1)t^2$ expand $M(t)$

$= 1 + \frac{1}{2}M''(t_1)t^2 = 1 + \frac{1}{2}t^2 + \frac{1}{2}t^2[M''(t_1) - 1]$

$\Rightarrow E[e^{tW}] = [M(\frac{t}{\sqrt{n}})]^n = [1 + \frac{1}{2}\frac{t^2}{n} + \frac{1}{2}\frac{t^2}{n}[M''(t_1) - 1]]^n, \ (|\frac{t}{\sqrt{n}}| < h, \ t_1 \in (0, \frac{t}{\sqrt{n}}))$ $() \to t \to t_1$

$\lim_{n \to \infty} M''(t_1) - 1 = 1 - 1 = 0$ $t_1 = \frac{t}{\sqrt{n}}$

then $\lim_{n \to \infty} E(e^{tW}) = \lim_{n \to \infty} [1 + \frac{1}{2}\frac{t^2}{n} + \frac{1}{2}\frac{t^2}{n}[M''(t_1) - 1]]^n = e^{\frac{t^2}{2}} \longrightarrow$ mgf of $N(0,1)$

prove right ✌