Mathematical Statistics II

Review of Mathematical Statistics I (Ch1-Ch5.2)

Jungsoon Choi

jungsoonchoi@hanyang.ac.kr

Table of Contents

Ch1. Probability

1.1 Basic Concept

Sample Space

The set, $\mathcal{S}$, of all possible outcomes of a particular experiment is called the *sample space* for the experiment.

Event

An *event* is a part of the collection of possible outcomes of an experiment, that is, any subset of $\mathcal{S}$ (including $\mathcal{S}$ itself).

Random Variable

A *random variable* is a function from a sample space $\mathcal{S}$ into the real numbers.

1.2 Disjoint/Partition

Disjoint

Two events A and B are *disjoint* (or mutually exclusive) if $A \cap B = \varnothing$. The events $A_1, A_2, \ldots, A_k$ are *pairwise disjoint* (or mutually exclusive) if $A_i \cap A_j = \varnothing$ for all $i \neq j$.

Partition

If $A_1, A_2, \ldots, A_k$ are pairwise disjoint and $\cup_{i=1}^{k} A_i = \mathcal{S}$, then the collection $A_1, A_2, \ldots, A_k$ forms a *partition* of $\mathcal{S}$.

1.3 Probability

Definition 1

Let $\mathcal{S}$ be a finite sample space. Assume that all the outcomes in $\mathcal{S}$ are equally likely. Suppose that $\mathcal{S} = \{s_1, \cdots, s_N\}$ and $P(\{s_i\}) = 1/N$.

$$P(A) = \sum_{s_i \in A} P(\{s_i\}) = \sum_{s_i \in A} \frac{1}{N} = \frac{\#\text{ of elements in } A}{\#\text{ of elements in } S}$$

Definition 2 (Axioms of Probability)

Let A be an event in the sample space $\mathcal{S}$. Probability is a real-valued set function P from sigma algebra (Borel field; collection of subsets of $\mathcal{S}$) to $[0, 1]$, satisfying

- $P(A) \geq 0$
- $P(\mathcal{S}) = 1$
- If $A_1, A_2, \ldots$ are pairwise disjoint events, then
$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

1.4 Conditional Probability

Conditional Probability

If A and B are events in $\mathcal{S}$, and $P(B) > 0$, then the *conditional probability of A given B*, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

1.5 Independent event

Independent event

Two events, A and B, are (statistically) independent if

$$P(A \cap B) = P(A)P(B).$$

Otherwise, A and B are dependent.

If A and B are independent events,

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

1.6 Bayes' Theorem

Bayes' Theorem

Let $A_1, A_2, \ldots, A_m$ be a partition of the sample space $\mathcal{S}$, and let B be any set. Then, for each $i = 1, 2, \ldots, m$

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{m} P(A_j \cap B)} \\
&= \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{m} P(B|A_j)P(A_j)}
\end{aligned}
$$

Ch2 & 3. Discrete/Continuous Distributions

2.1 Probability mass function (pmf)

Probability mass function (pmf)

The *probability mass function (pmf)* $f_X(x)$ of a discrete random variable X is is a function defined by

$$f_X(x) = P(X = x)$$

and it satisfies the following properties:

(a) $f_X(x) > 0, \quad x \in \mathcal{X},$

(b) $\displaystyle\sum_{x \in \mathcal{X}} f_X(x) = 1,$

(c) $P(X \in \mathbf{A}) = \displaystyle\sum_{x \in \mathbf{A}} f_X(x)$, where $\mathbf{A} \subset \mathcal{X}$.

2.2 Probability density function (pdf)

Probability density function (pdf)

The *probability density function (pdf)* of a continuous random variable X, with space $\mathcal{X}$ that is an interval or union of intervals, is an integrable function $f_X(x)$ satisfying the following conditions:

- $f_X(x) \geq 0, \quad x \in \mathcal{X}$
- $\displaystyle\int_{\mathcal{X}} f_X(x)dx = 1$
- If $(a, b) \subseteq \mathcal{X}$,

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

2.3 (Cumulative) Distribution Function

Definition

The *cumulative distribution function (cdf)* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P(X \leq x), \text{ for } x \in \mathbb{R}.$$

- discrete r.v. X: $F_X(x) = \sum_{t \leq x} f_X(t)$
- continuous r.v. X: $F_X(x) = \int_{-\infty}^{x} f(t)dt$

2.4 Expectation

Expectation

- discrete r.v. X:

$$E[g(X)] = \sum_{x \in \mathcal{X}} g(x) f_X(x)$$

- continuous r.v. X:

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

Thm.

Let X be a r.v. and a, b, and c be constants. For any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

- $E(c) = c$
- $E[ag_1(X)] = aE[g_1(X)]$
- $E[ag_1(X) + bg_2(X)] = aE[g_1(X)] + bE[g_1(X)]$

2.5 Variance and Standard Deviation

Definition

- variance:

$$\sigma_X^2 = Var(X) = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2, \quad \mu_X = E(X)$$

- standard deviation (sd):

$$\sigma_X = \sqrt{\sigma_X^2}$$

2.6 Moment

Definition

For each positive integer r, the rth *moment* of X, μ'_r is

$$\mu'_r = EX^r.$$

The rth *central moment* of X, μ_r, is

$$\mu_r = E(X - \mu_X)^r,$$

where $\mu_X = \mu'_1 = EX$.

2.7 Moment generating function (mgf)

Definition of mgf

The *moment generating function (mgf)* of r.v X (or F_X), denoted by $M_X(t)$, is

$$M_X(t) = E_X[e^{tX}] = \sum_{x \in \mathcal{X}} e^{tx} f_X(x), \quad \text{if } X \text{ is discrete,}$$
$$= \int_{-\infty}^{\infty} e^{tx} f_X(x)dx, \quad \text{if } X \text{ is continuous,}$$

if the expectation exists for t in some neighborhood of 0.

* Mgf is unique and completely determines the distribution of X.

** For each positive integer r, $E(X^r) = M_X^{(r)}(0)$.

r^{th} moment.

$X \sim f_X(x)$.

└ pmf or pdf.

$F_X(x) = P(X \leq x)$

$M_X(t) = E[e^{tX}]$

└ mgf.

2.8 Discrete Distributions

1) Bernoulli distribution

- $X \sim \text{Ber}(p)$, p: success probability
- *pmf*

$$f(x) = p^x(1-p)^{1-x}, \quad x = 0, 1; \quad 0 \leq p \leq 1$$

- Mean: $E(X) = p$
- Variance: $\mathrm{Var}(X) = p(1-p)$

2) Binomial distribution, $B(n, p)$

- Let r.v X be the total number of successes in n Bernoulli trials with the success probability p.
- $X = \sum_{i=1}^{n} X_i,\ X_i \sim \mathrm{Ber}(p)$.
- *pmf*

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

- Mean: $E(X) = np$
- Variance: $\mathrm{Var}(X) = np(1-p)$
- mgf: $M(t) = \left\{(1-p) + pe^t\right\}^n$

3) Negative Binomial Distribution $\mathrm{NB}(r, p)$

- Let r.v X be the number of failures before the rth success with the success probability p.
- *pmf*

$$f(x) = \binom{r+x-1}{x} p^r (1-p)^x, \quad x = 0, 1, \cdots; \quad 0 \leq p \leq 1$$

- Mean: $E(X) = \dfrac{r(1-p)}{p}$
- Variance: $\mathrm{Var}(X) = \dfrac{r(1-p)}{p^2}$
- mgf: $M(t) = \left(\dfrac{p}{1-(1-p)e^t}\right)^r, \quad t < -\log(1-p)$

4) Geometric Distribution $Geo(p)$

- Let r.v X be the number of trials until the first success with the success probability p.
- *pmf*

$$f(x) = p(1-p)^{x-1}, \quad x = 1, 2, \cdots,; \quad 0 < p < 1$$

- Mean: $E(X) = \dfrac{1}{p}$
- Variance: $\mathrm{Var}(X) = \dfrac{1-p}{p^2}$
- mgf: $M(t) = \dfrac{pe^t}{1-(1-p)e^t}, \quad t < -log(1-p)$

5) Poisson distribution, $\mathrm{Poi}(\lambda)$

- Let r.v X be the number of occurrences in a given time interval.
- *pmf*

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \cdots; \quad 0 \le \lambda < \infty$$

- Mean: $E(X) = \lambda$
- Variance: $\mathrm{Var}(X) = \lambda$
- mgf: $M(t) = \exp\left\{\lambda(e^t - 1)\right\}$

2.9 Continuous Distributions

1) Uniform distribution, $\mathrm{Unif}(a, b)$

- *pdf*

$$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

- Mean: E(X)= $\dfrac{a+b}{2}$
- Variance:Var(X)= $\dfrac{(b-a)^2}{12}$
- mgf:

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

2) Exponential distribution, $Exp(\beta)$

- Let r.v X be the waiting time until the first event.
- *pdf*

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 \leq x < \infty, \; 0 < \beta.$$

- Mean: E(X)= β
- Variance:Var(X)= β^2
- mgf:

$$M(t) = \int_0^\infty \frac{e^{tx}}{\beta} \exp\left(-\frac{x}{\beta}\right) dx = \frac{1}{1 - \beta t}, \quad t < \frac{1}{\beta}.$$

3) Gamma distribution, $\mathrm{Gamma}(\alpha, \beta)$

- Let r.v. X be the waiting time until the αth event occurs.
- *pdf*

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 \leq x < \infty, 0 < \alpha, \beta,$$

- Mean: E(X)$= \alpha\beta$
- Variance:Var(X)$= \alpha\beta^2$
- mgf:

$$M(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}, \quad t < \frac{1}{\beta}.$$

- $Exp(\beta) \equiv Gamma(1, \beta)$
- $\chi^2(r) \equiv Gamma(r/2, 2)$

4) Normal distribution, N(μ, σ^2)

- pdf:

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

- $X \sim N(\mu, \sigma^2)$
- $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$
- mgf:

$$M(t) = \exp\left(\mu t + \frac{\sigma^2}{2} t^2\right), \quad -\infty < t < \infty.$$

Standard Normal distribution, N(0, 1)

- pdf:

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

- $Z \sim N(0, 1)$.
- $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$
- $\Phi(z) = Pr(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$
- If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.
- If X is $N(\mu, \sigma^2)$, then $V = (X - \mu)^2/\sigma^2 = Z^2$ is $\chi^2(1)$.

Ch4. Bivariate Distributions

4.1 Joint probability mass function

Definition

Let X and Y be two discrete random variables. Let $\mathcal{X}$ denote the corresponding two-dimensional space of X and Y. The joint probability mass function *(joint pmf)* of X and Y is

$$f_{XY}(x, y) = P(X = x, Y = y)$$

Properties

1. $0 \leq f_{XY}(x, y) \leq 1$.
2. $\displaystyle\sum_x \sum_y f_{XY}(x, y) = 1$.
3. $\displaystyle P[(x, y) \in A] = \sum_x \sum_y f_{XY}(x, y)$, where $A \subset \mathcal{X}$.

4.2 Joint probability density function

Definition

Let X and Y be two continuous random variables. Then $f_{XY}(x,y)$ is called the joint probability density function *(joint pdf)* of X and Y, satisfying

1. $0 \leq f_{XY}(x,y)$.
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$.
3. $\displaystyle P[(x,y) \in A] = \int\int_A f_{XY}(x,y)\,dx$

4.3 Marginal distribution

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

Definition for discrete r.v (marginal pmf)

Let X and Y have joint probability mass function $f_{XY}(x, y)$ with space $\mathbb{R}^2$. The probability mass function of X alone, which is called the marginal probability mass function of X, is defined by

$$f_X(x) = \sum_y f_{XY}(x, y) = Pr(X = x), \quad x \in \mathbb{R}.$$

Similarly, the marginal probability mass function of Y is defined by

$$f_Y(y) = \sum_x f_{XY}(x, y) = Pr(Y = y), \quad y \in \mathbb{R}.$$

Definition for continuous r.v (marginal pdf)

Let X and Y have joint probability density function $f_{XY}(x, y)$ with space $\mathbb{R}^2$. The probability density function of X alone, which is called the marginal probability density function of X, is defined by

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy = f_X(x), \quad x \in \mathbb{R}.$$

Similarly, the marginal probability density function of Y is defined by

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx = f_Y(y), \quad y \in \mathbb{R}.$$

4.4 Expectation

Definition for joint distribution

Let X and Y be random variables and $g(X, Y)$ be a function of these two random variables. The expected value of the function $g(X, Y)$, $E[g(X, Y)]$, is defined by

$$E[g(X,Y)] = \begin{cases} \sum_x \sum_y g(x,y) f_{XY}(x,y) \\ \qquad\qquad \text{if X and Y are discrete,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y)\, dx\, dy \\ \qquad\qquad \text{if X and Y are continuous,} \end{cases}$$

if it exists.

4.5 Covariance and Correlation

Definition

- The covariance of X and Y is defined by

$$Cov(X,Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right] = E[XY] - \mu_X\mu_Y$$

- The correlation of X and Y is defined by

$$Corr(X,Y) = \rho_{XY} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

where ρ_{XY} is called the correlation coefficient.

4.6 Conditional Distributions

Definition

The conditional probability mass/density function of X, given that $Y = y$, is defined by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)},$$

provided that $f_Y(y) > 0$.

Similarly, the conditional probability mass/density function of Y, given that $X = x$, is defined by

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)},$$

provided that $f_X(x) > 0$.

4.7 Conditional Moments

Definition of conditional expectation

The conditional expectation of $g(X)$ given $Y = y$ is

$$E[g(X)|y] = \int_{-\infty}^{\infty} g(x)f(x|y)\, dx \quad \text{or} \quad \sum_x g(x)f(x|y).$$

Definition of conditional variance

The conditional variance of X given $Y = y$ is

$$Var[X|y] = E(X^2|y) - \{E(X|y)\}^2$$

Linear Conditional Mean

If $u(x) = E[Y|x] = a + bx$ is a linear function of x, then

$$a = \mu_Y - \rho\frac{\sigma_Y}{\sigma_X}\mu_X$$
$$b = \rho\frac{\sigma_Y}{\sigma_X}.$$

Let $Var[Y|x] = K(x)$. Then

$$E[K(x)] = \sigma_Y^2(1-\rho^2)$$

4.8 Independence

Definition

X and Y are called independent random variables if, for every $x \in \mathbb{R}$ and $y \in \mathbb{R}$,

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

Thm.

If X and Y are independent random variables, then $Cov(X, Y) = 0$ and $\rho_{XY} = 0$.

Ch5. Distributions of Functions of Random Variables

5.1.1 Functions of One Discrete Random Variable

Change of variable technique for discrete case

Let X be a discrete random variable with pmf $f_X(x)$, $x \in S_x = \{c_1, c_2, \ldots\}$. Let $Y = u(X)$ be a one-to-one transformation with inverse $X = v(Y)$ and $y \in S_y = \{u(c_1), u(c_2), \ldots\}$. The pmf of Y is

$$P(Y = y) = P[u(X) = y] = P[X = v(y)], \quad y \in S_y$$

5.1.2 Functions of One Continuous Random Variable

Concept

Let X be a continuous r.v. If we consider a function of X, $Y = u(X)$, Y has its own distribution:

$$F_Y(y) = P(Y \leq y) = P(u(X) \leq y)$$

Change of variable technique for continuous case

The pdf of $Y = u(X)$ is

$$f_Y(y) = f_X(v(y))|v'(y)|, \quad y \in S_y$$

where $X = v(Y)$ is the inverse function of u, and S_y is the support of Y found by mapping the support of X.

5.2 Transformations of Two Random Variables

Change of variable technique for discrete case

Let (X, Y) be a bivariate random vector with a joint distribution $f_{XY}(x, y)$. Consider a new bivariate random vector (U, V) defined by $U = g_1(X, Y)$ and $V = g_2(X, Y)$.

indep ($X_1 \sim Gamma(\alpha, \theta)$
$X_2 \sim$ " (β, θ)
$Y_1 = \frac{X_1}{X_1+X_2} \sim Beta$

- (X, Y): discrete bivariate random vector

$$f_{UV}(u, v) = P(U = u, V = v) = P((X, Y) \in A_{uv})$$
$$= \sum_{(x,y) \in A_{uv}} f_{XY}(x, y)$$

$\underbrace{U \sim \chi^2(r_1) \quad V \sim \chi^2(r_2)}_{\text{indep}}$.

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

where $A_{uv} = \{(x, y) \in A | g_1(x, y) = u, g_2(x, y) = v\}$

Change of variable technique for continuous case

The inverse functions of g_1 and g_2 functions are defined as $x = h_1(u, v)$ and $y = h_2(u, v)$. We assume the transformation is one-to-one.

- (X, Y): continuous bivariate random vector

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v))|J|$$

$$J = \begin{vmatrix} \frac{dx}{du} & \frac{dx}{dv} \\ \frac{dy}{du} & \frac{dy}{dv} \end{vmatrix} = \frac{dx}{du}\frac{dy}{dv} - \frac{dx}{dv}\frac{dy}{du}$$

where J is the determinant of a matrix of partial derivatives.