Chapter 4

Review of Probability Theory

4.1 Probability

In real world, there are lots of events occur randomly. In statistics, such events are called experiments. The set of all possible outcomes is the sample space corresponding to an experiment. The sample space is denoted by S. An event is a subset of the sample space S. For example, Φ, S.

Basic axiom of probability:

1. $0 \leq P(A) \leq 1$;
2. $P(\Phi) = 0, P(S) = 1$;
3. $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ for disjoints event A_i's.

Example 4.1.1. *Flip a coin once and record the side faces up.*

Example 4.1.2. *Flip a coin three times and record every time the side faces up.*

Example 4.1.3. *Mary went to a shopping mall. The time she spent there.*

4.1.1 Conditional Probability

Screening test is a commonly seen test in biomedical studies. Patients or subjects receive screening tests to pre-diagnose whether she/he has the disease of interest. Suppose

1. D = person has disease of interest;
2. N = person does not have disease;
3. T + = person gives positive test response;
4. T = person gives negative test response.

Example 4.1.4. *Random sample 1000 people known to have diabetes, and 1000 known to not have diabetes: Question: what is* $P(N|T-)$ *and* $P(T|N)$*?*

Test Result	Present (D)	Absent (N)	Total
T+	950	10	960
T-	50	990	1040
Total	1000	1000	2000

Definition 4.1.1 (Conditional probability)**.** *Let* A *and* B *be two events with* $P(B) \neq 0$*. The conditional probability of* A *given* B *is defined to be*

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

Definition 4.1.2 (Multiplication law)**.** *Let* A *and* B *be two events and* $P(B) \neq 0$*, then*

$$P(A \cap B) = P(A|B)P(B). \tag{4.1}$$

Definition 4.1.3 (Independence)**.** A *and* B *are said to be independent if* $P(A \cap B) = P(A)P(B)$*.*

4.1.2 Bayes Theorem

Theorem 4.1.1 (Bayes theorem)**.** *Let $B_1, \cdots, B_n$ be a partition of S. Then for any event A,*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)}. \tag{4.2}$$

Example 4.1.5. *A company manufactures MP3 players at two factories. Factory I produces 60% of the MP3 players and Factory II produces 40%. Two percent of the MP3 players produced at Factory I are defective, while 1% of Factory II's are defective. An MP3 player is selected at random and found to be defective. What is the probability it came from Factory I?*

Example 4.1.6. *Due to inaccuracies in drug testing procedures in the medical field, the results of a drug test represent only one factor in a physician's diagnosis. Yet, when Olympic athelets are tested for illegal drug use (i.e. doping), the results of a single test are used to ban the athelete from competitions.*

In a population of 1,000 atheletes, suppose 10 are illegally using testosterone. Of the users, suppose 95% would test positive for testosterone. Of the nonusers, suppose 5% would test positive.

Given that the drug test yield a positive result, what is the probability that the athelete doesn't use testosterone.

4.1.3 What is statistical inference?

Bayes theorem gives the fundamental in statistical inference. We consider H be the hypothesis/model and D be the data. When we do an experiment, we typically measure or determine $P(D|H)$, which is the probability of observing the data when H is true. For example, if we are testing whether the means of two populations are the same, then we often take the null that the means are the same and calculate the probability that we observe the data. This is a probability model.

What is statistical inference? Give the data, what is the chance that the model is true. Bayes theorem gives us a partial result:

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}.$$

From the observable data, draw conclusions about the population model.

4.2 Random variables

4.2.1 Discrete r.v.

Probability mass function.

Bernoulli distribution

Toss an unfair coin with the probability p of getting a head. Define X be the count of heads. Then X follows a Bernoulli(p) distribution, denoted as $X \sim Ber(p)$.

$$P(X = 0) = 1 - p, P(X = 1) = p.$$

Binomial distribution

Toss the previous unfair coin n times, let X be the total number of heads. Then X follows a Binomial distribution, denoted as $X \sim Bin(n, p)$.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \cdots, n.$$

Poisson distribution

Consider the following examples,

- the number of calls to a suicide hotline on a given 24 hour period;
- the number of cells in a culture that exhibit a genetic mutation;
- the number of deaths during heatwaves in the US.

Let X be the number of independent events per unit time that occur for some rate λ. $X \sim Pois(\lambda)$, then

$$P(X = k) = \frac{\lambda^k}{k!} \exp(-\lambda), k = 0, 1, \cdots, \infty.$$

4.2.2 Continuous r.v.

Probability density function and cumulative distribution function.

Uniform distribution

Let $X \sim U(a, b)$. Then the probability density function on [a,b] is given as $f(x) = \frac{1}{b-a}$, for $a \leq x \leq b$; and $f(x) = 0$ otherwise.

Normal distribution Let $X \sim N(\mu, \sigma^2)$, a normal distribution with mean μ and variance σ^2. The pdf of $N(\mu, \sigma^2)$ is given as

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{(y-\mu)^2}{2\sigma^2}\}.$$

Normal pdf is a bell-shaped curve over the entire real line.

Exponential distribution

Let $X \sim Exp(\lambda)$. Then the pdf and cdf of X are

$$f(x) = \frac{1}{\lambda} \exp(-x/\lambda), x > 0,$$

and

$$F(x) = (1 - \exp(-x/\lambda)), x > 0.$$

Exponential distributions are skewed distributions on the positive real line. $1/\lambda$ is sometimes thought of as an instantaneous failure rate, or a "hazard". Exponential distribution is sometimes used to characterize failure time, e.g. the function time of a light bulb.

Gamma distribution

Let $X \sim \Gamma(\alpha, \lambda)$ where the pdf is given as

$$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-x/\lambda), x > 0.$$

Here α is called the shape parameter and λ is called the scale parameter. Some times, we use $\beta = \frac{1}{\lambda}$, the rate parameter.

If $X \sim \Gamma(\alpha, \lambda)$, then $Y = \frac{1}{X}$ follows a distribution, named as the inverse gamma distribution. This is widely used as the prior distribution of the variance.

4.2.3 Features of distributions

Expectations

Expectation is the weighted mean of all possible values of a r.v.

Definition 4.2.1 (Expectation). *For a discrete r.v.,* $EX = \sum_x xP(X = x)$*; for a continuous r.v., then* $EX = \int_x x f_X(x)dx$*.*

Example.

1. $X \sim Bin(n, p)$, then $EX = np$;
2. $X \sim Poisson(\lambda)$, then $EX = \lambda$;
3. $X \sim Exp(\lambda)$, then $EX = \frac{1}{\lambda}$;
4. $X \sim N(\mu, \sigma^2)$, then $EX = \mu$.

Properties:

1. $E(aX + bY) = aEX + bEY$;
2. $E(XY) = EXEY$ if X and Y are independent.

Example 4.2.1. *Suppose in a population of interest, men takes up 52% of the population. The height of men follows* $N(\mu = 174, \sigma^2 = 25)$*, and the height of women follows* $N(\mu = 162, \sigma^2 = 36)$*. What would be the expectation of the height in the population?*

Other central tendency parameters

Definition 4.2.2 (Median). *For a continuous r.v.* X*, the median* ν *is defined as a value such that*

$$P(X \geq \nu) = P(X \leq \nu) = \frac{1}{2}.$$

Definition 4.2.3 (Mode). *For a r.v.* X*, the population mode* γ *is defined as*

$$\gamma = \inf\{x : f_X(t) \leq f_X(x), \forall t\}.$$

Variance

Variance is a measure of spread of the distribution in the population.

Definition 4.2.4 (Variance).

$$Var(X) = E(X - EX)^2.$$

Properties

1. $V(X) = EX^2 - (EX)^2$;
2. $V(aX + b) = a^2V(X)$;
3. If Y_i are independent of each other, then $V(\sum Y_i) = \sum V(Y_i)$;
4. $V(X+Y) = V(X)+V(Y)+2Cov(X,Y)$ where the covariance $Cov(X,Y)$ is defined as

$$Cov(X,Y) = E(X - EX)(Y - EY) = E(XY) - EXEY.$$

It provides a description of the precision of a statistic. For two unbiased estimates (EX = , the parameter of interest) of the same parameter, a statistic with lower variance is more interesting than the one with a higher variance. In addition, sometime variance itself is of interest since it characterize the population.

Why are we interested in variance? It provides a description of the precision of a statistic. For two unbiased estimates ($EX = \theta$, the parameter of interest) of the same parameter, a statistic with lower variance is more interesting than the one with a higher variance. In addition, sometime variance itself is of interest since it characterize the population.