

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 & Standard deviation; Moment Generating Function; Characteristic function
- Common distributions
- Sample mean; Sample variance; Sample distribution
- Moment inequalities

- Information 1. Possible outcomes

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

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ◀ ↺ 🔍 ↻

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.

- 1 The empty set $\phi \in \mathcal{F}$
- 2 If $A \in \mathcal{F}$, then the complement $A^c \in \mathcal{F}$.
- 3 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}
- $\sigma(A) = \{\emptyset, A, A^c, \Omega\}$.
- Flip a coin, the result can be either head or tail $\Omega = \{H, T\}$,
 $\mathcal{F} = \{\dots\}$

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.

- 1 $0 \leq \nu(A) \leq \infty$, for any $A \in \mathcal{F}$
- 2 $\nu(\emptyset) = 0$
- 3 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 $\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i)$

- 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 $A \in (F)$.
 -

$$P(A) = \frac{|A|}{|\Omega|}$$

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. $\mathcal{X} = \{0, 1, 2, \dots, 50\}$

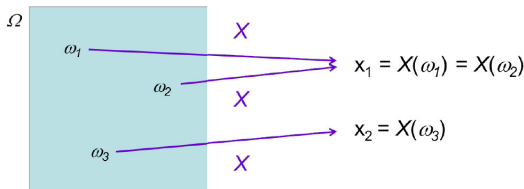
Random Variable

Random Variable

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

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

for any $B \subset R$.



Example: Random variable

Example 1.4.3

- 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

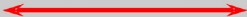
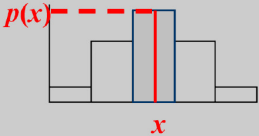

- $\mathcal{X} = \{0, 1, 2, 3\}$. The induced probability function on \mathcal{X} is given by

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

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

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

- 13 / 55

	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> $P(a < X < b) = \left[\begin{array}{c} \text{area} \\ \text{under the} \\ \text{graph of } f \\ \text{over } (a, b) \end{array} \right]$ 

- $$F_Y(y) = 1 - F_X(g^{-1}(y)).$$

17 / 55

Example 2.1.9

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 $(0, \infty)$.
- $\mathcal{Y} = (0, \infty)$.

$$A_0 = \{0\}$$

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

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

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 2.1.10 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

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

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

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

Independence

Theorem 4.2.10

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.

Expectation

- Definition:

$$\mathbf{E}(X) = \sum_x xp(x)$$

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

- Properties:

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

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

- Example 2.2.2

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

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

Then $\mathbf{E}[X] = \lambda$.

- Example 2.2.3

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

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

Then $\mathbf{E}[X] = np$.

- Example 2.2.4

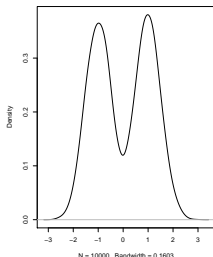
$$X \sim \text{Cauchy},$$

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

Then $\mathbf{E}[X]$ 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 <- 10000
3 comp <- sample(c(0, 1), size = n, prob = c(0.5, 0.5),
4               replace = T)
5 x <- rnorm(n, mean = ifelse(comp == 0, -1, 1),
6               sd = ifelse(comp == 0, 0.5, 0.5))
7 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.

(a) $\mathbf{E}[ag_1(X) + bg_2(X) + c] = a\mathbf{E}[g_1(X)] + b\mathbf{E}[g_2(X)] + c$

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

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

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

- **Example 2.2.6**

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

$$\min_b \mathbf{E}(X - b)^2 = \mathbf{E}[X - \mathbf{E}X]^2.$$

- **Homework**

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

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

Variance & Standard Deviation

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

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

- The covariance is

$$Cov(X, Y) = \mathbf{E}[(X - \mathbf{E}(X))(Y - \mathbf{E}(Y))] = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$$

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

$$\mathbf{E}[\mathbf{E}(X|Y)] = \mathbf{E}(X)$$

- Law of Total Variance

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

Theorem 4.4.3

If X and Y are any two r.v.s, then

$$\mathbf{E}(X) = \mathbf{E}[\mathbf{E}(X|Y)]$$

Proof:

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

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

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

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

Example 4.4.1 Hierarchical Model

$Y \sim$ 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

$$\begin{aligned} Y &\sim \text{Poisson}(\lambda) \\ X|Y &\sim \text{Binomial}(Y, p) \end{aligned}$$

So,

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

Example 4.4.5

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

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

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

Proof:

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

Theorem 4.4.7

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:

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

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 2.3.10

Consider the two pdfs given by

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

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

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

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

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

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

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

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

¹ $\sin(A+B) = \sin A \cos B + \sin B \cos A$

However, we have the following theorem.

Theorem 2.3.11

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 .

- **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
 - a) $\frac{\partial}{\partial \theta} h(\theta, x)$ is continuous in θ for each x ;
 - b) $\sum_{x=0}^{\infty} \frac{\partial}{\partial \theta} h(\theta, x)$ *converges uniformly on every closed bounded subinterval of (a, b) .*

Then

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

- **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 Distributions

- 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 also 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 distributions:

- 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.: $X \sim N(\mu, \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

- Theorem 4.7.3 (Cauchy-Schwarz Inequality)

For any two random variables X and Y ,

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

- **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

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

49 / 55

Some special cases: Markov's Inequality

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

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

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

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

for some constant $t > 0$



Thank you