ST2334 – Austin Santoso

Chapter 1: Probability

Basic Properties:

- $(A \cap B)' = A' \cup B'$
- $(A \cup B)' = A' \cap B'$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cup B = A \cup (B \cap A')$
- $A = (A \cap B) \cup (A \cap B')$
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$
- $P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A \cap B)$

Mutually exclusive events have these properties

$P(A_1 \cup A_2 \cup A_3 \ldots \cup A_n) = P(A_1) + P(A_2) \ldots + P(A_n)$

$P(A \cap B) = 0$

Independent events have these properties

$P(A.B) = P(A).P(B)$

Conditional Probability:
Probability of *'B'* given that *'A'*

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

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

Bayes Theorem:

Let B1, …, Bn be a partition of S. For any event A, and any $k \in 1, \ldots, n$

$P(A) = \sum_{i=1}^{n} P(B_i A) = \sum_{i=1}^{n} P(B_i)P(A \mid B_i)$

$= P(B_1)\, P(A \mid B_1) + \ldots + P(B_n)\, P(A \mid B_n)$

$$P(B_k \mid A) = \frac{P(B_k)P(A \mid B_k)}{P(B_1)P(A \mid B_1) + \cdots + P(B_n)P(A \mid B_n)}$$

$$= \frac{P(B_k)P(A \mid B_k)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)}$$

Chapter 2: Random Variables

Random variable: A random variable assigns a number to each outcome of a random circumstance

Mathematically, a random variable X is a mapping from the sample space S to the
set of real numbers R. That is
$X : S \rightarrow R$.

Ex: When we roll a pair of dice, let's say we are not interested in the numbers that are obtained on each die but we are only interested in the sum of the numbers.

There are two kinds of random variables, discrete and continuous.

Discrete: can only be one of finite or countably infinite number of values.
Probability mass function (**pmf**) simple bar graph

- $f(x_i) \geqslant 0$ for every x_i
- $\sum_{x_i} f(x_i) = 1$
- $P(X \in E) = \sum_{x_i \in E} f(x_i)$

Continuous: can assume one of a continuum of values, and the probability of each value is 0.

$P(a < X \leqslant b) = \int_a^b f_x(x)dx$, for $-\infty < a < b < \infty$

The function f_x is called the Probability density function (**pdf**).

- $f(x) \geqslant 0$ for all x
- The total area under the curve is 1, that is $\int_{-\infty}^{\infty} f(x)dx = 1$

Cumulative Distribution Function(cdf) :
The cdf of a random variable X is $F(x) = P(X \leqslant x)$

The CDF F(x) of a discrete random variable X, with pmf f(x) is

$F(x) = P(X \leqslant x) = \sum_{t \leq x} f(t) = \sum_{t \leq x} P(X = t)$
for $-\infty < x < \infty$

$P(a \leqslant X \leqslant b) = P(X \leqslant b) - P(X < a)$
$= F(b) - F(a^{-})$

The CDF F(x) of a continuous random variable X with pdf f(x) is

$F(x) = \int_{-\infty}^{x} f(t)dt$ if a derivative exists: we have $f(x) = \frac{d}{dx}F(x)$

$P(a \leqslant X \leqslant b) = P(a < X \leqslant b) = F(b) - F(a)$

Properties of a CDF for PDF

F(x) must satisfy the following conditions:

- F(x) is a non-decreasing function of x.
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
- F is right continuous with left limit at any x:
  $\lim_{h \downarrow 0} F(x+h) = F(x)$ *and* $\lim_{h \uparrow 0} F(x+h)$ *exist*.

If any function F(x) satisfies the above three conditions simultaneously, then it can be a cumulative distribution function of a random variable.

Mean average

Discrete:
$\mu_x = E(X) = \sum_{x_i} x_i P(X = x_i) = \sum_x x f(x)$

Continuous:
$\mu_x = E(X) = \int_{-\infty}^{\infty} x f(x)dx$

Properties of Expectation:
$E(a + bX) = a + b\,E(X)$
$E(aX) = a\,E(X)$
$E(X + b) = E(X) + b$

Sometimes we are interested in g(x) not just x, so we need to find E(g(x)). Given a pmf or pdf f_x

If X is discrete and provided the sum exists
$E(g(X)) = \sum_x g(x) f_x(x)$

If X is continuous and the integral exists
$E(g(X)) = \int_{-\infty}^{\infty} g(x) f_x(x)dx$

Special cases

- $g(x) = (X - \mu_x)^2 \rightarrow$ variance of random variable X
- $g(x) = x^k \rightarrow$ k-th moment of X

Variance: the average difference between each value and the mean.

Discrete random variable X
$\sigma_x^2 = V(X) = E[(X-\mu_x)^2] = \sum_x (X-\mu x)^2 f_x(x)$

Continuous random variable X
$\sigma_x^2 = V(X) = E[(X-\mu_x)^2] = \int_{-\infty}^{\infty} (X-\mu x)^2 f_x(x)dx$

Properties of Variance:

- $V(X) \geqslant 0$
- $V(X) = E(X^2) - [E(X)]^2$
- If $V(X) = 0$, $P(X = \mu_x) = 1$
- If a, b constants, $V(a + bX) = b^2V(X)$

Standard deviation
$\sigma_x = SD(X) = \sqrt{V(X)}$

Chebyshev's Inequality
If a random variable X has a mean μ and stand deviation σ, the probability of getting a value which deviates from μ by at least kσ is at most $1/k^2$

This means:

$P(|X-\mu| > k\sigma) \leq \frac{1}{k^2}$ or

$$P(|X-\mu| \leq k\sigma) \geq 1 - \frac{1}{k^2}$$

Chapter 3: Joint Distributions

Random Variable but is affected by 2 variables.

Joint Probability Mass Function

Let (X, Y) be a 2-dimensional discrete random variable defined on the sample space of an experiment. Their joint probability mass function is defined as:

$$f_{X,Y}(x,y) = P(X = x, Y = y)$$

Where x, y are possible values of X and Y respectively.

Properties of Joint PMF

- $f_{X,Y}(x,y) \geq 0$, for all $(x, y) \in R_{X,Y}$
- $\sum_x \sum_y f_{X,Y}(x,y) = \sum_x \sum_y P(X = x, Y = y) = 1$
- Let A be any set consisting of pairs of (x;y) values. Then the probability $P((X, Y) \in A))$ is defined by summing the joint probability mass function over pairs in A:
  - $P\big((X,Y) \in A\big) = \sum_{(x,y)\in A} f_{X,Y}(x,y)$

Joint Probability Density Function

Let (X, Y) be a 2-dimensional cont)inuous random variable assuming all values in some region R of the Euclidian plane,$\mathbb{R}^2$
The joint PDF of (X,Y) is a function $f_{X,Y}(x,y)$ such that

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x,y)dydx$$

Properties of Joint PDF

- $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in R_{X,Y}$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx = 1$

Marginal Distribution

Given a joint distribution function (X,Y), we call the distribution of X or Y alone the marginal distribution

Marginal Distribution: Discrete

The marginal probability mass function f_x of X:

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f_{X,Y}(x,y)$$

Marginal Distribution: Continuous

The marginal probability mass function f_x of X:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Conditional Probability Mass Function

The conditional probability mass (or density) function of X given Y=y is defined as

$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ provided $f_Y(y) > 0$

Independent Random Variables

Random variables X and Y are independent if and only if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Or

$$f_{X|Y}(x|y) = f_X(x)$$

For all x and y
Random variables that are not independent are dependent

Expectation of g(X,Y)

Discrete:

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) f_{X,Y}(x,y)$$

Continuous:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dydx$$

Covariance

Let

$$g(X,Y) = \big(X - E(X)\big)\big(Y - E(Y)\big)$$

$= (X - \mu_X)(Y - \mu_y)$

The expectation of E[g(X,Y)] leads to the definition of covariance

The covariance of (X,Y):

$$\sigma_{x,y} = cov(X,Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = E[(X - \mu_X)(Y - \mu_Y)]$$

Properties of covariance

- $Cov(X,Y) = E(XY)-E(X)E(Y) = E(XY)-\mu_X\mu_Y$
- $Cov(X,X) = V(X)$
- $Cov(X,Y) = cov(Y,X)$
- $Cov(aX+b, cY+d) = ab\ cov(X,Y)$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abcov(X,Y)$
- If X, Y are independent, $cov(X,Y) = 0$

Correlation Coefficient

The correlation coefficient of X and Y, denoted Cor(X,Y), $\rho_{X,Y}$ or ρ is

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

- $-1 \leq \rho_{X,Y} \leq 1$
- $\rho_{X,Y}$ is a measure of the degree of linear relationship between X and Y
- If X and Y are independent, the $\rho_{X,Y} = 0$. But $\rho_{X,Y} = 0$ does not imply independence

Chapter 4: Common Probability Distributions

Discrete Uniform Distribution

If a random variable X assumes the values $x_1, x_2, \ldots, x_k$ with equal probability, X is said to have a discrete uniform distribution and the probability mass function is given by

$$f_X(x) = P(X = x) = \begin{cases} 1/k, x = x_1, x_2 \ldots x_3 \\ 0, otherwise \end{cases}$$

The mean and variance of Discrete Uniform Distribution

- $\mu = E(X) = \sum x f_X(x) = \frac{1}{k}\sum_{i=1}^{k} x_i$
- $\sigma^2 = V(X) = \sum (x-\mu)^2 f_X(x) = \frac{1}{k}\sum_{i=1}^{k}(x_i - \mu)^2$
- $\sigma^2 = E(X^2) - \mu^2 = \frac{1}{k}\sum_{i=1}^{k} x_i^2 - \mu^2$

Random Variables Arising from Repeated Trials

Trials are repeated independently

Probability of success if p, failure is 1-p

Bernoulli Trials: an experiment with two outcomes success and failure.

Binomial Distribution

A random variable X is defined to have a Bernoulli distribution with parameter 0<p<1, when it has probability mas function given as

$f_X(x) = P(X = x) = p^x(1-p)^{1-x}$, for x = 0,1

E(X) = p and V(X) = p(1-p)

A random variable X is defined to have a binomial distribution with parameters $n \in \mathbb{Z}^+$ and 0<p<1 written as X~B(n,p), when it has probability mass function given as

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

For x=0, 1, 2, …, n

E(X) = np, V(X)=np(1-p)

Geometric Distribution

The number of trials required until the first success is achieved.

A random variable X is defined to have a geometric distribution with parameter 0<p<1, written as X~Geom(p), when it has probability mass function given as

$$f_X(x) = P(X = x) = (1-p)^{x-1} p$$

For x = 1, 2, …

E(X) = $\frac{1}{p}$ and V(X) = $\frac{1-p}{p^2}$

Negative Binomial random variables

We want the kth success and event number x

Counts the number of independent Bernoulli trials required in order to obtain k success

X~NB(k,p)

$$f_X(x) = P(X = x) = \binom{x-1}{k-1} p^k q^{x-k}$$

E(X) = $\frac{k}{p}$ and V(X)=$\frac{(1-p)k}{p^2}$

Poisson Distribution

Poisson Random Variable

The number of success X in a Poisson experiment is called a Poisson random variable. The probability mass function of the Poisson random variable X with parameter λ >0, denoted by X~Poisson(λ), is given by

$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

For x = 0, 1, 2…

E(X) = λ and V(X)=λ

Poisson Approximation to the Binomial

Let X be a Binomial random variable with parameters n and p.

Suppose that $n \to \infty$ and $p \to 0$ in such a way that λ = np remains a constant as $n \to \infty$.

Then X will have approximately a Poisson distribution with parameter np. That is

$$\lim_{\substack{n\to\infty \\ p\to 0}} P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

The approximation is good when $n \geq 20$ and $p \leq 0.05$ or if $n \geq 100$ and $np \leq 10$

Note:
Using Poisson Approximation when p is big

If p is close to 1, we can still use the Poisson distribution to approximate binomial probabilities by interchanging what we have defined to be a success and a failure so to change p to a value close to 0.

Continuous Uniform Distribution

A random variable X is said to follow a uniform distribution over the interval [a, b], denoted by X~U(a,b), if its probability density function is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a}, for\ a \leq x \leq b \\ 0, otherwise \end{cases}$$

E(X) = $\frac{a+b}{2}$ and var(x)=$\frac{(b-a)^2}{12}$

The distribution function of X~U(a,b) is

$$F_X(x) = \int_{-\infty}^{x} f_X(y)dy = \begin{cases} 0, & if\ x < a \\ \frac{x-a}{b-a}, & if\ a \leq x \leq b \\ 1, & if\ x \geq b \end{cases}$$

Exponential Distribution

A random variable X is said to be follow an exponential distribution with parameter $\lambda>0$, denoted by $X \sim Exp(\lambda)$, if its probability density function is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & for\ x > 0 \\ 0, & for\ x \le 0 \end{cases}.$$

$E(X) = \frac{1}{\lambda}$ and $var(X)=\frac{1}{\lambda^2}$

The cumulative distribution function of $X \sim Exp(\lambda)$ is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Memoryless Property

The exponential distribution satisfies the following memoryless property

$$P(X > s + t | X > s) = P(X > t)$$

For $s,t > 0$

Normal Distribution

A random variable X is said to follow a normal distribution with parameters $-\infty<\mu<\infty$ and $\sigma>0$, denoted $X \sim N(\mu, \sigma^2)$, if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), -\infty < x < \infty$$

$E(X)=\mu$ and $V(X)=\sigma^2$

Standard Normal

Let $Y \sim N(\mu, \sigma^2)$ then

$$P(a < Y \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right).$$

Properties of standard deviation

- $P(Z \geqslant 0) = P(Z \leqslant 0) = 0.5$
- $-Z \sim N(0,1)$
- $P(Z \leqslant x) = 1-P(Z>x)$ for $-\infty<x<\infty$
- $P(Z \leqslant -x) = P(Z \geqslant x)$ for $-\infty<x<\infty$
- If $Y \sim N(\mu, \sigma^2)$ then $X = \frac{Y-\mu}{\sigma} \sim N(0, 1)$
- If $X \sim N(0, 1)$ then $Y = aX+b \sim N(b, a^2)$ for $a, b \in \mathbb{R}$

Normal Approximation to the Binomial

If $n \to \infty$ and $p \to 0.5$, we can use normal distribution to approximate the binomial distribution. Even when n is small and p is not close to either 0 or 1. Rule of thumb: $np>5$ and $n(1-p)>5$

Suppose X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = np(1-p)$. Then as $n \to \infty$,

$Z = \frac{X-np}{\sqrt{np(1-p)}}$ is approximately distributed as N(0,1)

Continuity correction

When approximating a discrete random variable by a continuous random variable like the normal distribution, we need to "spread" its values over a continuous scale. It is an approximation in the interval sense. This makes sense when the interval is large.

For a small range, say $P(X = k)$, we do this by representing k by the interval from $k - {}^1\!/_2$ to $k + {}^1\!/_2$

$$P(X = k) \approx P\left(k - \frac{1}{2} < X < k + \frac{1}{2}\right)$$

$$P(X \ge k) \approx P\left(X \ge k - \frac{1}{2}\right)$$

$$P(X \le k) \approx P\left(X \le k + \frac{1}{2}\right)$$

Chapter 5: Sampling and Sampling Distributions

For random samples of size n taken from an infinite population or from a finite population wit replacement having mean μ_x and variance σ_x^2, the sampling distribution of a sample mean $\bar{X}$ has mean and variance given by

$\mu_{\bar{X}} = \mu_X$ and $\sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$

That is,

$E(\bar{X}) = E(X)$ and $V(\bar{X}) = \frac{V(X)}{n}$

Law of Large Number:
As the sample size increases, the probability that the sample mean differs from the population means goes to 0.

Let $e \in \mathbb{R}$

$$P(|\bar{X} - \mu| > e) \to 0\ as\ n \to \infty$$

Central Limit Theory (CLT)

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and finite variance σ^2. The sampling distribution of the sample mean $\bar{X}$ is approximately normal with mean μ and variance ${\sigma^2}/{n}$ if n is sufficiently large.

This means that

$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ approximately,

Or equivalently

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

Difference of two sample means

Consider:

1. $X_1, X_2, \ldots, X_{n1}$ is a random sample of size $n_1 \geqslant 30$ from a (large or infinite) population 1 with mean μ_1 and variance σ_1^2.
2. $Y_1, Y_2, \ldots, Y_{n2}$ is a random sample of size $n_2 \geqslant 30$ from a (large or infinite) population 2 with mean μ_2 and variance σ_2^2.
3. The two samples are independent.

Then the sampling distribution of the differences of means, $\bar{X}$ and $\bar{Y}$, is approximately normally distributed with mean and variance given by

$E(\bar{X} - \bar{Y}) = \mu_{\bar{X}-\bar{Y}} = \mu_1 - \mu_2$ and

$$V(\bar{X} - \bar{Y}) = \sigma_{\bar{X}-\bar{Y}}^2 = V(\bar{X}) + V(\bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

When n_1 and n_2 are both large, that is, $n_1 \geqslant 30$ and $n_2 \geqslant 30$, the Central Limit Theorem implies that $\bar{X}$ and $\bar{Y}$ will both be normal approximately. Thus $\bar{X} - \bar{Y}$ will be normal approximately

$Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$ approximately.

Chi Square ($\chi^2 - Distribution$)

Gamma Function, $\Gamma(.)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Properties of the gamma function

- Via integration by parts
  - $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\Gamma(1) = \int_0^\infty e^{-y} dy = 1$
- For integrals values of $\alpha = n = 1, 2, 3, \ldots$ $\Gamma(n) = (n-1)!$

Chi Square ($\chi^2 - Distribution$)

Let Y be a random variable with probability density function

$f_Y(y) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}$, for $y > 0$

Then Y is defined to have a chi-square distribution with n degree of freedom, denoted by $\chi^2(n)$, where n is a positive integer, and Γ(.) is the gamma function.

Properties of the χ^2-Distribution

- All χ^2 values are nonnegative
- The χ^2distribution is a family of curves, each determined by the degrees o freedom n. All the density functions have a long right tail.
- If $Y\sim\chi^2(n)$, then E(Y) = 2 and V(Y) =2n
- For large n, $\chi^2\sim N(n,2n)$ approximately
- If $Y_1, Y_2, \ldots, Y_k$ are independent chi-square random variables with $n_1, n_2, \ldots, n_k$ degrees of freedom respectively, then $Y_1+Y_2+\ldots+Y_k$ has a chi-square distribution with $n_1+n_2+\ldots+n_k$ degrees of freedom

$$\sum_{i=1}^{k} Y_i \sim\chi^2\left(\sum_{i=1}^{k} n_i\right)$$

Theorem 3: Chi Square χ^2

- If $Z\sim N(0,1)$, then $Z^2\sim\chi^2(1)$
- Let $\chi\sim N(\mu,\sigma^2)$, then $\left(\frac{X-\mu}{\sigma}\right)^2\sim\chi^2(1)$
- Let $X_1, \ldots, X_n$ be a random sample from a normal population with mean μ and variance σ^2. Then

$$\mathrm{Y} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$$

$\chi^2(10;0.9)$ means

- $\Pr\big(Y\geq\chi^2(10;0.9)\big) = 0.9$ or
- $\Pr\big(Y\ \leq\chi^2(10;0.9)\big) = 0.1$

Sampling Distribution Related to the Sample Variance

Sample Variance

Let $X_1, \ldots, X_n$ be a random sample from a population distributed with E(X) = μ and V(X) = σ^2. The sample variance is defined as

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

Theorem 2

Let S^2 be the sample variance of a random sample of size n taken from a normal population with E(X) = μ and V(X) = σ^2. Then the random variable

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$$

Follows a χ^2 distribution with n-1 degree of freedom

t-distribution

Given a normal distribution, we have by CLT

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\approx Z\sim N(0,1)$$

If we do not know σ, we can estimate σ by the sample deviation (S)

$$S = \sqrt{\frac{1}{n-1}\sum_{i}^{n}(X_i-\bar{X})^2}$$

Let Z be a standard normal variable and U a χ^2 random variable with n degrees of freedom. If Z and U are independent, then T is given by

$$T = \frac{Z}{\sqrt{U/n}}$$

Is said to be a t-distribution with n degrees of freedom.

Properties of t-Distribution

- The t-distribution (also called the Student's t) is denoted by t(n) and the shape of its density function is similar to that of the normal distribution.
- If $T\sim t(n)$, then E(T) = 0 and $V(T) = \frac{n}{n-2}$ for n > 2
- n is the degrees of freedom, and the t-distribution approaches N(0,1) as the parameter n →∞. That is, $\lim_{n\to\infty} f_r(t) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$
- In practice, when n ⩾ 30, we can replace t(n) with N(0,1)
- The density function of t-distribution is bell shaped, centered and symmetrical at 0

If the X_i's are normal, then

$$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$$

Is a random variable having t-distribution with n-1 degree of freedom.

F-distribution

Let U and V be independent random variables χ^2 distribution with n_1 and n_2 degrees of freedom, respectively. The distribution of the random variable,

$$F = \frac{U/n_1}{V/n_2}$$

Is called a F distribution with (n_1, n_2) degree of random.

Properties of the F-distribution

- If $F\sim F(n,m)$, then $1/F\sim F(m,n)$
- Values of the F-distribution can be found in the F statistical tables.
  - F(5,4;0.05) = 6.26 means that P(F>6.26) = 0.05, where F~F(5,4)
- $F(n_1,n_2;\alpha) = 1/F\,(n_2,n_1;1-\alpha)$

Chapter 6: Estimation based on Normal Distribution

Point Estimation

Based on sample data, a single value is calculated. The formula that describes is the point estimator, the resulting number is the point estimate.

A **Statistic** is a function of the random sample which does not depend on any unknown parameter.

For example: let

$$W = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

W is static if μ is known, else W is non static

Unbiased Estimator:

Let $\widehat{\Theta}$ be an estimator of θ. If $E(\widehat{\Theta}) = \theta$, we call $\widehat{\Theta}$ an unbiased estimator for θ

Examples:

- $\bar{X}$ is an unbiased estimator for μ. $E(\bar{X}) = \mu$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, E($S^2$)=$\sigma^2$

Example of biased

- A biased estimator of σ^2 is $T = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$. It can be shown that $E(T) = \frac{n-1}{n}\sigma^2\neq\sigma^2$

Interval Estimation

Based on sample data, two numbers are calculated to form an interval within which the parameter is expected to lie. In the form:

$$\widehat{\theta_L} < \theta < \widehat{\theta_U}$$

$\widehat{\theta_L}$, Lower Confidence Limit

$\widehat{\theta_U}$, Upper Confidence Limit

Suppose We Seek a random interval $(\widehat{\theta_L}, \widehat{\theta_U})$ containing θ with a given probability 1-α. That is

$$P(\widehat{\theta_L} < \theta < \widehat{\theta_U}) = 1 - \alpha$$

- The interval $\widehat{\theta_L} < \theta < \widehat{\theta_U}$, computed from the selected sample is called a $(1-\alpha)100\%$ **confidence interval** for θ
- The fraction (1-α) is called the **confidence coefficient** or **degree of confidence,**
- The end points $\widehat{\theta_L}$ *and* $\widehat{\theta_U}$ are called **lower and upper confidence limits respectively.**

Confidence Interval for the Mean

If the population is normal, or when n is large (n⩾30)

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ or } Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Thus,

$$P\left(-z_{\alpha/2} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le z_{\alpha/2}\right) = 1 - \alpha$$

Confidence Interval with known σ: Normal Population or Big n

$$\bar{X} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Sample Size of Estimating μ
For a given margin of error e, the sample size n:

$$n \ge \left(\frac{Z_{\alpha/2}.\sigma}{e}\right)^2$$

Confidence Interval with unknown σ: normal population & small n

Let

$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$, Where S^2 is the sample variance, then $T \sim t_{n-1}$.

If $\bar{X}$ and S are the sample mean and standard deviation of a random sample of size n < 30 from an approximate normal population with unknown variance σ^2, a (1-α)100% confidence interval for μ is given by:

$$\bar{X} \pm t_{n-1;\alpha/2} \frac{S}{\sqrt{n}}$$

If n > 30, the t distribution is the same as N(0,1)

$$\bar{X} \pm Z_{\alpha/2} \frac{S}{\sqrt{n}}$$

Confidence Intervals for the Difference between Two Means

If we have two populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2 respectively, then $\bar{X}_1 - \bar{X}_2$, is the point estimator for $\mu_1 - \mu_2$

Confidence Interval for $\mu_1 - \mu_2$ with known $\sigma_1^2 \neq \sigma_2^2$: Normal population or Big n:

$$(\overline{X_1} - \overline{X_2}) \pm Z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Confidence Interval for $\mu_1 - \mu_2$ with unknown $\sigma_1^2 \neq \sigma_2^2$: Big n (>30):

We replace σ_1^2 and σ_2^2 by their estimates, S_1^2 and S_2^2:

$$(\overline{X_1} - \overline{X_2}) \pm Z_{\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Confidence Interval for $\mu_1 - \mu_2$ with unknown $\sigma_1^2 = \sigma_2^2$: Normal population & small n

Let Pooled Sample Variance be

$$S_p^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{i=1}^{n_2}(Y_i i - \bar{Y})^2}{n_1 + n_2 - 2} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

Given a Confidence interval for $\mu_1 - \mu_2$:

$$(\overline{X_1} - \overline{X_2}) \pm t_{n_1+n_2-2;\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence Interval for $\mu_1 - \mu_2$ with unknown $\sigma_1^2 = \sigma_2^2$: Big n We can use normal distribution

$$(\overline{X_1} - \overline{X_2}) \pm Z_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Confidence Intervals for difference of means for paired(dependent) data

If we run a text on n individuals, and compare their initial scores X_i and final scores Y_i. Observations are made on the same individual are related so they form a pair. We usually consider the difference $D_i = X_i - Y_i$

These differences are values of the random sample $D_1, D_2, \ldots, D_n$ with mean μ_n and unknown variance σ_D^2.

The point estimate of μ_D is

$$\bar{D} = \frac{1}{n}\sum_{i=1}^{n} D_i = \frac{1}{n}\sum_{i=1}^{n} (X_i - Y_i)$$

The point estimate of σ_D^2 is

$$S_D^2 = \frac{1}{n-1}\sum_{i=1}^{n} (D_i - \bar{D})^2$$

The confidence Interval for $\mu_D = \mu_1 - \mu_2$ can be:

$$P\left(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}\right) = 1 - \alpha$$

Where $T = \frac{\bar{D}-\mu_D}{S_D/\sqrt{n}} \sim t_{n-1}$

Continuous Interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{D} \pm t_{n-1;\alpha/2} \frac{S_D}{\sqrt{n}}$$

For big values of n (>30) we can use normal distribution

$$\bar{D} \pm Z_{\alpha/2} \frac{S_D}{\sqrt{n}}$$

Confidence Intervals for Variances

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution. Then the sample variance is

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)$$

Is a point estimate for σ^2

Confidence Interval for σ^2: Normal distribution, known μ.
A (1-α)100% confidence interval for σ^2 from a N(μ, σ^2) distribution with known μ is

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$$

Confidence Interval for σ^2: Normal distribution, unknown μ.
A (1-α)100% confidence interval for σ^2 from a N(μ, σ^2) distribution with unknown μ is

$$\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

To find the confidence interval for σ, square root the inequalities above

Confidence Intervals for Ratio of Variances

Confidence Intervals for ${\sigma_1^2}/{\sigma_2^2}$: Normal population, unknown μ_1 and μ_2

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;1-\alpha/2}}$$

Or

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\alpha/2}$$

Confidence interval for ${\sigma_1}/{\sigma_2}$, just sqrt the inequality above

Chapter 7: Hypothesis Testing: Normal Distribution

Introduction

We have a statistical hypothesis and we need to either accept or reject it.

Null Hypothesis (H_0) is a hypothesis that we formulate with the hope of rejecting, usually has one strict value. The alternate hypothesis (H_1) can usually have multiple values.

Truth / Decision	H_0 is true	H_0 is false
Reject H_0	Type I error Serious Error α	Correct Decision 1-β
Not Reject H_0	Correct Decision 1-α	Type II error β

The probability of making a type I error is called level of significance $\alpha = P(reject\ H_0|H_0)$

Let β be the probability of making a type II error. The power of the test is 1-β.

$$\beta = P(do\ no\ reject\ H_0|H_1)$$

Hypothesis Concerning one Mean

Try to find μ with known variance σ^2 and normal population(n⩾30)

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

For two-sided test, set level of significance to 0.05

$$H_0: \mu = \mu_0, H_1: \mu \neq \mu_0$$

Rejection region: We reject when $\bar{X}$ is too large or too small when compared to μ_0.

p-value = $2P(Z < -|z|)$. If p-value > α, do not reject H_0, otherwise reject H_0.

$$P\left(|Z| > Z_{\alpha/2}\right) = \alpha$$

$$Z < -z_{(\alpha/2)} or\ Z > z_{\alpha/2}$$

For one sided

$$H_0: \mu = \mu_0, H_1: \mu < \mu_0\ or\ H_1: \mu > \mu_0$$

p-value: $P(Z > |z_\alpha|)$ or $P(Z < -|z_\alpha|)$

Rejection range and p-value

H_1	Rejection Region	p-value
$\mu > \mu_0$	$z > z_\alpha$	$P(Z > \lvert z\rvert)$
$\mu < \mu_0$	$z < -z_\alpha$	$P(Z < -\lvert z\rvert)$
$\mu \neq \mu_0$	$z < -z_{\frac{\alpha}{2}}$ or $z > z_{\frac{\alpha}{2}}$	$2P(Z > \lvert z\rvert)$

Hypothesis on μ with unknown σ

Given a normal population

Two-sided test:

$H_0: \mu = \mu_0$vs $H_1: \mu \neq \mu_0$

The test statistic is given

$$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$$

Where S^2 is the sample variance.

Rejection region and p-value

H_1	Rejection Region	p-value
$\mu > \mu_0$	$t > t_\alpha$	$P(T > \lvert t\rvert)$
$\mu < \mu_0$	$t < -t_\alpha$	$P(T < -\lvert t\rvert)$
$\mu \neq \mu_D$	$t < -t_{\frac{\alpha}{2}}$ or $t > t_{\frac{\alpha}{2}}$	$2P(T > \lvert t\rvert)$

Hypothesis testing vs confidence intervals

Confidence intervals can be used to perform two sided tests

Given a confidence interval $(1-\alpha)100\%$, α is level of significance. If the confidence interval does not contain μ_0, then H_0 will be rejected

Hypothesis concerning difference of two different means

Test statistic (known σ_1^2, σ_2^2): normal population or Big n's

$$H_0: \mu_1 - \mu_2 = \delta_0$$

If H_0 is true, we have the test statistic

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - \delta_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Rejection range and p-values

H_1	Rejection Region	p-value
$\mu_1-\mu_2>\delta_0$	$z>z_\alpha$	$P(Z>\lvert z\rvert)$
$\mu_1-\mu_2<\delta_0$	$z<-z_\alpha$	$P(Z<-\lvert z\rvert)$
$\mu_1-\mu_2\neq\delta_0$	$z<-z_{\frac{\alpha}{2}}$ or $z>z_{\frac{\alpha}{2}}$	$2P(Z>\lvert z\rvert)$

Test statistic on $\mu_1-\mu_2$ with unknown σ_1^2,σ_2^2: big n

When H_0 is true, we have test statistic

$$Z=\frac{(\overline{X_1}-\overline{X_2})-\delta_0}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim N(0,1)$$

Test statistic on $\mu_1-\mu_2$ with unknown $\sigma_1^2=\sigma_2^2$: normal population and small n's (n<30)

When H_0 is true, we have test statistic

$$T=\frac{(\overline{X_1}-\overline{X_2})-\delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

When

$$S_p^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Rejection range and p-values

H_1	Rejection Region	p-value
$\mu_1-\mu_2>\delta_0$	$t>t_\alpha$	$P(T>\lvert t\rvert)$
$\mu_1-\mu_2<\delta_0$	$t<-t_\alpha$	$P(T<-\lvert t\rvert)$
$\mu_1-\mu_2\neq\delta_0$	$t<-t_{\frac{\alpha}{2}}$ or $t>t_{\frac{\alpha}{2}}$	$2P(T>\lvert t\rvert)$

Test statistic: paired data

If n<30 is small and difference D_i are normally distributed, we have test statistic

$$T=\frac{\overline{D}-\mu_{D,0}}{S_D/\sqrt{n}}\sim t(n-1)$$

When H_0 is true

If n⩾30, we have test statistic when H_0 is true

$$T=\frac{\overline{D}-\mu_{D,0}}{S_D/\sqrt{n}}\sim N(0,1)$$

Rejection and p-values: paired data

When T~N(0,1)

H_1	Rejection Region	p-value
$\mu_D>\mu_{D,0}$	$z>z_\alpha$	$P(Z>\lvert z\rvert)$
$\mu_D<\mu_{D,0}$	$z<-z_\alpha$	$P(Z<-\lvert z\rvert)$
$\mu_D\neq\mu_{D,0}$	$z<-z_{\frac{\alpha}{2}}$ or $z>z_{\frac{\alpha}{2}}$	$2P(Z>\lvert z\rvert)$

When T~t(n-1)

H_1	Rejection Region	p-value
$\mu_D>\mu_{D,0}$	$t>t_\alpha$	$P(T>\lvert t\rvert)$
$\mu_D<\mu_{D,0}$	$t<-t_\alpha$	$P(T<-\lvert t\rvert)$
$\mu_D\neq\mu_{D,0}$	$t<-t_{\frac{\alpha}{2}}$ or $t>t_{\frac{\alpha}{2}}$	$2P(T>\lvert t\rvert)$

Hypothesis test on σ^2

To test

$$H_0:\sigma^2=\sigma_0^2$$

We can use

$$\chi^2=\frac{(n-1)S^2}{\sigma_0^2}\sim\chi^2(n-1)$$

The rejection region for the following alternatives

H_1	Rejection Region
$\sigma^2>\sigma_0^2$	$\chi^2>\chi^2_{n-1;\alpha}$
$\sigma^2<\sigma_0^2$	$\chi^2<\chi^2_{n-1;1-\alpha}$
$\sigma^2\neq\sigma_0^2$	$\chi^2<\chi^2_{n-1;1-\frac{\alpha}{2}}$ or $\chi^2>\chi^2_{n-1;\frac{\alpha}{2}}$

Hypothesis testing ration of variances

Hypothesis test on σ_1^2,σ_2^2

To test

$$H_0:\sigma_1^2=\sigma_2^2$$

We can use the test statistic

$$F=\frac{S_1^2}{S_2^2}\sim F(n_1-1;n_2-1)$$

The rejection regions

H_1	Rejection Region
$\sigma_1^2>\sigma_2^2$	$F>F_{n_1-1,n_2-1;\alpha}$
$\sigma_1^2<\sigma_2^2$	$F<F_{n_1-1,n_2-1;1-\alpha}$
$\sigma_1^2\neq\sigma_2^2$	$F<F_{n_1-1,n_2-1;1-\frac{\alpha}{2}}$ or $F>F_{n_1-1,n_2-1;\frac{\alpha}{2}}$