ST2334 Full Help-sheet

Basic probability concepts

Observation

Any recording of information, whether it is numerical or categorical.

Statistical Experiment

Any procedure that generates a set of data (observations).

Sample Space

The set of all possible outcomes of a statistical experiment is called the **sample space** and it is represented by the symbol S.

Sample Point

Every outcome in a sample space is called an element of the sample space or simply a sample point.

Event

An event is a subset of a sample space.

Simple Event

An event is said to be simple if it consists of exactly one outcome (i.e. one sample point)

Compound Event

An event is said to be compound if it consists of more than one outcomes (or sample points).

1. The sample space is itself an event and is usually called a sure event.
2. A subset of S that contains no elements at all is the empty set, denoted by $\emptyset$, and is usually called a null event.

Operations of Events

Union

The union of two events A and B, denoted by $A \cup B$, is the event containing all the elements that belong to A or B or to both. That is,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

Intersection

The intersection of two events A and B, denoted by $A \cap B$ or simply AB, is the event containing all elements that are common to A and B. That is

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

Complement

The complement of event A with respect to S, denoted by A' or A^C, is the set of all elements of S that are not in A. That is

$$A' = \{x : x \in S \text{ and } x \notin A\}$$

Mutually Exclusive Events

Two events A and B are said to be mutually exclusive or mutually disjoint if $A \cap B = \emptyset$, that is, if A and B have no elements in common.

Union of n Events

The union of n events $A_1, A_2, \cdots, A_n$, denoted by

$$A_1 \cup A_2 \cup \ldots \cup A_n$$

is the event containing all the elements that belong to one or more of the events A_1, A_2, or ..., or A_n. That is

$$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \cdots \cup A_n = \{x : x \in A_1 \text{ or } \cdots \text{ or } x \in A_n\}$$

Intersection of n Events

The intersection of n events $A_1, A_2, \cdots, A_n$, denoted by

$$A_1 \cap A_2 \cap \ldots \cap A_n$$

is the event containing all the elements that are common to all of the events A_1, A_2, or ..., or A_n. That is

$$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \cdots \cap A_n = \{x : x \in A_1 \text{ and } \cdots \text{ and } x \in A_n\}$$

Counting

Permutation

A permutation is an arrangement of r objects from a set of n objects, where $r \leq n$. (Note that the **order is taken into consideration** in permutation.)

$${}_nP_r = n(n-1)(n-2)\cdots(n-(r-1)) = n!/(n-r)!$$

When not all objects are distinct,

$$nP_{n_1,n_2,\cdots,n_k} = \frac{n!}{n_1!n_2!\cdots n_k!}$$

When in circle: $(n-1)!$

Combination

the number of ways of selecting r objects from n objects **without regard to the order**.

$$\left(\begin{array}{c} n \\ r \end{array}\right) = \frac{n!}{r!(n-r)!}$$

Axioms of Probability

Axiom 1

$0 \leq Pr(A) \leq 1$

Axiom 2

$Pr(S) = 1$

Axiom 3

If $A_1, A_2, \cdots$ are mutually exclusive (disjoint) events, then

$$\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$$

Inclusion-Exclusion Principle

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr(A_i \cap A_j) + \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} \Pr(A_i \cap A_j \cap A_k) - \cdots\cdots + (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n)$$

Conditional Probability

The conditional probability of B given A is defined as

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)}, \quad \text{if } \Pr(A) \neq 0$$

Multiplication Rule of Probability

In general,

$$\Pr(A_1 \cap \cdots \cap A_n) = \Pr(A_1) \Pr(A_2|A_1) \Pr(A_3|A_1 \cap A_2) \cdots \Pr(A_n|A_1 \cap \cdots \cap A_{n-1})$$
provided that $\Pr(A_1 \cap \cdots \cap A_{n-1}) > 0$

The Law of Total Probability

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S. That is $A_1, A_2, \cdots, A_n$ are mutually exclusive and exhaustive events such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{n} A_i = S$.
Then for any event B

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap A_i) = \sum_{i=1}^{n} \Pr(A_i) \Pr(B|A_i)$$

Bayes' Theorem

Let $A_1, A_2, \cdots, A_n$ be a partition of the sample space S. Then

$$\Pr(A_k|B) = \frac{\Pr(A_k) \Pr(B|A_k)}{\sum_{i=1}^{n} \Pr(A_i) \Pr(B|A_i)}$$

for $k = 1, \cdots, n$. Or

$$\Pr(A_k|B) = \frac{\Pr(A_k) \Pr(B|A_k)}{\Pr(B)}$$

Independent Events

Two events A and B are independent iff

$$\Pr(A \cap B) = \Pr(A) \Pr(B)$$

Pairwise Independence

A set of events $A_1, A_2, \cdots, A_n$ are pairwise independent iff

$$\Pr(A_i \cap A_j) = \Pr(A_i) \Pr(A_j)$$

for $i \neq j$ and $i, j = 1, \cdots, n$

Mutual Independence

A set of events $A_1, A_2, \cdots, A_n$ are mutually independent iff for any subset $\{A_{i1}, A_{i2}, \cdots, A_{ik}\}$ of $A_1, A_2, \cdots, A_n$,

$$\Pr(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \Pr(A_{i_1}) \Pr(A_{i_2}) \cdots \Pr(A_{i_k})$$

Note: their complements are also mutually independent.

Concepts of Random Variables

Random Variable

Let S be a sample space associated with the experiment, E. A function X, which assigns a number to every element $s \in S$, is called a random variable.

Discrete Random Variable

If the number of possible values of X(i.e.,R_X,the range space) is **finite or countably infinite**, we call X a discrete random variable.

Probability (Mass) Function

The probability of $X = x_i$ denoted by $f(x_i)$ (i.e. $f(x_i) = \Pr(X = x_i)$, must satisfy the following two conditions.
(1) $f(x_i) \geq 0$ for all x_i.
(2)$\sum_{i=1}^{\infty} f(x_i) = 1$

Continuous Random Variable

The range space R_x is an interval or a range of intervals.

Probability Density Function

Let X be a **continuous** random variable.

1. $f(x) \geq 0$ for all $x \in R_X$
2. $\int_{R_X} f(x)dx = 1$ or $\int_{-\infty}^{\infty} f(x)dx = 1$
   since $f(x) = 0$ for x not in R_X
3. For any c and d such that $c < d$, (i.e. $(c, d) \subset \boldsymbol{R}_X$) ,
   $\Pr(c \leq X \leq d) = \int_c^d f(x)dx$

Cumulative Distribution Function

We define $F(x)$ to be the **cumulative distribution function** of the random variable X (abbreviated as c.d.f.) where

$$F(x) = \Pr(X \leq x)$$

If X is a **discrete** random variable, then its c.d.f is a step function.

$$F(x) = \sum_{t \leq x} f(t)$$
$$= \sum_{t \leq x} \Pr(X = t)$$

If X is a **continuous** random variable, then

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

For a **continuous** random variable X,

$$f(x) = \frac{dF(x)}{dx}$$

if the derivative exists.

Mean

If X is a **discrete** random variable, taking on values $x_1, x_2, \cdots$ with probability function $f(x)$, then the mean or expected value of X, denoted by $E(X)$, is defined by

$$\mu_X = E(X) = \sum_i x_i f(x_i) = \sum_x x f(x)$$

If X is a **continuous** random variable with probability density function $f(x)$, then the mean is defined by

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

For any function $g(X)$,
(a) $E[g(X)] = \sum_x g(x) f_X(x)$
(b) $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)dx$
Property:

$$E(aX + b) = aE(X) + b$$

In general,

$$E[a_1 g_1(X) + a_2 g_2(X) + \cdots + a_k g_k(X)]$$
$$= a_1 E[g_1(X)] + a_2 E[g_2(X)] + \cdots + a_k E[g_k(X)]$$

Variance

$$\sigma_X^2 = V(X) = E\left[(X - \mu_X)^2\right]$$
$$= \begin{cases} \sum_x (x - \mu_X)^2 f_X(x), & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)dx, & \text{if } X \text{ is continuous.} \end{cases}$$

Remarks:

(a) $V(X) \geq 0$
(b) $V(X) = E(X^2) - [E(X)]^2$

Property:

$$V(aX + b) = a^2 V(X)$$

Standard Deviation

The **positive square root** of the variance.

Moment

The **k-th moment** of X is defined by $E(X^k)$.

Chebyshev's Inequality

Let X be a random variable (discrete or continuous) with $E(X) = \mu$ and $V(X) = \sigma^2$. For any positive number k,

$$\Pr(|X - \mu| \geq k\sigma) \leq 1/k^2$$

That is, the probability that the value of X lies at least k standard deviation from its mean is at most $\frac{1}{k^2}$.
Alternatively,

$$\Pr(|X - \mu| < k\sigma) \geq 1 - 1/k^2$$

This is true for **all** distributions with finite mean and variance.

Two-dimensional Random Variables

Definition of 2D RV

Let E be an experiment and S a sample space associated with E.
Let X and Y be two functions each assigning a real number to each $s \in S$.
We call (X, Y) a **two-dimensional random variable**.
(Sometimes called a **random vector**).
The above definition can be extended to n random variables.

Range Space

$$R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$$

The above definition can be extended to more than two random variables.

Discrete vs. Continuous

Discrete: (X, Y) is a two-dimensional **discrete** random variable if the possible values of $(X(s), Y(s))$ are **finite or countably infinite**.
Continuous: (X, Y) is a two-dimensional **continuous** random variable if the possible values of $(X(s), Y(s))$ can assume all values in some region of the Euclidean plane $\mathbb{R}^2$.

Joint Probability Function

Let (X, Y) be a 2-dimensional **discrete** random variable defined on the sample space of an experiment. With each possible value (x_i, y_i), we associate a number $f_{X,Y}(x_i, y_i)$ representing $\Pr(X = x_i, Y = y_i)$ and satisfying the following conditions:
1. $f_{X,Y}(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in R_{X,Y}$.
2.

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$$

The function $f_{X,Y}(x, y)$ is called the **joint probability function** for (X, Y).

$$\Pr((X, Y) \in A) = \underbrace{\sum\sum}_{(x,y) \in A} f_{X,Y}(x, y)$$

Joint Probability Density Function

Let (X, Y) be a 2-dimensional **continuous** random variable assuming all values in some region R of the Euclidean plane $\mathbb{R}^2$. $f_{X,Y}(x, y)$ is called a **joint probability density function** if it satisfies the following conditions:

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

Marginal Probability Distribution

Discrete:

$$f_X(x) = \sum_y f_{X,Y}(x, y) \quad \text{and} \quad f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Continuous:

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

Conditional Distribution

Then the conditional distribution of Y given that $X = x$is given by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}, \quad \text{if } f_X(x) > 0$$

for each x within the range of X. Vice versa for X given Y.

Independent Random Variables

Random variables X and Y are independent iff $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ for all x and y. This can be extended to n random variables where $n \geq 2$.

Expectation

$$E[g(X, Y)] = \begin{cases} \sum_x \sum_y g(x, y) f_{X,Y}(x, y), & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)dxdy, & \text{for cont. RVs} \end{cases}$$

Covariance

Let (X, Y) be a bivariate random vector with joint p.f. (or p.d.f.) $f_{X,Y}(x, y)$ then the covariance of (X, Y) is defined as

$$Cov(X, Y) = E[(x - \mu_x)(y - \mu_y)]$$
$$= \begin{cases} \sum_x \sum_y (x - \mu_x)(y - \mu_y) f_{X,Y}(x, y), & \text{for discrete RVs} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f_{X,Y}(x, y)dxdy, & \text{for cont. RVs} \end{cases}$$

Remarks: 1. $Cov(X, Y) = E(XY) - \mu_x \mu_y$
2. If X and Y are independent, $Cov(X, Y)$ is 0. The converse may not be true.
3. $Cov(aX + b, cY + d) = acCov(X, Y)$
4. $V(aX + bY) = a^2 V(X) + b^2 V(Y) + 2abCov(X, Y)$

Correlation Coefficient

It measures the degree of linear relationship between X and Y.
$-1 \leq \rho_{X,Y} \leq 1$.

The correlation coefficient of X and Y, denoted by $\mathrm{Cor}(X, Y), \rho_{X,Y}$ or ρ, is defined by
$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

Special Probability Distributions

Discrete Uniform Distribution

$$f_X(x) = \frac{1}{k}, for x = x_1, x_2, ..., x_k,$$

and 0 otherwise.

Bernoulli Experiment

A Bernoulli experiment is a random experiment with only two possible outcomes, say "success" or "failure".

Bernoulli Distribution

A random variable X is defined to have a Bernoulli distribution if the probability function of X is given by

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

and 0 otherwise. $Pr(X = 1) = p$ and $Pr(X = 0) = 1 - p = q$.
Mean: $\mu = E(X) = p$
Variance: $\sigma^2 = V(X) = p(1-p) = pq$

Binomial Distribution

A random variable X is defined to have a binomial distribution with two parameters n and p, (i.e. $X \sim B(n,p)$), if the probability function of X is given by

$$\Pr(X = x) = f_X(x) = \left(\begin{array}{c} n \\ x \end{array} \right) p^x(1-p)^{n-x} = \left(\begin{array}{c} n \\ x \end{array} \right) p^x q^{n-x}$$

for $x = 0, 1, ..., n$ and $0 < p < 1$ and $q = 1 - p$.
X is the number of successes that occur in n independent Bernoulli trials.
Mean: $\mu = E(X) = np$
Variance: $\sigma^2 = V(X) = np(1-p) = npq$

Negative Binomial Distribution

Let X be a random variable represents the number of trials to produce the k successes in a sequence of independent Bernoulli trials. The random variable X is said to follow a Negative Binomial distribution with parameters k and p (i.e. $NB(k,p)$). The probability function of X is given by

$$\Pr(X = x) = f_X(x) = \left(\begin{array}{c} x-1 \\ k-1 \end{array} \right) p^k q^{x-k}$$

for $x = k, k+1, k+2, \cdots$
Mean: $\mu = E(X) = \frac{k}{p}$
Variance: $\sigma^2 = V(X) = \frac{(1-p)k}{p^2}$

Geometric Distribution

The number of trials that are required to have the first success is known to follow a special case of negative binomial distribution called **geometric distribution**. Let X be the number of attempts necessary for the first success. Therefore X follows a Negative Binomial Distribution with parameters k = 1 and p. (or X follows a Geometric Distribution with p = 0.05).
That is $X \sim NB(1,p)$ or $X \sim Geom(p)$.

Poisson Distribution

$$f_X(x) = \Pr(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \cdots$$

Mean: $\mu = E(X) = \lambda$
Variance: $\sigma^2 = V(X) = \lambda$

Poisson approx. to Binomial Dist.

Let $X \sim B(n,p)$. Suppose that $n \to \infty$ and $p \to 0$ such that $\lambda = np$ stays constant. Then $X \sim P(np)$ approximately. That is

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

Remark: If p is close to 1, we can still use Poisson distribution to approximate binomial by interchanging success and failure to make p close to zero.

Continuous Uniform Distribution

$X \sim U(a,b)$ over an interval $[a,b]$ if

$$f_X(x) = \frac{1}{b-a}, \quad \text{for } a \le x \le b$$

and 0 otherwise. (a.k.a. rectangular distribution)
Mean: $\mu = E(X) = \frac{a+b}{2}$
Variance: $\sigma^2 = V(X) = \frac{1}{12}(b-a)^2$

Exponential Distribution

$$f_X(x) = \alpha e^{-\alpha x}$$

parameter $\alpha > 0$ and $x > 0$, 0 otherwise. $X \sim Exp(\alpha)$
(frequently used as a model for the distribution of times between the occurrence of successive events)
Mean: $\mu = E(X) = \frac{1}{\alpha}$
Variance: $\sigma^2 = V(X) = \frac{1}{\alpha^2}$
OR:

$$f_X(x) = \frac{1}{\mu}e^{-x/\mu}, \quad \text{for } x > 0$$

Mean: $\mu = E(X) = \mu$
Variance: $\sigma^2 = V(X) = \mu^2$
No Memory Property: for any two positive numbers s and t,

$$\Pr(X > s + t | X > s) = \Pr(X > t)$$

Upper-tailed cdf:

$$\Pr(X > x) = e^{-ax}, \quad \text{for } x > 0$$

Normal Distribution

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

denoted by $N(\mu, \sigma^2)$

Standardised Normal Distribution

If $X \sim N(\mu, \sigma^2)$, and if $Z = \frac{(X-\mu)}{\sigma}$, then $Z \sim N(0,1)$.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

* Linear Interpolation (e.g.)

Let $Pr(Z > a) = 0.12$. From the normal table, we have $Pr(Z \ge 1.17) = 0.121$ and $Pr(Z \ge 1.18) = 0.119$. Hence, $\frac{a-1.17}{1.18-1.17} = \frac{0.12-0.121}{0.119-0.121} \Rightarrow a = 1.175$

Normal approx. to Binomial Dist.

When $n \to \infty$ and $p \to 1/2$. **Rule of thumb**: use this only when $np > 5$ and $n(1-p) > 5$.
If $X \sim B(n,p)$ $(\mu = np, \sigma^2 = np(1-p))$, as $n \to \infty$,

$$Z = \frac{X - np}{\sqrt{npq}} \text{ is approximately } \sim N(0,1)$$

Continuity Correction

(for norm. approx. to binom.)
(a) $Pr(X = k) \approx Pr(k - 1/2 < X < k + 1/2)$
(b) $\Pr(a < X \le b) \approx \Pr(a + 1/2 < X < b + 1/2$
(c) $\Pr(X \le c) = \Pr(0 \le X \le c) \approx \Pr(-1/2 < X < c + 1/2)$
(d) $\Pr(X > c) = \Pr(c < X \le n) \approx \Pr(c + 1/2 < X < n + 1/2)$

Sampling

Statistic and Sampling Distribution

A function of a random sample $(X_1, X_2, ..., X_n)$ is called a statistic (e.g. $\bar{X}$). The probability distribution of a statistic is called a sampling distribution.

Sample Mean

If $(X_1, X_2, ..., X_n)$ represent a random sample of size n, then the sample mean is defined by the statistic

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Sampling Distribution of Sample Mean

For random samples of size n taken from an infinite population or from a finite population with replacement having population mean μ and population standard deviation σ, the sampling distribution of the sample mean $\bar{X}$ has its mean and variance given by

$$\mu_{\bar{X}} = \mu_X \quad \text{and} \quad \sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n}$$

Law of Large Numbers

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2. Then for any $\epsilon \in \mathbb{R}$,

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

Central Limit Theorem

Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a population having any distribution with mean μ and finite population variance σ^2. The sampling distribution of the sample mean $\bar{X}$ is approximately normal with mean μ and variance $\frac{\mu^2}{n}$ if n is sufficiently large (rule of thumb: at least 30). $\bar{X} \sim N(\mu, \frac{\mu^2}{n})$

Sampling distribution of the difference of two sample means

If independent samples of sizes $n_1(\ge 30)$ and $n_2(\ge 30)$ are drawn from two populations, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, then the sampling distribution of the differences of the sample means, $\bar{X}_1$and $\bar{X}_2$, is approximately normally distributed with mean and standard deviation given by

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \text{ and } \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Chi-square Distribution

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

and 0 otherwise. $Y \sim \chi^2(n)$ with n degrees of freedom (n is a positive integer).
Mean: $\mu = E(Y) = n$
Variance: $\sigma^2 = V(Y) = 2n$
(1) For large n, $\chi^2(n)$ approx $\sim N(n, 2n)$

(2) If $Y_1, Y_2, ..., Y_k$ are **independent** chi-square random variables with $n_1, n_2, ..., n_k$ degrees of freedom respectively, then

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

Theorem regarding Chi-square and random sample

1. If $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$.
2. Let $X \sim N(\mu, \sigma^2), then [(x-\mu)/\sigma]^2 \sim \chi^2(1)$.
3. Let $(X_1, X_2, ..., X_n)$ be a random sample of size n from a normal population with mean μ and variance σ^2. Define

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

Then $Y \sim \chi^2(n)$

Sample Variance

Let $X_1, X_2, ..., X_n$ be a random sample from a population. Sample variance:

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

Sampling distribution of $(n-1)S^2/\sigma^2$

If S^2 is the variance of a random sample of size n taken from a **normal** population having the variance σ^2, then

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

t-distribution

Suppose $Z \sim N(0,1)$, and $U \sim \chi^2(n)$. If Z and U are independent, then

$$\text{let } T = \frac{Z}{\sqrt{U/n}} \sim t(n)$$

(t-distribution with n degrees of freedom)
Its p.d.f. is given by:

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}, \quad -\infty < t < \infty$$

Mean: $\mu = E(T) = 0$
Variance: $\sigma^2 = V(T) = n/(n-2)$ for $n > 2$
Remark: If the random sample was selected from a normal population, then

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

and

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

It can be shown that $\bar{X}$ and S^2 are independent, and so are Z and U. Therefore,

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}}$$
$$= \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1}$$

T (t-value) has a t-distribution with $n-1$ degrees of freedom.

F-distribution

Let U and V be independent random variables having $\chi^2(n_1)$ and $\chi^2(n_2)$, respectively, then the distribution of the random variable,

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

is called an F-distribution with (n_1, n_2) degrees of freedom. Its p.d.f. $f_F(x) > 0$ for $x > 0$ and 0 otherwise.
Mean: $\mu = E(X) = n_2/(n_2 - 2)$ with $n_2 > 2$
Variance: $\sigma^2 = V(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)}$ with $n_2 > 4$
Remarks:
(1) Suppose that random samples of sizes n_1 and n_2 are selected from two normal populations with variances σ_1^2 and σ_2^2 respectively.

$$U = \frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1 - 1)$$
$$V = \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2 - 1)$$

are independent random variables. Therefore,

$$F = \frac{U/(n_1-1)}{V/(n_2-1)} = \frac{\frac{(n_1-1)S_1^2/\sigma_1^2}{(n_1-1)}}{\frac{(n_2-1)S_2^2/\sigma_2^2}{(n_2-1)}}$$
$$= \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

(2) If $F \sim F(n, m)$, then $1/F \sim F(m, n)$.
(3) $F(n_1, n_2; 1-\alpha) = 1/F(n_2, n_1; \alpha)$ (useful for statistical table)

Estimation based on Normal Distribution

Point Estimation

Point estimation is to use the value of some statistic, say $\hat{\theta} = \widehat{\Theta}(X_1, X_2, \cdots, X_n)$, to estimate the unknown parameter θ; such a statistic $\widehat{\Theta}$ is called a **point estimator**. (Note: a **statistic** does not depend on any unknown parameters)
The statistic that one uses to obtain a point estimate is called an **estimator**. e.g. $\bar{X}$ is an estimator for μ.

Interval Estimation

Interval estimation is to define two statistics, say, $\widehat{\Theta}_L$ and $\widehat{\Theta}_U$ where $\widehat{\Theta}_L < \widehat{\Theta}_U$, so that $(\widehat{\Theta}_L, \widehat{\Theta}_U)$ constitutes a random interval for which the probability of containing the unknown parameter θ can be determined.

Unbiased Estimator

A statistic $\widehat{\Theta}$ is said to be an unbiased estimator of the parameter θ if $E(\widehat{\Theta}) = \theta$. For example:
$E(\bar{X}) = \mu$
$E(S^2) = \sigma^2$

Margin of Error e

We want $\Pr(|\bar{X} - \mu| \leq e) \geq 1 - \alpha$.

$$e \geq z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

Hence, for a given margin of error e, the sample size is given by

$$n \geq \left(z_{\alpha/2}\frac{\sigma}{e}\right)^2$$

Confidence Interval

Suppose $(\widehat{\Theta}_L, \widehat{\Theta}_U)$ and $Pr(\widehat{\Theta}_L < \theta < \widehat{\Theta}_U) = 1 - \alpha$. Then this interval is called a $(1-\alpha)100\%$ **confidence interval** for θ. $(1-\alpha)$ is called **confidence coefficient** or **degree of confidence**. The end points $\widehat{\Theta}_L$ and $\widehat{\Theta}_U$ are called lower and upper **confidence limits** respectively.

3 conditions:
A. normal distributions
B. other parameters known
C. large sample sizes

One sample; CI for pop. mean μ

(1) B(known σ^2) AND (A OR C):
Test statistic: $Z = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}} \sim N(0,1)$
$(1-\alpha)100\%$(same below) CI: $\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
(2) $\neg B$(unknown σ^2) AND A:
Test statistic: $T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}} \sim t(n-1)$
CI:$\bar{X} \pm t_{n-1;\alpha/2}\frac{S}{\sqrt{n}}$
(3) $\neg B$(unknown σ^2) AND C:
Test statistic: $T = \frac{(\bar{X}-\mu)}{S/\sqrt{n}}$ approx $\sim N(0,1)$
CI: $\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$

Two indep. samples; CI for diff. b/w 2 means

(1) B(known and unequal σ_1^2, σ_2^2) AND (A OR C):
Test statistic: $T = \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1)$
CI: $\bar{X}_1 - \bar{X}_2 \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
(2) $\neg B$(unknown σ_1^2, σ_2^2) AND C:
Same as (1), just replace σ_1^2, σ_2^2 with S_1^2, S_2^2
(3) $\neg B$(unknown but EQUAL σ_1^2, σ_2^2) AND A:
Test statistic: $T = \frac{\bar{X}_1-\bar{X}_2-(\mu_1-\mu_2)}{\sqrt{S_p^2(1/n_1+1/n_2)}} \sim t(n_1+n_2-2)$
where σ_2 can be estimated by the pooled sample variance:
$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$
CI: $\bar{X}_1 - \bar{X}_2 \pm t_{n_1+n_2-2;\alpha/2}\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$ (note: for large sample sizes, replace $t_{n_1+n_2-2;\alpha/2}$ by $z_{\alpha/2}$.

Two paired /dep. samples; CI for diff. b/w 2 means

[Just treat the differences as a sample itself]
(1) $\neg B$ AND A:
Test statistic: $T = \frac{\bar{X}_D-\mu_D}{S_D/\sqrt{n}} \sim t(n-1)$
CI: $\bar{X}_D \pm t_{n-1;\alpha/2}\frac{S_D}{\sqrt{n}}$
(2) $\neg B$ AND C:
Test statistic: $T = \frac{\bar{X}_D-\mu_D}{S_D/\sqrt{n}}$ approx $\sim N(0,1)$
CI: $\bar{X}_D \pm z_{\alpha/2}\frac{S_D}{\sqrt{n}}$

One sample; CI for pop. variance σ^2

[note: for σ, just take square roots on both ends of CI]
(1) B(known μ) AND A:
Test statistic: $T = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$
CI: $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$
(2) $\neg B$(unknown μ) AND A:

Test statistic: $T = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
CI: $\frac{(n-1)S^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$

Two indep. samples; CI for ratio of 2 variances

[note: for σ_1/σ_2, just take square roots on both ends of CI]
(1) $\neg B$(unknown means) AND A:
Test statistic: $T = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$
CI: $\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\alpha/2}$

Hypothesis Testing

Type I Error

Pr(reject H_0 given H_0 is true) $= \alpha$.
i.e. Pr(Type I error) = **level of significance**

Type II Error

Pr(**do not** reject H_0 given H_0 is false) = Pr(Accept H_0 given H_1) $= \beta$.
[Note: Pr(**reject** H_0 given H_0 is false) $= 1-\beta$ is called the **power** of the test].

Steps of Hypo-testing

Step 1: Let μ be the mean of ...
Test $H_0: \mu = x$ against $H_1: \mu \neq x$.
Step 2: Set $\alpha = 0.05$
Step 3: State the test statistic used e.g. $Z = \frac{(\bar{X}-\mu_0)}{\sigma/\sqrt{n}}$
$z_{\alpha/2} = z_{0.025} = 1.96$
State the critical region(s).
Step 4: Substitute in values and calculate the value of the test statistic. z = ... = y.
[OR: H_0 is accepted if confidence interval covers μ_0]
[OR: p-value approach: H_0 is rejected if p-value $< \alpha$]
Step 5: Conclusion: Since the observed z value = y falls inside the critical region, H_0 is rejected at 5% level of significance.

p-value

Probability of obtaining a test statistic more extreme than the observed sample value **given** H_0 **is true**. Also called observed level of significance. (Remember to multiply probability by 2 for two-tailed tests)

Prepared by Zechu, AY2019/2020 Semester 1