This lecture note was done by Choongrak Kim(Professor, Department of Statistics, Pusan National University), based on "Introduction to Mathematical Statistics" (Authors : Hogg, McKean, Craig ; Publisher : Prentice Hall, 6th edition). I thank my graduate students, Junho Jeong and Soyeong Noh, for typing tex file.

I prepared this note only for students taking Mathematical Statistics I(Department of Statistics, Pusan National University) and its OCW version, not for any other purpose. I declare that this lecture note is for the academic purpose only, and should not be used for commercial purpose. The use of this lecture note is free, and any attempts for commercial use is illegal.

January 1, 2019
Choongrak Kim, Professor
Department of Statistics, Pusan National University
Updated on January 1, 2020

이 강의노트는 부산대학교 통계학과 김충락 교수가 다음의 교재에 근거하여 작성하였습니다.

교재명 : Introduction to Mathematical Statistics
저자 : Hogg, McKean, Craig
출판사: Prentice Hall (6th edition)

노트를 Tex파일로 작성하는 과정에 부산대학교 통계학과 대학원에 재학 중인 정준호, 노소영에게 감사드립니다.

이 강의노트는 부산대학교 통계학과 김충락 교수가 강의하는 수리통계학 I(Mathematical Statistic I)과 그 강의의 동영상 버전인 OCW강의를 듣는 사람들을 위한 것입니다. 이 강의노트는 그러한 학문적 목적으로만 사용되어야 하며 상업적 목적의 사용을 금합니다. 이 강의노트의 사용은 무료이며, 그 어떤 상업적 목적의 사용은 불법임을 알려드립니다.

2019년 1월 1일
부산대학교 통계학과 교수 김충락
2020년 1월 1일에 갱신

1 Probability and Distributions

1.1 Introduction

- *Statistical(random) experiment*: the outcome cannot be predicted with certainty prior to the performance of the experiment.
- *Sample space*: collection of every possible outcome from the random experiment, and denoted by $\mathscr{C}$.
- *Event*: subset of sample, and denoted by A, B, C.

Example 1.1.1. Consider tossing a coin, then $\mathscr{C} = \{H, T\}$.

Example 1.1.2. Consider tossing two die (one red, the other white), then $\mathscr{C} = \{(1,1), \cdots, (1,6), (2,1), \cdots, (6,6)\}$.

Example 1.1.3. Let C denote an event of sum seven when tossing two die, then $\mathscr{C} = \{(1,6), (2,5), \cdots, (6,1)\}$.

Remark 1.1.1. Two types of probability

(i) Relative frequancy approach.

(ii) Personal or subjective approach.

1.2 Set Theory

Definition 1.2.1. If each element of set C_1 is also an element of set C_2, then C_1 is called *subset* of C_2, and denoted by $C_1 \subset C_2$.

Definition 1.2.2. If a set C has no elements, C is called the *null(empty)* set, and denoted by $C = \phi$.

Definition 1.2.3. The set of all elements that belong to at least one of C_1 and C_2 is called the *union* of C_1 and C_2, and denoted by $C_1 \cup C_2$ and it can be generalized to any number of sets. For example, $C_1 \cup C_2 \cup \cdots \cup C_n = \cup_{k=1}^{\infty} C_k$.

Example 1.2.1. Let $C_k = \left\{x : \frac{1}{k+1} \leq x \leq 1\right\}$, then $\bigcup\limits_{k=1}^{\infty} C_k = \{x : 0 < x \leq 1\}$.

Definition 1.2.4. The set of all elements that belong to each of C_1 and C_2 is called the *intersection* of C_1 and C_2. and denoted by $C_1 \cap C_2$, and it can be generalized to any number of sets $C_1 \cap C_2 \cap \cdots := \cap_{k=1}^{\infty} C_k$

Example 1.2.2. Let $C_k = \left\{x : 0 < x < \frac{1}{k}\right\}$, then $\bigcap\limits_{k=1}^{\infty} C_k = \phi$.

Definition 1.2.5. Let C be a subset of $\mathscr{C}$, then the set that consists of all elements of $\mathscr{C}$ that are not elements of C is called *complement* of C, and denoted by C^c or $\overline{C}$.

- A function is called *point* or *set function* if its domain is point or set, respectively.

Example 1.2.3. point function: $f(x) = 2x$, $f(1) = 2$
set function: $Q(A) =$ # of positive integers in A
$A = \{x : -\infty < x < 6\} \Rightarrow Q(A) = 5$

- The symbol
$$\int_C f(x)dx$$
means the ordinary Riemann integral of $f(x)$ over a one-dimensional set C, the symbol
$$\int\int_C g(x,y)dxdy$$
means the Riemann integral of $g(x,y)$ over a two-dimensional set C. Similarly, one or two-dimensional sum is
$$\sum_C f(x), \sum\sum_C g(x,y).$$

Example 1.2.4. Let $Q(C) = \int_C \cdots \int dx_1 dx_2 \cdots dx_n$. If $C = \{(x_1, x_2, \cdots, x_n) : 0 \le x_1 \le x_2 \le \cdots \le x_n \le 1\}$, then
$$Q(C) = \int_0^1 \int_0^{x_n} \cdots \int_0^{x_3} \int_0^{x_2} dx_1 dx_2 \cdots dx_{n-1} dx_n = \frac{1}{n!}.$$

1.3 The Probability Set Function

Definition 1.3.1. *(σ-field)* Let $\mathscr{B}$ be a collection of subsets of $\mathscr{C}$. We say $\mathscr{B}$ is a *σ-field* if

(i) $\phi \in \mathscr{B}$.

(ii) $C \in \mathscr{B} \Rightarrow C^c \in \mathscr{B}$ (closed under complement).

(iii) $C_1, C_2, \cdots \in \mathscr{B} \Rightarrow \bigcup_{i=1}^{\infty} C_i \in (B)$ (closed under countable union).

- Example of σ-field

1. $\mathscr{B} = \{\phi, C, C^c, \mathscr{C}\}$.
2. $\mathscr{B}$ is the power set of $\mathscr{C}$, i.e. the collection of all subsets of $\mathscr{C}$.
3. $\mathscr{B} = \bigcap_{i=1}^{\infty} \{\varepsilon_i : \mathscr{D} \subset \varepsilon_i,\ \varepsilon_i \text{ is a } \sigma\text{-field } \}$. This is the smallest σ-field which containing $\mathscr{D}$, and it is called the σ-field generated by $\mathscr{D}$.
4. Let $\mathscr{I}$ be the set of all open intevals in $\mathbb{R}$ (set of real numbers), then the σ-field generated by $\mathscr{I}$ is called the Borel σ-field.

Definition 1.3.2. (probability) Let $\mathscr{C}$ be a sample space, $\mathscr{B}$ be a σ-field on $\mathscr{C}$. Let P be a real-valued function defined on $\mathscr{B}$. Then P is called a *probability set function* if it satisfies the following three conditions

(i) $P(C) \geq 0$, $^{\forall}C \in \mathscr{B}$ (*non-negativity*).

(ii) $P(\mathscr{C}) = 1$ (*normality*).

(iii) $C_1, C_2, \cdots \in \mathscr{B}$ s.t. $C_m \cup C_n = \phi$, $^{\forall}m \neq n$,
then $P\left(\bigcap_{i=1}^{\infty} C_n\right) = \sum_{i=1}^{\infty} P(C_i)$ (*countable additivity*).

Theorem 1.3.1. $P(C) = 1 - P(C^c)$, $^{\forall}C \in \mathscr{B}$.

(pf) Since $C \cup C^c = \mathscr{C}$ and $C \cap C^c = \phi$,

$$\begin{aligned} 1 &= P(\mathscr{C}) \\ &= P(C) + P(C^c). \end{aligned}$$

Theorem 1.3.2. $P(\phi) = 0$.

(pf) By taking $C = \phi$, we have $C^c = \mathscr{C}$, then by Thm.1.3.1,

$$P(\phi) = 1 - P(\mathscr{C}) = 0.$$

Theorem 1.3.3. $C_1 \subset C_2 \Rightarrow P(C_1) \leq P(C_2)$.

(pf) $C_2 = C_1 \cup (C_1^c \cap C_2) \Rightarrow P(C_2) = P(C_1) + P(C_1^c \cap C_2) \geq P(C_1)$.

Theorem 1.3.4. $0 \le P(C) \le 1$, ${}^{\forall}C \in \mathscr{B}$.

(pf) $\phi \subset C \subset \mathscr{C} \Rightarrow P(\phi) \le P(C) \le P(\mathscr{C}) \Rightarrow 0 \le P(C) \le 1$.

Theorem 1.3.5. $P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$.

(pf) $C_1 \cup C_2 = C_1 \cup (C_1^c \cap C_2) \Rightarrow P(C_1 \cup C_2) = P(C_1) + P(C_2 \cap C_1^c)$

$C_2 = (C_1 \cap C_2) \cup (C_1^c \cap C_2) \Rightarrow P(C_2) = P(C_1 \cap C_2) + P(C_2 \cap C_1^c)$

Hence, we have

$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$.

Remark 1.3.1. (inclusion-exclusion formula)
For 3 sets C_1, C_2, C_3, it is not difficult to show that

$$P(C_1 \cup C_2 \cup C_3) = p_1 - p_2 + p_3$$

where $p_1 = P(C_1) + P(C_2) + P(C_3)$,
$p_2 = P(C_1 \cap C_2) + P(C_1 \cap C_3) + P(C_2 \cap C_3)$,
$p_3 = P(C_1 \cap C_2 \cap C_3)$.
In general,

$$P(C_1 \cup C_2 \cup \cdots \cup C_k) = p_1 - p_2 + p_3 - \cdots + (-1)^{k-1} p_k$$

where p_i is sum of probability of all possibe intersections of i sets.

- $C_1, C_2, \cdots$ are called mutually exclusive if $C_i \cap C_j = \phi$, ${}^{\forall}i \ne j$
- Mutually exclusive sets $C_1, C_2, \cdots$ are called exhaustive if $\bigcup_{i=1}^{\infty} C_i = \mathscr{C}$
- Notation:

$$\lim_{n\to\infty} C_n = \begin{cases} \bigcup_{n=1}^{\infty} C_n & \text{for increasing sequence} \\ \bigcap_{n=1}^{\infty} C_n & \text{for decreasing sequence} \end{cases}$$

Theorem 1.3.6. Let $\{C_n\}$ be a increasing sequence of events. Then

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcup_{n=1}^{\infty} C_n\right).$$

Let $\{C_n\}$ be a decreasing sequence of events. Then

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcap_{n=1}^{\infty} C_n\right).$$

(pf) Assume $\{C_n\}$ is increasing sequence, and let $R_1 = C_1,\ R_n = C_n \cap C_{n-1}^c$

$$\begin{aligned}
P(\lim_{n\to\infty} C_n) &= P\left(\bigcup_{n=1}^{\infty} C_n\right) \\
&= P\left(\bigcup_{n=1}^{\infty} R_n\right) \\
&= \sum_{n=1}^{\infty} P(R_n) \\
&= \lim_{n\to\infty} \sum_{j=1}^{n} P(R_j) \\
&= \lim_{n\to\infty} \left\{ P(R_1) + \sum_{j=2}^{n} P(R_j) \right\} \\
&= \lim_{n\to\infty} \left[P(C_1) + \sum_{j=2}^{n} \{P(C_j) - P(C_{j-1})\} \right] \\
&= \lim_{n\to\infty} [P(C_1) + \{P(C_2) - P(C_1)\} + \{P(C_3) - P(C_2)\} + \cdots + \{P(C_n) - P(C_{n-1})\} \\
&= \lim_{n\to\infty} P(C_n).
\end{aligned}$$

Theorem 1.3.7. (*Boole's Inequality*) Let $\{C_n\}$ be an arbitrary sequence of events. Then

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) \leq \sum_{n=1}^{\infty} P(C_n).$$

(pf) Let $D_n = \bigcup_{i=1}^{n} C_i$, then $\{D_n\}$ is increasing sequence of sets.
Since $D_j = D_{j-1} \cup C_j$

$$\begin{aligned} P(D_j) &= P(D_{j-1}) + P(C_j) - P(D_{j-1} \cap C_j) \\ &\leq P(D_{j-1}) + P(C_j) \end{aligned}$$

i.e. $P(D_j) - P(D_{j-1}) \leq P(C_j)$.
Now,

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} C_n\right) &= P\left(\bigcup_{n=1}^{\infty} D_n\right) \\ &= \lim_{n\to\infty} P(D_n) \\ &= \lim_{n\to\infty} \left[P(D_1) + \sum_{j=2}^{n} \{P(D_j) - P(D_{j-1})\}\right] \\ &\leq \lim_{n\to\infty} \left\{P(D_1) + \sum_{j=2}^{n} P(D_j)\right\} \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} P(C_j) \\ &= \sum_{n=1}^{\infty} P(C_n). \end{aligned}$$

1.4 Conditional Probability and Independence

Let $C_1, C_2 \subset \mathscr{C}$, then the *conditional probability of* C_2 *given* C_1 is defined as

$$P = (C_2 \mid C_1) = \frac{P(C_2 \cap C_1)}{P(C_1)}, \text{ if } P(C_1) > 0$$

Note that the conditional probability satisfies 3 conditions of probability

(i) $P(C_2 \mid C_1) \geq 0$ (non-negativity).

(ii) $P\left(\bigcup_{i=2}^{\infty} C_i \mid C_1\right) = \sum_{i=2}^{\infty} P(C_i \mid C_1)$ if $C_2, C_3, \cdots$ are mutually disjoint (countable additivity).

(iii) $P(C_1 \mid C_1) = 1$ (normality).

Example 1.4.1. Consider drawing cards successively from a deck, at random and without replacement. Find thd probability that the third spade appears on the sixth draw. ($\spadesuit$: *spade*, $\diamondsuit$: *diamond*, $\clubsuit$: *clover*, $\heartsuit$: *heart*)

(sol) C_1 : two spades in the first five draws.
C_2 : a spade on the sixth draw.
We need to compute $P(C_1 \cap C_2)$, and use $P(C_1 \cap C_2) = P(C_2 \mid C_1)P(C_1)$

$$P(C_1) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = 0.2743,\ P(C_2 \mid C_1) = 11/47 = 0.234 \Rightarrow P(C_1 \cap C_2) = 0.064$$

From the definition of conditional probability, we have $P(C_1 \cap C_2) = P(C_2 \mid C_1)P(C_1)$ which is called the multiplication rule. For 3 events,

$$P(C_2 \mid C_1 \cap C_2) = P(C_3 \cap C_1 \cap C_2)/P(C_1 \cap C_2)$$

$\Rightarrow P(C_1 \cap C_2 \cap C_3) = P(C_3 \mid C_1 \cap C_2)P(C_1 \cap C_2) = P(C_3 \mid C_1 \cap C_2)P(C_2 \mid C_1)P(C_1)$.

In general,

$$P(C_1 \cap C_2 \cap C_3 \cap \cdots) = P(C_1)P(C_2 \mid C_1)P(C_3 \mid C_1 \cap C_2)P(C_4 \mid C_1 \cap C_2 \cap C_3) \cdots .$$

Bayes Theorem: Let $C_1, C_2, \cdots, C_k$ be mutually exclusive and exhaustive events, s.t. $P(C_i) > 0$, $i = 1, \cdots, k$. Then,

$$P(C_j \mid C) = \frac{P(C_j)P(C \mid C_j)}{\sum_{i=1}^k P(C_i)P(C \mid C_i)}, \ j = 1, \cdots, k.$$

(pf) Since $C = (C \cap C_1) \cup (C \cap C_2) \cup \cdots \cup (C \cap C_k)$

$$\begin{aligned} \Rightarrow P(C) &= P(C \cap C_1) + \cdots + P(C \cap C_k) \\ &= P(C_1)P(C \mid C_1) + \cdots + P(C_k)P(C \mid C_k) \\ &= \sum_{i=1}^k P(C_i)P(C \mid C_i) : \text{law of total probability} \end{aligned}$$

Now,

$$P(C_j \mid C) = \frac{P(C_j \cap C)}{P(C)} = \frac{P(C_j)P(C \mid C_j)}{\sum_{i=1}^k P(C_i)P(C \mid C_i)}.$$

Remark 1.4.1. $P(C_j)$: *prior* probability, $P(C_j \mid C)$: *posterior* probability.

Definition 1.4.1. Two events C_1 and C_2 are independent if $P(C_1 \mid C_2) = P(C_1)$, i.e.

$$P(C_1 \mid C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)} = P(C_1) \Rightarrow P(C_1 \cap C_2) = P(C_1)P(C_2).$$

In general, $C_1, \cdots, C_n$ are called independent iff for every collection of k events $(2 \leq k \leq n)$,

$$P(C_{i_1} \cap C_{i_2} \cap \cdots \cap C_{i_k}) = P(C_{i_1}) \cdots P(C_{i_k}).$$

1.5 Random Variables

Definition 1.5.1. A function X is called a *random variable*(r.v.) if it assigns to each element $c \in \mathscr{C}$ one and only one number $X(c) = x$. The space or range of X is $\mathscr{D} = \{x : x = X(c), c \in \mathscr{C}\}$.

The r.v. X is called *discrete* r.v. or *continuous* r.v. if $\mathscr{D}$ is countable set or an interval of real numbers, respectively.

The probability function P is defined on $\mathscr{B}$. Now, we define a probability function P_X defined on $\mathscr{F}$, and P_X is often called induced probability function by r.v. X. i.e.

$$P(C),\ C \in \mathscr{B},\ P_X(B),\ B \in \mathscr{F}.$$

i.e.

$$P_X(B) = P[c \in \mathscr{C} : X(c) \in B],\ B \in \mathscr{F}.$$

Let X is discrete r.v. with $\mathscr{D} = \{d_1, \cdots, d_m\}$, then

$$P_X(d_i) = P(X = d_i),\ i = 1, \cdots, m$$

is called probability mass function(pmf) of X.

Example 1.5.1. Consider tossing two fair die and let X be the sum of upfaces. Then, $\mathscr{C} = \{(1,1), (1,2), \cdots, (6,6)\}$ and $\mathscr{D} = \{2, 3, \cdots, 12\}$. The probability of sum 4 is

$$P\left((1,3) \cup (2,3) \cup (3,1)\right) = P_X(4) = 3/36.$$

Definition 1.5.2. (Cumulative Distribution Function) The cumulative distribution function(cdf) of r.v. X is defined as

$$F_X(x) = P_X((-\infty, x]) = P(X \leq x).$$

Example 1.5.2. Let X be the upface of tossing a fair dice, then the cdf of X is

Example 1.5.3. Let X be a real number chosen at random from the interval $(0,1)$. Then, it is reasonable to assign

$$P_X[(a,b)] = b - a \text{ for}, 0 < a < b < 1.$$

Want to obtain cdf of r.v. X. Let $x < 0$, then $P(X \leq x) = 0$. Let $x > 1$, then $P(X \leq x) = 1$. Let $0 < x < 1$, then $P(X \leq x) = P(0 < X \leq x) = x - 0 = x$. Hence, the cdf of X is

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Theorem 1.5.1. (Properties of cdf)

(a) $F(a) \leq F(b), ^{\forall} a < b$ (nondecreasing).

(b) $\lim\limits_{x \to -\infty} F(x) = 0$.

(c) $\lim\limits_{x \to \infty} F(x) = 1$.

(d) $\lim\limits_{x \downarrow x_0} F(x) = F(x_0)$ (right continuous).

(pf) (a) $\{X \leq a\} \subset \{X \leq b\} \Rightarrow P(X \leq a) \leq P(X \leq b)$ by Thm.1.3.3

(b) $\lim\limits_{x \to -\infty} \{X \leq x\} = \phi \Rightarrow \lim\limits_{x \to -\infty} P(X \leq x) = 0$ by Thm.1.3.2

(c) $\lim\limits_{x \to \infty} \{X \leq x\} = \mathscr{C} \Rightarrow \lim\limits_{x \to \infty} P(X \leq x) = 1$

(d) Let $\{X_n\}$ be nay sequence s.t. $x_n \downarrow x_0$, and let $C_n = \{X \leq x_n\}$. Then, $\{C_n\}$ is decreasing and $\bigcap\limits_{n=1}^{\infty} C_n = \{X \leq x_0\}$. Hence, by Thm.1.3.6, $\lim\limits_{n \to \infty} F(x_n) = P\left(\bigcap\limits_{n=1}^{\infty} C_n\right) = F(x_0)$.

Theorem 1.5.2. $P(a < X \leq b) = F_X(b) - F_X(a)$, $^\forall a < b$.

(pf) $\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}$.

Theorem 1.5.3. $P(X = x) = F_X(x) - F_X(x-)$, $F_X(x-) = \lim\limits_{z \uparrow x} F_X(z)$, i.e. left limit.

(pf) $^\forall x \in R$, $\{x\} = \bigcap\limits_{n=1}^{\infty} \left(x - \dfrac{1}{n}, x\right]$, therefore by Thm.1.3.6,

$$\begin{aligned} P(X = x) &= P\left[\bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]\right] = \lim_{n\to\infty} P\left(x - \frac{1}{n} < X \leq x\right) \\ &= \lim_{n\to\infty} P\left[F_X(x) - F_X\left(x - \frac{1}{n}\right)\right] = F_X(x) - F_X(x-). \end{aligned}$$

1.6 Discrete Random Variables

Definition 1.6.1. (Discrete Random Variable) A r.v. is called *discrete* if its space is either finite or countable.

Definition 1.6.2. (Probability Mass Function) The *probability mass function*(pmf) of a discrete r.v. X with space $\mathscr{D}$ is given by

$$P_X(x) = P(X = x),\ x \in \mathscr{D}$$

- The support of a discrete r.v. X is the points where $P_X(x) > 0$.

Example 1.6.1. Consider tossing a fair coin. Let X be the number of flips need to obtain the first head. Find the the pmf of X.

(sol) We must have a string of $x-1$ tails followed by a head, i.e. $T \cdots TH$. Hence, by independence of each flip,

$$P(X = x) = \left(\frac{1}{2}\right)^{x-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{x},\ x = 1, 2, 3, \cdots .$$

Example 1.6.2. An urn contains 100 balls, 20 white and 80 black. Let X be the number of white balls when we draw 5 ball. Find the pmf of X.

(sol)

$$P_X(x) = \begin{cases} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}}, & x = 0, 1, 2, 3, 4, 5 \\ 0, & \text{otherwise.} \end{cases}$$

We are interested in computing the pmf of $Y = g(X)$ where the pmf of X is known and g is 1-1.

$$P_Y(y) = P(Y = y) = P[g(X) = y] = P(X = g^{-1}(y)) = P_X(g^{-1}(y)).$$

Example 1.6.3. Find pmf of $Y = X - 1$ when $P_X(x) = \left(\frac{1}{2}\right)^x$, $x = 1, 2, \cdots$.

(sol) $g(x) = x - 1 \Rightarrow g^{-1}(y) = y + 1$

$$\therefore P_Y(y) = P_X(y+1) = \left(\frac{1}{2}\right)^{y+1}, \; y = 0, 1, 2, \cdots : \textit{geometric distribution.}$$

1.7 Continuous Random Variables

Definition 1.7.1. (Continuous Random Variables) A r.v. X is called *continuous* if its cdf $F_X(x)$ is continuous, $^\forall x \in R$.

When we write cdf as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

then $f_X(t)$ is called the probability density function(pdf) of a continuous r.v. X.

$$f_X(x) = \frac{d}{dx}F_X(x)$$

Note that

$$P(X = x) = F_X(x) - F_X(x-) = 0$$

for conti. r.v. also,

$$P(a < X \leq b) = F_X(b) - F_X(a) = \int_{a}^{b} f_X(t)dt$$

and

$$P(a < X \leq b) = P(a \leq X \leq b) = P(a \leq X < b) = P(a < X < b).$$

By the properties of $F_X(x)$, we have

(i) $f_X(x) \geq 0 \leftarrow F_X(x)$ is nondecreasing

(ii) $\displaystyle\int_{-\infty}^{\infty} f_X(t)dt = 1 \leftarrow F_X(\infty) = 1$

Example 1.7.1. Consider selecting a point at random in the interior of a circle of radius 1. Find the pdf of X, where X denote the distance of the selected point from the origin.

(sol) Note that $0 \le x \le 1$

$$F_X(x) = P(X \le x) = \frac{\pi x^2}{\pi 1^2} = x^2$$

and $P(X \le 0) = 0, P(X \le 1) = 1$. Therefore,

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \le x \le 1 \\ 1, & x \ge 1 \end{cases}$$

$$\therefore f_X(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Example 1.7.2. Find the pdf of $Y = X^2$ in Ex.1.7.1.

(sol)

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(X^2 \le y),\ y > 0 \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= P(0 \le X \le \sqrt{y}) \\ &= F_X(\sqrt{y}) \\ &= (\sqrt{y})^2 \\ &= y,\ 0 < y < 1 \end{aligned}$$

$$\therefore f_Y(y) = I(0 < y < 1)$$

Example 1.7.3. Find the pdf of $Y = X^2$ when $f_X(x) = \frac{1}{2}I(-1 < x < 1)$.

(sol)

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(X^2 \le y),\, y > 0 \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}dx \\ &= \sqrt{y} \end{aligned}$$

$$\therefore F_Y(y) = \begin{cases} 0, & y < 0 \\ \sqrt{y}, & 0 \le y \le 1 \\ 1, & y \ge 1 \end{cases}$$

$$\therefore f_Y(y) = \frac{1}{2\sqrt{y}} I(0 \le y \le 1)$$

Theorem 1.7.1. Let X be a continuous random variable with pdf $f_X(x)$ and support S_X. Let $Y = g(X)$, where $g(x)$ is a one-to-one differentiable function, on the support of X, S_X. Then the pdf of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dx}{dy} \right|,\, y \in S_Y,$$

where the support of Y is the set $S_Y = \{y = g(x) : x \in S_X\}$.

(pf) Since g is one-to-one and continuous, it is either increasing or decreassing. First, assume it is increasing.

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

$$\therefore f_Y(y) = \frac{d}{dy}F_Y(y) = f_X(g^{-1}(y))\frac{dx}{dy}$$

g is decreasing, then

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

$$\therefore f_Y(y) = -f_X(g^{-1}(y))\frac{dx}{dy}$$

Therefore,

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$$

Example 1.7.4. Find pdf of $Y = -2\log X$, where $f_X(x) = I(0 < x < 1)$

(sol) $g^{-1}(y) = e^{-y/2}$, $dx/dy = -\frac{1}{2}e^{-y/2}$

$$\therefore f_Y(y) = \frac{1}{2}e^{-y/2},\ y > 0$$

Example 1.7.5.

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}(x+1), & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

$$P\left(-3 < X \le \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F(-3) = \frac{3}{4} - 0 = \frac{3}{4}$$

$$P(X = 0) = F(0) - F(0-) = \frac{1}{2} - 0 = \frac{1}{2}$$

1.8 Expectation of a Random Variable

Definition 1.8.1. The expectation of r.v. X is defined as

$$E(x) = \begin{cases} \int_{-\infty}^{\infty} x f_X(x)dx & \text{if } \int_{-\infty}^{\infty} |x| f(x)dx < \infty \text{ (conti.)} \\ \sum_{x \in S_X} x p_X(x) & \text{if } \sum |x| p(x) < \infty \text{ (discrete)} \end{cases}$$

Theorem 1.8.1. The expectation of $Y = g(X)$ is given by

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)dx & \text{if } \int_{-\infty}^{\infty} |g(x)| f(x)dx < \infty \text{ (conti.)} \\ \sum_{x \in S_X} g(x) p_X(x) & \text{if } \sum |g(x)| p(x) < \infty \text{ (discrete)} \end{cases}$$

(pf) discrete case only

$$\begin{aligned} \sum_{x \in S_X} g(x) p_X(x) &= \sum_{g \in S_Y} \sum_{\{x \in S_X, g(x) = y\}} g(x) p_X(x) \\ &= \sum_{y \in S_Y} y \sum_{\{x \in S_X, g(x) = y\}} p_X(x) \\ &= \sum_{y \in S_Y} y p_Y(y) \\ &= E(Y) \end{aligned}$$

Theorem 1.8.2. $E[k_1 g_1(X) + k_2 g_2(X)] = k_1 E[g_1(X)] + k_2 E[g_2(X)]$ if $E[g_1(X)]$ and $E[g_2(X)]$ exist.

(pf) We are only to show $\int |k_1 g_1(x) + k_2 g_2(x)| f_X(x)dx < \infty$. By triangular inequality($|\boldsymbol{a} + \boldsymbol{b}| \leq |\boldsymbol{a}| + |\boldsymbol{b}|$)

$$\int |k_1 g_1(x) + k_2 g_2(x)| f_X(x)dx \leq |k_1| \int |g_1(x)| f_X(x)dx + |k_2| \int |g_2(x)| f_X(x)dx < \infty$$

1.9 Some Special Expectations

Definition 1.9.1. $\mu = E(X)$: *mean* of r.v. X

Definition 1.9.2. $\sigma^2 = E(X-\mu)^2$: *variance* of r.v. X and $\sigma = \sqrt{(}\sigma^2)$: standard deviation

Definition 1.9.3. X: r.v. s.t. $E(e^{tX}) < \infty$, $|t| < h$ for some $h > 0$. Then, $M(t) = E(e^{tX})$ is called the *moment generating function*(mgf) of r.v. X.

Theorem 1.9.1. X, Y: r.v. with mgf $M_X(t)$ and $M_Y(t)$, respectively. Then, $F_X(z) = F_Y(z)$, ${}^\forall z \in R$ iff $M_X(t) = M_Y(t)$, ${}^\forall t \in (-h, h)$, $h > 0$.(uniqueness of mgf)

Remark 1.9.1. (1) mgf may not exist. For example, let X be r.v. with pdf $f(x) = \frac{1}{x^2} I(x > 1)$, then

$$\begin{aligned} M_X(t) &= \int_1^\infty e^{tX} \frac{1}{x^2} dx \\ &= \lim_{b\to\infty} \int_1^b \left(1 + tx + \frac{t^2x^2}{2} + \cdots \right) \frac{1}{x^2} dx \\ &= \lim_{b\to\infty} \int_1^b \left[-\frac{1}{x} + t\log x + \frac{t^2x^2}{2} + \cdots \right]_1^b = \infty \end{aligned}$$

(2) Sometimes can find the pdf from the mgf. Let

$$M_x(t) = \frac{1}{10}e^t + \frac{2}{10}e^{2t} + \frac{3}{10}e^{3t} + \frac{4}{10}e^{4t}.$$

Now,

$$M_X(t) = \sum e^{tx} p(x) = p(1)e^t + p(2)e^{2t} + p(3)e^{3t} + p(4)e^{4t}.$$

By the uniqueness of polynomial coeff., we must have

$$p(x) = \frac{x}{10},\ x = 1, 2, 3, 4$$

(3) Can compute $E(X^m)$, $m = 1, 2, \cdots$ using the mgf. By Taylor expansion,

$$\begin{aligned} M_X(t) &= E(e^{tX}) \\ &= E\left[1 + tX + \frac{t^2X^2}{2!} + \frac{t^3X^3}{3!} + \cdots\right] \\ &= 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \cdots \end{aligned}$$

$$\therefore M_X^{(m)}(0) = E(X^m).$$

(4) *characteristic function(ch.f)*

$$\varphi(t) = E(e^{itX}) : \text{ch.f of r.v.} X.$$

Note that ch.f always exist. why?

$$|\varphi(t)| = \left|\int e^{itx} f(x)dx\right| \le \int \left|e^{itx} f(x)\right| dx$$

Now,

$$|e^{itx}| = |\cos tx + i \sin tx| = \sqrt{\cos^2 tx + \sin^2 tx} = 1$$

$$\therefore |\varphi(t)| \le 1.$$

Also, can show $E(X) = -i\varphi'(0)$, $E(X^2) = -\varphi''(0)$

(5) *cumulant generating function*(cgf)

$$\psi(t) = \log M(t) : \text{cgf of r.v. } X.$$

Relation between moment and cumulant. Recall that

$$M(t) = 1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \cdots, \ \mu_m = E(X^m),$$

assume that

$$\begin{aligned}
\psi(t) &= \kappa_0 + \kappa_1 t + \frac{\kappa_2 t^2}{2!} + \frac{\kappa_3 t^3}{3!} + \cdots, \ \kappa_m : \text{m-th cumulant} \\
&= \log\left(1 + \mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots\right) \\
&= \left(\mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots\right) - \frac{1}{2}\left(\mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots\right)^2 + \frac{1}{3}\left(\mu_1 t + \frac{\mu_2 t^2}{2!} + \cdots\right)^3 - \cdots \\
&= \mu_1 t + \frac{1}{2}(\mu_2 - \mu_1^2)t^2 + \frac{1}{6}(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3) + \cdots
\end{aligned}$$

$\therefore \kappa_0 \equiv 0$, $\kappa_1 = \mu_1$, $\kappa_2 = \mu_2 - \mu_1^2 \equiv \sigma^2$, $\kappa_3 = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 = E(X-\mu)^3 = \mu_3'$.

(6) skewness and kurtosis

$$\rho_3 = E[(X-\mu)^3]/\sigma^3 : \textit{skewness}.$$

$$\rho_4 = E[(X-\mu)^4]/\sigma^4 : \textit{kurtosis}.$$

1.10 Important Inequalities

Theorem 1.10.1. If $E(X^m)$ exists then $E(X^k)$ exists for $k \leq m$.

(pf) We are only to prove $\int |x|^k f(x)dx < \infty$

$$\begin{aligned}
\int_{-\infty}^{\infty} |x|^k f(x)dx &= \int_{|x|\leq 1} |x|^k f(x)dx + \int_{|x|>1} |x|^k f(x)dx \\
&\leq \int_{|x|\leq 1} f(x)dx + \int_{|x|>1} |x|^m f(x)dx \\
&\leq \int_{-\infty}^{\infty} f(x)dx + \int_{-\infty}^{\infty} |x|^m f(x)dx \\
&= 1 + E|X|^m < \infty
\end{aligned}$$

Theorem 1.10.2. (*Markov's Inequality*). $u(X)$: nonnegative function of r.v. X. Assume $E[u(X)]$ exists. Then, $^\forall c > 0$, $P[u(X) \geq c] \leq E[u(X)]/c$.

(pf)Let $A = \{x : u(x) \geq c\}$. Then,

$$\begin{aligned}
E[u(X)] &= \int u(x)f(x)dx \\
&= \int_A u(x)f(x)dx + \int_{A^c} u(x)f(x)dx \\
&\geq \int_A u(x)f(x)dx \\
&\geq c\int_A f(x)dx \\
&= cP(u(X) \geq c).
\end{aligned}$$

Theorem 1.10.3. (*Chebyshev Inequality*). $P(|X-\mu| \geq k\sigma) \leq 1/k^2$, ${}^{\forall}k > 0$.

(pf) Let $u(X) = (X-\mu)^2$ and $c = k^2\sigma^2$, then

$P((X-\mu)^2 \geq k^2\sigma^2) \leq E[(X-\mu)^2]/k^2\sigma^2 \;\Rightarrow\; P(|X-\mu| \geq k\sigma) \leq 1/k^2$.

Definition 1.10.1. ϕ: function defined on (a,b), $-\infty \leq a < b \leq \infty$. ϕ is said to be *convex* if for all x, y in (a,b) and $0 < \gamma < 1$,

$$\phi[\gamma x + (1-\gamma)y] \leq \gamma\phi(x) + (1-\gamma)\phi(y).$$

ϕ is said to be strictly convex if the inequality is strict.

Theorem 1.10.4. Assume ϕ is differentiable on (a,b), then

(a) ϕ: convex iff $\phi'(x) \leq \phi'(y)$, ${}^{\forall}a < x < y < b$

(b) ϕ: strictly convex iff $\phi'(x) < \phi'(y)$, ${}^{\forall}a < x < y < b$.
If ϕ is twice differentiale on (a,b), then

(c) ϕ: convex iff $\phi''(x) > 0$, ${}^{\forall}a < x < b$

(d) ϕ: strictly convex iff $\phi''(x) > 0$, ${}^{\forall}a < x < b$

Theorem 1.10.5. (*Jensen's Inequality*). ϕ: convex on an open interval I. X: r.v. with support $S \subset I$ and $E(X) < \infty \rightarrow \phi[E(X)] \leq E[\phi(X)]$.

(pf)Let ξ is between x and μ, then

$$\begin{aligned} \phi(x) &= \phi(\mu) + \phi'(\mu)(x-\mu) + \frac{1}{2}\phi''(\xi)(x-\mu)^2 \\ &\geq \phi(\mu) + \phi'(\mu)(X-\mu) \Rightarrow \text{Take expectation on both sides.} \end{aligned}$$

Example 1.10.1. $\{a_1, \cdots, a_n\}$: set of positive numbers.
Let X be a r.v. s.t. $P(X = a_i) = 1/n$, $i = 1, \cdots, n$.

(i) $E(X) = \sum_{i=1}^{n} a_i \frac{1}{n} = \bar{a}$: arithmetic mean(AM)

(ii) Since $-\log x$ is convex, we have by Jensen's ineq.,

$$\begin{aligned} -\log[E(X)] &= -\log(\bar{a}) \\ &\leq E[-\log X] \\ &= -\frac{1}{n}\sum \log a_i \\ &= -\log(a_1 \cdots a_n)^{1/n} \end{aligned}$$

i.e. $(a_1 \cdots a_n)^{1/n}$: geometric mean(GE) $\leq \bar{a} = \frac{1}{n}\sum a_i$

(iii) Replace a_i by $1/a_i$, then

$$\left(\frac{1}{a_1 \cdots a_n}\right)^{1/n} \leq \frac{1}{n}\sum \frac{1}{a_i}$$

i.e. $(a_1 \cdots a_n)^{1/n} \geq \frac{1}{\frac{1}{n}\sum \frac{1}{a_i}}$: harmonic mean(HM)

We have shown the relationship HM $\leq$ GM $\leq$ AM.

2 Multivariate Distributions

2.1 Distributions of Two Random Variables

Definition 2.1.1. (X_1, X_2) is called *random vector* if X_1, X_2 are random variables which assign to each element c of $\mathscr{C}$ one and only one ordered pair of numbers $X_1(c) = x_1$, $X_2(c) = x_2$. The space of (X_1, X_2) is $\mathscr{D} = \{(x_1, x_2) : x_1 = X_1(c), x_2 = X_2(c), c \in \mathscr{C}\}$.

- will use the vector notation $\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = (X_1, X_2)'$
- The cdf of $\boldsymbol{X} = (X_1, X_2)'$ is
$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$
and can easily show
$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) =$$
$$F_{X_1,X_2}(b_1, b_2) - F_{X_1,X_2}(a_1, b_2) - F_{X_1,X_2}(b_1, a_2) + F_{X_1,X_2}(a_1, a_2)$$
- The joint prob. mass function of $\boldsymbol{X} = (X_1, X_2)'$ is
$$p_{X_1,X_2}(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$
if $\boldsymbol{X}$ is discrete random vector.
- For the continuous random vector, $f_{X_1,X_2}(x_1, x_2)$ satisfying
$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_{X_1,X_2}(w_1, w_2) dw_1 dw_2$$
is called the joint pdf, and we have
$$\frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1 \partial x_2} = f_{X_1,X_2}(x_1, x_2)$$

- $F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1, x_2)$: marginal cdf of X_1
- $p_{X_1}(x_1) = \sum_{x_2} p_{X_1,X_2}(x_1, x_2)$: marginal pmf of X_1
- $f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_2$: marginal pdf of X_1

Example 2.1.1. $f(x_1, x_2) = x_1 + x_2$, $0 < x_1 < 1$, $0 < x_2 < 1$ jpdf of X_1 and X_2 compute $P(X_1 \leq 1/2)$ and $P(X_1 + X_2 \leq 1)$.

(sol)

(i) $P(X_1 \leq 1/2) = \int_0^{1/2} f_1(x_1) dx_1$

$$
\begin{aligned}
f_{X_1}(x_1) &= \int f(x_1, x_2) dx_2 \\
&= \int_0^1 (x_1 + x_2) dx_2 \\
&= \left[x_1 x_2 + \frac{1}{2} x_2^2 \right]_0^1 \\
&= x_1 + \frac{1}{2}
\end{aligned}
$$

$$
\therefore P\left(X_1 \leq \frac{1}{2}\right) = \int_0^{1/2} (x_1 + \frac{1}{2}) dx_1 = \left[\frac{x_1^2}{2} + \frac{x_1}{2} \right]_0^{1/2} = \frac{3}{8}
$$

(ii)

$$
\begin{aligned}
P(X_1 + X_2 \leq 1) &= \int_0^1 \int_0^{1-x_2} f(x_1, x_2) dx_1 dx_2 \\
&= \int_0^1 \int_0^{1-x_2} (x_1 + x_2) dx_1 dx_2 \\
&= \frac{1}{3}
\end{aligned}
$$

- $$E[g(X_1,X_2)] = \begin{cases} \int\int g(x_1,x_2)f(x_1,x_2)dx_1dx_2 & \text{if} \int\int |g(x_1,x_2)|f(x_1,x_2)dx_1dx_2 < \infty \\ \sum_{x_1}\sum_{x_2} g(x_1x_2)p(x_1,x_2) & \text{if} \sum\sum |g(x_1,x_2)|p(x_1,x_2) < \infty \end{cases}$$

Theorem 2.1.1. $E[k_1g_1(X_1,X_2)+k_2g_2(X_1,X_2)] = k_1E[g_1(X_1,X_2)]+k_2E[g_2(X_1,X_2)]$: *linearity property of expectation.*

Example 2.1.2. $f(x_1,x_2) = 8x_1x_2I(0 < x_1 < x_2 < 1)$ compute $E(X_1X_2^2)$, $E(X_2)$, and $E[7X_1X_2^2+5X_2]$.

(sol)

(i) $E(X_1X_2^2) = \int_0^1 \int_0^{x_2} x_1x_2^2 8x_1x_2dx_1dx_2 = \frac{8}{21}$

(ii) $E(X_2) = \int_0^1 x_2 f_{X_2}(x_2)dx_2 = \int_0^1 x_2 \left[\int_0^{x_2} 8x_1x_2dx_1\right] dx_2 = \frac{4}{5}$

(iii) $E[7X_1X_2^2+5X_2] = 7\frac{8}{21}+5\frac{4}{5} = \frac{20}{3}$

Definition 2.1.2. Let $\boldsymbol{X} = (X_1, X_2)'$ be a random vector.

$$\begin{aligned} M_{\boldsymbol{X}}(\boldsymbol{t}) &= E\left[e^{\boldsymbol{t}'\boldsymbol{X}}\right], \ \boldsymbol{t} = (t_1, t_2)', \ ||\boldsymbol{t}|| < h, \ h > 0 \\ &= M_{X_1,X_2}(t_1, t_2) \\ &= E\left[e^{t_1 X_1 + t_2 X_2}\right] : \text{mgf of } \boldsymbol{X} = (X_1, X_2)' \end{aligned}$$

Note that $M_{X_1,X_2}(t_1, t_2) = \int\int e^{t_1 x_1 + t_2 x_2} f_{X_1,X_2}(x_1, x_2) dx_1 dx_2$

$$\begin{aligned} M_{X_1,X_2}(t_1, 0) &= \int\int e^{t_1 x_1} f(x_1, x_2) dx_1 dx_2 \\ &= \int\int e^{t_1 x_1} f(x_1, x_2) dx_2 dx_1 \\ &= \int e^{t_1 x_1} \left\{ \int f(x_1, x_2) dx_2 \right\} dx_1 \\ &= \int e^{t_1 x_1} f_{X_1}(x_1) dx_1 \\ &= E[e^{t_1 x_1}] \\ &= M_{X_1}(t_1) : \text{marginal mgf of } X_1 \end{aligned}$$

Similarly, $M_{X_1,X_2}(0, t_2) = M_{X_2}(t_2)$: marginal mgf of X_2

Example 2.1.3. $f(x, y) = e^{-y} I(0 < x < y < \infty)$: jpdf of (X, Y)

(sol) $M(t_1, t_2) = \int_0^\infty \int_x^\infty e^{t_1 x + t_2 y} e^{-y} dy dx = \dfrac{1}{(1 - t_1 - t_2)(1 - t_2)}$

$$M(t_1, 0) = \frac{1}{1 - t_1} : \text{mgf of } X,$$

$$M(0, t_2) = \frac{1}{(1 - t_2)^2} : \text{mgf of } Y.$$

2.2 Transformations: Bivariate R.V.'s

Want to find the distribution of $Y = g(X_1, X_2)$ when jpdf of X_1 and X_2 is known. Two methods are possible. First, find the cdf of Y and take derivative. Secondly, use transformation technique.

(1) discrete case

(X_1, X_2): discrete random vector with jpmf $p_{X_1,X_2}(x_1, x_2)$ and support S
$y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$: 1-1 transformation from S to $\mathscr{T}$.

$$(X_1, X_2) \xrightarrow{u_1,u_2} (Y_1, Y_2)$$

$x_1 = w_1(y_1, y_2)$, $x_2 = w_2(y_1, y_2)$: inverse function
$\Rightarrow p_{Y_1,Y_2}(y_1, y_2) = p_{X_1,X_2}(w_1(y_1, y_2), w_2(y_1, y_2))$, $(y_1, y_2) \in \mathscr{T}$.

Example 2.2.1. $p_{X_1,X_2}(x_1, x_2) = \mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1-\mu_2}/x_1!x_2!$, $x_1 = 0, 1, 2, \cdots$, $x_2 = 0, 1, 2, \cdots$ Find the pdf of $Y_1 = X_1 + X_2$.

(sol) need to define Y_2 s.t. $(x_1, x_2) \to (y_1, y_2)$ is 1-1.
Let $Y_2 = X_2$, then $y_1 = x_1 + x_2$ and $y_2 = x_2$ represent 1-1 transformation.

$$S = \{(x_1, x_2) : x_1 = 0, 1, 2, \cdots, x_2 = 0, 1, 2, \cdots\}$$

$$\to \mathscr{T} = \{(y_1, y_2) : y_1 = 0, 1, 2, \cdots, y_2 = 0, 1, \cdots, y_1\}$$

i.e. $x_1 = y_1 - y_2$, $x_2 = y_2$. So, the jpdf of Y_1 and Y_2 is

$$p_{Y_1,Y_2}(y_1, y_2) = \frac{\mu_1^{y_1-y_2}\mu_2^{y_2}e^{-\mu_1-\mu_2}}{(y_1 - y_2)!y_2!}, \ (y_1, y_2) \in \mathscr{T}$$

$$
\begin{aligned}
\therefore p_{Y_1}(y_1) &= \sum_{y_2=0}^{y_1} p_{Y_1,Y_2}(y_1,y_2) \\
&= \frac{e^{-\mu_1-\mu_2}}{y_1!}\sum_{y_2=0}^{y_1}\frac{y_1!}{(y_1-y_2)!y_2!}\mu_1^{y_1-y_2}\mu_2^{y_2} \\
&= \frac{(\mu_1+\mu_2)^{y_1}e^{-\mu_1-\mu_2}}{y_1!},\ y_1=0,1,2,\cdots
\end{aligned}
$$

(2) continuous case

Let X_1, X_2 be conti. r.v.'s, and $X = (X_1, X_2)'$ be random vector with jpdf $f_{X_1,X_2}(x_1,x_2)$ and support S. Consider a transformation $(x_1,x_2) \to (y_1,y_2)$ s.t. $y_1 = u_1(x_1,x_2)$ and $y_2 = u_2(x_1,x_2)$ be 1-1 and let $x_1 = w_1(y_1,y_2)$ and $x_2 = w_2(y_1,y_2)$ be inverse function with the Jacobian

$$
J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}
$$

Then , the jpdf of Y_1 and Y_2 is given by

$$
f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(w_1(y_1,y_2), w_2(y_1,y_2))|J|,\ (y_1,y_2) \in \mathscr{T}.
$$

Example 2.2.2. $f_{X_1,X_2}(x_1,x_2) = I(0 < x_1 < 1,\, 0 < x_2 < 1)$ Find the pdf of $Y_1 = X_1 + X_2$

(sol) Two methods are possible

(i) cdf technique

$$F_{Y_1}(y_1) = P(Y_1 \le y_1) = P(X_1 + X_2 \le y_1)$$

$$\therefore F_{Y_1}(y_1) = \begin{cases} 0 & , y_1 < 0 \\ \int_0^{y_1}\int_0^{y_1 - x_1} dx_2 dx_1 & , 0 \le y_1 < 1 \\ 1 - \int_{y_1-1}^{1}\int_{y_1-x_1}^{1} dx_2 dx_1 & , 1 \le y_1 < 2 \\ 1 & , y_1 \ge 2 \end{cases}$$

$$= \begin{cases} 0 & , y_1 < 0 \\ y_1^2/2 & , 0 \le y_1 < 1 \\ 1 - (2-y_1)^2/2 & , 1 \le y_1 < 2 \\ 1 & , y_1 \ge 2 \end{cases}$$

$$\therefore f_{Y_1}(y) = \begin{cases} y_1 & , 0 < y_1 < 1 \\ 2 - y_1 & , 1 \le y_1 < 2 \\ 0 & , \text{o.w.} \end{cases}$$

(ii) transformation technique

Need to define Y_2 s.t. $(x_1, x_2) \to (y_1, y_2)$ be 1-1.
Let $Y_2 = X_2$, then $y_1 = x_1 + x_2$, $y_2 = x_2$ represent 1-1 and $x_1 = y_1 - y_2$, $x_2 = y_2$ are inverse function. Jacobian is $J = 1$.

$$\therefore f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(y_1 - y_2, y_2)|J| = 1,\ 0 < y_2 < 1,\ y_2 < y_1 < 1 + y_2$$

$$\therefore f_{Y_1}(y) = \int f_{Y_1,Y_2}(y_1,y_2) dy_2$$

$$= \begin{cases} \int_0^{y_1} dy_2 = y_1 & , 0 < y_1 < 1 \\ \int_{y_1-1}^{1} dy_2 = 2 - y_1 & , 1 \le y_1 < 2 \end{cases}$$

Example 2.2.3. $f_{X_1,X_2}(x_1,x_2) = \frac{1}{4}exp\left[-\frac{x_1+x_2}{2}\right], 0 < x_1 < \infty, 0 < x_2 < \infty$
Find the pdf of $Y_1 = \frac{1}{2}(X_1 - X_2)$.

(sol) Let $Y_2 = X_2$, then $y_1 = \frac{1}{2}(x_1 - x_2), y_2 = x_2$ is 1-1 and $x_1 = 2y_1 + y_2$, $x_2 = y_2$ are inverse function with $J = 2$.

$$\mathscr{T} = \{(y_1, y_2) : -\infty < y_1 < \infty,\ y_2 > 0,\ -2y_1 < y_2\}$$

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= f_{X_1,X_2}(2y_1 + y_2, y_2)|J| \\ &= \frac{2}{4}exp\left[-\frac{1}{2}(2y_1 + y_2) - \frac{1}{2}y_2\right] \\ &= \frac{1}{2}e^{-y_1-y_2},\ -\infty < y_1 < \infty,\ y_2 > 0,\ -2y_1 < y_2 \end{aligned}$$

$$\therefore f_{Y_1}(y_1) = \begin{cases} \int_{-2y_1}^{\infty} \frac{1}{2}e^{-y_1-y_2}dy_2 = \frac{1}{2}e^{y_1} & ,-\infty < y_1 < 0 \\ \int_{0}^{\infty} \frac{1}{2}e^{-y_1-y_2}dy_2 = \frac{1}{2}e^{-y_1} & ,y_1 \geq 0 \end{cases}$$

$$= \frac{1}{2}e^{-|y_1|},\ -\infty < y_1 < \infty : \text{double exponential or Laplace pdf}$$

2.3 Conditional Distribution and Expectation

- $p_{X_2|X_1}(x_2|x_1) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)}$: conditional pmf of X_2 given $X_1 = x_1$
- $f_{X_2|X_1}(x_2|x_1) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)}$: conditional pdf of X_2 given $X_1 = x_1$

(e.g)

$$P(a < X_2 < b|X_1 = x_1) = \int_a^b f(x_2|x_1)dx_2$$

$$E[u(X_2)|x_1] = \int_{-\infty}^{\infty} u(x_2)f(x_2|x_1)dx_2 \text{ : conditional mean of } u(X_2) \text{ given } X_1 = x_1$$

$$Var(X_2|x_1) = E[\{X_2 - E(X_2|x_1)\}^2|x_1] = E(X_2^2|x_1) - E^2(X_2|x_1) :$$
$$\text{conditional var. of } X_2 \text{ given } X_1 = x_1$$

Example 2.3.1. Find $E(X_1|x_2)$ and $Var(X_1|x_2)$ when
$f(x_1, x_2) = 2I(0 < x_1 < x_2 < 1)$
(sol)

$$f_2(x_2) = \int_0^{x_2} 2dx_1 = 2x_2 I(0 < x_2 < 1)$$

$$\therefore f(x_1|x_2) = \frac{2}{2x_2} = \frac{1}{x_2} I(0 < x_1 < x_2 < 1)$$

$$\therefore E(X_1|x_2) = \int_0^{x_2} x_1 \frac{1}{x_2} dx_1 = \frac{x_2}{2} I(0 < x_2 < 1)$$

$$\therefore Var(X_1|x_2) = \int_0^{x_2} \left(x_1 - \frac{x_2}{2}\right)^2 \frac{1}{x_2} dx_1 = \frac{x_2^2}{12} I(0 < x_2 < 1)$$

Theorem 2.3.1. (a) $E[E(X_2|X_1)] = E(X_2)$

(b) $Var[E(X_2|X_1)] \leq Var(X_2) = Var[E(X_2|X_1)] + E[Var(X_2|X_1)]$

(pf)

(a)

$$\begin{aligned}
E[E(X_2|X_1)] &= \int \left\{ \int x_2 f_{X_2|X_1}(x_2|x_1) dx_2 \right\} f_{X_1}(x_1) dx_1 \\
&= \int \int x_2 \frac{f(x_1, x_2)}{f_{X_1}(x_1)} f_{X_1}(x_1) dx_2 dx_1 \\
&= \int \int x_2 f(x_1, x_2) dx_1 dx_2 \\
&= \int x_2 \left\{ \int f(x_1, x_2) dx_1 \right\} dx_2 \\
&= \int x_2 f_{X_2}(x_2) dx_2 \\
&= E(X_2)
\end{aligned}$$

(b)

$$\begin{aligned}
Var(X_2) &= E[(X_2 - \mu_2)^2],\ \mu_2 = E(X_2) \\
&= E[\{X_2 - E(X_2|X_1) + E(X_2|X_1) - \mu_2\}^2] \\
&= E[\{X_2 - E(X_2|X_1)\}^2] + E[\{E(X_2|X_1) - \mu_2\}^2] \\
&+ 2E[\{X_2 - E(X_2|X_1)\}\{E(X_2|X_1) - \mu_2\}]
\end{aligned}$$

Now,

$$E[\{X_2 - E(X_2|X_1)\}^2] = E[E[\{X_2 - E(X_2|X_1)\}^2|X_1]] = E[Var(X_2|X_1)]$$

$$E[\{E(X_2|X_1) - \mu_2\}^2] = E[\{E(X_2|X_1) - E(E(X_2|X_1))\}^2]$$

Therefore,

$$\begin{aligned}
Var(X_2) &= E[\{X_2 - E(X_2|X_1)\}^2] + E[\{E(X_2|X_1) - \mu_2\}^2] \\
&= E[Var(X_2|X_1)] + Var[E(X_2|X_1)] \\
&\geq Var[E(X_2|X_1)]
\end{aligned}$$

2.4 The Correlation Coefficient

X: r.v. with $\mu_1 = E(X), \sigma_1^2 = Var(X)$.
Y: r.v. with $\mu_2 = E(Y), \sigma_2^2 = Var(Y)$.

$Cov(X,Y) := E[(X_1 - \mu_1)(Y - \mu_2)] = E(XY) - E(X)E(Y)$: *covariance* between X and Y.

$$\rho := \frac{Cov(X,Y)}{\sigma_1\sigma_2} = \frac{E[(X-\mu_1)(Y-\mu_2)]}{\sigma_1\sigma_2} \text{ : \textit{corr. coef.} of } X \text{ and } Y.$$

Theorem 2.4.1. If $E(X|Y)$ is linear in X, then $E(Y|X) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$ and $E[Var(Y|X)] = \sigma_2^2(1-\rho^2)$.

(pf) Let $E(Y|X) = a + bX$, by taking expectation on both side

$$\begin{aligned} E[E(Y|X)] &= E[a+bX] \\ E(Y) &= a + bE(X) \\ \mu_2 &= a + b\mu_1 \end{aligned}$$

By multiplying X on both sides of $E(Y|X) = a + bX$

$$\begin{aligned} XE(Y|X) &= aX + bX^2 \\ E[XE(Y|X)] &= E[aX + bX^2] \\ E[E(XY|X)] &= aE(X) + bE(X^2) \\ E(XY) &= a\mu_1 + b(\sigma_1^2 + \mu_1^2) \\ \rho\sigma_1\sigma_2 + \mu_1\mu_2 &= a\mu_1 + b(\sigma_1^2 + \mu_1^2) \end{aligned}$$

$$\Rightarrow\ a = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1,\ b = \rho\frac{\sigma_2}{\sigma_1}$$

$$E(Y|X) = a + bX = \mu_2 - \rho\frac{\sigma_2}{\sigma_1}\mu_1 + \rho\frac{\sigma_2}{\sigma_1}X = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X - \mu_1)$$

$$
\begin{aligned}
E[Var(Y|X)] &= \int \left\{ \int (y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1))^2 f_{Y|X}(y|x)dy \right\} f_X(x)dx \\
&= \int\int \left\{ y-\mu_2-\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \right\}^2 f_{X,Y}(x,y)dydx \\
&= \int\int \left\{ (y-\mu_2)^2-2\rho(y-\mu_2)\frac{\sigma_2}{\sigma_1}(x-\mu_1)+\rho^2\frac{\sigma_2^2}{\sigma_1^2}(x-\mu_1)^2 \right\} f_{X,Y}(x,y)dydx \\
&= Var(Y)-2\rho\frac{\sigma_2}{\sigma_1}Cov(X,Y)+\rho^2\frac{\sigma_2^2}{\sigma_1^2}Var(X) \\
&= \sigma_2^2-2\rho\frac{\sigma_2}{\sigma_1}\rho\sigma_1\sigma_2+\rho^2\frac{\sigma_2^2}{\sigma_1^2}\sigma_1^2 \\
&= \sigma_2^2(1-\rho^2)
\end{aligned}
$$

Since

$$\frac{\partial^{k+m}M(t_1,t_2)}{\partial t_1^k \partial t_2^m} = \int\int x^k y^m e^{t_1x+t_2y} f(x,y)dxdy,$$

we have

$$\left.\frac{\partial^{k+m}M(t_1,t_2)}{\partial t_1^k \partial t_2^m}\right]_{t_1=t_2=0} = E(X^kY^m)$$

2.5 Independent Random Variables

Definition 2.5.1. X and Y: *indep.* iff $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

Theorem 2.5.1. X and Y: indep. iff $f_{X,Y}(x,y) = g(x)h(y)$, where $g(x)$ is function of x only and $g(y)$ is function of y only.

(pf) ($\Rightarrow$) If X and Y are indep., then $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, so that take $g(x) = f_X(x)$, $h(y) = f_Y(y)$.

($\Leftarrow$) Assume $f_{X,Y}(x,y) = g(x)h(y)$, then

$$\begin{aligned} f_X(x) &= \int f(x,y)dy \\ &= \int g(x)h(y)dy \\ &= g(x)\int h(y)dy \\ &= c_1 g(x),\ c_1 = \int h(y)dy \end{aligned}$$

$$\begin{aligned} f_Y(y) &= \int f(x,y)dx \\ &= \int g(x)h(y)dx \\ &= h(y)\int g(x)dx \\ &= c_2 h(y),\ c_2 = \int g(x)dx \end{aligned}$$

Also, $1 = \int\int g(x)h(y)dxdy = c_1c_2$

$\therefore f(x,y) = g(x)h(y) = c_1g(x)c_2h(y) = f_X(x)f_Y(y)$

Example 2.5.1. $f_{X,Y}(x,y) = (x+y)I(0 < x < 1, 0 < y < 1)$: jpdf of X and Y. Are X and Y indep.?

(sol) Note that we cannot express $f_{X,Y}(x,y)$ as a product pf $g(x)$ and $h(y)$. Hence, X and Y are not indep.

Theorem 2.5.2. X and Y: indep. iff $F_{X,Y}(x,y) = F_X(x)F_Y(y)$.

(pf) ($\Rightarrow$)

$$\begin{aligned} F_{X,Y}(x,y) &= \int_{-\infty}^{x}\int_{-\infty}^{y} f(t,w)dwdt \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_X(t)f_Y(w)dwdt \\ &= \int_{-\infty}^{x} f_X(t)dt \int_{-\infty}^{y} f_Y(w)dw \\ &= F_X(x)F_Y(y) \end{aligned}$$

($\Leftarrow$)

$$\begin{aligned} f_{X,Y}(x,y) &= \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} \\ &= \frac{\partial^2 F_X(x)F_Y(y)}{\partial x \partial y} \\ &= f_X(x)f_Y(y) \end{aligned}$$

Theorem 2.5.3. X and Y: indep. iff $P(a < X \leq b, c < Y \leq d) = P(a < X \leq b)P(c < Y \leq d)$.

(pf) ($\Rightarrow$)

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) &= F(b,d) - F(a,d) - F(b,c) + F(a,c) \\ &= F_x(b)F_Y(d) - F_X(a)F_Y(d) - F_X(d)F_Y(c) + F_X(a)F_Y(c) \\ &= \{F_X(b) - F_X(a)\}\{F_Y(d) - F_Y(c)\} \\ &= P(a < X \leq b)P(c < Y \leq d) \end{aligned}$$

($\Leftarrow$) trivial

Theorem 2.5.4. X and Y: indep. $\Rightarrow E[u(X)v(Y)] = E[u(X)]E[v(Y)]$.

(pf)

$$\begin{aligned} E[u(X)v(Y)] &= \int\int u(x)v(y)f_{X,Y}(x,y)dxdy \\ &= \int\int u(x)v(y)f_X(x)f_Y(y)dxdy \\ &= \int u(x)f_X(x)dx \int v(y)f_Y(y)dy \\ &= E[u(X)]E[v(Y)] \end{aligned}$$

Theorem 2.5.5. X and Y: indep. iff $M(t_1,t_2) = M(t_1,0)M(0,t_2)$.

(pf) ($\Rightarrow$)

$$\begin{aligned} M(t_1,t_2) &= E\left[e^{t_1X+t_2Y}\right] \\ &= E\left[e^{t_1X}e^{t_2Y}\right] \\ &= E\left[e^{t_1X}\right]E\left[e^{t_2Y}\right] \\ &= M(t_1,0)M(0,t_2) \end{aligned}$$

($\Leftarrow$)

$$\begin{aligned} M(t_1,0)M(0,t_2) &= \int e^{t_1x}f_X(x)dx \int e^{t_2y}f_Y(y)dy \\ &= \int\int e^{t_1x+t_2y}f_X(x)f_Y(y)dxdy \\ &= \int\int e^{t_1x+t_2y}f_{X,Y}(x,y)dxdy \end{aligned}$$

By the uniqueness of mgf, we must have $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

2.6 Extension to Several Random Variables

Definition 2.6.1. $\boldsymbol{X} = (X_1, \cdots, X_n)'$: n-dim random vector, X_i's: r.v.'s

- $F_{\boldsymbol{X}}(\boldsymbol{x}) = P(\boldsymbol{X} \le \boldsymbol{x}) = P(X_1 \le x_1, \cdots, X_n \le x_n)$: joint cdf.

- $Y = u(X_1, \cdots, X_n) \Rightarrow E(Y) = \int \cdots \int u(x_1, \cdots, x_n) f_{X_1, \cdots, X_n}(x_1, \cdots, x_n) dx_1 \cdots dx_n$

$$f_{X_1}(x_1) = \int \cdots \int f_{X_1, \cdots, X_n}(x_1, \cdots, x_n) dx_2 \cdots dx_n$$

$$f_{X_1, X_3}(x_1, x_3) = \int \cdots \int f_{X_1, \cdots, X_n}(x_1, \cdots, x_n) dx_2 dx_4 \cdots dx_n$$

$$f_{2, \cdots, n|1}(x_2, \cdots, x_n | x_1) = \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n)}{f_1(x_1)}$$

$$f_{1|2, \cdots, n}(x_1 | x_2, \cdots, x_n) = \frac{f_{X_1, \cdots, X_n}(x_1, \cdots, x_n)}{f_{X_2, \cdots X_n}(x_2, \cdots, x_n)}$$

Remark 2.6.1. mutually indep. $\underset{\times}{\overset{\circ}{\rightleftharpoons}}$ pairwise indep.
(counter example)

$$f(x_1, x_2, x_3) = \frac{1}{4}, \ (x_1, x_2, x_3) \in \{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$$

$$f_{ij}(x_i, x_j) = \frac{1}{4}, \ (x_i, x_j) \in \{(0,0), (1,0), (0,1), (1,1)\}$$

$$f_i(x_i) = \frac{1}{2}, \ x_i = 0, 1$$

$$\therefore f_{ij}(x_i, x_j) = f_i(x_i) f_j(x_j) \text{ but } f(x_1, x_2, x_3) \ne f_1(x_1) f_2(x_2) f_3(x_3)$$

- $X_1, \cdots, X_n$ are called iid(independent and identically distributed) if $X_1, \cdots, X_n$ are mutually indep and have the same distribution.

- $E(\boldsymbol{X}) = (E(X_1), \cdots, E(X_n))'$
  $E(\boldsymbol{W}) = [E(W_{ij})]$, where $\boldsymbol{W}$ is $m \times n$ matrix of random variables.

Theorem 2.6.1. $\boldsymbol{W}_1$, $\boldsymbol{W}_2$: $m \times n$ matrices of r.v.'s.
$\boldsymbol{A}_1$, $\boldsymbol{A}_2$: $k \times m$ matrices of constants. $\boldsymbol{B}$: $n \times l$ matrix of constants. Then

$$E[\boldsymbol{A}_1\boldsymbol{W}_1 + \boldsymbol{A}_2\boldsymbol{W}_2] = \boldsymbol{A}_1E[\boldsymbol{W}_1] + \boldsymbol{A}_2E[\boldsymbol{W}_2]$$

$$E[\boldsymbol{A}_1\boldsymbol{W}_1\boldsymbol{B}] = \boldsymbol{A}_1E[\boldsymbol{W}_1]\boldsymbol{B}$$

- $\boldsymbol{\mu} = E(\boldsymbol{X})$: mean of $\boldsymbol{X}$.

- $Cov(\boldsymbol{X}) = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] = E[\boldsymbol{X}\boldsymbol{X}'] - \boldsymbol{\mu}\boldsymbol{\mu}' = [\sigma_{ij}]$: variance-covariance matrix.

  $Cov(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}Cov(\boldsymbol{X})\boldsymbol{A}'$

  Variance-covariance matrix $Cov(\boldsymbol{X})$ is p.s.d. i.e. $\boldsymbol{a}'Cov(\boldsymbol{X})\boldsymbol{a} \geq 0$. why?
  Let $Y = \boldsymbol{a}'\boldsymbol{X}$, then $0 \leq Var(Y) = Var(\boldsymbol{a}'\boldsymbol{X}) = \boldsymbol{a}'Cov(\boldsymbol{X})\boldsymbol{a}$.

2.7 Transformation of Random Vectors

Consider transforming n random variales $X_1, \cdots, X_n$ to n random variables $Y_1, \cdots, Y_n$ s.t. $y_1 = u_1(x_1, \cdots, x_n), \cdots, y_n = u_n(x_1, \cdots, x_n)$.

(1) one-to-one transformation case

$S \to \mathscr{T}$ is 1-1 s.t. $x_1 = w_1(y_1, \cdots, y_n), \cdots, x_n = w_n(y_1, \cdots, y_n)$

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix}, \ f(x_1, \cdots, x_n) : \text{jpdf of } X_1, \cdots, X_n$$

Then, the jpdf of $Y_1, \cdots, Y_n$ is

$$g(y_1, \cdots, y_n) = |J| f(w_1(y_1, \cdots, y_n), \cdots, w_n(y_1, \cdots, y_n))$$

Example 2.7.1. $f(x_1, x_2, x_3) = 48x_1x_2x_3 I(0 < x_1 < x_2 < x_3 < 1)$, jpdf of $Y_1 = X_1/X_2$, $Y_2 = X_2/X_3$, $Y_3 = X_3$

(sol) $x_1 = y_1y_2y_3$, $x_2 = y_2y_3$, $x_3 = y_3$ (1-1 transf.)

$$J = \begin{vmatrix} y_2y_3 & y_1y_3 & y_1y_2 \\ 0 & y_3 & y_2 \\ 0 & 0 & 1 \end{vmatrix} = y_2y_3^2$$

$0 < y_1 < 1, 0 < y_2 < 1, 0 < y_3 < 1$

$\therefore g(y_1, y_2, y_3) = 48(y_1y_2y_3)(y_2y_3)(y_3)|y_2y_3^2| = 48y_1y_2^3y_3^5$, $0 < y_i < 1$, $i = 1, 2, 3$

(2) many-to-one transformation case

$S \to \mathscr{T}$ is $k-1$.

Let $A_1, \cdots, A_k$ be exhaustive sets s.t. $\bigcup_{i=1}^{k} A_i = S$ and $A_i \cap A_j = \phi$, and $A_i \to \mathscr{T}$ is 1-1 for each $i = 1, \cdots, k$. Then, we apply the same method to each $A_i \to \mathscr{T}$. i.e.

$$g(y_1, \cdots, y_n) = \sum_{i=1}^{k} |J_i| g(w_{1i}(y_1, \cdots, y_n), \cdots, w_{ni}(y_1, \cdots, y_n))$$

Example 2.7.2. $f(x_1, x_2) = \frac{1}{\pi} I(0 < x_1^2 + x_2^2 < 1)$. Find the jpdf of $Y_1 = X_1^2 + X_2^2$, $Y_2 = \frac{X_1^2}{(X_1^2 + X_2^2)}$.

(sol) $y_1 y_2 = x_1^2$, $x_2^2 = y_1(1-y_2)$, $0 < y_1 < 1$, $0 < y_2 < 1$, i.e. $x_1 = \pm\sqrt{y_1 y_2}$, $x_2 = \pm\sqrt{y_1(1-y_2)}$.

$$A_1,\ x_1 = \sqrt{y_1 y_2},\ x_2 = \sqrt{y_1(1-y_2)}$$

$$A_2,\ x_1 = -\sqrt{y_1 y_2},\ x_2 = \sqrt{y_1(1-y_2)}$$

$$A_3,\ x_1 = -\sqrt{y_1 y_2},\ x_2 = -\sqrt{y_1(1-y_2)}$$

$$A_1,\ x_1 = \sqrt{y_1 y_2},\ x_2 = -\sqrt{y_1(1-y_2)}$$

$$J = \begin{vmatrix} \frac{1}{2}\sqrt{\frac{y_2}{y_1}} & \frac{1}{2}\sqrt{\frac{y_1}{y_2}} \\ \frac{1}{2}\sqrt{\frac{(1-y_2)}{y_1}} & -\frac{1}{2}\sqrt{\frac{y_1}{(1-y_2)}} \end{vmatrix} = -\frac{1}{4}\frac{1}{\sqrt{y_2(1-y_2)}}$$

similarly, $J_2 = J_3 = J_4 = J_1$

$$\begin{aligned}
\therefore g(y_1, y_2) &= \sum_{i=1}^{4} |J_i| f(w_{1i}(y_1, y_2), w_{2i}(y_1, y_2)) \\
&= \frac{4}{\pi}\frac{1}{4}\frac{1}{\sqrt{y_2(1-y_2)}} \\
&= \frac{1}{\pi\sqrt{y_2(1-y_2)}} I(0 < y_1 < 1,\ 0 < y_2 < 1)
\end{aligned}$$

3 Some Special Distributions

3.1 The Binomial and Related Distributions

① binomial distribution

(i) binomial equation

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} b^x a^{n-x}$$

(ii) Bernoulli trial

A seq. of experiment is called *Bernoulli trials* if each outcome is either success or failure, and each trial is indep. $X_1, \cdots, X_n$ are called Bernoulli r.v.'s if $X_1, \cdots, X_n$ are indep. and $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$, $0 \leq p \leq 1$. We denote that $X_i \sim B(n, p)$. Note that

$$E(X_i) = \sum_{x_i=0}^{1} X_i f(x_i) = 0(1-p) + p = p$$

$$Var(X_i) = E(X_i^2) - E^2(X_i) = 0^2(1-p) + 1^2 \times p - p^2 = p - p^2 = p(1-p)$$

(iii) pmf of binomial distriution

Let $X_1, \cdots, X_n$ be bernoulli trials with prob. of success p. i.e. $X_i \sim B(1,p)$, X_i's are indep. Then, $X = \sum_{i=1}^n X_i$ is the number of success out of n trials, and X is called to have *binomial distribution* with pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 0, 1, \cdots, n$$

(iv) mgf

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_{x=0}^n e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} \\ &= [(1-p) + pe^t]^n \end{aligned}$$

(v) μ and σ^2

$$\begin{aligned} M'(t) &= n[(1-p) + pe^t]^{n-1} pe^t \\ M''(t) &= n(n-1)[(1-p) + pe^t]^{n-2} p^2 e^{2t} + n[(1-p) + pe^t]^{n-1} pe^t \end{aligned}$$

$\therefore \mu = M'(0) = np$, $\sigma^2 = M''(0) - M'(0)^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p)$

(vi) $X_i \sim B(n_i, p)$, $i = 1, \cdots, m$. X_i's are indep. $\Rightarrow Y = \sum_{i=1}^{m} X_i \sim B\left(\sum_{i=1}^{m} n_i, p\right)$

(pf) Use the uniqueness of mgf, i.e. 1-1 correspondence between pdf and mgf

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t\sum_{x=1}^{m} X_i}] \\
&= E[e^{tX_1+\cdots+tX_m}] \\
&= E[e^{tX_1} \cdots e^{tX_m}] \\
&= \prod_{i=1}^{m} E[e^{tX_i}] \\
&= \prod_{i=1}^{m} M_{X_i}(t) \\
&= \prod_{i=1}^{m} [(1-p)+pe^t]^{n_i} \\
&= [(1-p)+pe^t]^{\sum_i^m n_i} : \text{mgf of } B\left(\sum_{i=1}^{m} n_i, p\right)
\end{aligned}
$$

Example 3.1.1. (*WLLN: Weak Law of Large Numbers*)
$Y \sim B(n,p) \Rightarrow P\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) \to 0$ as $n \to \infty$

(pf) Will use Tchebyshev's inequality

$$
\begin{aligned}
P\left(\left|\frac{Y}{n} - p\right| \geq \varepsilon\right) &= P(|Y-np| \geq n\varepsilon),\ E(Y) = np,\ Var(Y) = np(1-p) = \sigma^2 \\
&= P\left(|Y-\mu| \geq \varepsilon\sqrt{\frac{n}{p(1-p)}}\sqrt{np(1-p)}\right) \\
&= P\left(|Y-\mu| \geq \varepsilon\sqrt{\frac{n}{p(1-p)}}\sigma\right) \\
&= P(|Y-\mu| \geq k\sigma),\ k = \varepsilon\sqrt{\frac{n}{p(1-p)}} \\
&\leq \frac{1}{k^2} = \frac{p(1-p)}{n\varepsilon^2} \to 0 \text{ as } n \to \infty
\end{aligned}
$$

Example 3.1.2. X_1, X_2, X_3: indep. with the same pdf $f(x)$ and cdf $F(x)$. Find the pdf of $Y = mid(X_1, X_2, X_3)$.

(sol) First, find the cdf of Y

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(mid(X_1, X_2, X_3) \leq y) \\ &= P(\text{at least two of } X_1, X_2, X_3 \text{ are } \leq y) \end{aligned}$$

Let $\{X_i \leq y\}$ be success, and Y be the number of successes out of 3.
i.e. $Y \sim B(3, p)$, $p = P(X_i \leq y) = F(y)$.
Now,

$$\begin{aligned} G(y) &= P(Y \geq 2) \\ &= P(Y = 2) + P(Y = 3) \\ &= \binom{3}{2}[F(y)]^2[1 - F(y)] + [F(y)]^3 \end{aligned}$$

$$\therefore g(y) = G'(y) = 6F(y)[1 - F(y)]f(y)$$

② negative binomial distribution

(i) definition

Consider a seq. of indep. Bernoulli trials $B(1, p)$. Let Y be the number of failures before the r-th success, then Y is called to have the *negative binomial distribution*. The pmf of Y is

$$p(y) = \binom{y + r - 1}{r - 1} p^r (1 - p)^y,\ y = 0, 1, 2, \cdots$$

and denoted by $Y \sim NB(r, p)$.

(ii) mgf

$$\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= \sum_{y=0}^{\infty} e^{ty} \binom{y+r-1}{r-1} p^r (1-p)^y \\
&= p^r \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} [(1-p)e^t]^y \\
&= p^r \left\{ 1x^0 + rx + \frac{r(r+1)}{2} x^2 + \cdots \right\}, \ (1-p)e^t = x
\end{aligned}$$

Now, we consider

$$g(x) = (1-x)^{-r}$$

By using the Taylor expansion of $g(x)$ w.r.t. $x = 0$,

$$\begin{aligned}
g(x) &= g(0) + (x-0)g'(0) + \frac{1}{2}(x-0)^2 g''(0) + \cdots \\
g(0) &= (1-0)^{-r} = 1 \\
g'(x) &= -r(1-x)^{-r-1}(-1) \Rightarrow g'(0) = r \\
g''(x) &= -r(-r-1)(1-x)^{-r-2} \Rightarrow g''(0) = r(r+1)
\end{aligned}$$

$$\begin{aligned}
\therefore g(x) &= (1-x)^{-r} \\
&= 1 + rx + \frac{r(r+1)}{2} x^2 + \cdots
\end{aligned}$$

$$\therefore M_Y(t) = p^r (1-x)^{-r} = p^r [1-(1-p)e^t]^{-r}$$

(iii) μ and σ^2.

$$\begin{aligned}
M'(t) &= p^r(-r)[1-(1-p)e^t]^{-r-1}[-(1-p)e^t] \\
M''(t) &= p^r[-r(r+1)][1-(1-p)e^t]^{-r-2}[-(1-p)^2 e^{2t}] \\
&\quad + p^r(-r)[1-(1-p)e^t]^{-r-1}[-(1-p)e^t]
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mu &= M'(0) \\
&= p^r r p^{-r-1}(1-p) \\
&= \frac{r(1-p)}{p}
\end{aligned}$$

$$\begin{aligned}
\sigma^2 &= M''(0) - M'(0)^2 \\
&= \frac{p^r r(r+1)p^{-r-2}(1-p)^2 + p^r p^{-r-1} r(1-p) - r^2(1-p)^2}{p^2} \\
&= \frac{r(1-p)}{p^2}
\end{aligned}$$

(iv) geometric distribution

$Y \sim NB(1, p)$ is called the geometric distribution, i.e.

$$\begin{aligned}
p(y) &= \binom{y+r-1}{r-1} p^r (1-p)^y,\ y = 0, 1, 2, \cdots \\
&= \binom{y}{0} p^1 (1-p)^y,\ y = 0, 1, 2, \cdots \\
&= p(1-p)^y,\ y = 0, 1, 2, \cdots
\end{aligned}$$

which is called *geometric distribution*.

③ trinomial distribution

(i) pmf

The jpdf of the random vector (X, Y) is

$$f(x, y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y p_3^{n-x-y}$$

which is called *trinomial* pmf and denoted by $(X, Y) \sim T(n, p_1, p_2)$.

(ii) mgf

$$\begin{aligned}
M_{X,Y}(t_1,t_2) &= E[e^{t_1X+t_2Y}] \\
&= \sum_{x=0}^{n}\sum_{y=0}^{n-x} e^{t_1x+t_2y}\frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y p_3^{n-x-y} \\
&= \sum_{x=0}^{n} e^{t_1x}\frac{n!}{x!(n-x)!}p_1^x\left\{\sum_{y=0}^{n-x}\frac{(n-x)!e^{t_2y}}{y!(n-x-y)!}p_2^y p_3^{n-x-y}\right\} \\
&= \sum_{x=0}^{n}\binom{n}{x}(p_1e^{t_1})^x\sum_{y=0}^{n-x}\binom{n-x}{y}(p_2e^{t_2})^y p_3^{n-x-y} \\
&= \sum_{x=0}^{n}\binom{n}{x}(p_1e^{t_1})^x(p_2e^{t_2}+p_3)^{n-x} \\
&= (p_1e^{t_1}+p_2e^{t_2}+p_3)^n
\end{aligned}$$

(iii) marginal pmf

The marginal pmf of X is

$$\begin{aligned}
f_X(x) &= \sum_{y=0}^{n-x} f_{X,Y}(x,y) \\
&= \sum_{y=0}^{n-x}\frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y(1-p_1-p_2)^{n-x-y} \\
&= \frac{n!}{x!(n-x)!}p_1^x\sum_{y=0}^{n-x}\frac{(n-x)!}{y!(n-x-y)!}p_2^y(1-p_1-p_2)^{n-x-y} \\
&= \frac{n!}{x!(n-x)!}p_1^x(1-p_1)^{n-x} : \text{pmf of } B(n,p_1)
\end{aligned}$$

i.e. $X \sim B(n,p_1)$.
Similarly, the marginal pmf of Y is $B(n,p_2)$.

(iv) conditional pmf

The conditional pmf of Y given X is

$$
\begin{aligned}
f_{Y|X}(y|x) &= \frac{f_{X,Y}(x,y)}{f_X(x)} \\
&= \frac{\frac{n!}{x!y!(n-x-y)!}p_1^x p_2^y (1-p_1-p_2)^{n-x-y}}{\frac{n!}{x!(n-x)!}p_1^x (1-p_1)^y} \\
&= \frac{(n-x)!}{y!(n-x-y)!}\frac{p_2^y (1-p_1-p_2)^{n-x-y}}{(1-p_1)^{n-x-y+y}} \\
&= \binom{n-x}{y}\left(\frac{p_2}{1-p_1}\right)^y \left(\frac{1-p_1-p_2}{1-p_1}\right)^{n-x-y} \\
&= \binom{n-x}{y}\left(\frac{p_2}{1-p_1}\right)^y \left(1-\frac{p_2}{1-p_1}\right)^{n-x-y} \sim B\left(n-x, \frac{p_2}{1-p_1}\right)
\end{aligned}
$$

i.e. conditional pmf of Y given $X = x$ is $B\left(n-x, \frac{p_2}{1-p_1}\right)$.
Can easily show $X|Y \sim B\left(n-y, \frac{p_1}{1-p_2}\right)$.

④multinomial distribution

(i) pmf

The pmf of random vector $\boldsymbol{X} = (X_1, \cdots, X_{k-1})$ is

$$f(x_1, \cdots, x_{k-1}) = \frac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}p_2^{x_2}\cdots p_k^{x_k},$$

where

$x_k = n - x_1 - \cdots - x_{k-1}$, $p_k = 1 - p_1 - \cdots - k_{k-1}$, $0 \leq x_1 + \cdots + x_{k-1} \leq n$

and denoted by $\boldsymbol{X} \sim \mathcal{M}(n, p_1, \cdots, p_{k-1})$.

(ii) mgf

$$M(t_1, \cdots, t_{k-1}) = (p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k)^n$$

(iii) Each one-variable marginal pmf is binomial, each two-variables marginal pmf is trinomial, and so on.

3.2 The Poisson Distribution

① pmf

motivation: Consier a Taylor expansion of $g(m) = e^m$ about $m = 0$, i.e.

$$\begin{aligned} g(m) &= g(0) + \frac{g'(0)}{1!}(m-0)' + \frac{g''(0)}{2!}(m-0)^2 + \cdots \\ &= 1 + m + \frac{m^2}{2!} + \cdots \\ &= \sum_{x=0}^{\infty} \frac{m^x}{x!} = e^m \end{aligned}$$

A r.v. is said to have *Poisson distriution* with parameter m if its pmf is given by

$$p(x) = \frac{e^{-m} m^x}{x!},\ x = 0, 1, 2, \cdots$$

and it is denoted by $X \sim \mathcal{P}(m)$.

② mgf

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_{x=0}^{\infty} e^{tx} \frac{e^{-m} m^x}{x!} \\ &= e^{-m} \sum_{x=0}^{\infty} \frac{(me^t)^x}{x!} \\ &= e^{-m} e^{me^t} = exp[m(e^t - 1)] \end{aligned}$$

$$\begin{aligned} M'(t) &= me^t exp[m(e^t - 1)] \\ M''(t) &= me^t e^{m(e^t-1)} + me^t me^t e^{m(e^t-1)} \end{aligned}$$

$$\begin{aligned} \mu &= M'(0) = me^0 e^{m(e^0-1)} = m \\ \sigma^2 &= M''(0) - M'(0)^2 = (m + m^2) - m^2 = m \end{aligned}$$

③ property

$X_i \sim \mathcal{P}(m_i)$, X_i's are indep. $\Rightarrow Y = \sum_{i=1}^{n} X_i \sim \mathcal{P}\left(\sum_{i=1}^{n} m_i\right)$.

(pf)

$$
\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[e^{t\sum X_i}] \\
&= E[e^{tX_1}e^{tX_2}\cdots e^{tX_n}] \\
&= \prod_{i=1}^{n} E[e^{tX_i}] \\
&= \prod_{i=1}^{n} M_{X_i}(t) \\
&= \prod_{i=1}^{n} exp[m_i(e^t-1)] \\
&= exp\left[\sum m_i(e^t-1)\right]
\end{aligned}
$$

3.3 The Γ, χ^2, and β Distributions

① gamma distribution

(i) *gamma function*

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy,\ \alpha > 0$$

(ii) properties of gamma function

a. For $\alpha > 1$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$

(pf)

$$\begin{aligned}
\Gamma(\alpha) &= \int_0^\infty y^{\alpha-1}e^{-y}dy \\
&= [y^{\alpha-1}(-e^{-y})]_0^\infty - \int_0^\infty (\alpha-1)y^{\alpha-2}(-e^{-y})dy \\
&= 0 + (\alpha-1)\int_0^\infty y^{\alpha-2}e^{-y}dy \\
&= (\alpha-1)\Gamma(\alpha-1)
\end{aligned}$$

b. If α is positive integer, then $\Gamma(\alpha) = (\alpha-1)!$

(pf)

$$\begin{aligned}
\Gamma(\alpha) &= (\alpha-1)\Gamma(\alpha-1) \\
&\vdots \\
&= (\alpha-1)(\alpha-2)\cdots 1\Gamma(1)
\end{aligned}$$

Now,

$$\Gamma(1) = \int_0^\infty y^{1-1}e^{-y}dy = 1$$

$$\therefore \Gamma(1) = (\alpha-1)!$$

c. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

(pf)

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty y^{\frac{1}{2}} e^{-y} dy$$

Let $y = \frac{x^2}{2}$, $x > 0$, then

$$\begin{aligned}
\Gamma\left(\frac{1}{2}\right) &= \int_0^\infty \left(\frac{x^2}{2}\right)^{-\frac{1}{2}} e^{-\frac{x^2}{2}} x dx \\
&= \sqrt{2} \int_0^\infty e^{-\frac{x^2}{2}} dx \\
&= \sqrt{2} \frac{\sqrt{2\pi}}{2} \\
&= \sqrt{\pi}
\end{aligned}$$

For example, $\Gamma\left(\frac{7}{2}\right) = \dfrac{5}{2}\Gamma\left(\dfrac{5}{2}\right) = \dfrac{5}{2}\dfrac{3}{2}\dfrac{1}{2}\sqrt{\pi} = \dfrac{15}{8}\sqrt{\pi}$

(iii) pdf

The continuous r.v. X is called to have the *gamma distribution* with parameters $\alpha > 0$ and $\beta > 0$ if its pdf is given by

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} I(x > 0)$$

and denoted by $X \sim \Gamma(\alpha, \beta)$.

(idea) By letting $y = \dfrac{x}{\beta}$ in $\Gamma(\alpha)$, we have

$$\Gamma(\alpha) = \int_0^\infty \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \frac{1}{\beta} dx$$

$$\therefore 1 = \int_0^\infty \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} dx$$

(iv) mgf

$$
\begin{aligned}
M(t) &= E[e^{tX}] \\
&= \int_0^\infty e^{tx} \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} dx \\
&= \int_0^\infty \frac{x^{\alpha-1}e^{-x\left(-t+\frac{1}{\beta}\right)}}{\Gamma(\alpha)\beta^\alpha} dx \\
&= \int_0^\infty \frac{\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha}{\Gamma(\alpha)\beta^\alpha} \frac{x^{\alpha-1}e^{-x/\left(\frac{\beta}{1-\beta t}\right)}}{\Gamma(\alpha)\left(\frac{\beta}{1-\beta t}\right)^\alpha} dx \\
&= \frac{\frac{\beta^\alpha}{(1-\beta t)^\alpha}}{\beta^\alpha} \\
&= (1-\beta t)^{-\alpha}
\end{aligned}
$$

$$
\begin{aligned}
M'(t) &= (-\alpha)(1-\beta t)^{-\alpha-1}(-\beta) = \alpha\beta(1-\beta t)^{-\alpha-1} \\
M''(t) &= \alpha\beta(-\alpha-1)(1-\beta t)^{-\alpha-2}(-\beta) = \alpha\beta^2(1-\beta t)^{-\alpha-2}(\alpha+1)
\end{aligned}
$$

$$
\begin{aligned}
\mu &= M'(0) = \alpha\beta \\
\sigma^2 &= M''(0) - M'(0)^2 = \alpha^2\beta^2 + \alpha\beta^2 - (\alpha\beta)^2 = \alpha\beta^2
\end{aligned}
$$

(v) sum of indep. gamma

$$X_i \sim \Gamma(\alpha_i, \beta),\ X_i\text{'s are indep.} \Rightarrow Y = \sum_{i=1}^{n} X_i \sim \Gamma\left(\sum \alpha_i, \beta\right)$$

(pf)

$$\begin{aligned} M_Y(t) &= E[e^{t\sum X_i}] \\ &= E[e^{tX_1}\cdots e^{tX_n}] \\ &= \prod_{i=1}^{n} E[e^{tX_i}] \\ &= \prod (1-\beta t)^{-\alpha_i} \\ &= (1-\beta t)^{-\sum \alpha_i} \end{aligned}$$

(vi) relationship with Poisson distribution

W: time needed to obtain k changes(or deaths)

$$G(w) = P(W \le w) = 1 - P(W > w)$$

Now, $\{W > w\}$ is equivalent to "less than k changes in an interval of length w". i.e.

$$P(W > w) = \sum_{x=0}^{k-1} P(X = x) = \sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!},$$

where, X: number of changes in an interval of length w. Now, it can be shown

$$\sum_{x=0}^{k-1} \frac{(\lambda w)^x e^{-\lambda w}}{x!} = \int_{\lambda w}^{\infty} \frac{z^{k-1} e^{-z}}{(k-1)!} dz$$

$$\therefore G(w) = \int_0^{\lambda w} \frac{z^{k-1} e^{-z}}{\Gamma(k)} dz.$$

Let $z = \lambda y$, then

$$G(w) = \int_0^{w} \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)} dy \Rightarrow g(w) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{\Gamma(k)} \sim \Gamma\left(k, \frac{1}{\lambda}\right)$$

②χ^2-distribution

(i) definition

If $X \sim \Gamma\left(\frac{r}{2}, 2\right)$, then X is called to have *chi-square distribution* with d.f. r, and denoted by $X \sim \chi^2(r)$.

(ii) pdf

$$f(x) = \frac{x^{\frac{r}{2}-1}e^{-\frac{x}{2}}}{\Gamma\left(\frac{r}{2}\right)2^{\frac{r}{2}}}, \, x > 0$$

(iii) mgf

$$M(t) = (1-2t)^{-r/2}$$

(iv) μ and σ^2.

$$\mu = r, \, \sigma^2 = 2r$$

(v) property

$X_i \sim \chi^2(r_i), i = 1, \cdots, n$. X_i's: indep. $\Rightarrow Y = \sum X_i \sim \chi^2(\sum r_i)$.

③ beta distribution

(i) pdf

A r.v. X is said to have *beta distribution* with parameters α and β if its pdf is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1},\ 0 < x < 1$$

and denoted by $X \sim Beta(\alpha, \beta)$.

(ii) mgf

$$M_X(t) = E[e^{tX}] = \int_0^1 e^{tx} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx$$

In fact, $M_X(t)$ does not have a closed(analytic) form.
Hence, to compute mean and variance, use the definition of expectation, i.e.

$$\begin{aligned}
E(X) &= \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} dx \\
&= \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)} \frac{\Gamma(\alpha+1+\beta)}{\Gamma(\alpha+1)\Gamma(\beta)} x^{\alpha+1-1}(1-x)^{\beta-1} dx \\
&= \frac{\Gamma(\alpha+1)\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

Similarly, can easily show

$$Var(X) = E(X^2) - E^2(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

(iii) derivation from gamma distribution

$X_1 \sim \Gamma(\alpha, 1), X_2 \sim \Gamma(\beta, 1), X_1$ and X_2 are indep. $\Rightarrow \dfrac{X_1}{X_1 + X_2} \sim Beta(\alpha, \beta)$.
(pf) Let

$$Y_1 = X_1 + X_2,\ Y_2 = \frac{X_1}{X_1 + X_2}$$

then it is 1-1 transformation, and the inverse function is

$$x_1 = y_1 y_2,\ x_2 = y_1 - y_1 y_2 = y_1(1 - y_2)$$

Also,

$$S = \{(x_1, x_2) : 0 < x_1 < \infty, 0 < x_2 < \infty\}$$

and

$$\mathscr{T} = \{(y_1, y_2) : 0 < y_1 < \infty, 0 < y_2 < 1\}$$

$$J = \begin{vmatrix} y_2 & y_1 \\ 1 - y_2 & -y_1 \end{vmatrix} = -y_1 y_2 - y_1(1 - y_2) = -y_1$$

Therefore, the jpdf pf (Y_1, Y_2) is

$$g(y_1, y_2) = f(y_1 y_2, y_1(1 - y_2))|J|$$

where

$$\begin{aligned} f(x_1, x_2) &= \frac{x_1^{\alpha-1} e^{-x_1/1}}{\Gamma(\alpha) 1^\alpha} \frac{x_2^{\beta-1} e^{-x_2/1}}{\Gamma(\beta) 1^\beta} \\ &= \frac{x_1^{\alpha-1} x_2^{\beta-1} e^{-x_1 - x_2}}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

$$\begin{aligned} \therefore g(y_1, y_2) &= \frac{(y_1 y_2)^{\alpha-1} \{y_1(1 - y_2)\}^{\beta-1} e^{-y_1 y_2 - y_1(1 - y_2)}}{\Gamma(\alpha)\Gamma(\beta)} |-y_1| \\ &= \frac{y^{\alpha+\beta-2+1} y_2^{\alpha-1} (1 - y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} e^{-y_1} \end{aligned}$$

Finally, the pdf of $Y_2 = \dfrac{X_1}{X_1 + X_2}$ is

$$
\begin{aligned}
g_2(y_2) &= \int g(y_1, y_2) dy_1 \\
&= \int_0^\infty \frac{y_2^{\alpha-1}(1-y_2)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \Gamma(\alpha+\beta) \frac{1}{\Gamma(\alpha+\beta)} y_1^{\alpha+\beta-1} e^{-y_1} dy_1 \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y_2^{\alpha-1}(1-y_2)^{\beta-1} I(0 < y_2 < 1) \\
&\sim Beta(\alpha, \beta)
\end{aligned}
$$

④ *Dirichlet distribution*

$X_i \sim \Gamma(\alpha_i, 1)$, X_i's are indep. $i = 1, 2, \cdots, k+1$.
$Y_i = \dfrac{X_i}{(X_1 + \cdots + X_{k+1})}$, $i = 1, \cdots, k$ and $Y_{k+1} = X_1 + \cdots + X_{k+1}$.
Then, the jpf of $Y_1, \cdots, Y_k$ is called the Dirichlet distribution with pdf

$$
g(y_1, \cdots, y_k) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_{k+1})}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{k+1})} y_1^{\alpha_1 - 1} y_2^{\alpha_2 - 1} \cdots y_k^{\alpha_k - 1} (1 - y_1 - \cdots - y_k)^{\alpha_{k+1} - 1}
$$

(pf) cf. If $k = 1$, the Dirichlet distribution reduces to the Beta distriution. i.e., the Dirichlet distriution is a multivariate extension of Beta distribution.

$$
y_1 = \frac{x_1}{\sum_{i=1}^{k+1} x_i},\ y_2 = \frac{x_2}{\sum_{i=1}^{k+1} x_i}, \cdots, y_k = \frac{x_k}{\sum_{i=1}^{k+1} x_i},\ y_{k+1} = \sum_{i=1}^{k+1} x_i
$$

which is 1-1 from $(x_1, \cdots, x_{k+1})$ to $(y_1, \cdots, y_{k+1})$.
Inverse functions are

$$
x_1 = y_1 y_{k+1},\ x_2 = y_2 y_{k+1}, \cdots, x_k = y_k y_{k+1},\ x_{k+1} = y_{k+1}(1 - y_1 - y_2 - \cdots - y_k)
$$

$$
S = \{(x_1, \cdots, x_{k+1}) : 0 < x_i < \infty,\ i = 1, \cdots, k+1\}
$$

$$
\mathscr{T} = \{(y_1, \cdots, y_{k+1}) : 0 < y_i < 1,\ i = 1, \cdots, k;\ 0 < y_{k+1} < \infty\}
$$

$$J = \begin{vmatrix} y_{k+1} & 0 & 0 & \cdots & 0 & y_1 \\ 0 & y_{k+1} & 0 & \cdots & 0 & y_2 \\ 0 & 0 & y_{k+1} & \cdots & 0 & y_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & y_{k+1} & y_k \\ -y_{k+1} & -y_{k+1} & -y_{k+1} & \cdots & -y_{k+1} & 1-\sum_{i=1}^{k} y_i \end{vmatrix} = y_{k+1}^k$$

Also, the jpdf of $(X_1, \cdots, X_{k+1})$ is

$$\begin{aligned} f(x_1, \cdots, x_{k+1}) &= \frac{x_1^{\alpha_1-1}e^{-x_1/1}}{\Gamma(\alpha_1)1^{\alpha_1}} \cdots \frac{x_{k+1}^{\alpha_{k+1}-1}e^{-x_{k+1}/1}}{\Gamma(\alpha_{k+1})1^{\alpha_{k+1}}} \\ &= \frac{x_1^{\alpha_1-1} \cdots x_{k+1}^{\alpha_{k+1}-1}e^{-(x_1+\cdots+x_{k+1})}}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})} \end{aligned}$$

Hence, the jpdf of $(Y_1, \cdots, Y_{k+1})$ is

$$\begin{aligned} g(y_1, \cdots, y_{k+1}) &= f(y_1 y_{k+1}, \cdots, y_k y_{k+1}, y_{k+1}(1-y_1-\cdots-y_k))|y_{k+1}^k| \\ &= \frac{(y_1 y_{k+1})^{\alpha_1-1}\cdots(y_k y_{k+1})^{\alpha_k-1}\{y_{k+1}(1-y_1-\cdots-y_k)\}^{\alpha_{k+1}-1}}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})} e^{-y_{k+1}} y_{k+1}^k \end{aligned}$$

Finally, the jpdf of $(Y_1, \cdots, Y_k)$ is

$$\begin{aligned} g(y_1, \cdots, y_k) &= \int_0^\infty g(y_1, \cdots, y_{k+1})dy_{k+1} \\ &= \frac{\Gamma(\alpha_1+\cdots+\alpha_{k+1})}{\Gamma(\alpha_1)\cdots\Gamma(\alpha_{k+1})} y_1^{\alpha_1-1}\cdots y_k^{\alpha_k-1}(1-y_1-\cdots y_k)^{\alpha_{k+1}-1} \end{aligned}$$

3.4 The Normal Distribution

① derivation

First, want to compute

$$I = \int_{-\infty}^{\infty} e^{-y^2/2} dy$$

consider

$$\left(\sum_{i=1}^{n} a_i\right)^2 = \left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{n} a_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j$$

Similarly, consider

$$I^2 = \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right)^2 = \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy\right)\left(\int_{-\infty}^{\infty} e^{-z^2/2} dz\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+z^2)/2} dydz$$

use polar coordinate system, i.e. $y = rcos\theta, z = rsin\theta$

$\{(y,z) : -\infty < y < \infty, -\infty < z < \infty\} \to \{(r,\theta) : 0 < r < \infty, 0 < \theta < 2\pi\}$

1-1 correspondence between (y,z) and (r,θ)

$$J = \begin{vmatrix} \frac{dy}{dr} & \frac{dy}{d\theta} \\ \frac{dz}{dr} & \frac{dz}{d\theta} \end{vmatrix} = \begin{vmatrix} cos\theta & -rsin\theta \\ sin\theta & rcos\theta \end{vmatrix} = rcos^2\theta - (-r)sin^2\theta = r$$

Therefore,

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(y^2+z^2)/2} dydz \\ &= \int\int e^{-r^2/2} rdrd\theta \\ &= \int_0^{2\pi} \left[-e^{-r^2/2}\right]_0^{\infty} d\theta \\ &= \int_0^{2\pi} 1 d\theta \\ &= 2\pi \end{aligned}$$

$$\therefore I = \sqrt{2\pi}$$

i.e.

$$1 = \int \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

Also, let

$$y = \frac{x-\mu}{\sigma},$$

then

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2/2} \frac{1}{\sigma} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \end{aligned}$$

② pdf

The continuous r.v. X is said to have *normal distribution* with mean μ and variance σ^2 if its pdf is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \infty < x < \infty$$

and denoted by $X \sim N(\mu, \sigma^2)$.

As a special case, if $\mu = 0$ and $\sigma^2 = 1$, then it is called standard normal(Gaussian) distriution with mean 0, variance 1

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

③ mgf

$$\begin{aligned}
M(t) &= E[e^{tX}] \\
&= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{1}{2\sigma^2}\{-2\sigma^2 tx + x^2 - 2\mu x + \mu^2\}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{1}{2\sigma^2}\{x^2 - 2(\sigma^2 t + \mu)x + (\sigma^2 t + \mu)^2 - (\sigma^2 t + \mu)^2 + \mu^2\}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[\frac{1}{2\sigma^2}\{(x - (\sigma^2 t + \mu))^2 - \sigma^4 t^2 - 2\sigma^2 t\mu\}\right] dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{\{x - (\sigma^2 t + \mu)\}^2}{2\sigma^2} + \left(t\mu + \frac{\sigma^2 t^2}{2}\right)\right] dx \\
&= exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} exp\left[-\frac{\{x - (\sigma^2 t + \mu)\}^2}{2\sigma^2}\right] dx \\
&= exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)
\end{aligned}$$

$$\begin{aligned}
M'(t) &= (\mu + \sigma^2 t) exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \\
M''(t) &= \sigma^2 exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right] + (\mu + \sigma^2 t^2) exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)
\end{aligned}$$

$$\begin{aligned}
\mu &= M'(0) = \mu \\
\sigma^2 &= M''(0) - M'(0)^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2
\end{aligned}$$

④ higher order moments

Let $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X-\mu}{\sigma^2} \sim N(0,1)$ and $M_Z(t) = e^{t^2/2}$.
Using the mgf of Z, we can get $E(X^k)$, $k = 1, 2, \cdots$
Recall that

$$M_Z(t) = 1 + E(Z)t + \frac{E(Z^2)}{2!}t^2 + \frac{E(Z^3)}{3!}t^3 + \cdots$$

also,

$$\begin{aligned} e^{t^2/2} &= 1 + \frac{t^2}{2} + \frac{1}{2!}\left(\frac{t^2}{2}\right)^2 + \frac{1}{3}\left(\frac{t^2}{2}\right)^3 + \cdots \\ &= 1 + \frac{t^2}{2!} + \frac{3 \times 1}{4!}t^4 + \cdots + \frac{(2k-1)\cdots(3)(1)}{(2k)!}t^{2k} + \cdots \end{aligned}$$

Therefore,

$$\begin{cases} E(Z^{2k}) = (2k-1)\cdots(3)(1) = \dfrac{(2k)!}{2^k k!} \\ E(Z^{2k+1}) = 0 \end{cases}$$

Now,

$$\begin{aligned} E(X^k) &= E[(\mu + \sigma z)^k] \\ &= E\left[\sum_{j=0}^{k} \binom{k}{j} (\sigma z)^j \mu^{k-j}\right] : \text{binomial eq.} \\ &= \sum_{j=0}^{k} \binom{k}{j} \sigma^j E(Z^j) \mu^{k-j} \end{aligned}$$

⑤ properties

(i) $Z \sim N(0,1) \Rightarrow Z^2 \sim \chi^2(1)$

(pf) Let $V = Z^2$, then the cdf of V is

$$\begin{aligned} F(v) &= P(V \le v) \\ &= P(Z^2 \le v) \\ &= P(-\sqrt{v} \le Z \le \sqrt{v}),\ v > 0 \\ &= 2P(0 \le Z \le \sqrt{v}) \\ &= 2\int_0^{\sqrt{v}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \end{aligned}$$

Let $y = z^2$, then $dy = 2zdz$, i.e. $dz = \dfrac{1}{2\sqrt{y}}dy$

$$\therefore F(v) = 2\int_o^{v} \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} dy$$

Therefore, the pdf of V is

$$\begin{aligned} f(v) &= F'(v) \\ &= \frac{1}{\sqrt{2\pi}} e^{-v/2} \frac{1}{\sqrt{v}} \\ &= \frac{v^{-1/2} e^{-v/2}}{\sqrt{\pi}\sqrt{2}} \\ &= \frac{v^{\frac{1}{2}-1} e^{-v/2}}{\Gamma\left(\frac{1}{2}\right) 2^{1/2}} : \text{pdf of } \chi^2(1) \end{aligned}$$

(ii) $X_i \sim N(\mu_i, \sigma_i^2)$, X_i's are indep. $\Rightarrow Y = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum a_i\mu_i, \sum a_i^2\sigma_i^2\right)$

(pf)

$$\begin{aligned}
M_Y(t) &= E[e^{tY}] \\
&= E[exp(t\sum a_i X_i)] \\
&= E[e^{ta_1X_1}e^{ta_2X_2}\cdots e^{ta_nX_n}] \\
&= \prod_{i=1}^{n} E[e^{ta_iX_i}] \\
&= \prod_{i=1}^{n} M_{X_i}(ta_i) \\
&= \prod_{i=1}^{n} exp\left(\mu t a_i + \frac{1}{2}\sigma_i^2 t^2 a_i^2\right) \\
&= exp\left[t\sum a_i\mu_i + \frac{1}{2}t^2\sum a_i^2\sigma_i^2\right] : \text{mgf of } N\left(\sum a_i\mu_i, \sum a_i^2\sigma_i^2\right)
\end{aligned}$$

⑥ contaminated normals

$Z \sim N(0,1)$, $I_\varepsilon \sim B(1, 1-\varepsilon)$, Z and I_ε are indep.
Want to find the pdf of

$$W = I_\varepsilon Z + (1 - I_\varepsilon)\sigma_c Z$$

$$\begin{aligned}
F_W(w) &= P(W \le w) \\
&= P(W \le w, I_\varepsilon = 1) + P(W \le w, I_\varepsilon = 0) \\
&= P(W \le w | I_\varepsilon = 1)P(I_\varepsilon = 1) + P(W \le w | I_\varepsilon = 0)P(I_\varepsilon = 0) \\
&= (1-\varepsilon)P(Z \le w) + \varepsilon P(Z \le w/\sigma_c) \\
&= (1-\varepsilon)\Phi(w) + \varepsilon\Phi\left(\frac{w}{\sigma_c}\right)
\end{aligned}$$

$$\therefore f_W(w) = (1-\varepsilon)\phi(w) + \frac{\varepsilon}{\sigma_c}\phi\left(\frac{w}{\sigma_c}\right)$$

3.5 The Multivariate Normal Distribution

① derivation

(i) standard normal case

$\boldsymbol{Z} = (Z_1, \cdots, Z_n)'$, $Z_i \sim N(0,1)$, Z_i's are indep.
Then, the pdf of $\boldsymbol{Z}$ is

$$\begin{aligned} f(\boldsymbol{z}) &= \prod_{i=1}^{n} f(Z_i) \\ &= \prod_{i=1}^{n} (2\pi)^{-1/2} exp\left[-\frac{1}{2} z_i^2\right] \\ &= (2\pi)^{-n/2} exp\left(-\frac{1}{2}\sum_{i=1}^{n} z_i^2\right) \\ &= (2\pi)^{-n/2} exp\left(-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right) \end{aligned}$$

which is called the standard multivariate normal distribution with $E(\boldsymbol{Z}) = \boldsymbol{0}$, $Cov(\boldsymbol{Z}) = I_n$ and denoted by $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, I_n)$.

Now, the mgf of $\boldsymbol{Z}$ is

$$\begin{aligned}
M_{\boldsymbol{Z}}(\boldsymbol{t}) &= E[e^{\boldsymbol{t}'\boldsymbol{Z}}] \\
&= E[e^{t_1 Z_1} \cdots e^{t_n Z_n}] \\
&= \prod_{i=1}^{n} E[e^{t_i Z_i}] \\
&= \prod_{i=1}^{n} M_{Z_i}(t_i) \\
&= \prod_{i=1}^{n} exp\left[0t_i + \frac{1}{2}1^2 t_i^2\right] \\
&= \prod_{i=1}^{n} exp\left(\frac{1}{2}t_i^2\right) \\
&= exp\left(\frac{1}{2}\boldsymbol{t}'\boldsymbol{t}\right)
\end{aligned}$$

(ii) spectral decomposition

Theorem 3.5.1. spectral decomposition theorem
Let A be $n \times n$ symmetric matrix, then $\exists$ an orthogonal matrix Γ s.t. $A = \Gamma'\Lambda\Gamma$, then $\Lambda = diag(\lambda_1, \cdots, \lambda_n)$ and λ_i's are eigenvalues of A and corresponding eigenvectors are column vectors of Γ.

(iii) general normal case

Let Σ be $n \times n$ symmetric and positive definite matrix. By spectral decomposition,

$$\Sigma = \Gamma'\Lambda\Gamma$$

where $\lambda = diag(\lambda_1, \cdots, \lambda_n)$ s.t. $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$.

Then, $\exists\, \Lambda^{1/2} = diag(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n})$ and

$$\begin{aligned}
\Sigma &= \Gamma'\Lambda\Gamma \\
&= \Gamma'\Lambda^{1/2}\Lambda^{1/2}\Gamma \\
&= \Gamma'\Lambda^{1/2}\Gamma\Gamma'\Lambda^{1/2}\Gamma \\
&= \Sigma^{1/2}\Sigma^{1/2}
\end{aligned}$$

where $\Sigma^{1/2} = \Gamma'\Lambda^{1/2}\Gamma$

$$\begin{aligned}
\therefore (\Sigma^{1/2})^{-1} &= (\Gamma'\Lambda^{1/2}\Gamma)^{-1} \\
&= (\Gamma)^{-1}(\Lambda^{1/2})^{-1}(\Gamma')^{-1} \\
&= \Gamma'\Lambda^{-1/2}\Gamma
\end{aligned}$$

Let $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, I_n)$ and let

$$\boldsymbol{X} = \Sigma^{1/2}\boldsymbol{Z} + \boldsymbol{\mu}$$

i.e. $\Sigma^{1/2}\boldsymbol{Z} = \boldsymbol{X} - \boldsymbol{\mu} \Rightarrow \boldsymbol{Z} = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})$

$$J = \left|\frac{dz}{dx}\right| = |\Sigma^{-1/2}| = |\Sigma|^{-1/2}$$

$$g(\boldsymbol{z}) = (2\pi)^{-n/2} exp\left[-\frac{1}{2}\boldsymbol{z}'\boldsymbol{z}\right] : \text{pdf of } N_n(\boldsymbol{0}, I_n)$$

Hence, the pdf of $\boldsymbol{X}$ is

$$\begin{aligned}
f(\boldsymbol{x}) &= g(\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}))|J| \\
&= (2\pi)^{-n/2} exp\left[-\frac{1}{2}\{\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})\}'\{\Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu})\}\right]|\Sigma|^{-1/2} \\
&= (2\pi)^{-n/2}|\Sigma|^{-1/2} exp\left[-\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu})\right]
\end{aligned}$$

which is called the multivariate normal pdf with mean μ and variance-covariance matrix Σ, and denoted by $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$.

Also, the mgf of $\boldsymbol{X}$ is

$$\begin{aligned}
M_{\boldsymbol{X}}(\boldsymbol{t}) &= E[e^{\boldsymbol{t'X}}] \\
&= E[exp\{\boldsymbol{t}'(\Sigma^{1/2}\boldsymbol{z}+\boldsymbol{\mu})\}] \\
&= E[exp(\boldsymbol{t'}\Sigma^{1/2}\boldsymbol{z}+\boldsymbol{t'\mu})] \\
&= exp(\boldsymbol{t'\mu})E[exp\{(\Sigma^{1/2}\boldsymbol{t})'\boldsymbol{z}\}] \\
&= exp(\boldsymbol{t'\mu})M_{\boldsymbol{Z}}(\Sigma^{1/2}t) \\
&= exp(\boldsymbol{t'\mu})exp\left[\frac{1}{2}(\Sigma^{1/2}\boldsymbol{t})'(\Sigma^{1/2}\boldsymbol{t})\right] \\
&= exp\left[\boldsymbol{t'\mu}+\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right]
\end{aligned}$$

② bivariate normal distribution

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\begin{aligned}
f(x,y) &= (2\pi)^{-1}|\Sigma|^{-1/2}exp\left[-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X}-\boldsymbol{\mu})\right] \\
&= \frac{1}{2\pi\sigma_2\sigma_2\sqrt{1-\rho^2}}exp[-\frac{1}{2(1-\rho^2)}\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2-2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) \\
&\quad +\left(\frac{y-\mu_2}{\sigma_2}\right)^2\}]
\end{aligned}$$

③ linear transformation

$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma) \Rightarrow \boldsymbol{Y} = A\boldsymbol{X} + \boldsymbol{b} \sim N_m(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A')$

(pf) First, recall that

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = exp\left[\boldsymbol{t'\mu} + \frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right]$$

Next, compute mgf of $\boldsymbol{Y}$

$$\begin{aligned} M_{\boldsymbol{Y}}(\boldsymbol{t}) &= E[e^{\boldsymbol{t'Y}}] \\ &= E[exp\{\boldsymbol{t}'(A\boldsymbol{X} + \boldsymbol{b})\}] \\ &= E[exp\{(A'\boldsymbol{t})'\boldsymbol{X}\}e^{\boldsymbol{t'b}}] \\ &= e^{\boldsymbol{t'b}}exp\left[(A'\boldsymbol{t})'\boldsymbol{\mu} + \frac{1}{2}(A'\boldsymbol{t})'\Sigma(A'\boldsymbol{t})\right] \\ &= exp\left(\boldsymbol{t'b} + \boldsymbol{t}'A\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}'A\Sigma A'\boldsymbol{t}\right) \\ &= exp\left(\boldsymbol{t'}(A\boldsymbol{\mu} + \boldsymbol{b}) + \frac{1}{2}\boldsymbol{t}'A\Sigma A'\boldsymbol{t}\right) \end{aligned}$$

which is mgf of $N_m(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A')$

④ marginal and conditional distribution

Assume that

$$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$$

Decompose $\boldsymbol{X}$ s.t.

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

(i) marginal distiribution

Let $A = (I_m : O)$, then

$$AX = (I_m : O)\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = X_1$$

i.e. X_1 is linear transformation of X

$$\begin{aligned} E(X_1) &= E(AX) \\ &= A\mu \\ &= (I_m : O)\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \\ &= \mu_1 \end{aligned}$$

$$\begin{aligned} Cov(X_1) &= Cov(AX) \\ &= ACov(X)A' \\ &= (I_m : O)\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I_m \\ O \end{pmatrix} \\ &= \Sigma_{11} \end{aligned}$$

$$\therefore X_1 \sim N_n(\mu_1, \Sigma_{11})$$

Similarly,

$$X_2 \sim N_n(\mu_2, \Sigma_{22})$$

(ii) indpendence of X_1 and X_2

$X \sim N_m(\mu, \Sigma)$, X_1 and X_2 are indep. iff $\Sigma_{12} = O$.

(pf) The joint mgf of $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ is

$$\begin{aligned}
M_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{t}_1,\boldsymbol{t}_2) &= E[exp(\boldsymbol{t_1'X_1}+\boldsymbol{t_2'X_2})] \\
&= E[e^{\boldsymbol{t'X}}] \\
&= exp\left(\boldsymbol{t'\mu}+\frac{1}{2}\boldsymbol{t}'\Sigma\boldsymbol{t}\right) \\
&= exp\left[(\boldsymbol{t}_1'\boldsymbol{t}_2')\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}+\frac{1}{2}(\boldsymbol{t}_1'\boldsymbol{t}_2')\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}\right] \\
&= exp\left[\boldsymbol{t_1'\mu_1}+\boldsymbol{t_2'\mu_2}+\frac{1}{2}\boldsymbol{t}_1'\Sigma_{11}\boldsymbol{t}_1+\frac{1}{2}\boldsymbol{t}_1'\Sigma_{12}\boldsymbol{t}_2+\frac{1}{2}\boldsymbol{t}_2'\Sigma_{21}\boldsymbol{t}_1+\frac{1}{2}\boldsymbol{t}_2'\Sigma_{22}\boldsymbol{t}_2\right]
\end{aligned}$$

Now,

$$M_{\boldsymbol{X}_1}(\boldsymbol{t}_1)M_{\boldsymbol{X}_2}(\boldsymbol{t}_2) = exp\left[\boldsymbol{t_1'\mu_1}+\frac{1}{2}\boldsymbol{t}_1'\Sigma_{11}\boldsymbol{t}_1\right]exp\left[\boldsymbol{t_2'\mu_2}+\frac{1}{2}\boldsymbol{t}_2'\Sigma_{22}\boldsymbol{t}_2\right]$$

Hence, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are indep. iff $M_{\boldsymbol{X}_1,\boldsymbol{X}_2}(\boldsymbol{t}_1,\boldsymbol{t}_2) = M_{\boldsymbol{X}_1}(\boldsymbol{t}_1)M_{\boldsymbol{X}_2}(\boldsymbol{t}_2)$
iff $\Sigma_{12} = O$.

(iii) conditional distribution of $\boldsymbol{X}_1$ given $\boldsymbol{X}_2$

$\boldsymbol{X}_1|\boldsymbol{X}_2 \sim N_m(\boldsymbol{\mu}_1+\Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}_2-\boldsymbol{\mu}),\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$

(pf) Let

$$\boldsymbol{W} = \boldsymbol{X}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}_2$$

and consider jpdf of $\boldsymbol{W}$ and $\boldsymbol{X}_2$

$$\begin{pmatrix}\boldsymbol{W}\\ \boldsymbol{X}_2\end{pmatrix} = \begin{pmatrix}I_m & -\Sigma_{12}\Sigma_{22}^{-1}\\ O & I_{n-m}\end{pmatrix}\begin{pmatrix}\boldsymbol{X}_1\\ \boldsymbol{X}_2\end{pmatrix} = A\boldsymbol{X}$$

i.e.

$$\begin{pmatrix}\boldsymbol{W}\\ \boldsymbol{X}_2\end{pmatrix} = A\boldsymbol{X} \sim N_n(A\mu, A\Sigma A')$$

where

$$\begin{aligned}
A\mu &= \begin{pmatrix}I & -\Sigma_{12}\Sigma_{22}^{-1}\\ O & I\end{pmatrix}\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix} \\
&= \begin{pmatrix}\boldsymbol{\mu}_1-\Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2\\ \boldsymbol{\mu}_2\end{pmatrix}
\end{aligned}$$

$$
\begin{aligned}
A\Sigma A' &= \begin{pmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ O & I \end{pmatrix} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \begin{pmatrix} I & O \\ -\Sigma_{22}^{-1}\Sigma_{21} & I \end{pmatrix} \\
&= \begin{pmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & O \\ O & \Sigma_{22} \end{pmatrix}
\end{aligned}
$$

Hence, $\boldsymbol{W}$ and $\boldsymbol{X}_2$ are indep.
Therefore,

$$\boldsymbol{W} \sim \boldsymbol{W}|\boldsymbol{X}_2 \sim N_m(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})$$

Now,

$$
\begin{aligned}
\boldsymbol{X}_1|\boldsymbol{X}_2 &\sim \boldsymbol{W} + \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}_2 \\
&\sim N_m(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\mu}_2 + \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{X}_2, \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma 21) \\
&\sim N_m(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\boldsymbol{X}_2 - \boldsymbol{\mu}), \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21})
\end{aligned}
$$

⑤ relationship with χ^2-distribution

$$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma) \;\Rightarrow\; \boldsymbol{W} = (\boldsymbol{X} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(n)$$

(pf)

$$\boldsymbol{Z} = \Sigma^{-1/2}(\boldsymbol{X} - \boldsymbol{\mu}) \sim N_n(\boldsymbol{0}, I_n)$$
$$\boldsymbol{W} = \boldsymbol{Z}'\boldsymbol{Z} = \sum z_i^2 \sim \chi^2(n)$$

3.6 t- and F-Distributions

① t-distribution

(i) definition

$W \sim N(0,1)$, $V \sim \chi^2(r)$, W and V are indep. Define

$$T = \frac{W}{\sqrt{V/r}}$$

then the r.v. T is called to have a *t-distribution* with degree of freedom r, and denoted by $T \sim t(r)$.

(ii) derivation of pdf

Consider a transformation

$$(w,v) \;\rightarrow\; (t,u)$$

where

$$t = \frac{w}{\sqrt{v/r}},\; u = v$$

then it is 1-1 transformation with inverse function

$$w = t\sqrt{u}/\sqrt{r},\; v = u$$

and

$$J = \begin{vmatrix} \frac{dw}{dt} & \frac{dw}{du} \\ \frac{dv}{dt} & \frac{dv}{du} \end{vmatrix} = \begin{vmatrix} \sqrt{u}/\sqrt{r} & t/2\sqrt{ur} \\ 0 & 1 \end{vmatrix} = \sqrt{u}/\sqrt{r}$$

Now, the joint pdf of (w,v) is

$$f(w,v) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} \frac{v^{\frac{r}{2}-1} e^{-v/2}}{\Gamma\left(\frac{r}{2}\right) 2^{r/2}} = \frac{e^{-w^2/2} v^{\frac{r}{2}-1} e^{-v/2}}{\sqrt{2\pi}\Gamma\left(\frac{r}{2}\right) 2^{r/2}}$$

Then, the joint pdf of (t, u) is

$$
\begin{aligned}
g(t,u) &= f\left(\frac{t\sqrt{u}}{\sqrt{r}}, u\right)|J| \\
&= \frac{1}{\sqrt{2\pi}\Gamma\left(\frac{r}{2}\right)2^{r/2}} exp\left(-\frac{t^2u}{2r}\right) u^{\frac{r}{2}-1} e^{-u/2} \frac{u^{1/2}}{r^{1/2}} \\
&= \frac{1}{\sqrt{2\pi}\Gamma\left(\frac{r}{2}\right)2^{r/2}} u^{\frac{r+1}{2}-1} exp\left\{-\frac{1}{2}\left(1+\frac{t^2}{r}\right)u\right\}\frac{1}{r^{1/2}}
\end{aligned}
$$

Hence, the pdf of t is

$$
\begin{aligned}
g(t) &= \int_0^\infty g(t,u)du \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\Gamma\left(\frac{r}{2}\right)2^{r/2}\sqrt{r}}\Gamma\left(\frac{r+1}{2}\right)\left(\frac{2}{1+\frac{t^2}{r}}\right)^{(r+1)/2} \frac{u^{\frac{r+1}{2}-1}e^{-u/\left(\frac{2}{1+\frac{t^2}{r}}\right)}}{\Gamma\left(\frac{r+1}{2}\right)\left(\frac{2}{1+\frac{t^2}{r}}\right)^{(r+1)/2}}du \\
&= \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\Gamma\left(\frac{r}{2}\right)}\left(1+\frac{t^2}{r}\right)^{-(r+1)/2}, -\infty < t < \infty
\end{aligned}
$$

(iii) μ and σ^2

$$
\begin{aligned}
E(T) &= E\left[\frac{W}{\sqrt{V/r}}\right] \\
&= E\left[W\frac{1}{\sqrt{V/r}}\right] \\
&= E(W)E\left(\frac{1}{\sqrt{V/r}}\right) \\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
Var(T) &= E(T^2) - E^2(T) \\
&= E(T^2) \\
&= E\left[\frac{W^2}{V/r}\right] \\
&= E(W^2)E\left(\frac{r}{V}\right)
\end{aligned}
$$

Now, $E(W^2) = Var(W) + E^2(W) = 1 + 0^2 = 1$

$$
\begin{aligned}
E\left(\frac{r}{V}\right) &= rE(V^{-1}) \\
&= r\int_0^\infty v^{-1}\frac{v^{\frac{r}{2}-1}e^{-v/2}}{\Gamma\left(\frac{r}{2}\right)2^{r/2}}dv \\
&= r\int_0^\infty v^{-1}\frac{\Gamma\left(\frac{r-2}{2}\right)2^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)2^{r/2}}\frac{v^{\frac{r-2}{2}-1}e^{-v/2}}{\Gamma\left(\frac{r-2}{2}\right)2^{\frac{r-2}{2}}}dv \\
&= r\frac{\Gamma\left(\frac{r}{2}-1\right)}{\Gamma\left(\frac{r}{2}\right)2} \\
&= r\frac{\Gamma\left(\frac{r}{2}-1\right)}{\left(\frac{r}{2}-1\right)\Gamma\left(\frac{r}{2}-1\right)} \\
&= \frac{r}{r-2}
\end{aligned}
$$

② F-distribution

(i) definition

$U \sim \chi^2(r_1)$, $V \sim \chi^2(r_2)$, U and V are indep. Then

$$W = \frac{U/r_1}{V/r_2}$$

is called to have *F-distribution* with degrees of freedom r_1 and r_2, and denoted by $W \sim F(r_1, r_2)$.

(ii) derivation of pdf

Consider a transformation

$$(u, v) \rightarrow (w, z)$$

where

$$w = \frac{u/r_1}{v/r_2}, \; z = v$$

then it is 1-1 transformation with inverse function

$$u = \frac{r_1}{r_2} wz, \; v = z$$

and

$$J = \begin{vmatrix} \frac{du}{dw} & \frac{du}{dz} \\ \frac{dv}{dw} & \frac{dv}{dz} \end{vmatrix} = \begin{vmatrix} \frac{r_1}{r_2} z & \frac{r_1}{r_2} w \\ 0 & 1 \end{vmatrix} = \frac{r_1}{r_2} z$$

Now, the joint pdf of (U, V) is

$$\begin{aligned} f(u,v) &= f_U(u) f_V(v) : U \text{ and } V \text{ are indep.} \\ &= \frac{u^{\frac{r_1}{2}-1} e^{-u/2}}{\Gamma\left(\frac{r_1}{2}\right) 2^{r_1/2}} \frac{v^{\frac{r_1}{2}-1} e^{-v/2}}{\Gamma\left(\frac{r_2}{2}\right) 2^{r_2/2}} \\ &= \frac{u^{\frac{r_1}{2}-1} v^{\frac{r_1}{2}-1} e^{-(u+v)/2}}{\Gamma\left(\frac{r_1}{2}\right) \Gamma\left(\frac{r_2}{2}\right) 2^{(r_1+r_2)/2}} \end{aligned}$$

Then, the joint pdf of (W, Z) is

$$\begin{aligned} g(w,z) &= f\left(\frac{r_1}{r_2} wz, z\right) |J| \\ &= \frac{\left(\frac{r_1}{r_2} wz\right)^{\frac{r_1}{2}-1} z^{\frac{r_2}{2}-1} e^{-\left(\frac{r_1}{r_2} wz + z\right)/2}}{\Gamma\left(\frac{r_1}{2}\right) \Gamma\left(\frac{r_2}{2}\right) 2^{(r_1+r_2)/2}} \frac{r_1}{r_2} z \\ &= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}-1+1} w^{\frac{r_1}{2}-1} z^{\frac{r_1}{2}-1+\frac{r_2}{2}-1+1} e^{-z\left(\frac{r_1}{r_2} w + 1\right)/2}}{\Gamma\left(\frac{r_1}{2}\right) \Gamma\left(\frac{r_2}{2}\right) 2^{(r_1+r_2)/2}} \\ &= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} w^{\frac{r_1}{2}-1} z^{\frac{r_1+r_2}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right) \Gamma\left(\frac{r_2}{2}\right) 2^{(r_1+r_2)/2}} exp\left[-z/\left(\frac{2}{\frac{r_1}{r_2} w + 1}\right)\right] \end{aligned}$$

Hence, the pdf of W is

$$\begin{aligned}
g(w) &= \int_0^\infty g(w,z)dz \\
&= \frac{\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} w^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)2^{(r_1+r_2)/2}} \int_0^\infty \Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{2}{\frac{r_1}{r_2}w+1}\right)^{(r_1+r_2)/2} \\
&\quad \frac{z^{\frac{r_1+r_2}{2}-1} e^{-z/\left(\frac{2}{\frac{r_1}{r_2}w+1}\right)}}{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{2}{\frac{r_1}{r_2}w+1}\right)^{(r_1+r_2)/2}} dz \\
&= \frac{\Gamma\left(\frac{r_1+r_2}{2}\right)\left(\frac{r_1}{r_2}\right)^{\frac{r_1}{2}} w^{\frac{r_1}{2}-1}}{\Gamma\left(\frac{r_1}{2}\right)\Gamma\left(\frac{r_2}{2}\right)\left(\frac{r_1}{r_2}w+1\right)^{(r_1+r_2)/2}}
\end{aligned}$$

(iii) μ and σ^2

$$F = \frac{U/r_1}{V/r_2} \sim F(r_1, r_2)$$

$$\begin{aligned}
E(F) &= \frac{r_2}{r_1} E\left(\frac{U}{V}\right) \\
&= \frac{r_2}{r_1} E(U) E\left(\frac{1}{V}\right) \\
&= \frac{r_2}{r_1} E(U) E(V^{-1}) \\
&= \frac{r_2}{r_1} r_1 \frac{1}{r_2-2} \\
&= \frac{r_2}{r_2-2}
\end{aligned}$$

$$\begin{aligned}
Var(F) &= E(F^2) - E^2(F) \\
&= \frac{r_2^2}{r_1^2} E\left(\frac{U^2}{V^2}\right) - \left\{\frac{r_2}{r_1} E(U)E(V^{-1})\right\}^2 \\
&= \frac{r_2^2}{r_1^2} E(U^2)E(V^{-2}) - \left\{\frac{r_2}{r_1} E(U)E(V^{-1})\right\}^2 \\
&= r_2^2 \frac{2(r_1 + r_2 - 2)}{r_1(r_2-2)^2(r_2-4)}
\end{aligned}$$

In general,

$$\begin{aligned}
E(V^{-k}) &= \int_0^\infty v^{-k} \frac{v^{\frac{r_2}{2}-1} e^{-v/2}}{\Gamma\left(\frac{r_2}{2}\right) 2^{r_2/2}} dv \\
&= \int_0^\infty \frac{\Gamma\left(\frac{r_2}{2}-k\right) 2^{\frac{r_2}{2}-k}}{\Gamma\left(\frac{r_2}{2}\right) 2^{r_2/2}} \frac{v^{\left(\frac{r_2}{2}-k\right)-1} e^{-v/2}}{\Gamma\left(\frac{r_2}{2}-k\right) 2^{\frac{r_2}{2}-k}} dv \\
&= \frac{\Gamma\left(\frac{r_2}{2}-k\right)}{\Gamma\left(\frac{r_2}{2}\right) 2^k}, \ k = 1, 2, 3, \cdots
\end{aligned}$$

3 Student's theorem

Theorem 3.6.1. $X_1, \cdots, X_n$: iid $N(\mu, \sigma^2)$, $\bar{X} = \frac{1}{n}\sum X_i$, $s^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$
Then

(a) $\bar{X} \sim N(\mu, \sigma^2/n)$

(b) $\bar{X}$ and s are indep.

(c) $\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$

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

(pf)

(a) Let

$$\boldsymbol{X} = (X_1, \cdots, X_n)', \ \mathbf{1} = (1, \cdots, 1)'$$

Now, let

$$\boldsymbol{a} = \frac{1}{n}\mathbf{1} = \left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$$

then

$$\boldsymbol{a'X} \sim N(\boldsymbol{a}'(\mu\mathbf{1}), \boldsymbol{a}'Cov(\boldsymbol{X})\boldsymbol{a})$$

because

$$\boldsymbol{X} \sim N_n(\mu\mathbf{1}, \sigma^2 I)$$

i.e.

$$\bar{X} = \boldsymbol{a'X} \sim N(\mu, \sigma^2/n)$$

(b) Let

$$\boldsymbol{Y} = (X_1 - \bar{X}, \cdots, X_n - \bar{X})'$$

Consider

$$\boldsymbol{W} = \begin{pmatrix} \bar{X} \\ \boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\mathbf{1} \\ I - \frac{1}{n}\mathbf{11'} \end{pmatrix} \boldsymbol{X} = A\boldsymbol{X}$$

First, will show $\bar{X}$ and $\boldsymbol{Y}$ are indep.
Recall that if both $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are normally distributed then,

$$Cov(\boldsymbol{X_1}, \boldsymbol{X_2}) = 0$$

implies that $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are indep.
Since $\bar{X}$ and $\boldsymbol{Y}$ are normal to show independence between $\bar{X}$ and $\boldsymbol{Y}$

we need to show $Cov(\bar{X}, \boldsymbol{Y}) = \boldsymbol{0}$

$$\begin{aligned}
Cov(\boldsymbol{W}) &= Cov\begin{pmatrix} \bar{X} \\ \boldsymbol{Y} \end{pmatrix} \\
&= \begin{pmatrix} Var(\bar{X}) & Cov(\bar{X}, \boldsymbol{Y}) \\ Cov(\boldsymbol{Y}, \bar{X}) & Cov(\boldsymbol{Y}) \end{pmatrix} \\
&= Cov(A\boldsymbol{X}) \\
&= ACov(\boldsymbol{X})A' \\
&= \begin{pmatrix} \frac{1}{n}\boldsymbol{1} \\ I - \frac{1}{n}\boldsymbol{11'} \end{pmatrix} \boldsymbol{X}\sigma^2 I \begin{pmatrix} \frac{1}{n}\boldsymbol{1} & I - \frac{1}{n}\boldsymbol{11'} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \frac{1}{n^2}\boldsymbol{1'1} & \frac{1}{n}\boldsymbol{1}'\left(I - \frac{1}{n}\boldsymbol{11'}\right) \\ \left(I - \frac{1}{n}\boldsymbol{11'}\right)\frac{1}{n}\boldsymbol{1} & \left(I - \frac{1}{n}\boldsymbol{11'}\right)^2 \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \frac{1}{n} & \frac{1}{n}\left(\boldsymbol{1}' - \frac{1}{n}\boldsymbol{1'11'}\right) \\ \frac{1}{n}\left(\boldsymbol{1}' - \frac{1}{n}\boldsymbol{11'1}\right) & I - \frac{1}{n}\boldsymbol{11'} \end{pmatrix} \\
&= \sigma^2 \begin{pmatrix} \frac{1}{n} & \boldsymbol{0} \\ \boldsymbol{0} & I - \frac{1}{n}\boldsymbol{11'} \end{pmatrix}
\end{aligned}$$

Therefore,

$$Cov(\bar{X}, \boldsymbol{Y}) = \boldsymbol{0}$$

i.e. $\bar{X}$ and $\boldsymbol{Y}$ are indep.
Finally, note that

$$s^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 = \frac{1}{n-1}\boldsymbol{Y'Y},\ \boldsymbol{Y} = (X_1 - \bar{X}, \cdots, X_n - \bar{X})^2$$

i.e. s^2 is function of $\boldsymbol{Y}$. Therefore, $\bar{X}$ and $\boldsymbol{Y}$ are indep.

(c) Note that

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \bar{X})(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

Therefore,

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 = \frac{\sum(X_i-\bar{X})^2}{\sigma^2} + \frac{n(\bar{X}-\mu)^2}{\sigma^2}$$

Now,

$$\frac{X_i-\mu}{\sigma} \sim N(0,1) \Rightarrow \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(1) \Rightarrow \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$$

$$\frac{n(\bar{X}-\mu)}{\sigma^2} = \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$$

Apply mgf technique on both sides, i.e.

$$\begin{aligned} M_A(t) &= E[e^{tA}] \\ &= E[e^{t(B+C)}] \\ &= E[e^{tB}e^{tC}] \\ &= E[e^{tB}]E[e^{tC}] \end{aligned}$$

$$(1-2t)^{-n/2} = E[e^{tB}](1-2t)^{-1/2}$$

$\therefore E(t^{tB}) = (1-2t)^{-n/2}(1-2t)^{-1/2} = (1-2t)^{-(n-1)/2}$: mgf of $\chi^2(n-1)$

(d)

$$\frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}} = \frac{Z}{\sqrt{V/(n-1)}} \sim t(n-1)$$

where, $Z \sim N(0,1)$, $V \sim \chi^2(n-1)$, Z and V are indep.

4 Unbiasedness, Consistency and Limiting Distribution

4.1 Expectation of Functions

① definitions

(i) $X_1, X_2, \cdots, X_n$ are called *random sample* (r.s.) if they are iid.

(ii) T is called a *statistic* if T is a fuction of random sample only.

(iii) $\bar{X} = \sum \frac{X_i}{n}$: *sample mean*, $\quad s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$: *sample variance*

② expectations

$\boldsymbol{x} = (X_1, \cdots, X_n)', \quad \boldsymbol{y} = (Y_1, \cdots, Y_n)'$: ramdom vector

$\boldsymbol{a} = (a_1, \cdots, a_n)', \quad \boldsymbol{b} = (b_1, \cdots, b_n)'$: constant vector

Let $T = \boldsymbol{a}'\boldsymbol{x}, \quad W = \boldsymbol{b}'\boldsymbol{y}$ be statistics

(i) $E(T) = \boldsymbol{a}'E(\boldsymbol{x})$

(ii) $Var(T) = \boldsymbol{a}'Cov(\boldsymbol{x})\boldsymbol{a}$

(iii) $Cov(T, W) = Cov(\boldsymbol{a}'\boldsymbol{x}, \boldsymbol{b}'\boldsymbol{y}) = \boldsymbol{a}'Cov(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{b}$

③ unbiasedness

$X_1, \cdots, X_n$: random sample from $f(\boldsymbol{x} : \theta)$, $\theta \in \Omega$ is parameter

$T = T(X_1, \cdots, X_n)$: statistic, T is called *unbiased* if $E(T) = \theta$, $^{\forall}\theta \in \Omega$.

4.2 Convergence in Probability

① definitions

Definition 4.2.1. Let $\{X_n\}$ be a seq.of r.v.'s and X be a r.v.
We say X_n *converges in probability* to X if $\,^{\forall}\varepsilon > 0,\;\; P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$, and denoted by $X_n \xrightarrow{P} X$.

② properties

(i) WLLN

Theorem 4.2.1. $\{X_n\}$: seq. of *iid* r.v.'s with mean μ, variance $\sigma^2 < \infty$.
Then, $\bar{X}_n \xrightarrow{P} \mu$.

(pf) Can be shown easily by Chebyshev's ineq.

(ii)

Theorem 4.2.2. $X_n \xrightarrow{P} X,\;\; Y_n \xrightarrow{P} Y \;\;\Rightarrow\;\; X_n + Y_n \xrightarrow{P} X + Y$

(pf) By triangular ineq.
we have $|X_n - X| + |Y_n - Y| \geq |(X_n + Y_n) - (X + Y)|$

$$\begin{aligned}\therefore P(|(X_n + Y_n) - (X+Y) \geq \varepsilon) &\leq P(\{|X_n - X| + |Y_n - Y|\} \geq \varepsilon) \\ &\leq P(|(X_n - X) \geq \varepsilon/2) + P(|Y_n - Y| \geq \varepsilon/2)\end{aligned}$$

(iii)

Theorem 4.2.3. $X_n \xrightarrow{P} X, \quad , a: \text{const.} \Rightarrow aX_n \xrightarrow{P} aX$

(pf) $P(|aX_n - aX| \geq \varepsilon) = P(|a||X_n - X| \geq \varepsilon) = P(|X_n - X| \geq \varepsilon/|a|)$

(iv)

Theorem 4.2.4. $X_n \xrightarrow{P} a, \quad g: \text{conti. function at } a \ \Rightarrow g(X_n) \xrightarrow{P} g(a)$

(pf) Let $\varepsilon > 0$. Since g is conti. at a, $\exists\ \delta > 0$ s.t. if $|x - a| < \delta$,
then $|g(x) - g(a)| < \varepsilon$. Thus, $|g(x) - g(a)| \geq \varepsilon$ implies $|x - a| > \delta$.
Therefore, $P(|g(X_n) - g(a)| \geq \varepsilon) \leq P(|X_n - a| \geq \delta) \to 0$

(v) (Remark)

Theorem 4.2.5. $X_n \xrightarrow{P} X, \quad g: \text{conti.} \Rightarrow g(X_n) \xrightarrow{P} g(X)$

③ consistency

(i) definition : a statistic T_n is called *consistent* est. of θ if $T_n \xrightarrow{P} \theta$.

(ii)

Example 4.2.1. $X_1, \cdots, X_n$: r.s. from $dist^n$ with mean μ, variance σ^2.
Then, $s^2 = \sum(X_i - \bar{X})^2/(n-1)$ is consistent est. of σ^2

(pf) $s^2 = \frac{n}{n-1}(\frac{1}{n}\sum X_i^2 - \bar{X}^2) \xrightarrow{P} 1 \cdot [E(X_1^2) - \mu^2] = \sigma^2$

4.3 Convergence in Distribution

① convergence in distributions

(i) def

Definition 4.3.1. $\{X_n\}$: seq of r.v.'s with cdf F_{X_n}, X : r.v. with cdf F_X. We say X_n *converges in distribution* to X if $\lim F_{X_n}(x) = F_X(x)$, $^\forall x$ in which F_X is conti., and denoted by $X_n \xrightarrow{D} X$.
X is also called *limiting distribution* of X_n or asymptotic distribution of X_n.

(ii)

Example 4.3.1. Let X_n have the cdf

$$\begin{aligned} F_{X_n}(x) &= \int_{-\infty}^{x} \frac{1}{\sqrt{1/n}\sqrt{2\pi}} e^{-nw^2/2} dw \\ &= \int_{-\infty}^{\sqrt{n}x} \frac{1}{2\pi} e^{-v^2/2} dv \end{aligned}$$

by changing variable $v = \sqrt{n}w$.

$$\therefore \lim_{x\to\infty} F_{X_n}(x) = \begin{cases} 0, & x < 0 \\ 1/2, & x = 1 \\ 1, & x > 1. \end{cases}$$

So, take

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

We call X is degenerate at $x = 0$.

(iii)

Example 4.3.2. Let $X_1, \cdots, X_n$ be r.s. from $u(0,\theta)$, and let $Z_n = n(\theta - Y_n)$, where $Y_n = max(X_1, \cdots, X_n)$. Find the limiting distribution of Z_n.

(sol) The cdf of Z_n is

$$\begin{aligned} F_{Z_n}(t) &= P(Z_n \le t) = P(Y_n \ge \theta - \frac{t}{n}) = 1 - P(Y_n \le \theta - \frac{t}{n}) \\ &= 1 - (\frac{\theta - t/n}{\theta})^n = 1 - (1 - \frac{t/\theta}{n})^n \Rightarrow 1 - e^{-t/\theta} \\ &: \ cdf \ \ of \ \ \varepsilon(\theta), \quad i.e. \quad Z_n \xrightarrow{D} \varepsilon(\theta). \end{aligned}$$

(iv)

Theorem 4.3.1. $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{D} X$

(pf) Let x be a continuous point in $F_X(x)$ and $\varepsilon > 0$.

$$\begin{aligned} F_{X_n}(x) &= P(X_n \le x) = P(\{X_n \le x\} \cap \{|X_n - X| < \varepsilon\}) + P(\{X_n \le x\} \cap \{|X_n - X| \ge \varepsilon\} \\ &\le P(X \le x + \varepsilon) + P(|X_n - X| \ge \varepsilon) \end{aligned}$$

$$\therefore \ \overline{\lim} \ F_n(x) \le F_X(x + \varepsilon)$$

now,

$$\begin{aligned} P(X_n > x) &= P(\{X_n > x\} \cap \{|X_n - X| < \varepsilon\}) + P(\{X_n > x\} \cap \{|X_n - X| \ge \varepsilon\}) \\ &\le P(X > x - \varepsilon) + P(|X_n - X| \ge \varepsilon) \end{aligned}$$

i.e. $\quad 1 - P(X_n > x) \ge 1 - P(X \ge x - \varepsilon) - P(|X_n - X| \ge \varepsilon)$

i.e. $\quad F_{X_n}(x) \ge P(X \le x - \varepsilon) - P(|X_n - X| \ge \varepsilon)$

$$\therefore \ \underline{\lim} \ F_{X_n}(x) \ge F_X(x - \varepsilon)$$

Conclusively, $\ F_X(x - \varepsilon) \le \underline{\lim} \ F_{X_n}(x) \le \overline{\lim} \ F_{X_n}(x) \le F_X(x + \varepsilon)$
By letting $\varepsilon \to 0$, we have $\lim F_{X_n}(x) = F_X(x)$.

(v)

Theorem 4.3.2. The converse of Thm 4.3.1 does not hold. i.e.,
$X_n \xrightarrow{D} X \nRightarrow X_n \xrightarrow{P} X$.
However, if X is degenerate at c, then it is true. i.e.,
$X_n \xrightarrow{D} c \Rightarrow X_n \xrightarrow{P} c$.

(pf) $\lim P(|X_n - c| \leq \varepsilon) = \lim\{F_{X_n}(c+\varepsilon) - F_{X_n}(c-\varepsilon)\} = F_X(c+) - F_X(c-) = 1 - 0 = 1$.

(vi)

Theorem 4.3.3. $X_n \xrightarrow{D} X,\ Y_n \xrightarrow{P} 0 \ \Rightarrow X_n + Y_n \xrightarrow{D} X$

(vii)

Theorem 4.3.4. $X_n \xrightarrow{D} X$, g : conti. ftn. $\Rightarrow g(X_n) \xrightarrow{D} g(X)$

(viii)

Theorem 4.3.5. (Slutzky Thm) $X_n,\ X,\ Y_n,\ Z_n$: r.v's $\quad k_1,\ k_2$: const.
s.t. $X_n \xrightarrow{D} X,\ Y_n \xrightarrow{P} k_1,\ Z_n \xrightarrow{P} k_2 \ \Rightarrow\ Y_n + Z_n X_n \xrightarrow{D} k_1 + k_2 X$.

② bounded in probability

(i) Landau's Big Oh and little oh

When we write $x_n \to 0$ as $n \to \infty$, what is the rate of convergence?

Let $\{r_n\} \subset (0, \infty)$ be the rate of convergence (*e.g.* $r_n = n^{-P},\ P > 0$)

- $x_n = o(r_n)$ iff $\dfrac{x_n}{r_n} \to 0$ as $n \to \infty$
- $x_n = O(r_n)$ iff $\underline{\lim} \dfrac{x_n}{r_n} < \infty$

  iff $\exists\ M \in (0, \infty),\ \exists\ N$ s.t. ${}^{\forall} n \geq N,\ |\frac{x_n}{r_n}| \leq M.$

We can extend this notation when X_n is r.v.

- $X_n = o_p(r_n)$ iff $\dfrac{X_n}{r_n} \xrightarrow{P} 0$ iff ${}^{\forall}\varepsilon > 0,\ P(|\frac{X_n}{r_n}| > \varepsilon) \to 0$
- $X_n = O_p(r_n)$ iff ${}^{\forall}\varepsilon > 0, \exists\, M$ s.t. $P(|\frac{X_n}{r_n}| > M) < \varepsilon$

  if $X_n = O_p(1)$, then $\{X_n\}$ is called bounded in prob.

(ii) Taylor expansion

If $g(x)$ is k-times differentiable at $x = x_0$,
we have $g(x) = \displaystyle\sum_{j=0}^{k} \frac{1}{j!}\ g^{(j)}(x_0) \cdot (x - x_0)^j + o(|x - x_0|^k)$ as $|x - x_0| \to 0$

(iii)

Theorem 4.3.6. $X_n \xrightarrow{D} X \;\Rightarrow\; X_n = O_P(1)$

(pf) Let η be continuous point in $F_X(x)$, then

$$P(|X_n| \leq \eta) = F_{X_n}(\eta) - F_{X_n}(-\eta^-) \to F_X(\eta) - F_X(-\eta) \quad \cdots (*)$$

now, can choose η_1 and η_2 s.t. for a given $\varepsilon > 0$,

$$F_X(x) < \frac{\varepsilon}{2} \;\; for \;\; x \leq \eta_1 \;\;\&\;\; F_X(x) > 1 - \frac{\varepsilon}{2} \;\; for \;\; x \geq \eta_2$$

Take $\eta = max(|\eta_1|, |\eta_2|)$, then

$$P(|X| \leq \eta) = F_X(\eta) - F_X(-\eta^-) \geq 1 - \frac{\varepsilon}{2} - \frac{\varepsilon}{2} = 1 - \varepsilon$$

By taking limit in (*), we have $\lim P(|X_n| \leq \eta) \geq 1 - \varepsilon$.

(iv)

Theorem 4.3.7. $X_n = O_P(1),\; Y_n \xrightarrow{P} 0 \;\Rightarrow\; X_n Y_n \xrightarrow{P} 0$

(pf)

$$P(|X_n Y_n| \geq \varepsilon) = P(|X_n Y_n| \geq \varepsilon,\; |X_n| \leq M) + P(|X_n Y_n| \geq \varepsilon,\; |X_n| > M)$$

$$\begin{aligned} \therefore \lim P(|X_n Y_n| \geq \varepsilon) &= \lim P(|X_n Y_n \geq \varepsilon,\; |X_n| \leq M) \\ &\leq \lim P(|Y_n| \geq \varepsilon / M) = 0 \end{aligned}$$

③ Δ-method

(i)

Theorem 4.3.8. $Y_n = Op(1),\ X_n = o_p(Y_n) \Rightarrow\ X_n \xrightarrow{P} 0$

(pf)

$$\begin{aligned} P(|X_n| \geq \varepsilon) &= P(|X_n| \geq \varepsilon,\ |Y_n| \leq M) + P(|X_n| \geq \varepsilon,\ |Y_n| > M) \\ &\leq P(|\frac{X_n}{Y_n}| \geq \frac{\varepsilon}{M}) + P(|Y_n| > M) \end{aligned}$$

Take limit on both sides.

$$\begin{aligned} \lim P(|X_n| \geq \varepsilon) &\leq \lim P(|\frac{X_n}{Y_n}| \geq \frac{\varepsilon}{M}) + \lim P(|Y_n| > M) \\ &= 0 \end{aligned}$$

(ii) Δ-method

Theorem 4.3.9. Assume $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$ and $g(x)$ is diff. at $x = \theta$, $g'(\theta) \neq 0$. Then, $\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2 g'(\theta)^2)$

(pf) By Taylor expansion,

$$g(X_n) = g(\theta) + g'(\theta)(X_n - \theta) + o_p(|X_n - \theta|)$$

$$i.e.\quad \sqrt{n}(g(X_n) - g(\theta)) = g'(\theta) \cdot \sqrt{n}(X_n - \theta) + o_p(\sqrt{n}|X_n - \theta|)$$

now, $\sqrt{n}(X_n - \theta) = O_p(1)$, so that by Thm 4.3.8., $o_p(\sqrt{n}|X_n - \theta|) = o_p(|O_p(1)|) \xrightarrow{P} 0$.

(iii) Example

Assume $\sqrt{n}(\bar{X} - \mu) \xrightarrow{D} N(0, \sigma^2)$.

Find the limiting distribution of $\sqrt{n}(\bar{X}^2 - \mu^2)$.

(sol) $\quad g(x) = x^2, \quad g'(x) = 2x \qquad \therefore g'(\mu)^2 = 4\mu^2$

$$\therefore\ \sqrt{n}\ (\bar{X}^2 - \mu^2) \xrightarrow{D} N(0,\ 4\mu^2\sigma^2)$$

④ mgf technique

(i)

Theorem 4.3.10. $\{X_n\}$: seq. of r.v's with mgf $M_{X_n}(t)$.
X : r.v. with mgf $M_X(t)$. If $\lim M_{X_n}(t) = M_X(t)$, then $X_n \xrightarrow{D} X$.

(ii) useful result for limit
If $\lim \psi(n) = 0$, then $\lim(1 + \frac{b}{n} + \frac{\psi(n)}{n})^{cn} = e^{bc}$

(iii)

Example 4.3.3. $Z \sim \chi^2(n)$. Show that $Y = \frac{Z-n}{\sqrt{2n}} \xrightarrow{D} N(0,\ 1)$

(pf)

$$\begin{aligned} M_Y(t) &= E\ [\exp\ \{t \cdot (\frac{Z-n}{\sqrt{2n}})\}] = e^{-tn/\sqrt{2n}} \cdot E\ [e^{tZ/\sqrt{2n}}] \\ &= \exp\ [-(t\sqrt{\frac{2}{n}})\frac{n}{2}\ ] \cdot (1 - 2 \cdot \frac{t}{\sqrt{2n}})^{-n/2} \\ &= (e^{-t/\sqrt{2/n}} - t\ \sqrt{\frac{2}{n}}\ e^{\ t\sqrt{2/n}})^{-\frac{n}{2}} \end{aligned}$$

now, $e^{t\sqrt{2/n}} = 1 + t\sqrt{\frac{2}{n}} + \frac{1}{2}(t\sqrt{\frac{2}{n}})^2 + \frac{1}{6}(t\sqrt{\frac{2}{n}})^3 + o(n^{-3/2})$

$$\begin{aligned} \therefore\ \ e^{\ t\sqrt{2/n}} - t\sqrt{\frac{2}{n}}\ e^{t\sqrt{2/n}} &= (1 + t\sqrt{\frac{2}{n}} + \frac{t^2}{n} + \frac{\sqrt{2}t^3}{3n^{3/2}} + o(n^{-3/2}) - t\sqrt{\frac{2}{n}} - \frac{2t^2}{n} - \frac{\sqrt{2}t^3}{n^{3/2}} + \\ &= (1 - \frac{t^2}{n} + \frac{\psi(n)}{n})\ , \qquad \psi(n) = -\frac{2\sqrt{2}t^3}{3n^{1/2}} + O(n^{-1/2}) \end{aligned}$$

$$\therefore\ \ M_Y(t) \to e^{t^2/2} : mgf\ \ of\ \ N(0,\ 1)$$

4.4 Central Limit Theorem

Theorem 4.4.1. $X_1, \cdots, X_n$: r.s. from a distribution with mean μ, variance σ^2. Then, $\sqrt{n}(\bar{X}-\mu)/\sigma \xrightarrow{D} N(0,\ 1)$

(pf) Let $m(t) = E[e^{t(x-\mu)}]$: mgf of $Y = X - \mu$.

$$\begin{aligned} m(t) &= m(o) + m'(o)t + \frac{m''(o)}{2}t^2 + \frac{m'''(o)}{6}t^3 + \cdots \\ &= 1 + \frac{\sigma^2}{2}t^2 + \frac{m'''(o)}{6}t^3 + \cdots \end{aligned}$$

Now, consider mgf of $Z = \sqrt{n}(\bar{X}-\mu)/\sigma$

$$\begin{aligned} M_Z(t) &= E\ [\exp(t \cdot \frac{\sum X_i - n\mu}{\sigma\sqrt{n}})] = E\ [\exp(t \cdot \frac{X_1-\mu}{\sigma\sqrt{n}}) \cdots \exp(t \cdot \frac{X_n-\mu}{\sigma\sqrt{n}})] \\ &= [E\{\exp(t \cdot \frac{X_1-\mu}{\sigma\sqrt{n}})\}]^n = \{m(\frac{t}{\sigma\sqrt{n}})\}^n \\ &= [1 + \frac{\sigma^2}{2} \cdot \frac{t^2}{\sigma^2 n} + \frac{m'''(o)}{6} \cdot \frac{t^3}{\sigma^2 n^{3/2}} + \cdots]^n = (1 + \frac{t^2}{2n} + \frac{\psi(n)}{n})^n \to e^{t^2/2} \end{aligned}$$

Example 4.4.1. $X_1, \cdots, X_n$: r.s. from $B(1,\ ,p)$. Then, by CLT, $\sqrt{n}(\bar{X}-p)/\sqrt{p(1-p)} \xrightarrow{D} N(0,\ 1)$

Example 4.4.2. Find h s.t. $\sqrt{n}(h(\bar{X}) - h(p)) \ \to\ N(0,\ c^2)$, c : const.

(sol)

$\sqrt{n}(\bar{X}-p) \to N(0,\ p(1-p)) \Rightarrow \sqrt{n}(h(\bar{X}) - h(p)) \to N(0,\ h'(p)^2 p(1-p))$

$$\therefore h'(p)^2 p(1-p) = c^2 \quad \Rightarrow \quad h'(p) = \sqrt{c^2/p(1-p)}$$
$$h(p) = (2c) \cdot arc\ sin(\sqrt{p})$$

This kind of transformation called the variance stabilizing transformation.

4.5 Asymptotics for Multivariate Distributions

① Euclidean norm

(i) definition : $\boldsymbol{v} = (v_1, \cdots, v_p)' \in R^p$, $||\boldsymbol{v}|| = (\sum v_i^2)^{1/2}$: Euclidean norm.

(ii) properties

(a) $||\boldsymbol{v}|| \leq 0$. Equality holds when $\boldsymbol{v} = \mathbf{0}$

(b) $^\forall a \in R'$, $||a\boldsymbol{v}|| = |a||\boldsymbol{v}|$

(c) $||\boldsymbol{u} + \boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$: triangular inequality.

(iii) basis

$\boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_p$: basis for R_p

$\boldsymbol{v} = \sum_{i=1}^{p} \boldsymbol{v}_i \boldsymbol{e}_i$

(iv)

Lemma 4.5.1. $|v_j| \leq ||v|| \leq \sum_{i=1}^{p} |v_i|$, $j = 1, \cdots, p$

(pf) $v_j^2 \leq \sum_{i=1}^{p} v_i^2 = ||\boldsymbol{v}||^2 \Rightarrow |v_j| \leq ||\boldsymbol{v}||$

also, $||\boldsymbol{v}|| = ||\sum v_i \boldsymbol{e}_i|| \leq \sum |v_i|||e_i|| = \sum |v_i|$.

② Convergence in probability

(i) definition : $\{\boldsymbol{X}_n\}$ converges in prob. to $\boldsymbol{X}$ if $P(||\boldsymbol{X}_n - \boldsymbol{X}|| \geq \varepsilon) \to 0$, and denoted by $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$.

(ii)

Theorem 4.5.1. $\boldsymbol{X}_n \xrightarrow{P} \boldsymbol{X}$ iff $X_{n_j} \xrightarrow{P} X_j,\ j = 1, \cdots, p$

(pf)

($\Rightarrow$) By Lemma 4.5.1, $|X_{n_j} - X_j| \leq ||\boldsymbol{X}_n - \boldsymbol{X}||$
($\Leftarrow$) By Lemma 4.5.1, $\sum_{i=1}^{p} |X_{n_j} - X_j| \geq ||\boldsymbol{X}_n - \boldsymbol{X}||$

$$\therefore P(||\boldsymbol{X}_n - \boldsymbol{X}|| \geq \varepsilon) \leq P(\sum |X_{n_j} - X_j| \geq \varepsilon) \leq \sum_{i=1}^{p} P(|X_{n_j} - X_j| \geq \varepsilon/p)$$

(iii) Examples

(i) $X_1, \cdots, X_n$: r.s. from a distribution with mean μ and variance σ.
We know that $\bar{X} \xrightarrow{P} \mu,\ s^2 \xrightarrow{P} \sigma^2$, by Thm 4.5.1, $(\bar{X},\ s^2) \xrightarrow{P} (\mu,\ \sigma^2)$.

(ii) $\boldsymbol{X}_1, \cdots, \boldsymbol{X}_n$: r.s. from a distribution with mean $\boldsymbol{\mu}$ and var-cov. $\boldsymbol{\Sigma}$.
We know that $\bar{X}_j \xrightarrow{P} \mu_j,\ j = 1, \cdots, p,$ then $\bar{\boldsymbol{X}} \xrightarrow{P} \boldsymbol{\mu}$
also, $s_j^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)^2 \xrightarrow{P} \sigma_j^2$ and $s_{jk} = \frac{1}{n-1} \sum_{i=1}^{n} (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_j) \xrightarrow{P} \sigma_{jk}$, we have $S \xrightarrow{P} \Sigma$.

③ Convergence in distribution

(i) definition : $\{X_n\}$ converges in distribution to $\boldsymbol{X}$ if $F_{X_n}(\boldsymbol{X}) \to F_X(\boldsymbol{X})$ for all points $\boldsymbol{x}$ at which $F_X(\boldsymbol{x})$ is conti, and denoted by $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$.

(ii)

Theorem 4.5.2. $\boldsymbol{X} \xrightarrow{D} \boldsymbol{X}$, g: conti. $\Rightarrow$ $g(\boldsymbol{X}_n) \xrightarrow{D} g(\boldsymbol{X})$

(iii)

Theorem 4.5.3. $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$ iff $M_n(\boldsymbol{t}) \to M(\boldsymbol{t})$

④ CLT

(i) multivariate CLT

Theorem 4.5.4. $\{\boldsymbol{X}_n\}$: seq. of iid random vectors with mean μ, var-cov. Σ $\Rightarrow$ $\boldsymbol{Y}_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\boldsymbol{X}_i - \boldsymbol{\mu}) = \sqrt{n}\ (\bar{\boldsymbol{X}} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\mathbf{0},\ \Sigma)$

(pf)

$$\begin{aligned} M_n(\boldsymbol{t}) &= E\ [\exp\{t' \cdot \frac{1}{\sqrt{n}}\sum(X_i - \mu)\}] = E\ [\exp\{\frac{1}{\sqrt{n}}\sum t'(X_i - \mu)\}] \\ &= E\ [\exp\{\frac{1}{\sqrt{n}}\sum W_i\}],\ \ W_i = t'(X_i - \mu) \end{aligned}$$

now, $W_1, \cdots, W_n$ are iid with mean 0, variance $t'\Sigma t$.
Then, by CLT, $Z = \frac{1}{\sqrt{n}}\sum W_i \xrightarrow{D} N(0,\ t'\Sigma t)$
now, $M_n(t) = E\ [\exp\{\frac{1}{\sqrt{n}}\sum W_i\}] = E\ [e^{1 \cdot Z}]$, i.e. mgf evaluated at $t = 1$.
Therefore, $M_n(t) \rightarrow \exp(0 \cdot 1 + \frac{1}{2}t'\Sigma t \cdot 1^2) = \exp(t'\Sigma t/2)$ which is mgf of
$N_p(\mathbf{0},\ \Sigma)$.

(ii)

Theorem 4.5.5. $X_n \xrightarrow{D} N_p(\boldsymbol{\mu},\ \Sigma)$. A : m×p, $\boldsymbol{b}$: m×1 $\Rightarrow$
$A\boldsymbol{X}_n + \boldsymbol{b} \xrightarrow{D} N_m(A\boldsymbol{\mu} + \boldsymbol{b},\ A\Sigma A')$

(iii)

Theorem 4.5.6. $\{\boldsymbol{X}_n\}$: seq. of p-dim random vector.
$\sqrt{n}\ (X_n - \mu) \xrightarrow{D} N_p(\mathbf{0},\ \Sigma)$. $g(\boldsymbol{X}) = (g_1(X), \cdots, g_k(X))' : R^p \rightarrow R^k$
$B = (\dfrac{\partial g_i}{\partial \mu_j})$: $k \times p$ matrix. Then, $\sqrt{n}(g(\boldsymbol{X}_n) - g(\boldsymbol{\mu})) \xrightarrow{D} N_k(\mathbf{0},\ B\Sigma B')$

CHAPTER 5. Some Elementary Statistical Inference

5.1 Sampling and Statistic

- sampling with (without) replacement.
- random sample, statistic

5.2 Order Statistic

① definition

$X_1, \cdots, X_n$: r.s. from a pdf $f(x)$ and cdf $F(x)$.
Let Y_1 be the smallest of X_i's, Y_2 be the 2nd smallest of X_i's, $\cdots$, and Y_n be the largest of X_i's. Then, $Y_1 < Y_2 < \cdots < Y_n$ is called the order statistics of $X_1, \cdots, X_n$.

② pdf

(i) joint pdf of $Y_1, \cdots, Y_n$ (Thm 5.2.1)

$$g(y_1, \cdots, y_n) = n!\, f(y_1) \cdots f(y_n), \qquad y_1 < y_2 < \cdots < y_n$$

(pf) Consider transformation $(x_1, \cdots, x_n) \to (y_1, \cdots, y_n)$.
Then, there are $n!$ methods, and Jacobian is ± 1.
Therefore, $g(y_1, \cdots, y_n) = \sum_{i=1}^{n!} |J_i|\, f(y_1) \cdots f(y_n) = n! \prod_{i=1}^{n} f(y_i)$.

(ii) marginal pdf of Y_k

$$g_k(y_k) = \frac{n!}{(k-1)!(n-k)!} (F(y_k))^{k-1} (1 - F(y_k))^{n-k} f(y_k)$$

(iii) joint pdf of Y_i and Y_j $(i < j)$

$$g_{ij}(y_i,\, y_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(y_i))^{i-1} (F(y_j) - F(y_i))^{j-i-1}$$
$$(1 - F(y_j))^{n-j} f(y_i) f(y_j)$$

(iv) **Example 5.2.3.**

X_1, X_2, X_3 : r.s. from $u(0,\ 1)$, $Y_1 < Y_2 < Y_3$: order stat. Find the pdf of the sample range $Z_1 = Y_3 - Y_1$.

(sol) $z_1 = y_3 - y_1,\ z_2 = y_3 \longrightarrow y_1 = z_2 - z_1,\ y_3 = z_2,\ \ |J| = 1$.
The jpdf of Y_1 and Y_3 is

$$g_{13}(y_1, y_3) = 6(y_3 - y_1)\ ,\ \ 0 < y_1 < y_3 < 1.$$

$$\therefore\ h(z_1,\ z_2) = 6z_1\ ,\ \ 0 < z_1 < z_2 < 1.$$

$$\therefore\ h_1(z_1) = \int_{z_1}^{1} 6z_1\, dz_2 = 6z_1(1 - z_1),\ \ 0 < z_1 < 1.$$

③ quantiles

(i) definition

X : r.v. with conti. cdf $F(x)$. $\xi_p = F^{-1}(p)$: p th quantile of X

(ii) estimator of ξ_p.

Let $Y_1 < \cdots < Y_n$ be order statistic, and consider Y_k, where $k = [p(n+1)]$, as an estimator of ξ_p. For Y_k to be a good estimator of ξ_p.

$$\begin{aligned} E[F(Y_k)] &= \int F(y_k) g_k(y_k)\, dy_k \\ &= \int F(y_k) \frac{n!}{(k-1)!(n-k)!} (F(y_k))^{k-1}(1-F(y_k))^{n-k} f(y_k)\, dy_k \end{aligned}$$

Let $z = F(y_k)$, then $dz = f(y_k)dy_k$, so that

$$E[F(Y_k)] = \int \frac{n!}{(k-1)!(n-k)!} z^k (1-z)^{n-k} dz = \frac{k}{n+1} \simeq p.$$

Y_k is called the p th sample quantile or $100p - th$ percentile.

(iii) five number summary (J. Tukey)

$$Y_1,\ \ Y_{[.25(n+1)]},\ Y_{[.5(n+1)]},\ Y_{[.75(n+1)]},\ Y_n$$

(iv) boxplot

Boxplot is based on the five number summary.
To exhibit potential outliers, define the lower and upper fence (LF, UF), LF$= Q_1 - h$, UF$= Q_3 + h$, $h = 1.5(Q_3 - Q_1)$.
Points that lie outside the fence (LF, UF) are called potential outlies.

(v) **Example 5.2.4.**

data : 56, 70, 89, 94, 96, $\cdots$, 110, 113, 116 $(n = 15)$

$Y_1 = 56,\ Q_1 = y_4 = 94,\ Q_2 = y_8 = 102,\ Q_3 = y_{12} = 108,\ y_{15} = 116,$
$h = 1.5(108 - 94) = 21 \rightarrow$ (LF, UF)$= (73,\ 129)$
$\therefore$ 56, 70 are potential outliers.

(vi) q-q plot

X : r.v. from a location-scale family with cdf $F(\frac{x-a}{b})$, where $F(x)$ is known, but a and b are unknown. Let $\xi_{z,p}$ be the pth quantile of $z = \frac{x-a}{b}$ now,

$$p = P(X \leq \xi_{x,p}) = P(Z \leq \frac{\xi_{x,p} - a}{b}) = P(Z \leq \xi_{z,p})$$

$\therefore \quad \xi_{x,p} = b\, \xi_{z,p} + a \qquad (\ \xi_{z,p}$: known, $\xi_{x,p}$: unknown)

Now, Y_k is estimator of ξ_{x,p_k}, where $p_k = k/(n+1)$. The plot of Y_k vs ξ_{z,p_k} is called q-q plot. If X is distributed as $F = (\frac{x-a}{b})$, then the q-q plot should be linear.

④ confidence intervals of quantiles

Y_k : point est. of ξ_p, where $k = [(n+1)p]$.

As a C.I. for ξ_p, consider $(Y_i,\ Y_j)$ s.t. $i < [(n+1)p] < j$.

When we say $(Y_i,\ Y_j)$ as 100γ% C.I., what is γ?

Need to compute $\gamma = P(Y_i < \xi_p < Y_j)$.

$\{Y_i < \xi_p\} \Leftrightarrow \{$at least i of X values are less than $\xi_p\}$

$\{Y_j > \xi_p\} \Leftrightarrow \{$fewer than j of X values are less than $\xi_p\}$

now, consider this problem as Bernoulli trial, i.e., if $X < \xi_p$, then success, and if $X \geq \xi_p$, then failure. Also, the prob. of success is $P(X < \xi_p) = F(\xi_p) = p$.

$\therefore \quad P(Y_i < \xi_p < Y_j) = \sum_{k=i}^{j-1} \binom{n}{k} p^k (1-p)^{n-k} \equiv \gamma.$

We call $(Y_i,\ Y_j)$ as 100γ% C.I. for ξ_p.

5.3 Tolerance Limits for Distributions

① definition

$X_1, \cdots, X_n$: r.s. from a $dist^n$ with cdf $F(x)$, $Y_1 < \cdots < Y_n$ is order stat. Then, $(y_i,\ y_j)$ s.t. $\gamma = P[F(Y_j) - F(Y_i) \geq p]$ is called $100\gamma\%$ tolerance limits for $100p\%$ of the prob. for the $dist^n$ of X.

② computation of γ

(i) jpdf of $Z_1 = F(Y_n), \cdots, Z_n = F(Y_n)$

Note that $Z = F(X) \sim u(0,1)$ because $G(z) = P(Z \leq z) = P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = \delta$. Hence, $Z_1, \cdots, Z_n$ are order stat. from $u(0,1)$, so that the jpdf of $Z_1, \cdots, Z_n$ is

$$h(\delta_1, \cdots, \delta_n) = n!\ I(0 < Z_1 < \cdots < Z_n < 1)$$

(ii) computation of γ

To compute γ, note that

$$\gamma = P(Z_j - Z_i \geq p) = \int_0^{1-p} \int_{p+z_i}^{1} h_{ij}(z_i,\ z_j)\ dz_j dz_i$$

where

$$h(Z_i, Z_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} Z_i^{i-1} (Z_j - Z_i)^{j-i-1} (1 - Z_j)^{n-j}$$

This computation is quite tedious. So, we use an alternative way to compute γ. Consider transforming

$$W_1 = Z_1, W_2 = Z_2 - Z_1, W_3 = Z_3 - Z_2, \cdots, W_n = Z_n - Z_{n-1}$$

Then $Z_i = \sum_{j=1}^{i} W_j$ and $|J| = 1$. Therefore, the jpdf of $W_1, \cdots, W_n$ is

$$k(w_1, \cdots, w_n) = n!,\ 0 < w_i,\ i = 1, \cdots, n,\ w_1 + \cdots + w_n < 1$$

Now, the jpdf of $W_1, \cdots, W_n$ is symmetric in $w_1, \cdots, w_n$, and hence, the pdf of $W_{i+1} + W_{i+2} + \cdots + W_j$ is the same as that of $W_1 + W_2 + \cdots + W_{j-i}$. Note that $W_{i+1} + \cdots + W_j = Z_j - Z_i$ and $W_1 + \cdots + W_{j-i} = Z_{j-i}$. Therefore,

$$\gamma = P(Z_j - Z_i \geq p) = P(Z_{j-i} \geq p)$$

where the pdf of Z_k is

$$h_k(v) = \frac{n!}{(k-1)!(n-k)!} v^{k-1}(1-v)^{n-k}$$

Hence,

$$\gamma = \int_p^1 h_{j-i}(v)dv$$

(iii) **Example 5.3.1**

Let $Y_1 < \cdots < Y_6$ be the order statistic from conti. distribution. Compute γ when we use (y_1, y_6) as a tolerance limit for 80% of the distribution.

$$\gamma = P(F(Y_6) - F(Y_1) \geq 0.8) = 1 - \int_0^{0.8} 30v^4(1-v)dv = 0.34$$

5.4 More on Confidence Intervals

① approximate C.I.

When $Z = (X-\mu)/\sigma \sim N(0,1)$, then $100(1-\alpha)\%$ C.I. for μ is $(X - Z_{\alpha/2}\sigma,\ X + Z_{\alpha/2}\sigma)$ if σ is known. When we known the asymptotic distribution only by CLT such as $\sqrt{n}(T-\theta) \to N(0,\sigma^2)$ then $(T - Z_{\alpha/2}\sigma/\sqrt{n},\ T + Z_{\alpha/2}\sigma/\sqrt{n})$ is called approximate $100(1-\alpha)\%$ C.I. for θ.

② examples

(i) C.I. for μ

$X_1, \cdots, X_n$: r.s. from a dist. with mean μ and σ^2 (both unknown)
Find approximate $100(1-\alpha)\%$ C.I. for μ.

(sol) $\sqrt{n}(\bar{X}-\mu)/s \xrightarrow{\mathscr{D}} N(0,1)$

$$\therefore \bar{x} \pm Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

(ii) C.I. for p

$X_1, \cdots, X_n$: r.s. from $B(1,p)$. By CLT, $\sqrt{n}(\hat{p}-p) \to N(0, p(1-p))$
also, $\sqrt{n}(\hat{p}-p)/\sqrt{\hat{p}(1-\hat{p})} \to N(0,1)$

$$\therefore \hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

(iii) C.I. for μ under normality

$X_1, \cdots, X_n$: r.s. from $N(\mu, \sigma^2)$. $\sqrt{n}(\bar{X}-\mu)/s \sim t(n-1)$

$$\therefore \bar{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

(iv) C.I. for $\mu_1 - \mu_2$

$X_1, \cdots, X_{n_1}$: r.s. from a dist. with mean μ_1, variance σ_1^2 and
$Y_1, \cdots, Y_{n_2}$: r.s. from a dist. with mean μ_2, variance σ_2^2 are indep.
Want to obtain approximate C.I. for $\mu_1 - \mu_2$.
Let $n = n_1 + n_2$, and assume $\frac{n_1}{n} \to \lambda_1$, $\frac{n_2}{n} = \lambda_2$. By CLT,
$\sqrt{n_1}(\bar{X} - \mu)/\sigma_1 \to N(0,1)$.

$$\therefore \sqrt{n}(\bar{X} - \mu)/\sigma_1 = \sqrt{\frac{n}{n_1}}\sqrt{n_1}(\bar{X} - \mu)/\sigma_1$$
$$\to \sqrt{\frac{1}{\lambda_1}}N(0,1) = N(0,1/\lambda_1).$$

Similarly, $\sqrt{n}(\bar{Y} - \mu_2)/\sigma_2 \to N(0,1/\lambda_2)$.

$$\therefore \sqrt{n}[(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)] \to N\left(0, \frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2}\right),$$

Also,

$$n\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right) \xrightarrow{P} \frac{\sigma_1^2}{\lambda_1} + \frac{\sigma_2^2}{\lambda_2}$$

$$\therefore \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \to N(0,1)$$

$$\Rightarrow (\bar{X} - \bar{Y}) \pm Z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

(v) C.I. for $\mu_1 - \mu_2$ under normality

$X_1, \cdots, X_{n_1}$: r.s. from $N(\mu_1, \sigma^2)$ and
$Y_1, \cdots, Y_{n_2}$: r.s. from $N(\mu_2, \sigma^2)$ are indep.(common variances)
$\bar{X} \sim N(\mu_1, \sigma^2/n_1)$, $\bar{Y} \sim N(\mu_2, \sigma^2/n_2)$

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$

Now, $(n_1-1)s_1^2/\sigma^2 \sim \chi^2(n_1-1)$ and $(n_2-1)s_2^2/\sigma^2 \sim \chi^2(n_2-1)$

$$\Rightarrow \frac{1}{\sigma^2}\{(n_1-1)s_1^2+(n_2-1)s_2^2\} \sim \chi^2(n_1+n_2-2)$$

$$\therefore T = \frac{Z}{\sqrt{V/(n_1+n_2-2)}} \sim t(n_1+n_2-2),\ s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$$

(vi) C.I. for $p_1 - p_2$

$X_1, \cdots, X_{n_1}$: r.s. from $B(1, p_1)$ and
$Y_1, \cdots, Y_{n_2}$: r.s. from $B(1, p_2)$ are indep.
By the same argument as in (iv),

$$(\hat{p}_1 - \hat{p}_2) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

5.5 Introduction to Hypothesis Testing

① definitions

(i) null and alternative hypothesis

$X_1, \cdots, X_n$: r.s. from a dist. with pdf $f(x:\theta), \theta \in \Omega$.
statistical hypothesis: $H_0 : \theta \in \omega_0$ vs $H_1 : \theta \in \omega_1$, $\omega_0 \cup \omega_1 = \Omega$, $\omega_0 \cap \omega_1 = \phi$

$$H_0 : \text{null hypothesis, } H_1 : \text{alternative hypothesis}$$

(ii) two type of errors

test statistic: a statistic $T = T(X_1, \cdots, X_n)$ for testing H_0 vs H_1

rejection region(critical region): a set C where H_0 is rejected

type I error: an error caused by rejecting H_0 even when H_0 is true

type II error: an error caused by accepting H_0 when H_1 is true

a critical region C is of size α if $\alpha = max_{\theta \in \omega_0} P[(X_1, \cdots, X_n) \in C]$

a power of a test is $P_\theta[(X_1, \cdots, X_n) \in C]$, $\theta \in \omega_1$, i.e. power is $1 - P_\theta[\text{type II error}]$, $\theta \in \omega_1$

the power function of a critical region C is

$$\gamma_c(\theta) = P_\theta[(X_1, \cdots, X_n) \in C], \theta \in \omega_1$$

② examples

(i) test for p

$X_1, \cdots, X_n$: r.s. from $B(1,p)$. Want to make a test of size α for testing $H_0: p = p_0$ vs $H_1: p < p_0$.
test statistic: $S = \sum_{i=1}^n X_i$: # of successes
rejection region: reject H_0 if $S \le k$ s.t. $\alpha = P_{H_0}(S \le k)$
assume $n = 20$, $P_0 = 0.7$, $\alpha = 0.15$, $P_{H_0}(S \le 11) = .1133$, $P_{H_0}(S \le 12) = .2277$
Hence, a test of size .15 rejects H_0 if $S \le 11$.
Compare power function for $S \le 11$ and $S \le 12$.(Fig. 5.5.1)

(ii) large sample test for μ

$X_1, \cdots, X_n$: r.s. from a dist. with mean μ, variance σ^2.
test of size α for testing $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$
test stat: $\bar{X}$, rejection region : $\bar{X} \ge k$
By using $\sqrt{n}(\bar{X} - \mu_0)/s \to N(0,1)$ under H_0, reject H_0 if $\sqrt{n}(\bar{X} - \mu_0)/s \ge Z_\alpha$
Now, compute approximate power function

$$\begin{aligned}\gamma(\mu) &= P_\mu(Z \ge Z_\alpha) = P_\mu(\bar{X} \ge \mu_0 + Z_\alpha s/\sqrt{n}) \\ &= P_\mu\left(\frac{\bar{X} - \mu}{s/\sqrt{n}} \ge \frac{\mu_0 - \mu}{s/\sqrt{n}} + Z_\alpha\right) \simeq 1 - \Phi\left(\frac{\mu_0 - \mu}{s/\sqrt{n}} + Z_\alpha\right)\end{aligned}$$

(iii) test for μ under normality

$X_1, \cdots, X_n$: r.s. from $N(\mu, \sigma^2)$
test of size α for testing $H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$
test stat.: $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} > t_\alpha(n-1)$

5.6 Additional Comments about Statistical Tests

(i) Large sample two-sided test for μ

$X_1, \cdots, X_n$: r.s. from a dist with mean μ, variance σ^2
test of size α for testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$
Intuitively, we reject H_0 if $\bar{X} \leq h$ or $\bar{X} \geq k$ s.t.
$\alpha = P_{H_0}(\bar{X} \leq h \text{ or } \bar{X} \geq k) = P_{H_0}(\bar{X} \leq h) + P_{H_0}(\bar{X} \geq k)$
Now, it is reasonable to set $P_{H_0}(\bar{X} \leq h) = \alpha/2$ and $P_{H_0}(\bar{X} \geq k) = \alpha/2$
i.e. Reject H_0 if $\left|\dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}\right| \geq Z_{\alpha/2}$

(ii) Randomized test

$X_1, \cdots, X_{10}$ r.s. from $P(\theta)$.
test of size $\alpha = .05$ for testing $H_0 : \theta = 0.1$ vs $H_1 : \theta > 0.1$
test stat: $Y = \sum_{i=1}^{10} X_i$, critical region: $Y \geq k$.
Note that $Y \sim P(1)$, therefore $P(Y \geq 3) = .080$, $P(Y \geq 4) = .019$
Hence, size $\alpha = .05$ test is rejecting H_0 if $Y \geq 4$
This test is called a non-randomized test. To achieve $.05 = P_{H_0}(\text{reject } H_0)$, we need a randomized test.
Let W be a Bernoulli trial with prob. of succes
$P(W = 1) = \dfrac{0.050 - 0.019}{0.080 - 0.019} = \dfrac{31}{61}$, and let the rejection region be $\{\sum_{i=1}^{10} X_i \geq 4\}$ or $\{\sum X_i = 3 \text{ and } W = 1\}$, then

$$P(Y \geq 4) + P(Y = 3)\frac{31}{61} = .019 + (.080 - .019)\frac{.050 - .019}{.080 - .019} = .050$$

(iii) p-value(observed significance level)

$Y = u(X_1, \cdots, X_n)$: test stat.
rejection region: $Y \leq c$
If the observed test stat. is d, then $P_{H_0}(Y \leq d)$ is called the p-value of d. In general, p-value is defined as the minimum of prob. of type I error to reject H_0 for a given value of test stat.

5.7 Chi-square Tests

3 types of chi-square test:
goodness-of-fit(GOF) test, homogeneity test, independence test

① goodness-of-fit test

(i) derivation

$X_1 \sim B(n, p_1)$, $X_2 = n - X_1$, $p_2 = 1 - p_1$

$$\begin{aligned} Q_1 &= \frac{(X_1 - np_1)^2}{np_1(1-p_1)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)} \\ &= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} \xrightarrow{\mathscr{D}} \chi^2(1) \end{aligned}$$

In general, let $\boldsymbol{X} = (X_1, \cdots, X_{k-1})' \sim \mathscr{M}(n, p_1, \cdots, p_{k-1})$ and $X_k = n - (X_1 + \cdots + X_{k-1})$, $p_k = 1 - (p_1 + \cdots + p_{k-1})$, then

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \xrightarrow{\mathscr{D}} \chi^2(k-1)$$

(ii) **Example.5.7.1**

Want to test a die is fair by tossing 60 times.
$H_0 : p_{10} = \cdots = p_{60} = \frac{1}{6}$, p_{i0} =prob. of obtaining face i
data: $13, 19, 11, 8, 5, 4 \therefore np_i = 60 \cdot \frac{1}{6} = 10$

$$Q_5 = \frac{(13-10)^2}{10} + \cdots + \frac{(4-10)^2}{10} = 15.6 > \chi^2_{.05}(5) = 11.1$$

(iii) computation of degree of freedom

$H_0 : p_1 = p_{10}, \cdots, p_k = p_{k0}$
Where, p_{i0} is not completely specified, for example

$$p_i = \int_{A_i} \frac{1}{\sqrt{2\pi}\sigma} exp[-(y-\mu)^2/2\sigma^2]dy,\ i = 1, \cdots, k$$

In this case, $Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \xrightarrow{\mathscr{D}} \chi^2(k-1-2)$
i.e. 2 d.f. are lost to estimate μ and σ^2

② homogeneity test

Consider two indep. multinomial dist.
$\boldsymbol{X}_1 = (X_{11}, \cdots, X_{k1}) \sim \mathscr{M}(n_1, p_{11}, \cdots, p_{k1})$,
$\boldsymbol{X}_2 = (X_{12}, \cdots, X_{k2}) \sim \mathscr{M}(n_2, p_{12}, \cdots, p_{k2})$
$H_0 : p_{11} = p_{12}, \cdots, p_{k1} = p_{k2}$
test stat: $Q = \sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(X_{ij} - n_j \hat{p}_{ij})^2}{n_j \hat{p}_{ij}},\ \hat{p}_{ij} = \frac{X_{i1} + X_{i2}}{n_1 + n_2},\ {}^{\forall} j = 1, 2$
d.f.: $(k-1) + (k-1) - (k-1) = k-1$
In general, for the $r \times c$ contingency table, d.f. is $(r-1) \times (c-1)$

③ independence test

Consider two categorical variales A and B. A has a categories $A_1, \cdots, A_a$ and B has b categories $B_1, \cdots, B_b$.
Let $P_{ij} = P(A_i \cap B_j),\ i = 1, \cdots, a,\ j = 1, \cdots, b$.
H_0: two variables are indep.
test stat.: $Q = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - n\hat{p}_{ij})^2}{n\hat{p}_{ij}},\ n = \sum_j \sum_i X_{ij}$
$\hat{p}_{ij} = \hat{p}_{i.}\hat{p}_{.j} = \frac{X_{i.}}{n}\frac{X_{.j}}{n},\ X_{i.} = \sum_j X_{ij},\ X_{.j} = \sum_i X_{ij}$
d.f.: $(ab-1) - \{(a-1) + (b-1)\} = ab - a - b + 1 = (a-1)(b-1)$

5.8 The Method of Monte Carlo

① random number generation

(i) **Thm.5.8.1**

$U \sim u(0,1)$, F: conti. d.f. $\Rightarrow\ X = F^{-1}(U) \sim F$

(pf) $P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$

(ii) Ex.(generation of $\mathscr{E}(1)$, i.e. $F(x) = 1 - e^{-x}$)

$\therefore F^{-1}(u) = -\log(1-u)$, $0 < u < 1$, then $X = -\log(1-U) \sim \mathscr{E}(1)$

(iii) Ex.(estimation of π)

U_1, U_2: i.i.d. $u(0,1)$

$$X = \begin{cases} 1, & U_1^2 + U_2^2 < 1 \\ 0, & O.W. \end{cases}$$

$$\therefore E(X) = \pi/4 \Rightarrow \pi = 4E(X)$$

$$\therefore \text{ Monte Carlo estimation of } \pi \text{ is } \hat{\pi} = 4 \cdot \frac{1}{n}\sum_{i=1}^{n} X_i$$

② Monte Carlo integration

(i) Monte Carlo integration

$$\int_a^b g(x)dx = (b-a)\int_a^b g(x)\frac{1}{b-a}dx = (b-a)E[g(X)],\ X \sim u(a,b)$$

Therefore, the Monte Carlo integration of $g(x)$ is

$$\int_a^b g(x)dx \simeq (b-a)\frac{1}{n}\sum_{i=1}^{n} g(X_i),\ X_i \sim u(a,b)$$

(ii) estimation of π by Monte Carlo integration

Let $g(x) = 4\sqrt{1-x^2}$, $0 < x < 1$,
then $\pi = \int_0^1 g(x)dx = E[g(X)]$, $X \sim u(0,1)$

$$\therefore \hat{\pi} \simeq \frac{1}{n}\sum_{i=1}^{n} g(X_i)$$

③ Box-Muller transformation

Y_1, Y_2: i.i.d. $u(0,1)$, $X_1 = (-2\log y_1)^{1/2}\cos(2\pi y_2)$, $X_2 = (-2\log y_1)^{1/2}\sin(2\pi y_2)$
$\therefore y_1 = exp[-(x_1^2+x_2^2)/2]$, $y_2 = \frac{1}{2}\arctan(\frac{x_2}{x_1})$

$$J = \begin{vmatrix} (-x_1)exp[-(x_1^2+x_2^2)/2] & (-x_2)exp[-(x_1^2+x_2^2)/2] \\ \frac{-x_2/x_1^2}{(2\pi)(1+x_2^2/x_1^2)} & \frac{1/x_1}{(2\pi)(1+x_2^2/x_1^2)} \end{vmatrix} = \frac{-1}{2\pi}exp\left[-\frac{x_1^2+x_2^2}{2}\right]$$

$\therefore X_1$, X_2: indep $N(0,1)$

④ accept-reject generation algorithm

(i) algorithm: Y: r.v. with pdf $g(y)$. $U \sim u(0,1)$, Y, U: indep.
$f(x)$: pdf s.t. $f(x)/g(x) \leq M$. Then, the following alforithm generate r.v. X with pdf $f(x)$.
(1) generate Y and U
(2) If $U \leq f(y)/Mg(y)$, then take $X = Y$. Else return to (1)

(pf)

$$
\begin{aligned}
P(X \le x) &= P[Y \le x | U \le f(y)/Mg(y)] \\
&= \frac{P[Y \le x, U \le f(y)/Mg(y)]}{P(U \le f(y)/Mg(y)]} \\
&= \frac{\int_{-\infty}^{x} \{\int_0^{f(y)/Mg(y)} du\} g(y) dy}{\int_{-\infty}^{\infty} \{\int_0^{f(y)/Mg(y)} du\} g(y) dy} \\
&= \int_{-\infty}^{x} f(y) dy
\end{aligned}
$$

(ii) example: $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $g(x) = \pi^{-1}(1+x^2)^{-1}$

easy to generate since its inverse cdf is known.
$\frac{f(x)}{g(x)} = \sqrt{\frac{\pi}{2}} e^{-x^2/2}(1+x^2)$ is maximized at $x = \pm 1$

$$\therefore M = 1.52$$

5.9 Bootstrap Procedures

① definition

$X_1, \cdots, X_n$: r.s. from a dist. with pdf $f(x : \theta), \theta \in \Omega$.
$\hat{\theta} = \hat{\theta}(\boldsymbol{X}), \boldsymbol{X} = (X_1, \cdots, X_n)$
$\boldsymbol{X}^* = (X_1^*, \cdots, X_n^*)$ is called bootstrap sample if X_j^*, $j = 1, \cdots, n$ is drawn with replacement from $(X_1, \cdots, X_n)$, i.e. X_i is selected with prob. $1/n$.

② percentile bootstrap confidence interval

$\hat{\theta}_j^* = \hat{\theta}(\boldsymbol{X}_j^*)$, $\boldsymbol{X}_j^*$: j-th bootstrap sample, $j = 1, \cdots, B$, where B is bootstrap size which is usually larger than 3000.
$\hat{\theta}_{(1)}^* \leq \hat{\theta}_{(2)}^* \leq \cdots \leq \hat{\theta}_{(B)}^*$: order stat. for $\hat{\theta}_1^*, \cdots, \hat{\theta}_B^*$.
Then $100(1-\alpha)\%$ percentile bootstrap C.I. for θ is $(\hat{\theta}_{(m)}^*, \hat{\theta}_{(B+1-m)}^*)$, where $m = \left[\frac{\alpha}{2}B\right]$.

③ bootstrap testing

$X_1, \cdots, X_{n_1}$: r.s. from a dist. with cdf $F(x)$
$Y_1, \cdots, Y_{n_2}$: r.s. from a dist. with cdf $F(x - \Delta)$
$H_0 : \Delta = 0$ vs $H_1 : \Delta > 0$
$\boldsymbol{X}^* = (x_1^*, \cdots, x_{n_1}^*)$, $\boldsymbol{Y}^* = (y_1^*, \cdots, y_{n_2}^*)$: bootstrap sample.
Want to obtain bootstrap p-value.
Let $\bar{x} = \frac{1}{n_1}\sum x_i$, $\bar{y} = \frac{1}{n_2}\sum y_i$, and $\bar{x}_j^*$, $\bar{y}_j^*$ be sample mean of the j-th bootstrap samples, then the p-value of $H_0 : \Delta = 0$ is

$$\frac{1}{B}\sum_{j=1}^{B} I(\bar{y}_j^* - \bar{x}_j^* \geq \bar{y} - \bar{x})$$