Probability and Statistics

Luo Shan

School of Mathematical Sciences,
Shanghai Jiao Tong University

Tuesday, 12:55-15:40, 陈瑞球 216

About the course

- **Textbook**:
  1. **All of Statistics.** by Larry Wasserman, Springer.
- **Reference books**:
  1. **Engineering Statistics.** by Douglas C. Montgomery, George C. Runger, Norma Faris Hubele. Fifth Edition.
  2. **Probability and Statistics (For Engineers and Scientists).** by Anthony Hayter. Fourth Edition
- **Grading Policy**: Homework (30%), Term paper (30%), Final exam (40%)
- **Teaching Assistant**: Wu Jiaya (Phd student)

Chapter 0: Review of Probability Theory

Sample Spaces

- The sample space Ω is the set of possible outcomes of an experiment.
- Points ω in Ω are called sample outcomes or realizations.
- Events are subsets of Ω.

例 1

In the case of rolling a fair six-sided die,

- $\Omega = \{1, 2, 3, 4, 5, 6\}$, ω is each number on the die.
- Rolling an even number, then $A = \{2, 4, 6\}$.
- Rolling an odd number, then $B = \{1, 3, 5\}$.

Events

- Given an event A, let $A^c = \{\omega \in \Omega : \text{not } (\omega \in A)\}$ denote the complement of A.
- The union of events A and B is defined as $A \cup B = \{\omega \in \Omega : \omega \in A \text{ or } \omega \in B\}$.
- If $A_1, A_2, \ldots$ is a sequence of events, then

$$\cup_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for some } i\}.$$

- The intersection of A and B is $A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$.
- If $A_1, A_2, \ldots$ is a sequence of events, then

$$\cap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \text{ for all } i\}.$$

- Let $A - B = \{\omega \in \Omega : \omega \in A \text{ and not } (\omega \in B)\}$.
- If every element of A is contained in B we write $A \subset B$ or $B \supset A$.
- If A is a finite set, let $|A|$ denote the number of elements in A.

Events

- We say that $A_1, A_2, \ldots$ are disjoint or mutually exclusive if $A_i \cap A_j = \emptyset$ whenever $i \neq j$.
- A partition of Ω is a sequence of disjoint sets $A_1, A_2, \ldots$ such that $\cup_{i=1}^{\infty} A_i = \Omega$.
- Given an event A, define the indicator function of A by

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

Events

- A sequence of events $A_1, A_2, \ldots$ is monotone increasing if $A_1 \subset A_2 \subset \ldots$, and we define $\lim_{n\to\infty} A_n = \cup_{i=1}^{\infty} A_i$.
- A sequence of events $A_1, A_2, \ldots$ is monotone decreasing if $A_1 \supset A_2 \supset \ldots$ and then we define $\lim_{n\to\infty} A_n = \cap_{i=1}^{\infty} A_i$.
- In either case, we will write $A_n \to A$.

Basic Properties of Probability

A probability measure is a function P from each event A to the real numbers, satisfying

1. $0 \leq P(A) \leq 1$
2. $P(\emptyset) = 0$
3. $P(\Omega) = 1$
4. If $A_1, A_2 \ldots$ are disjoint, $P\left(\cup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k)$.

Elementary Theorems

- $P(A^c) = 1 - P(A)$
- If $A \subseteq B$ then $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (Inclusion-exclusion principle)

Elementary Theorems

- Law of total Probability: Let $A_1, A_2, \dots$ form a partition of the sample space Ω, and let B be any event. Then $P(B) = \sum_{k=1}^{\infty} P(A_k \cap B)$.
- Sub-additivity: Let $A_1, A_2, \dots$ be a collection of events, not necessarily disjoint. Then 、$P\left(\cup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k)$
- Continuity 1: Let $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ and let $A = \cup_{k=1}^{\infty} A_k$. Then $\lim_{k\to\infty} P(A_k) = P(A)$.
- Continuity 2: Let $A_1 \supseteq A_2 \supseteq A_3 \supseteq \dots$ and let $A = \cap_{k=1}^{\infty} A_k$. Then $\lim_{k\to\infty} P(A_k) = P(A)$.

Independent Events

- Two events A and B are independent if

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B).$$

  and we write $A \perp B$.
- A set of events $\{A_i : i \in I\}$ is independent if

$$\mathbb{P}\left(\cap_{i \in J} A_i\right) = \prod_{i \in J} \mathbb{P}(A_i)$$

  for every finite subset J of I.

Conditional Probability

- If $\mathbb{P}(B) > 0$ then the conditional probability of A given B is
$$\mathbb{P}(A|B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)}.$$
- $\mathbb{P}(\cdot|B)$ satisfies the axioms of probability, for fixed B. In general, $\mathbb{P}(A|\cdot)$ does not satisfies the axioms of probability for fixed A.
- In general, $\mathbb{P}(B|A) \neq \mathbb{P}(A|B)$.
- A and B are independent if and only if $\mathbb{P}(A|B) = \mathbb{P}(A)$.

Bayes' Theorem

定理 2 (The Law of Total Probability)

Let $A_1, \dots, A_k$ *be a partition of* Ω. *Then, for any event* B,

$$\mathbb{P}(B) = \sum_{i=1}^{k} \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

定理 3 (Bayes' Theorem)

Let $A_1, \dots, A_k$ *be a partition of* Ω *such that* $\mathbb{P}(A_i) > 0$ *for each* i. *If* $\mathbb{P}(B) > 0$, *then, for each* $i = 1, \dots, k$,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}.$$

We call $\mathbb{P}(A_i)$ *the prior probability of* A_i *and* $\mathbb{P}(A_i|B)$ *the posterior probability of* A_i.

Random Variables

- A random variable is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $X(\omega)$ to each outcome ω.
- Given a random variable X and a subset A of the real line, define $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$ and let

$$\begin{aligned} \mathbb{P}(X \in A) &= \mathbb{P}(X^{-1}(A)) = \mathbb{P}(\{\omega \in \Omega; X(\omega) \in A\}) \\ \mathbb{P}(X = x) &= \mathbb{P}(X^{-1}(x)) = \mathbb{P}(\{\omega \in \Omega; X(\omega) = x\}). \end{aligned}$$

Distribution Functions and Probability Functions

- The cumulative distribution function, CDF, $F_X : \mathbb{R} \to [0,1]$ of a random variable X is defined by

$$F_X(x) = \mathbb{P}(X \leq x).$$

- A function F mapping the real line to $[0,1]$ is a CDF for some probability measure $\mathbb{P}$ if and only if if satisfies the following three conditions:
  1. F is non-decreasing, i.e. $x_1 < x_2$ implies that $F(x_1) \leq F(x_2)$.
  2. F is normalized: $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$.
  3. F is right-continuous, i.e. $F(x) = F(x^+)$ for all x, where

$$F(x^+) = \lim_{y\to x, y>x} F(y).$$

Discrete Random Variables

X is discrete if it takes countably many values

$$\{x_1, x_2, ...\}.$$

We define the probability density function or probability mass function for X by

$$f_X(x) = \mathbb{P}(X = x).$$

The CDF of X is related to the PDF by

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f_X(x_i).$$

Continuous Random Variables

A random variable X is continuous if there exists a function f_X such that $f_X(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f_X(x)dx = 1$, and for every $a \leq b$,

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)dx.$$

The function f_X is called the probability density function. We have that

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

and $f_X(x) = F'_X(x)$ at all points x at which F_X is differentiable.

Quantile Function

Let X be a random variable with CDF F.

- The inverse CDF or quantile function is defined by

$$F^{-1}(q) = \inf\{x : F(x) \geq q\}$$

  for $q \in [0, 1]$. If F is strictly increasing and continuous then $F^{-1}(q)$ is the unique real number x such that $F(x) = q$.
- We call $F^{-1}(1/4)$ the first quartile, $F^{-1}(1/2)$ the median (or second quartile), and $F^{-1}(3/4)$ the third quartile.

Examples of Discrete Random Variables

- The Point Mass Distribution: $\mathbb{P}(X = a) = 1$.
- The Discrete Uniform Distribution on $\{1, \dots, k\}$:

$$f(x) = 1/k \text{ for } x = 1, \dots, k.$$

- The Bernoulli Distribution: $\mathbb{P}(X = 1) = p$ and $\mathbb{P}(X = 0) = 1 - p$ for some $p \in [0, 1]$.
- The Binomial Distribution:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, \dots, n.$$

- The Geometric Distribution Geom(p), if

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

- The Poisson Distribution:

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \geq 0$$

Examples of Continuous Random Variables

- The Uniform Distribution Uniform(a, b), if

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

  where $a < b$.

- Normal (Gaussian) $N(\mu, \sigma^2)$, if

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

  where $\mu \in \mathbb{R}$ and $\sigma > 0$.

- Exponential Distribution Exp(β), if

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x > 0$$

  where $\beta > 0$.

Examples of Continuous Random Variables (Cont.)

- t and Cauchy Distribution t_ν, if

$$f(x) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\left(1 + \frac{x^2}{\nu}\right)^{(\nu+1)/2}}.$$

  The Cauchy distribution is $t(1)$.

- The χ^2 Distribution χ_p^2, if

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

Independent Random Variables

Two random variables X and Y are independent if, for every A and B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

We write $X \perp Y$.

定理 4

Let X and Y have joint PDF $f_{X,Y}$. Then $X \perp Y$ if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all values x and y.

Homework 1: page 15: Exercise 13; page 16: Exercise 20; page 45: Exercise 9, Exercise 11.

例 5

Let X have probability density function,

$$f_X(x) = \begin{cases} 1/4 & \text{if } 0 < x < 1 \\ 3/8 & \text{if } 3 < x < 5 \\ 0 & \text{otherwise,} \end{cases}$$

(a) Find the cumulative distribution function of X.

(b) Let $Y = 1/X$. Find the probability density function $f_Y(y)$ for Y.

例 6

Let X be such that $\mathbb{P}(X = 2) = \mathbb{P}(X = 3) = 1/10$ and $\mathbb{P}(X = 5) = 8/10$. Plot the CDF F. Use F to find $\mathbb{P}(2 < X \leq 4.8)$ and $\mathbb{P}(2 \leq X \leq 4.8)$.

例 7

Let X and Y be independent and suppose that each has a Uniform$(0, 1)$ distribution. Let $Z = \min\{X, Y\}$. Find the density $f_Z(z)$ for Z.

Expectation of a Random Variable

The expected value, mean or first moment of X is defined to be

$$\mathbb{E}(X) = \int x \, dF(x) = \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int x f(x) \, dx & \text{if } X \text{ is continuous} \end{cases}$$

assuming that the sum (or integral) is well-defined.

定理 8 (The rule of the lazy statistician)

Let $Y = r(X)$, *then*

$$\mathbb{E}(Y) = \mathbb{E}(r(X)) = \int r(x) \, dF_X(x).$$

Variance and Covariance

- Let X be a random variable with mean μ. The variance of X is defined by

$$\mathbb{V}(X) = \sigma^2 = \mathbb{E}(X-\mu)^2 = \int (x-\mu)^2 \, dF(x),$$

assuming this expectation exists. $\sigma = \sqrt{\mathbb{V}(X)}$ is the standard deviation of X.

- Let X and Y be random variables with means μ_X and μ_Y and standard deviation σ_X and σ_Y. Define the covariance between X and Y by,

$$\mathsf{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)],$$

and the correlation by,

$$\rho = \rho_{X,Y} = \rho(X,Y) = \frac{\mathsf{Cov}(X,Y)}{\sigma_X \sigma_Y}.$$

例 9

Let $X_1, \dots, X_n \overset{i.i.d}{\sim}$ Uniform$(0,1)$ (i.i.d stands for "independent and identically distributed") and let $Y_n = \max\{X_1, \dots, X_n\}$. Find $\mathbb{E}(Y_n)$.

例 10

A particle starts at the origin of the real line and moves along the line in jumps of one unit. For each jump the probability is p that the particle will move one unit to the left and the probability is $1-p$ that the particle will jump one unit to the right. Let X_n be the position of the particle after n units. Find $\mathbb{E}(X_n)$ and $\mathbb{V}(X_n)$. (This is known as a random walk.)

例 11

Let $X \sim N(0,1)$ and let $Y = e^X$. Find $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

Types of convergence

- X_n converges to X in probability, written $X_n \xrightarrow{\text{P}} X$, if, for every $\epsilon > 0$,
$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$
as $n \to \infty$.
- X_n converges to X in distribution, written $X_n \rightsquigarrow X$, if
$$\lim_{n\to\infty} F_n(t) = F(t)$$
for all t for which F is continuous.
- X_n converges to X in quadratic mean, written $X_n \xrightarrow{\text{qm}} X$, if,
$$\mathbb{E}(X_n - X)^2 \to 0$$
as $n \to \infty$.

The Law of Large Numbers

定理 12 (The Weak Law of Large Numbers (WLLN))

If $X_1, X_2, \ldots, X_n$ *are i.i.d, then* $\overline{X}_n \xrightarrow{P} \mu$.

定理 13 (The Central Limit Theorem)

Let $X_1, X_2, \ldots, X_n$ *be i.i.d with mean* μ *and variance* σ^2. *Let* $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$. *Then*

$$Z_n \equiv \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \rightsquigarrow Z$$

where $Z \sim N(0,1)$. *In other words,*

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq z) = \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx.$$

The Delta Method

Suppose that

$$\frac{\sqrt{n}(Y_n - \mu)}{\sigma} \rightsquigarrow N(0,1)$$

and that g is a differentiable function such that $g'(u) \neq 0$. Then

$$\frac{\sqrt{n}(g(Y_n) - g(u))}{|g'(u)|\sigma} \rightsquigarrow N(0,1).$$

例 14

Let $X_1, \ldots, X_n$ be i.i.d with finite mean $\mu = \mathbb{E}(X_i)$ and finite variance $\sigma^2 = \mathbb{V}(X_i)$. Let $\overline{X}_n$ be the sample mean and let S_n^2 be the sample variance. i.e,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2.$$

Show that $\mathbb{E}(S_n^2) = \sigma^2$.

例 15

Suppose that the height of men has mean 68 inches and standard deviation 4 inches. We draw 100 men at random. Find (approximately) the probability that the average height of men in our sample will be at least 68 inches.

例 16

Let $X_1, \dots, X_n \overset{i.i.d}{\sim}$ Uniform$(0,1)$. Let $Y_n = \overline{X}_n^2$. Find the limiting distribution of Y_n.

Homework 2: Page 46: Exercise 18, Exercise 21; Page 60: Exercise 14; Page 83: Exercise 8.