

Lecture 1: Review of Basic Probability

Xuejun Ma

School of Mathematical Sciences

Soochow University

<https://xuejunma.github.io>

August 15, 2021

We would like to thank Professor [Tian Maozai \(RUC\)](#) and Professor [Wang Wanjie \(NUS\)](#) for sending slides and beneficial discussions.

- 1 Outline
- 2 Random variable
- 3 Transformation etc.
- 4 Ditribution
- 5 Moment inequalities

Outline

- Syllabus
- Brief review of basic probability and statistics
 - Why is a random variable?
 - Transformations; independence; expectation
 - Important distributions
 - Some statistics

Terms

- Sample space; Measure; Random variable
- Transformation; Independence; Expectation; Conditional expectation; Variance retur Standard deviation; Moment Generating Function; Characteristic function
- Common distributions
- Sample mean; Sample variance; Sample distribution
- Moment inequalities

Sample space and Measure

What do we mean by randomness?

- We construct an experiment, yet the result of the experiment has many possibilities.
 - Flip a coin, the result can be either head or tail
- Although we can not know the result beforehand, we do have some information about the result.
 - Approximately, there is equal chance for a head and a tail
- Randomness: the **uncertainty** of experiment results

Question: How to describe our **information**?

Sample space and Measure

- Information 1. Possible outcomes

Definition 1 (Sample space (Outcome space))

Let Ω be a sample space, which is a set containing all possible outcomes.

- Information 2. Probabilities for these possible outcomes
 - σ – field \mathcal{F} : a set of subsets of Ω which satisfies 3 rules.
 - Measurable space: (Ω, \mathcal{F})
 - Event(measurable sets): element of \mathcal{F}
 - Probability measure P : for any element in the σ -field, assign it a probability, indicating the chance this event will happen
- $(\Omega; \mathcal{F}; P)$ (Probability space, measure space) is our information about the possible outcomes of this experiment. In short, we write it as the sample space Ω with probability P , or just Ω if there is no confusion.

Definition 2 (σ -field)

Let \mathcal{F} be a collection of subsets of a sample space. \mathcal{F} is called a σ -field (or σ -algebra) if and only if it has the following properties.

- The empty set $\phi \in \mathcal{F}$.
- If $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$
- If $A_i \in \mathcal{F}, i = 1, 2, \dots$, then their union $\cup A_i \in \mathcal{F}$.

- **Measurable space:** (Ω, \mathcal{F})
- Event (measurable sets): element of \mathcal{F}
- $\sigma(A) = \{\phi, A, A^c, \Omega\}$.
- Flip a coin, the result can be either head or tail
 $\Omega = \{H, T\}, \mathcal{F} = \{\dots\}$

Definition 3 (Measure)

Let Measurable space (Ω, \mathcal{F}) , A be a measurable space. A set function ν defined on \mathcal{F} is called a measure if and only if it has the following properties.

- $0 \leq \nu(A) \leq \infty$, for any $A \in \mathcal{F}$
- $\nu(\emptyset) = 0$
- If $A_i \in \mathcal{F}$, $i = 1, 2, \dots$, and A_i 's are disjoint, i.e. $A_i \cap A_j = \emptyset$ for any $i \neq j$, then $\nu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$
- measure space: $(\Omega, \mathcal{F}, \nu)$
- probability measure $\nu(\Omega) = 1$. We usually denote it by P instead of ν , (Ω, \mathcal{F}, P) .
- Flip a coin, the result can be either head or tail
 $\Omega = \{H, T\}$, $\mathcal{F} = \{\dots\}$
 - $\nu(A) = |A|$ the number of elements in $A \in \mathcal{F}$.
 - $P(A) = \frac{|A|}{|\Omega|}$

Random Variables

What is of interest?

- Manufacturers Ω : all the combinations of good light bulbs and defective light bulbs. Need: proportion of defective light bulbs from a lot
- Market researchers Ω : survey results of all consumers for one product. Need: preference of all consumers about this product, with a scale 1-10.

Our interest:

- Not the details of Ω , but a special measurable characteristic of the outcomes!
- A random variable, is a mapping from Ω to R , which draws the measurable characteristic of interest

Example: an opinion poll. 50 people; 1: agree; 0 disagree:

- Ω has 2^{50} elements.
- interest: the number of people who agree out of 50. X = number of 1s recorded out of 50. $\chi = \{0, 1, 2, \dots, 50\}$

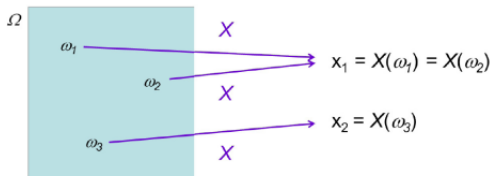
Random Variable

Definition 4 (Random Variable)

- Let (Ω, \mathcal{F}) and $(\mathcal{R}, \mathcal{B})$ (\mathcal{B} : Borel σ -field) be measurable spaces
- X is a function from Ω to \mathcal{R} . The function X is called a **random variable** (r.v.; measurable function) if and only if

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \subset \mathcal{F}$$

for any $B \in \mathcal{B}$.



[illegible]

- $$P(Y = y) = P(\{\omega \in \Omega : Y(\omega) = y\})$$

Example: Random variable

Example 1

- Consider the experiment of tossing a coin three times.
- H : Head; T : Tail.
- X : the number of heads obtained in the three tosses.

ω	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
$X(\omega)$	3	2	2	2	1	1	1	0

- $\chi = \{0, 1, 2, 3\}$. The induced probability function on χ is given by

x	0	1	2	3
$P_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$P_X(X = 1) = P(\{HTT, THT, TTH\}) = \frac{3}{8}$$

Cumulative Density Function

Definition 5 (Cumulative Density Function)

The cumulative distribution function (CDF) of a random variable is defined by

$$F(x) = P(X \leq x); -\infty < x < \infty$$

For all CDF's, there is

- $F(x)$ is right-continuous. At each x , $\lim_{n \rightarrow \infty} F(y_n) = F(x)$ for any sequence $y_n \rightarrow x$ with $y_n > x$.
- $F(x)$ is non-decreasing.
- $\lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow \infty} F(x) = 1$.

Any $F(x)$ satisfying Properties 1-3 is a CDF for some random variable.

Example: Logistic distribution.

Example 2

$$F_X(x) = \frac{1}{1 + e^{-x}}$$

- $\lim_{x \rightarrow -\infty} F_X(x) = 0$

- $\lim_{x \rightarrow \infty} F_X(x) = 1$

-

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1 + e^{-x})^2} > 0$$

Discrete v.s. Continuous r.v.

- If X is discrete, then its **probability mass function (pmf)** is


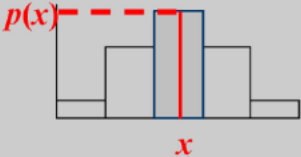

$$p_X(x) = p(x) = P(X = x)$$

- If X is continuous, then its **probability density function (pdf)** satisfies

$$P(X \in A) = \int_A f_X(x) dx = \int_A f(x) dx = \int_A dF(x)$$

and $f_X(x) = f(x) = F'(x)$.

- We say that X and Y have the same distribution (i.e. $X \stackrel{D}{=} Y$) if $P(X \in A) = P(Y \in A)$ for all A . $X \stackrel{D}{=} Y$ if and only if $F_X(t) = F_Y(t)$

RANDOM VARIABLE, X		
Type	Discrete	Continuous
Values	A finite/countable set of numbers x_1, x_2, x_3, \dots	All numbers in an interval 
Probability	Probability Mass Function, p <i>pmf</i> $P(X = x) = p(x)$ 	Probability Density Function, f <i>pdf</i> $P(a < X < b) = \left[\begin{array}{l} \text{area} \\ \text{under the} \\ \text{graph of } f \\ \text{over } (a, b) \end{array} \right]$ 

Example 3

Suppose X has a uniform distribution on the interval $(0, 2\pi)$, that is

$$f_X(x) = \begin{cases} 1/2\pi, & 0 < x < 2\pi, \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Consider $Y = \sin^2(X)$

$$\begin{aligned} P(Y \leq y) &= P(X \leq x_1) + P(x_2 \leq X \leq x_3) + P(X \geq x_4) \\ &= 2P(X \leq x_1) + 2P(x_2 \leq X \leq \pi) \end{aligned} \quad (3)$$

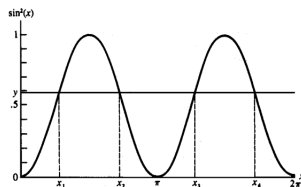


Figure 2.1.1. Graph of the transformation $y = \sin^2(x)$ of Example 2.1.2

Example 4

$X \sim f_X(x) = 1/(0 < x < 1)$, $F_X(x) = x$. $Y = g(x) = -\log x$, find its distribution.

Proof:

- $Y = g(x) = -\log x \Rightarrow x = e^{-y}, g^{-1}(y) = e^{-y}$
- g is decreasing function.

$$\frac{d}{dx}g(x) = \frac{d}{dx}(-\log x) = \frac{-1}{x} < 0, 0 < x < 1$$

•

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) = P_X(g(x) \leq y) \\ &= P_X(X \geq g^{-1}(y)) \\ &= 1 - P_X(X \leq g^{-1}(y)) \\ &= 1 - e^{-y} \end{aligned}$$

Example 5

Let

$$f_X(x) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}, 0 < x < \infty$$

be the Gamma pdf $Y = 1/X$. Find the pdf of Y

Proof. $g^{-1}(y) = 1/y$, $\mathcal{Y} = (0, \infty)$, $|\frac{d}{dy}g^{-1}(y)| = 1/y^2$. Therefore for all $y > 0$,

$$\begin{aligned} f_Y(y) &= f_X(g^{-1}(y)) \left| \frac{d}{dy}g^{-1}(y) \right| \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y}\right)^{n-1} e^{-1/(\beta y)} \frac{1}{y^2} \\ &= \frac{1}{(n-1)!\beta^n} \left(\frac{1}{y}\right)^{n+1} e^{-1/(\beta y)} \end{aligned} \quad (5)$$

- A special case of a pdf known as the inverted Gamma distribution.

Theorem 7

Let X have pdf $f_X(x)$, let $Y = g(X)$. Suppose there exists a partition A_0, A_1, \dots, A_k such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i .

$$P(X \in \cup_{i=1}^k A_i) = 1.$$

Futher, we have $g(\cdot)$ is monotone if restricted to $A_i, i = 1, 2, \dots, k$. Let

$$g_i^{-1}(y) = \{x \in A_i : g(x) = y\}$$

and assume $g_i^{-1}(y)$ has continuous derivative on \mathcal{Y} for each i . Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^k f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right|, & y \in \mathcal{Y}, \\ 0, & \text{otherwise} \end{cases}$$

- **Remark** Unfortunately, I found the above Theorem has very little practical use.

Example 6

Let $X \sim N(0, 1)$, $Y = X^2$, we may use the above theorem to find the pdf of Y .

Proof:

- $g(x) = x^2$ is monotone on $(-\infty, 0)$ and on $(0, \infty)$.
- $\mathcal{Y} = (0, \infty)$.

$$A_0 = \{0\}$$

$$A_1 = (-\infty, 0), g_1(x) = x^2, g_1^{-1}(y) = -\sqrt{y}$$

$$A_2 = (0, \infty), g_2(x) = x^2, g_2^{-1}(y) = \sqrt{y}$$

The pdf Y is

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \Phi(\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} + \Phi(-\sqrt{y}) \frac{1}{2} \frac{1}{\sqrt{y}} = \frac{1}{\sqrt{y}} \phi(\sqrt{y})$$

Probability integral transform

Theorem 8 (Probability integral transform)

Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then Y is uniformly distributed on $(0, 1)$, that is

$$P(Y \leq y) = y, 0 < y < 1.$$

- $F_X^{-1}(\tau) = \inf\{x : F_X(x) \geq \tau\}$
- Proof:

$$\begin{aligned} P_Y(Y \leq y) &= P_X(F_X(X) \leq y) \\ &= P_X(F_X^{-1}[F_X(X)] \leq F_X^{-1}(y)) \\ &= P_X(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) \\ &= y \end{aligned}$$

Theorem 9

Two r.v.'s X and Y are *independent* if and only if

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

for all A and B .

- $F(x, y) = F(x)F(y)$ for any x and y , $f(x, y) = f(x)f(y)$ or $p(x, y) = p(x)p(y)$
- When X and Y are independent, $h(X)$ and $g(Y)$ are also independent, if h and g are well-defined functions.

Expection

- Definition:

$$E(X) = \sum_x xp(x)$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

- Properties:

- $E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$
- $E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy$
- If X_1, \dots, X_n are independent, then

$$E\left(\prod_{i=1}^n X_i\right) = \prod_{i=1}^n E(X_i)$$

- Example 2.2.2

$X \sim \exp(\lambda),$

$$f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} x > 0.$$

Then $E[X] = \lambda.$

- Example 2.2.3

$X \sim \text{Binomial}(n, p),$

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad x = 0, 1, \dots.$$

Then $E[X] = np.$

- Example 2.2.4

$X \sim \text{Cauchy},$

$$f_X(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad -\infty < x < \infty.$$

Then $E[X]$ is not defined!(or do not exist).

• Mixed normal distribution

$$X = 0.5N(-1, 0.5^2) + 0.5N(1, 0.5^2)$$

_____ Mixed normal distribution _____

```
1 rm(list=ls())
2 n <- 1000
3 x <- rnorm(n, mean = ifelse(comp == 0, -1, 1),
4 sd = ifelse(comp == 0, 0.5, 0.5))
5 plot(density(x), main="")
```

- **Theorem 2.2.5** Let X be a r.v. and let a, b and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist.

$$(1) E[ag_1(X) + bg_2(X) + c] = aE[g_1(X)] + bE[g_2(X)] + c$$

$$(2) \text{ If } g_1(x) \geq 0 \text{ for all } x, \text{ then } E[g_1(X)] \geq 0.$$

$$(3) \text{ If } g_1(x) \geq g_2(x) \text{ for all } x, \text{ then } E[g_1(X)] \geq E[g_2(X)]$$

$$(4) \text{ If } a \leq g_1(x) \leq b \text{ for all } x, \text{ then } a \leq E[g_1(x)] \leq b$$

- **Example 2.2.6**

$E(X)$ is the "center" of a distribution (or its r.v.) in the sense that

$$\min_b E(X - b)^2 = E[X - EX]^2.$$

- Homework

$$\min_b E_{\rho_\tau}(X - b)$$

Remark: $\rho_\tau(t) = \tau t I(t \geq 0) + (\tau - 1) t I(t \leq 0).$

Variance & Standard Deviation

- Motivation: Describe the "spread" of r.v.
- Definition. $Var(x) = E[(x - \mu)^2]$, where $\mu = E(X)$, $sd(X) = \sqrt{Var(x)}$.
- Properties.
 - $Var(X) = E(X^2) - [E(X)]^2$
 - If X_1, \dots, X_n are independent, then

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

- The covariance is

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

and the **correlation coefficient** is

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

- For any two r.v.s with variance existed,

$$Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$$

Conditional Expectation

- Conditional Expectation of X when Y is given as y is that
 - $E(X|Y = y) = \sum_x x p_{X|Y}(X|Y)$ for discrete r.v.
 - $E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|Y) dx$ for cont.r.v.
 - Interpretation: Note that $X|Y = y$ is a new r.v., $E(X|Y = y)$ is the expectation on this r.v.
- Law of Total Expectation

$$E[E(X|Y)] = E(X)$$

- Law of Total Variance

$$\text{Var}(X) = \text{Var}[E(X|Y)] + E[\text{Var}(X|Y)]$$

Theorem 10

If X and Y are any two r.vs, then

$$E(X) = E[E(X|Y)]$$

Proof:

$$\begin{aligned} EX &= \int \int xf(x, y) dx dy \\ &= \int \left[\int xf(x|y) dx \right] f_Y(y) dy \\ &= \int E(X|y) f_Y(y) dy = E[E(X|Y)] \end{aligned}$$

In general, the conditional expectation $E[X|Y]$ can be defined as r.v. $g(Y)$ such that

$$E[(X - g(Y))^2] = \inf_{\text{among all reasonable function } h} E[(X - h(Y))^2]$$

or $E[X|Y]$ is the function of Y which is "closest" to X in terms of mean square error.

Theorem 11

$Y \sim$ Number of eggs lay by a mother fish, and $X \sim$ Number of survivors(young fish). On the average, how many eggs will survive?

Then it is reasonable to assume

$$Y \sim \text{Poisson}(\lambda)$$

$$X|Y \sim \text{Binomial}(Y, p)$$

So,

$$\begin{aligned} EX &= E[E(X|Y)] \\ &= E(pY) \\ &= p\lambda \end{aligned}$$

Example 7

$$X|Y \sim \text{Binomial}(Y, p)$$

$$Y|\Lambda \sim \text{Poisson}(\Lambda)$$

$$\Lambda \sim \text{exponential}(\beta)$$

Proof:

$$\begin{aligned} E[X] &= E[E(X|Y)] \\ &= pE(Y) \\ &= pE[E(Y|\Lambda)] \\ &= pE[\Lambda] \\ &= p\beta. \end{aligned}$$

Theorem 12

For any two random variables X and Y .

$$\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}[E(X|Y)]$$

provided that the expectation exist.

Proof:

$$\begin{aligned} \text{Var}(X) &= E[X - E(X|Y) + E(X|Y) - EX]^2 \\ &= E\{[X - E(X|Y)]^2 + [E(X|Y) - EX]^2 \\ &\quad + 2[X - E(X|Y)][E(X|Y) - EX]\} \\ &= E[X - E(X|Y)]^2 + E[E(X|Y) - EX]^2 \\ &= E[\text{Var}(X|Y)] + \text{Var}[(EX|Y)] \end{aligned}$$

Moment Generating Function and Characteristic Function

- Moment Generating Function(MGF)
 - Definition: $M_X(t) = E(e^{tX})$: a function of t , not r.v.
 - If $Y = aX + b$, $M_Y(t) = e^{bt} M_X(at)$
 - If X and Y are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$
- Characteristic Function
 - Definition: $\phi_X(t) = E[e^{itX}]$: a function of t ; $i = \sqrt{-1}$.
 - Bounded: $|\phi(t)| \leq 1$
 - If X and Y are independent, then $\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$.

An example of two distribution functions but with the same moments.

Example 8

Consider the two pdfs given by

$$f_1(x) = \frac{1}{\sqrt{2\pi x}} e^{-(\log x)^2/2}, 0 \leq x < \infty,$$

$$f_2(x) = f_1(x)[1 + \sin(2\pi \log x)], 0 \leq x < \infty.$$

Then it can be shown if $X_1 \sim f_1(x)$,

$$E[X_1^r] = e^{r^2/2}, r = 0, 1, \dots$$

Now suppose that $X_2 \sim f_2(x)$, we have for $r = 0, 1, \dots$

$$\begin{aligned} E[X_2^r] &= \int_0^\infty x^r f_1(x)[1 + \sin(2\pi \log x)] dx \\ &= E[X_1^r] + \int_0^\infty x^r f_1(x) \sin(2\pi \log x) dx \end{aligned}$$

$$\begin{aligned}
& \int_0^\infty x^r \frac{1}{\sqrt{2\pi \log x}} dx \\
&= \int_{-\infty}^\infty e^{(y+r)r} \frac{1}{\sqrt{2\pi}} e^{-(y+r)^2/2} \sin(2\pi(y+r)) dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(y^2-r^2)} \sin(2\pi y) dy \cdot \cos(2\pi r) \\
&\quad + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty e^{-\frac{1}{2}(y^2-r^2)} \cos(2\pi y) dy \cdot \sin(2\pi r) \\
&= 0 \quad r = 0, 1, \dots
\end{aligned}$$

since $e^{-\frac{1}{2}(y^2-r^2)} \sin(2\pi y)$ is an odd function.

However, we have the following theorem.

Theorem 13

Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist.

(a) If F_X and F_Y have bounded support, then $F_X(u) = F_Y(u)$ for all u iff $EX^r = EY^r$ for all $r = 0, 1, 2, \dots$

(b) If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u .

Differentiating Under An Integral Sign

If a, b are finite and $f(x, \theta)$ is differential with respect to θ . Then we have

$$\frac{d}{d\theta} \int_a^b f(x, \theta) dx = \int_a^b \frac{\partial}{\partial \theta} f(x, \theta) dx$$

But in statistics, we often need to evaluate $\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x, \theta) dx$, which may or may not be $\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} f(x, \theta) dx$.

Theorem 14

Suppose the function $h(x, y)$ is continuous at y_0 for each x , and there exists a function $g(x)$ satisfying

a) $|h(x, y)| \leq g(x)$, for all x and y ;

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

Then

$$\lim_{y \rightarrow y_0} \int_{-\infty}^{\infty} h(x, y) dx = \int_{-\infty}^{\infty} \lim_{y \rightarrow y_0} h(x, y) dx$$

Apply the above Theorem to the differentiation case, then we have

- **Theorem 2.4.3** Suppose $f(x, \theta)$ is differentiable at $\theta = \theta_0$, and there exists a function $g(x, \theta_0)$ and a constant $\delta > 0$ such that

a) $\left| \frac{f(x, \theta_0 + \Delta) - f(x, \theta_0)}{\Delta} \right| \leq g(x, \theta_0)$, for all x and $|\Delta| \leq \delta$;

b) $\int_{-\infty}^{\infty} g(x, \theta_0) dx < \infty$. Then

$$\frac{d}{d\theta} \int_{-\infty}^{\infty} f(x, \theta) dx \Big|_{\theta=\theta_0} = \int_{-\infty}^{\infty} \left[\frac{\partial}{\partial \theta} f(x, \theta) \Big|_{\theta=\theta_0} \right] dx \quad (*)$$

- **Corollary** Suppose that there exists $\delta > 0$ and function $g(x, \theta)$ such that $\left| \frac{\partial}{\partial \theta} f(x, \theta) \Big|_{\theta=\theta'} \right| \leq g(x, \theta)$, for all θ' with $|\theta' - \theta| < \delta$, and $\int_{-\infty}^{\infty} g(x, \theta) dx < \infty$. Then $(*)$ holds.

- **Remark** Finding bound $g(x, \theta)$ is cumbersome. We need to know that differentiating under the integral sign is not always automatic. In most situations, we just do it!!
- **Example 2.4.6** $X \sim N(\mu, 1)$,

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

$$\frac{d}{dt} M_X(t) = \frac{d}{dt} E(e^{tX}) = E\left(\frac{\alpha}{\alpha t} e^{tX}\right) = E(X e^{tX}).$$

For the exchange of operation of **differentiation** and **summation**, we have

- **Theorem 2.4.8** Suppose that the series $\sum_{x=0}^{\infty} h(\theta, x)$ converges for all θ in an interval (a, b) and
 - $\frac{\alpha}{\alpha\theta} h(\theta, x)$ is continuous in θ for each x ;
 - $\sum_{x=0}^{\infty} \frac{\alpha}{\alpha\theta} h(\theta, x)$ converges uniformly on every closed bounded subinterval of (a, b) .

Then

$$\frac{d}{d\theta} \sum_{x=0}^{\infty} h(\theta, x) = \sum_{x=0}^{\infty} \frac{\alpha}{\alpha\theta} h(\theta, x)$$

- **Theorem 2.4.10** Suppose that the series $\sum_{x=0}^{\infty} h(\theta, x)$ converges uniformly on $[a, b]$ and that, for each x , $h(\theta, x)$ is a continuous function of θ . Then

$$\int_a^b \sum_{x=0}^{\infty} h(\theta, x) d\theta = \sum_{x=0}^{\infty} \int_a^b h(\theta, x) d\theta$$

Important Distribution

- Discrete distributions:

- Bernoulli r.v.: $X \sim \text{Bernoulli}(p), p(1) = p, p(0) = 1 - p, p(x) = 0$ if $x \neq 0$ and $x \neq 1$. It can be written as $p^x(1-p)^{1-x}$ for $x = 0, 1$.

- Binomial r.v.: $X \sim \text{Binomial}(n, p), p(x) = \binom{n}{x} p^x q^{n-x}, x = 0, 1, 2, \dots, n$. Summation of n Bernoulli random variables.

- Poisson r.v.: $X \sim \text{Pois}(\lambda), p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, x = 0, 1, 2, \dots, n$.

- Continuous distribution

- Uniform r.v.: $X \sim \text{Unif}(a, b), f(x) = \frac{1}{b-a}, x \in (a, b)$
- Exponential r.v.: $X \sim \text{Exp}(\lambda), f(x) = \lambda e^{-\lambda x}$
- Normal r.v.: $X \sim N(\mu, \sigma^2), f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Multivariate Normal Distribution

- The d random vector $X \sim N(\mu, \Sigma)$,

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right)$$

- $AX + b \sim N(A\mu + b, A\Sigma^{-1}A^T)$
- Conditional distribution.

$$X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

then

$$X_1 | Y_2 \sim N(\mu_1 + \Sigma_{11} \Sigma_{22}^{-1} (Y_2 - \mu_2), \Sigma_{11} - \sigma_{12} \Sigma_{22}^{-1} \Sigma_{21})$$

Stastic

- Sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
- Sample variance: $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$
- Sampling distribution of \bar{X}_n : $G_n(t) = P(\bar{X}_n \leq t)$

When it is normal:

- If $X \sim N(\mu, \Sigma^2)$, then \bar{X}_n and S_n^2 are independent,

$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Moment inequalities

Lemma 15

Let a and b be any two positive numbers, and let p and q be any positive numbers satisfying

$$\frac{1}{p} + \frac{1}{q} = 1$$

. Then

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab$$

with equality holds if and only if $a^p = b^q$.

- **Proof:** Consider for fixed b (or a),

$$g(a) = \frac{1}{p}a^p + \frac{1}{q}b^q - ab$$

with equality if and only if $a^p = b^q$.

Theorem 16

Let X and Y be any two random variables. Let p and q be any positive numbers satisfying

$$\frac{1}{p} + \frac{1}{q} = 1$$

Then

$$|E(XY)| \leq E|XY| \leq (E|X|^p)^{\frac{1}{p}} (E|Y|^q)^{\frac{1}{q}}$$

Proof: In the inequality (1), let

$$a = \frac{|X|}{(E|X|^p)^{\frac{1}{p}}}, b = \frac{|Y|}{(E|Y|^q)^{\frac{1}{q}}}$$

After some simplification, and take expectation on the two sides of the inequality. The result can be obtained.

- Theorem 4.7.3(Cauchy-Schwarz Inequality)

For any two random variables X and Y ,

$$|E(XY)| \leq E|XY| \leq (E|X|^2)^{\frac{1}{2}}(E|Y|^2)^{\frac{1}{2}}$$

- Example 4.7.4 (Covariance Inequality)

If X and Y have means μ_X and μ_Y , and variances σ_X^2 and σ_Y^2 , respectively. We can apply the Cauchy-Schwarz Inequality to get

$$(\text{Cov}(X, Y))^2 \leq \sigma_X^2 \sigma_Y^2$$

- **Example**

Let $p > 1$, then apply Holders Inequality. For any random variables X ,

$$E|X| \leq \{E|X|^p\}^{\frac{1}{p}}$$

If $1 < r < s$, we have (Liapounovs Inequality)

$$(E|X|^r)^{\frac{1}{r}} \leq (E|X|^s)^{\frac{1}{s}}$$

- **Proof** Let q be such that $\frac{1}{p} + \frac{1}{q} = 1$, then

$$E|X| = E|X| \cdot 1 \leq (E|X|^p)^{\frac{1}{p}} (E1^q)^{\frac{1}{q}} = (E|X|^p)^{\frac{1}{p}}$$

- **Proof** Let s be such that $s = pr$, then $s > 1$.

$$E(|X|^r) \leq (E|X|^s)^{\frac{1}{p}}$$

Theorem 17 (Minkowskis Inequality)

Let X and Y be any two random variables. Then for $1 < p < \infty$

$$[E|X + Y|^p]^{\frac{1}{p}} \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}$$

Proof:

$$\begin{aligned} E|X + Y|^p &= E(|X + Y||X + Y|^{p-1}) \\ &\leq E(|X||X + Y|^{p-1}) + E(|Y||X + Y|^{p-1}) \end{aligned} \quad (6)$$

Using Holder's Inequality,

$$E(|X||X + Y|^{p-1}) \leq (E|X|^p)^{\frac{1}{p}} [E|X + Y|^{q(p-1)}]^{\frac{1}{q}} \quad (7)$$

where q is such that $\frac{1}{p} + \frac{1}{q} = 1$ or $\frac{1}{q} = 1 - \frac{1}{p}$, i.e. $q = \frac{p}{p-1}$ or $q(p-1) = p$. Similarly,

$$E(|Y||X + Y|^{p-1}) \leq (E|Y|^p)^{\frac{1}{p}} [E|X + Y|^{q(p-1)}]^{\frac{1}{q}} \quad (8)$$

So combine (6) and (7) with (8), divide through by $[E(|X + Y|^{q(p-1)})]^{\frac{1}{q}}$, we have

$$E|X + Y|^p \leq (E|X + Y|^p)^{\frac{p-1}{p}} [(E|X|^p)^{\frac{1}{p}} + (E|Y|^p)^{\frac{1}{p}}]$$

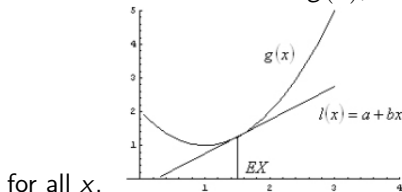
Theorem 18 (Jensens Inequality)

For any random variable X , if $g(x)$ is a convex function, then

$$Eg(X) \geq g(EX)$$

- Equality holds if and only if, for any line $a + bx$ that is tangent to $g(x)$ at $x = EX$, $P(g(X) = a + bX) = 1$.
- If $g(x)$ is linear, $g(EX) = a + bEX = Eg(X)$.

Remark For any twice differentiable function $g(x)$, it is convex if $g''(x) \geq 0$



for all x .

Example 9 (An inequality for means)

Let a_1, a_2, \dots, a_n be n non-negative numbers. Define

$$a_A = \frac{1}{n}(a_1 + a_2 + \dots + a_n)$$

$$a_G = [a_1 a_2 \dots a_n]^{1/n} a_H = \frac{1}{\frac{1}{n}(\frac{1}{a_1} + \dots + \frac{1}{a_n})}$$

An inequality relating these means is

$$a_H \leq a_G \leq a_A$$

Remark The above inequality gives a reason for Maximum Likelihood Estimation (MLE).

Proof: Let X be a random variable with range a_1, \dots, a_n , and $P(X = a_i) = 1/n$, $n = 1, \dots, n$. Since $\log x$ is a concave function, $E \log X \leq \log(EX)$, hence

$$\begin{aligned} \log a_G &= \frac{1}{n} \sum_{i=1}^n \log a_i = E \log X \leq \log(EX) \\ &= \log\left(\frac{1}{n} \sum_{i=1}^n a_i\right) = \log a_A \end{aligned}$$

So, $a_G \leq a_A$. Furthermore,

$$\begin{aligned} \log \frac{1}{a_H} &= \log\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{a_i}\right) = E \log \frac{1}{X} \geq E\left(\log \frac{1}{X}\right) = -\log(EX) \\ &= -\log a_G = \log\left(\frac{1}{a_G}\right). \end{aligned}$$

So, $a_G \geq a_H$.

Markovs Inequality

Theorem 19 (Markovs(Chebyshevs) Inequality)

- If g is strictly increasing and positive on $(0, \infty)$, $g(x) = g(-x)$.
- X is a r.v. such that $E[g(X)] < \infty$, then for each $a > 0$

$$P(|X| \geq a) \leq \frac{E[g(X)]}{g(a)}$$

Proof:

$$\begin{aligned} E[g(X)] &\geq E[g(X)I_{\{g(X) \geq g(a)\}}] \\ &\geq g(a)E[I_{\{g(X) \geq g(a)\}}] \\ &= g(a)E[I_{|X| \geq a}] \\ &= g(a)P(|X| \geq a) \end{aligned}$$

Some special cases: Markovs Inequality

$$g(x) = |x| \Rightarrow P(|X| \geq a) \leq \frac{E|X|}{a}$$

$$g(x) = x^p \Rightarrow P(|X| \geq a) \leq \frac{E|g(X^p)|}{a^p}$$

$$g(x) = x^2 \Rightarrow P(|X - EX| \geq a) \leq \frac{Var(X)}{a^2}$$

$$g(x) = e^{t|x|} \Rightarrow P(|X| \geq a) \leq \frac{E[e^{t|X|}]}{e^{ta}}$$

for some constant $t \geq 0$

Homework

- If $\mu = EX \geq 0$ and $0 \leq \lambda \leq 1$, then

$$P(X > \lambda\mu) \geq \frac{(1 - \lambda)^2 \mu^2}{EX^2}$$

Consequently, if $E|Y| = 1$, $P(|Y| > \lambda) \geq (1 - \lambda) \geq (1 - \lambda)^2 / EY^2$ (This gives a lower bound complementing Chebyshevs inequality.)



Thanks !