AMS 572 Data Analysis I
Review of Probability

Pei-Fen Kuan

Applied Math and Stats, Stony Brook University

Course Info

- Instructor: Pei-Fen Kuan, Ph.D, Office: Room 1-106 Math Tower, Phone: (631) 632-1419 (peifen.kuan@stonybrook.edu)
- Grader: Yicong Zhu (yicong.zhu@stonybrook.edu)

Course Info

Office Hours

- Pei-Fen Kuan: Thu 1:00-3:00 PM Math Tower 1-106
- Yicong Zhu: Mon/Wed 10:00-11:00 AM TBA Harriman 132

Course materials will be posted on Blackboard

Course Info

Textbook

- Statistics and Data Analysis, by Tamhane and Dunlop, Pearson
- Applied Statistics and SAS Programming Language, by Jeffrey K. Smith, Pearson

We will cover Chapters 5, 6, 7, 8, 9, 10, 12, 14 (part) (Tamhane and Dunlop) in lectures. Chapters 11, 13, 14, 15 will be covered in group projects. Chapters 1-4 will be covered briefly in class (Please read these chapters in details on your own).

Course Info

- The computer lab is located at Math Tower S-235 (math lab sinc site).
- You can also use SAS program from Virtual Sinc Site. f you already have the SAS Software installed, but need a new license file, you can download the license file from SoftWeb.
- You are also recommended to learn more SAS from its on-line resources:
  `http://support.sas.com/documentation/onlinedoc/stat/`
- We will also demonstrate using R programming for data analysis in lectures.

Course Objectives

To introduce students to basic statistical procedures. Survey of elementary statistical procedures such as the t-test and chi-square test. Procedures to verify that assumptions are satisfied. Extensions of simple procedures to more complex situations and introduction to one-way analysis of variance. Basic exploratory data analysis procedures.

Homework

- Homework Assignments: Given regularly.
- Homework Policy: You may discuss problems with other students in this class, but you must write up your HW completely on your own; if you do work with other student(s), you must declare this on the cover of your own HW, giving names of your collaborators. (Your writings must be independent: Do not look at another write up, either of classmate or of anything you found on internet when writing your own solution. To do otherwise is a case of Academic Dishonesty and is subject to University policy through CASA.)

Grading

- There will be two exams (Mid-term and final). The exams are CLOSED book and CLOSED notes.
- There is a group project which includes a presentation. The presentation will be held in the last 3 weeks of the semester (More information will be provided later).
- The homework will count for 10% of the grade, the project will count for 20% of the grade, midterm (30%) and final exam (40%) of the grade.

Exam dates

- Mid-term: October 26, 2017 (Thu) in class
- Final Exam: December 13, 2017 5:30-8:00 PM (Wed) (University final examination schedule)

Review of Probability

Discrete random variable

A random variable that can take at most a countable number of possible values is said to be discrete. The space of discrete random variable contains at most a countable number of points. If the space of a random variable contains an interval, the random variable is called continuous random variable.

For a discrete random variable X, we define the probability mass function (p.m.f. or prob distribution function, p.d.f.) $p(x)$ of X by

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

Cumulative distribution function

The cumulative distribution function (cdf) F of the random variable X is defined by $F(x) = P(X \leq x)$ for $-\infty < x < \infty$. For a discrete random variable X:

$$F(X) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

It is a non decreasing step function. That is, if the possible values of X are $x_1, x_2, \ldots$ where $x_1 < x_2 < \ldots$, the F is constant in the intervals (x_{i-1}, x_i) and then takes a step (or jump) of size $p(x_i)$ at x_i.

Definition

If X is a discrete random variable having p.d.f. $p(x)$, the expectation or the expected value of X, $E(X)$ is defined by

$$E(X) = \sum_i x_i p(x_i)$$

Also known as mean μ : weighted average of the possible values of X.

Properties:

1. $E(c) = c$, c : constant.
2. $E(cU(X)] = cE(U(X))$.
3. $E[c_1U_1(X) + c_2U_2(X)] = c_1E(U_1(X)) + c_2E(U_2(X))$.

Definition

If X is a random variable with mean μ, then the variance of X, $\text{Var}(X)$ is defined by

$$\text{Var}(X) = E[(X-\mu)^2] = \sum_i (x_i - \mu)^2 p(x_i)$$

which is the mean squared deviation with respect to μ.

Properties

1. $\text{Var}(c) = 0$, if c is constant.
2. $\text{Var}(aX + b) = a^2\text{Var}(X)$, a, b constant.
3. $\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$.

Definition

The standard deviation of X is

$$\text{sd}(X) = \sqrt{\text{Var}(X)}$$

Binomial Distribution

Binomial experiment is one that possesses the following properties

1. A Bernoulli (success-failure) trial is performed n times.
2. The trials are independent
3. The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of failure is $(1-p)=q$.
4. The random variable of interest is X : the number of successes observed during the n trials.

X is called the Binomial random variable. The p.d.f. of X is called Binomial distribution and denoted as

$$X \sim Bin(n, p)$$

The p.d.f. of X is

$$P(X = x) = p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \ldots, n$

Properties

1. $P(X = x) = p(x)$ is a p.d.f..
2. $E(X) = np$.
3. $\text{Var}(X) = np(1 - p)$.
4. When x goes from 0 to n, $p(x)$ first increases monotonically then decreases monotonically reaching its largest value when x is the largest integer less than or equal to $(n + 1)p$.

Example: There are three coins in a bag. When coin 1 is flipped, it lands on head with probability 0.3. When coin 2 is flipped, it lands on head with probability 0.8. When coin 3 is flipped, it lands on head with probability 0.6. One of these coins is randomly chosen and flipped 8 times.

(a) What is the probability that the coin lands on head exactly 3 out of the 8 flips?

(b) Given that the last 3 of these 8 flips lands on head, what is the conditional probability that exactly 6 out of the 8 flips lands on head?

Poisson Distribution

A random variable X taking on one of the values $0, 1, 2, \ldots$ is said to be a Poisson random variable with parameter λ if for some $\lambda > 0$, the p.d.f. of X is

$$P(X = x) = p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, 2, \ldots$$

X is said to have a Poisson distribution with parameter λ and is denoted as $X \sim \text{Poisson}(\lambda)$ or $X \sim Po(\lambda)$ or $X \sim P(\lambda)$.

Properties of $X \sim P(\lambda)$.

1. $E(X) = \lambda$.
2. $\text{Var}(X) = \lambda$.
   Note that $E(X) = \text{Var}(X) = \lambda$ for a Poisson distribution.
3. If $X \sim P(\lambda)$, then $P(X = x)$ increases monotonically and then decreases monotonically as x increases, reaching maximum when x is the largest integer not exceeding λ.

Example: Ten calls enter a college switchboard on the average of 2 every 3 minutes. What is the probability of 5 or more calls arriving in a 9 minute period.

Continuous Random Variable

Random variable whose set of possible values is an interval of union of intervals is said to be a continuous random variable (or a random variable of the continuous type).

The p.d.f. $f(x)$ is used to describe probability of events concerning the continuous random variable and satisfies the following conditions:

1. $f(x) \geq 0$.
2. $\int_R f(x)\,\mathrm{d}x = 1$, where R is the space of X.

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

3. The probability of event A is $\int_A f(x)\,\mathrm{d}x = P(X \in A)$

The c.d.f. of random variable X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

Expectation of X is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

Expectation of $U(X)$ is

$$\mu = E[U(X)] = \int_{-\infty}^{\infty} U(x) f(x)\,\mathrm{d}x$$

Moment generating function

Definition:

The moment generating function $M(t)$ of random variable of X is defined by

$$M(t) = E(e^{tX})$$

for $t \in \mathbb{R}$.

$M(t)$ gets its name because all the moments of X can be obtained by successfully differentially $M(t)$ and then evaluating the result at $t = 0$.

Moment generating function of X:

1.
2.
3.

C. Normal Distribution

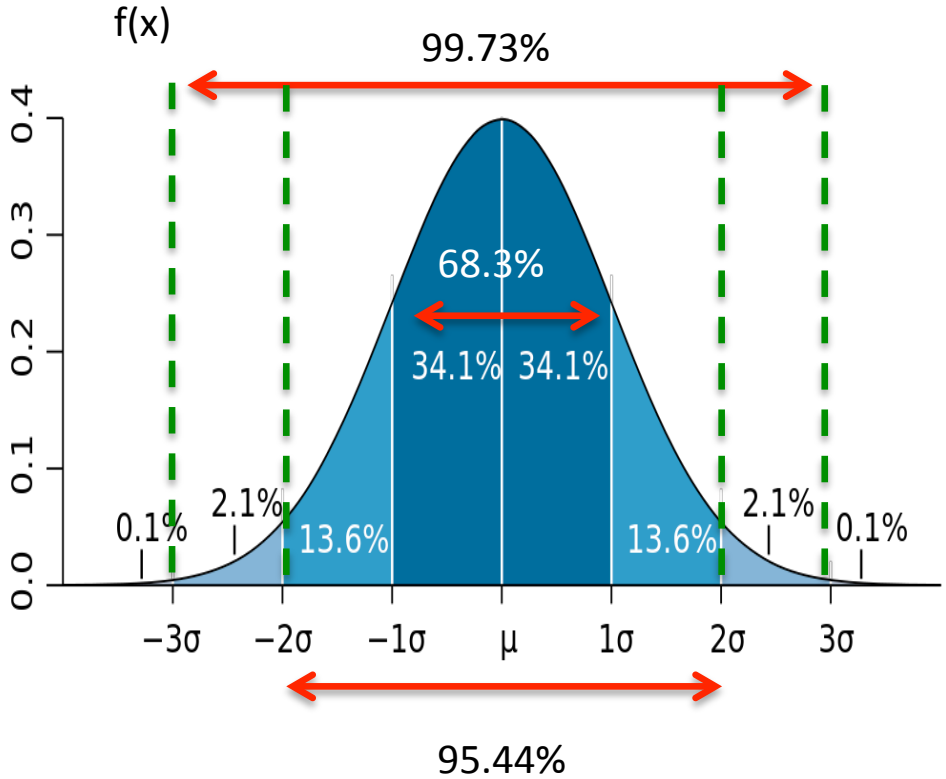
Normal distribution is the most important distribution in statistical applications because in practice, many measurements obey, at least approximately a normal distribution. The central limit theorem gives a theoretical base to this fact.

X is distributed as a normal random variable if its p.d.f. is

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

where σ^2 and μ are the parameters. Denote

$$X \sim N(\mu, \sigma^2)$$



$Z = (X - \mu)/\sigma \sim N(0, 1)$ Denote the c.d.f. of standard normal by $\Phi(x)$.

$$\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \, \mathrm{d}t$$

1. Moment generating function.
   Consider $Z \sim N(0,1)$ first

$$\begin{aligned} M(t) &= E(e^{tZ}) \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, \mathrm{d}z \end{aligned}$$

If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma$, $X = \sigma Z + \mu$.

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

Distribution of a function of a random variable

Many problems require finding the distribution of some function of X, say $Y = g(X)$ from the distribution of X. Suppose X has density $f_X(x)$, where a subscript now is used to distinguish densities of different random variables. We want to find $f_Y(y)$.

Method I: c.d.f. Method (works for every transformation)

Example:

If X is a continuous random variable with p.d.f. f_X and $Y = X^2$, find $f_Y(y)$.

Example: Let $X \sim N(0,1)$. Find the p.d.f. of $Y = X^2$.

Method II: Jacobian Method

Let X be a random variable with density $f_X(x)$ on the range (a, b). Let $Y = g(X)$ where g is either strictly increasing of strictly decreasing on (a, b). Range of Y is an interval with endpoints $g(a)$ and $g(b)$. Then

$$f_Y(y) = f_X(x) \left| \frac{\mathrm{d}x}{\mathrm{d}y} \right|$$

Example: Square root of an exponential random variable

Let $X \sim \text{Exp}(1)$, i.e., $f_X(x) = e^{-x}, x > 0$.

Find the density of $Y = \sqrt{X} = g(X)$.

Multivariate Distribution
Joint Distribution Function

Usually 2 or more random variable are of interest.
Examples:

(a) Math score X and statistics score Y of freshmen.

(b) Height X and weight Y of students.

(c) Classify n manufactured items into 3 or more categories, say X : number of good items, Y : number of okay items, $n - X - Y$: number of defectives.

How to describe the probability of events concerning 2 or more random variables?

Definition:

Let X and Y be 2 discrete random variables and $\mathbb{R}$ be the two dimensional space of X and Y. The joint probability distribution for X and Y is given by

$$p(x, y) = P(X = x, Y = y) \text{ for } (x, y) \in \mathbb{R}$$

The function $p(x, y)$ is called the joint probability density(or mass) function (joint p.d.f. or p.m.f.) and has the following properties:

1. $0 \leq p(x, y) \leq 1$.
2. $\sum_{(x,y)\in\mathbb{R}} p(x, y) = 1$.
3. $P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y)$.

Definition: Let X and Y have the joint probability mass function $p(x, y)$ with space $\mathbb{R}$. The p.m.f. of X alone, called the marginal probability mass function of X is defined by

$$p_X(x) = \sum_y p(x, y), \ x \in \mathbb{R}_X$$

The marginal p.m.f. of Y is

$$p_Y(y) = \sum_x p(x, y), \ y \in \mathbb{R}_Y$$

The random variables X and Y are independent if and only if

The notation of joint p.m.f. of 2 discrete random variables can be extended to a joint p.m.f. of n random variables of the discrete type. The joint p.m.f. of n random variables $X_1, X_2, \ldots, X_n$ is defined by

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

The marginal p.m.f. of one of the n discrete random variable, say X_k is

$$p_k(x_k) = \sum_{x_1} \cdots \sum_{x_{k-1}} \sum_{x_{k+1}} \cdots \sum_{x_n} p(x_1, x_2, \ldots, x_n)$$

which is obtained by summing $p(x_1, x_2, \ldots, x_n)$ over all x_i's except x_k.

Definition: $X_1, X_2, \ldots, X_n$ are mutually independent if and only if

Trinomial Distribution

Extend the binomial distribution from 2 classes to 3 classes. Let
X : number of observations in class 1.
Y : number of observations in class 2.
$n - X - Y$: number of observations in class 3.
$P(\text{class } 1) = p_1$, $P(\text{class } 2) = p_2$, $P(\text{class } 3) = 1 - p_1 - p_2$.

The joint p.m.f. of X and Y is given by

$$p(x, y) = P(X = x, Y = y) = \frac{n!}{x!y!(n - x - y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y}$$

where x, y are non-negative integers such that $x + y \leq n$.
We say X and Y have trinomial distribution with parameter n and p_1, p_2.

Multinomial Distribution

Multinomial experiment:

1. The experiment consists of n identical and independent trials.
2. The outcome of each trial falls into one of k class.
3. The probability that the outcome of a single trial will fall in a particular class, say class i is p_i, $(i = 1, \ldots, k)$ and remains the same from trial to trial

$$p_1 + p_2 + \ldots + p_k = 1$$

4. The random variable X_i is equal to the number of trials in which the outcome falls in class i, $i = 1, \ldots, k$. Note that

$$X_1 + X_2 + \ldots + X_k = n$$

The joint p.m.f. of $X_1, X_2, \ldots, X_k$ is given by

$$\begin{aligned} p(n_1, n_2, \ldots, n_k) &= P(X_1 = n_1, X_2 = n_2, \ldots, X_k = n_k) \\ &= \frac{n!}{n_1! n_2! \ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} \end{aligned}$$

where $\sum_{i=1}^k p_i = 1$ and $\sum_{i=1}^k n_i = n$.

The random variables $X_1, X_2, \ldots, X_k$ are said to have a multinomial distribution with parameters n and $p_1, p_2, \ldots, p_k$.

Note:

$$\sum \frac{n!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k} = (p_1 + p_2 + \ldots + p_k)^n = 1$$

where the summation is taken over the set of all non-negative integers $n_1, n_2, \ldots, n_k$ whose sum is n.

Properties:

1. The marginal p.m.f. of X_i is the binomial distribution $B(n, p_i)$, $i = 1, \ldots, k$.
2. The joint p.m.f. of X_i and X_j is a trinomial with parameters n and p_i, p_j.

Example: In Stony Brook, at 8:00 P.M., 30% of the TV viewing audience watch the news CNN, 25% watch sitcom FRIENDS, and the rest watch other programs. What is the probability that, in a statistical survey of seven randomly selected viewers, exactly three watch CNN and at least two watch FRIENDS?

Continuous Random Variables

The joint p.d.f. of two continuous random variables X and Y is an integrable function $f(x, y)$ with the following properties:

1. $f(x, y) \geq 0$.
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1$.
3. $P[(X, Y) \in A] = \int\int_{A} f(x, y)\, \mathrm{d}x\, \mathrm{d}y$.

The marginal p.d.f. of X is

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y$$

The marginal p.d.f. of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x$$

The random variables X and Y are independent if and only if

Expectation, Variance and Covariance

Let $X_1, \ldots, X_n$ be random variables with joint p.d.f. $f(x_1, x_2, \ldots, x_n)$.
If $U(X_1, \ldots, X_n)$ is a function of $X_1, \ldots, X_n$, then the expectation of $U(X_1, \ldots, X_n)$ is defined by

$$E[U(X_1, \ldots, X_n)] = \begin{cases} \sum_{(x_1,\ldots,x_n)} \cdots \sum U(x_1, \ldots, x_n) p(x_1, \ldots, x_n) \\ \quad \text{(discrete case)} \\ \int_{(x_1,\ldots,x_n)} \cdots \int U(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \, \mathrm{d}x_1 \ldots \mathrm{d}x_n \\ \quad \text{(continuous case)} \end{cases}$$

(a) If $U(X_1, \ldots, X_n) = X_i$, then

$$E[U(X_1, \ldots, X_n)] = E(X_i) = \mu_i$$

is called the mean of X_i, $i = 1, \ldots, n$.

(b) If $U(X_1, \ldots, X_n) = (X_i - \mu_i)^2$, then

$$E[U(X_1, \ldots, X_n)] = E[(X_i - \mu_i)^2] = \text{Var}(X_i) = \sigma_i^2$$

is called the variance of X_i, $i = 1, \ldots, n$.

(c) If $U(X_1, \ldots, X_n) = (X_i - \mu_i)(X_j - \mu_j)$ such that $i \neq j$, then

$$\begin{aligned} E[U(X_1, \ldots, X_n)] &= E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \mathrm{Cov}(X_i, X_j) = \sigma_{ij} \end{aligned}$$

is called the covariance of X_i and X_j.

(d)

$$\rho_{ij} = \frac{\mathrm{Cov}(X_i, X_j)}{\sqrt{\mathrm{Var}(X_i)\mathrm{Var}(X_j)}} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

is called the correlation coefficient of X_i and X_j.

Properties of $\mathrm{Cov}(X, Y)$ and ρ_{XY}

(a) $\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)$.

(b) $|\rho_{XY}| \leq 1$
$\mathrm{Cov}^2(X, Y) \leq \mathrm{Var}(X)\mathrm{Var}(Y)$
$\rho_{XY} = 0 \Rightarrow$ uncorrelated (no linear association between X and Y).
$\rho_{XY} = 1 \Rightarrow$ completely correlated (perfect linear association between X and Y).

(c) If X and Y are independent, then $\mathrm{Cov}(X, Y) = 0$.