ST2334 Finals
by Zong Xun

Events

Sample space is the set of all possible outcomes. An **event** is a subset of the sample space. The sample space is denoted as the **sure event**, and the empty set, $\emptyset$ is denoted as the **null event**.

Event Relationships

- **Mutually exclusive** if $A \cap B = \emptyset$.
- If $A \subset B$, then B **contains** A.
- If $B \subset A$ and $A \subset B$, then A and B are **equivalent**.

Important Properties

1. Distributive laws
2. $A \cup B = A \cup (B \cap A')$
3. De Morgan's laws
4. $A \cap B' = A - (A \cap B)$
5. If $(A \cap B') \cup (A' \cap B) = \emptyset \Rightarrow A = B$

Counting Methods

Multiplication is used when there are n **different** events(independent) to be performed **sequentially**.
Addition is used when the same event can be performed by k different procedure(mutually exclusive).

Permutation

The number of ways to arrange r objects out of n.

$$P_r^n = \frac{n!}{(n-r)!} = \binom{n}{r} \times P_r^r$$

- Permutations around a circle = $(n-1)!$
- Permutations when not all objects are distinct = $\frac{n!}{n_1!n_2!\ldots n_k!}$

Combination

A selection of r objects out of n, without regard to the order.

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \binom{n}{n-r} = \binom{n-1}{r} + \binom{n-1}{r-1}$$

Probability

$$Pr(A) = lim_{n\to\infty} f_A$$

An interpretation of probability is relative frequency.

Propositions

1. The probability of the empty set, $P(\emptyset) = 0$.
2. If $A_1, \ldots, A_n$ are mutually exclusive events ($A_i \cap A_j$ for any $i \neq j$), then
$$P(A_1 \cup \cdots \cup A_n) = P(A_1) + \cdots + P(A_n)$$
3. For any two events A and B,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
4. If A is contained by B, then $P(A) \leq P(B)$

Conditional Probability

For any two events A and B with P(A) > 0, the conditional probability of B given that A has occurred is defined by

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

Two events are said to be **independent** if and only if $P(A \cap B) = P(A)P(B)$. If A and B are independent, then

- their conditional probability, $P(A|B) = P(A)$.
- $A \perp B', A' \perp B$, and $A' \perp B'$.
- A and B cannot be mutually exclusive, if $P(A), P(B) > 0$
- All events are independent with S and $\emptyset$.

Law of total probability

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

Assuming that events $A_1, \ldots, A_n$ are mutually exclusive and exhaustive events.

Bayes' Theorem

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

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

Bayes Theorem can be derived using conditional probability, multiplication rule and the law of total probability.

Birthday Problem

P(>2 people have same birthday) = $Pr(n) = 1 - Pr(q)$, where $Pr(q)$ is the probability that every person in the room has a different birthday i.e $\frac{365(364)\ldots(365-n+1)}{365^n}$.

Inverse Birthday Problem

P(someone has the same birthday as you) = $Pr(n) = 1 - Pr(q) \geq 0.5$, where Pr(q) is the probability of someone having a different birthday i.e. $(\frac{364}{365})^n$

Monty Hall Problem

Suppose you chose to switch, W = {Win the car}
A = {Car is behind the door of initial pick}

$$P(W) = P(A)P(W|A) + P(A')P(W|A') = \frac{1}{3}(0) + \frac{2}{3}(1) \quad = \frac{2}{3}$$

Random Variables

A function X, which assigns a real number to every $s \in S$ is called a random variable. Range space,

$$R_x = \{x | x = X(s), s \in S\}$$

Likewise, the set $X \in A$, for A being a subset of $\mathbb{R}$, is also a subset of S: $\{s \in S : X(s) \in A\}$.

Discrete Probability Distribution

Each value of X has a certain probability, $f(x)$, and this function f is called the **probability mass function**. It must satisfy the following:

1. $0 \leq f(x_i) \leq 1$ for all $x_i \in R_x$, otherwise 0.
2. Sum of the probabilities, $\sum_{i=1}^{\infty} f(x_i) = 1$

Continuous Probability Distribution

R_x is an **interval** or a collection of intervals. The **probability density function** quantify the probability that X is in a certain range. It must satisfy the following:

1. $f(x) \geq 0$ for all $x \in R_x$, otherwise 0.
2. **No need** for $f(x) \leq 1$ (consider area under curve).
3. Pr(A) = 0 $\not\Longrightarrow$ A is $\emptyset$. Consider $A \notin R_x$ or (5).
4. Sum of probabilities, $\int_{Rx} f(x)dx = 1$
5. $P(a \leq X \leq b) = \int_a^b f(x)dx \implies P(X = a) = 0$.

Note: $P(A < X < B) = P(A \leq X \leq B)$

Cumulative Probability Distribution

For any random variable X, define its c.d.f by

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

The cumulative probabilities is ≥ 0 and ≤ 1.
For the **discrete** case, the c.d.f is a step function. For any two numbers $a < b$, we have

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

For the **continuous** case,

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

Some remarks:

- $F(X)$ is non-decreasing i.e. $x_1 < x_2, F(X_1) \leq F(X_2)$.
- p.f and c.d.f have a one-to-one correspondence.
- c.d.f have to be **right continuous**

Expectation and Variance

For discrete,

$$E(X) = \sum_{x_i \in R_x} x_i f(x_i) \quad Var(X) = E(g(x))$$

where $g(x) = (x - \mu_x)^2$. Similarly, in the continuous case,

$$E(X) = \int_{x_i \in R_x} x f(x)dx \quad Var(X) = E(g(x))$$

where $g(x) = (x - \mu_x)^2$.

Properties of Expectation and Variance

1. Let X be a positive integer-valued (excluding 0) random variable. (tut 4, qn 8)
$$E(X) = \sum_{k=1}^{\infty} P(X \geq k)$$
2. There exists probability distributions for which E(X) do not exists, i.e. $E(X) = \infty$
3. Variance = 0, if $P(X = E(X))$ = 1

Joint Distributions

Let (X, Y) be a 2-dimensional random variable, each assigning a real number to **every outcome** $\in$ **S**. Consider the case where both are **discrete**, or both are **continuous**.

Discrete case

The **joint probability function** is defined by

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

for $(x, y) \in R_{X,Y}$.
Properties:

- $f_{X,Y}(x, y) \geq 0$ for any $(x, y) \in R_{X,Y}$, otherwise 0.
- Sum of probabilities, $\sum\sum_{(x,y)\in R_{X,Y}} f(x, y) = 1$.
- Let A be any subset of $R_{X,Y}$, then
$$P((X, Y) \in A) = \sum\sum_{(x,y)\in A} f(x, y)$$

Continuous case

The **joint probability (density) function** is defined by

$$P((X, Y) \in D) = \int\int_{(x,y)\in D} f_{X,Y}(x, y)dydx$$

for any $D \subset \mathbb{R}^2$.
Properties:

- $f_{X,Y}(x, y) \geq 0$ for any $(x, y) \in R_{X,Y}$, otherwise 0.
- Sum of probabilities, $\int\int f(x, y)dxdy = 1$.

Marginal Probability Distributions

Let (X,Y) be a RV with joint probability function, $f_{X,Y}(x, y)$. We define the marginal distribution for X as:

- If Y is a **discrete** RV, then for any x,
$$f_X(x) = \sum_y f_{X,Y}(x, y)$$
- If Y is a **continuous** RV, then for any x,
$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

Note: We can view the marginal distribution as the distribution of X **ignoring the presence of Y**, $P(X = x)$.

Conditional Distribution

Let (X,Y) be a RV with joint p.f. $f_{X,Y}(x, y)$. Let $f_X(x)$ be the marginal p.f. for X. Then for any x such that $f_X(x) > 0$, the **conditional probability function of Y given X = x** is defined to be

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

Note:

1. $f_{Y|X}(y|x)$ is defined only for x such that $f_X(x) > 0$.
2. Considering y as the variable, $f_{Y|X}(y|x)$ is a p.f., so it should satisfy all the properties of p.f.
3. This is **not a p.f** for x.
4. $E(Y|X = x)$ is dependent on the conditional probability.

Independent Random Variables

X and Y are independent if and only if for any x and y

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

Conditions:

1. $R_{X,Y}$ needs to be a product space, $f_{X,Y}(x,y) > 0$
2. Can be factorized to

$$f_{X,Y}(x,y) = C \cdot g_1(x) \cdot g_2(y)$$

where $g_1(x)$ depends on x, $g_2(y)$ depends on y, and C is a constant. Note: g_1 and g_2 are **NOT necessarily** p.f.s.
If X is a **discrete** RV, its p.m.f is given by

$$f_X(x) = \frac{g_1(x)}{\sum_{t \in R_X} g_1(t)}$$

If X is a **continuous** RV, its p.d.f. is given by

$$f_X(x) = \frac{g_1(x)}{\int_{t \in R_X} g_1(t)}$$

Properties

Suppose X, Y are independent RVs,

1. If A and B are arbitrary subsets of $\mathbb{R}$, the events $X \in A$ and $Y \in B$ are independent events in S. Then,

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

2. $g_1(X)$ and $g_2(Y)$ are independent.
3. If $f_X(x) > 0$, then $f_{Y|X}(y|x) = f_Y(y)$

Covariance and Expectation

- if (X,Y) is a **discrete** RV,

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

- if (X,Y) is a **continuous** RV,

$$E(g(X,Y)) = \int\int g(x,y) f_{X,Y}(x,y) dy dx$$

For **covariance**, replace $g(X,Y)$ with

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

Properties

1. $cov(X,Y) = E(XY) - E(X)E(Y)$
2. If X and Y are independent, then $cov(X,Y) = 0$.
3. $cov(aX+b, cY+d) = ac \cdot cov(X,Y)$
4. $V(aX+bY) = a^2V(X) + b^2V(Y) + 2ab \cdot cov(X,Y)$

Special Probability Distributions

Discrete Distributions

If X assumes the values $x_1, \dots, x_k$ with **equal probability**, then X follows a **discrete uniform distribution**.

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

$$E(X) = \frac{1}{k}\sum_{i=1}^k x_i \quad V(X) = (\frac{1}{k}\sum_{i=1}^k x_i^2) - \mu_X^2$$

Bernoulli Distribution

A **Bernoulli Trial** has two possible outcomes. The number of successes in a Bernoulli trial, $X \sim Bernoulli(p)$.
A **Bernoulli process** consists of a sequence of repeatedly performed i.i.d Bernoulli trials.

$$f_X(x) = p^x(1-p)^{1-x}, \text{ for } x = 0 \text{ or } 1$$

$$E(X) = p \quad V(X) = p(1-p)$$

Remark: The collection of the distributions that are determined by one or more unknown **parameters** is called a family of probability distributions.

Binomial Distribution

Let X be number of successes in n **bernoulli trials** where

1. the probability of success for each trial is the same p,
2. the trials are **independent**

then X follows a binomial distribution, $X \sim B(n,p)$. Alternatively, we say

$$X = X_1 + X_2 + \cdots + X_n$$

with $X_1, \dots, X_n$ being i.i.d. $Bernoulli(p)$ RVs.

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

$$E(X) = np \quad V(X) = np(1-p)$$

Remark: Bernoulli is a special case of Binomial, where $n = 1$. If X, Y are i.i.d RV, then $X + Y \sim B(2n, p)$

Negative Binomial Distribution

Let X be number of i.i.d. Bernoulli trials until the kth success occurs, $X \sim NB(k,p)$

$$P(X=x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}, \text{ for } x = 0, 1, \dots, n$$

$$E(X) = k/p \quad V(X) = (1-p)k/p^2$$

Remark: Geometric is a special case of NB, where $k = 1$.

Poisson Distribution

Let X be number of occurrences in a **fixed period of time or fixed region**, $X \sim Poisson(\lambda)$ where $\lambda > 0$ is the expected number of occurrences.

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!} \quad E(X) = \lambda = V(X)$$

Poisson Process is a continuous time process. The properties of a *Poisson Process* with rate parameter α:

1. An interval of length T has $E(X) = \alpha T$
2. there are no simultaneous occurrences
3. the number of occurrences in disjoint time intervals are **independent**

Remark: If X,Y are i.i.d. RVs, then $X + Y \sim Poi(\lambda_1 + \lambda_2)$

Poisson approximation of Binomial Distribution
Suppose $n \to \infty$ and $p \to 0$ such that $\lambda = np$ remains a constant. Then, $X \sim Poisson(np)$.
This approximation is valid when $n \geq 20$ and $p \leq 0.05$, or if $n \geq 100$ and $np \leq 10$.

Continuous Distribution

A random variable X is said to follow a uniform distribution over the interval (a,b) if its p.d.f is given by

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

$$E(X) = \frac{a+b}{2} \quad V(X) = \frac{(b-a)^2}{12}$$

and it c.d.f is given by:

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

Exponential Distribution

Exponential distribution with parameter $\lambda > 0$ if its p.d.f. is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases}$$

It has $E(X) = \frac{1}{\lambda}$ and $V(X) = \frac{1}{\lambda^2} = E(X)^2$. The c.d.f of $X \sim Exp(\lambda)$ is given by

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

Properties

- **Memory-less:** If $s, t > 0$, then

$$P(X > x_2 | X > x_1) = P(X > x_2 - x_1)$$

- For any real numbers $x_2 > x_1 > 0$, we must have $P(X > x_1) > P(X > x_2) > 0$.

Normal Distribution

The p.d.f. of normal distribution is **positive** over the whole real line, **symmetric** about x = μ, and **bell-shaped**. Its p.d.f is given by

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

its c.d.f is given by

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx.$$

Standardization:

$$X \sim N(\mu, \sigma^2) \implies Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Properties:

1. $P(Z \geq 0) = P(Z \leq 0) = 0.5$
2. $P(Z \leq z) = P(Z \geq -z) = P(-Z \leq z) = 1 - F(-z)$
3. By symmetry, $-Z \sim N(0,1)$.
4. If $Z \sim N(0,1)$ then $\sigma Z + \mu \sim N(\mu, \sigma^2)$
5. $P(|Z| < z) = 2F(z) - 1$

Note: By (3), $E(X^{2k+1}) = 0$ for $k = 0, 1, \dots, n$ (T9Q1).

Normal approximation to Binomial Distribution

Let $X \sim B(n,p)$. $E(X) = np$ and $V(X) = np(1-p)$. Then as $n \to \infty$,

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} \sim N(0,1)$$

This approximation is good when $np > 5$ and $n(1-p) > 5$.

Continuity Correction

1. $P(X = k) \approx P(k - 1/2 < X < k + 1/2)$
2. $P(a \leq X \leq b) \approx P(a - 1/2 < X < b + 1/2)$
3. $P(a < X \leq b) \approx P(a + 1/2 < X < b + 1/2)$
4. $P(a \leq X < b) \approx P(a - 1/2 < X < b - 1/2)$
5. $P(a < X < b) \approx P(a + 1/2 < X < b - 1/2)$
6. $P(X \leq c) = P(0 \leq X \leq c) \approx P(-1/2 < X < c + 1/2)$
7. $P(X > c) = P(c < X \leq n) \approx P(c + 1/2 < X < n + 1/2)$

Sampling Distribution

The number of possible **observations** of an experiment is called a **population**. A **sample** is a subset of a population $\to$ **infinite** if the # of obs. is infinite.

Random Sampling

A random sample is **unbiased** and **models** the population.

(**Finite**) **Simple random sample**: every subset of n observations has the same probability of being selected, $\binom{N}{n}$ possible samples. Note: **sampling with replacement** is infinite and random only if (1) $P(selected)$ is same (2) draws are independent.

(**Infinite**) Let X be a random variable with certain probability distribution $f_X(x)$. Let $X_1, X_2, \dots, X_n$ be n independent random variables each having the same distribution as X. Then $(X_1, X_2, \dots, X_n)$ is called a random sample of size n. The *joint probability function* is given by:

$$f_{X1,\dots,X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \dots f_{X_n}(x_n)$$

where $f_X(x)$ is the probability function of the population.

Sample Distribution of Sample Mean

A function of $(X_1, \dots, X_n)$ is called a **statistic**. **There cannot be any unknown variables**. The sampling distribution of the sample mean has

$$E(\bar{X}) = \mu_x \quad V(\bar{X}) = \frac{\sigma_X^2}{n}$$

For an **infinite** population, when $n \to \infty$, $V(\bar{X}) \to 0$. The accuracy of $\bar{X}$ as an estimator of μ_X keeps improving. This is known as the **law of big numbers**, $P(|\bar{X} - \mu| > \epsilon) \to 0$ as n $\to \infty$. The **standard error** of $\bar{X}$ describes how much it varies from sample to sample of size n.

Central Limit Theorem

CLT does not make any assumptions about the underlying distribution, it only applies to the sampling distribution. CLT does not apply if $X_1, X_2, \dots, X_n$ are i.i.d. N(μ, σ^2) since $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ regardless of size n (assume σ is **known**).

Chi-squared Distribution

Let $Z_1, \ldots, Z_n$ be **i.i.d. standard normal** random variables. Then,

$$\sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$$

with n degrees of freedom.

Properties

1. If $Y \sim \chi^2(n)$, then $E(Y) = n$ and $V(Y) = 2n$.
2. For large n, $\chi^2(n)$ is approximately $N(n, 2n)$.
3. Linearity of freedom. $Y_1 + Y_2 \sim \chi^2(m+n)$
4. Right skewed.

Sampling Distribution of Sample Variance

Let S^2 be the variance of a random sample taken from a normal population with variance σ^2, then

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

has a χ^2 distribution with $n-1$ df. It can be shown that:

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

With the above, we derive $E(S^2) = \sigma^2$. (C5P17) Note that $E(S) \neq \sigma$. The derivation is out of scoped.

T Distribution

Suppose $Z \sim N(0,1)$ and $U \sim \chi^2(n)$ and are **independent**, then $T = \frac{Z}{\sqrt{U/n}} \sim t_n$.

$$E(T) = 0 \quad V(T) = \frac{n}{n-2} \text{ for } n > 2$$

.

If $X_1, \ldots, X_n$ are i.i.d RVs $\sim N(\mu, \sigma^2)$, then $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$ (σ is **unknown**). When $n \geq 30$, we can replace it with N(0, 1).

F Distribution

Suppose $U \sim \chi^2(m)$ and $V \sim \chi^2(n)$ are **independent**. Then the distribution of the random variable

$$F = \frac{U/m}{V/n}$$

is called a F-distribution with (m, n) degrees of freedom.

$E(X) = \frac{n}{n-2}$ for $n > 2$, $V(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ for $n > 4$.

Properties

- If $F \sim F(n, m)$, then $1/F \sim F(m, n)$
- $P(F > F(m, n; \alpha)) = \alpha$ where $F \sim F(m, n)$
- $F(m, n; 1-\alpha) = 1/F(n, m; \alpha)$ (note the swap in df)

Estimation of population parameters

A **point estimate** estimates the population parameter.
A **point estimator** is the method that describes this calculation e.g. $\bar{X}$ is the estimator and $\bar{x}$ = 5 is the estimate.
An **interval estimation** forms an interval in which the parameter is estimated to lie.

Unbiased estimator

Let $\hat{\theta}$ be an estimator of θ. If $E(\hat{\theta}) = \theta$, we say that $\hat{\theta}$ is an unbiased estimator of θ. As $V(\hat{\theta})$ decreases, the estimation gets better.

Maximum Error of Estimate

$E = \bar{X} - \mu$ measures the difference between the estimator and the parameter value. The maximum error, $E \sim N(0,1)$, where

$$P(|\bar{X} - \mu| \leq z_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}) = 1 - \alpha$$

Confidence Interval

An interval (a,b) in which you are fairly certain the parameter of interest lies in. This can be quantified by the **degree of confidence**, $P(a < \mu < b) = 1 - \alpha$. The general formula is given by $\bar{X} \pm E_0$.
Note: This is a probability statement about the **procedure** by which we compute the interval, not the value of μ itself.

Hypothesis Testing

Test statistic quantify how unlikely it is to observe the sample, given that the null hypothesis is true.
P-value is the probability of obtaining a test statistic at least as extreme as the observed sample value, given H_0 is true. It **does not** refer to a specific realization.
The outcome is either to **reject or not reject** H_0. In the latter case, we did not "prove" that H_0 is true.
Types of error

- **Type 1 error** - we reject H_0 when H_0 is true.
- **Type 2 error** - we do not reject H_0 when H_0 is false.

Type 1 error is also known as the **level of significance**, α = P(Type 1 error). Let β = P(Type 2 error). Then, $1-\beta$ is the **power of the test**, a.k.a. true negative rate. If the α decreases, likelihood of not rejecting H_0 increases $\implies$ power of the test decreases
Note: We can compute Type 1 error using the formula $P(\text{reject } H_0 \mid H_0 \text{ is true})$. We cannot compute Type 2 errors unless we have a specific realization of the alternative hypothesis, i.e. $F(H_1 = ?)$.

Single Sample Hypothesis Testing

Known variance

We consider the case where (1) n is sufficiently large, or (2) the underlying distribution is normal.

1. Perform standardization, $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$.
2. For two sided tests, reject H_0 if $2P(Z > |z_{\alpha/2}|) < \alpha$.

Unknown variance

We consider the case where the underlying distribution is **normal** and sample size is **small**.

1. Perform standardization, $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$.
2. For two sided tests, reject H_0 if $P(|T| > t_{n-1,\alpha/2}) < \alpha$

Note: if n ≥ 30, then $T \sim N(0,1)$

Two Sided Tests and Confidence Intervals

Two sided hypothesis test is equivalent to finding a $100(1-\alpha)\%$ confidence interval for μ. If the confidence interval contains μ_0, H_0 will not be rejected at level of significance, α.

Two independent samples

Known/Unknown but unequal variance

We consider the case where the underlying distribution is **normal**, or the value n_1, n_2 are **sufficiently large** (≥ 30).

$$E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2, var(\bar{X} - \bar{Y}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

1. Perform standardization, $Z = \frac{(\bar{X}-\bar{Y})-\delta_0}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$.
2. For two sided tests, reject H_0 if $Z > z_{\alpha/2}$

Note: Null hypothesis, $H_0: \mu_1 - \mu_2 = \delta_0$. When variance is unknown, replace σ_1, σ_2 with sample variance s_1, s_2 instead

Unknown but equal variance

We can roughly assumed the equal variance if the ratio, S_1/S_2, is between $1/2$ and 2. We consider the case where the underlying distribution is normal, but n_1, n_2 are **small** (≤ 30).

1. Perform standardization,

$$T = \frac{(\bar{X}-\bar{Y})-\delta_0}{S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

2. For two sided tests, reject H_0 if $T > t_{n_1+n_2-2;\alpha/2}$

Note: S_p refers to the pooled variance,

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

When $n1, n2 \geq 30$, we can replace t-distribution with normal. Replace S_p with σ when variance is known.

Paired/Dependent samples

We consider the case where n is small and **population** is normally distributed.

1. Perform standardization, $T = \frac{\bar{D}-\mu_D 0}{S_D/\sqrt{n}} \sim t_{n-1}$.
2. For two sided tests, reject H_0 if $Z > z_{\alpha/2}$

Note: if n ≥ 30, then $T \sim N(0,1)$

Tips

- Density manipulation technique for deriving mean and variances (C4P15).
- $A \cap B \neq \emptyset, A \cap C \neq \emptyset \centernot\implies A \cap B \cap C \neq \emptyset$.
- Given $E(X)$, derive $E(X(X-1))$ to derive $V(X)$.
- The normality of the sample distribution makes no assumptions about the distribution of the samples itself.
- $E(X) = E(X+1-1) = E(X+1) - 1$
- For any +ve integer x, $P(X \leq x) = 1 - q^x$ is the probability that at most x tries are needed, given that q is the probability of failure.

Summary

	Population	σ	n	Statistic	Error	Sample size
I	Normal	known	any	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$(\frac{z_{\alpha/2} \cdot \sigma}{E_0})^2$
II	any	known	large	$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$	$Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$	$(\frac{z_{\alpha/2} \cdot \sigma}{E_0})^2$
III	Normal	unknown	small	$T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	$T_{n-1;\alpha/2} \cdot \frac{S}{\sqrt{n}}$	$(\frac{t_{n-1;\alpha/2} \cdot s}{E_0})^2$
IV	any	unknown	large	$Z = \frac{\bar{X}-\mu}{S/\sqrt{n}}$	$Z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$	$(\frac{z_{\alpha/2} \cdot s}{E_0})^2$

Code Snippets

- **Binomial**
  `pbinom(x, n, p, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dbinom(x, n, p)` gives $P(X = x)$.
- **Negative Binomial**
  `pnbinom(x-k,k,p,lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dnbinom(x-k, k, p)` gives $P(X = x)$.
- **Poisson**
  `ppois(x, lambda, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dpois(x, lambda)` gives $P(X = x)$.
- **Exponential**
  `pexp(x, lambda, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dexp(x, lambda)` gives $P(X = x)$.
- **Normal**
  `pnorm(x, mu, sigma)` $\implies$ P(X $\leq x$).
  `dnorm(x, mu, sigma)` gives $P(X = x)$.
  `qnorm(quantile, mu, sigma)` $\implies$ $x_{1-\alpha}$.
- **Chi Squared**
  `pchisq(x, df, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dchisq(x, df)` gives $P(X = x)$.
  `qchisq(quantile, df)` $\implies$ $x_{1-\alpha}$.
- **T distribution**
  `pt(x, mu, sigma, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `dt(x, mu, sigma)` gives $P(X = x)$.
  `qt(quantile, mu, sigma)` $\implies$ $x_{1-\alpha}$.
- **F distribution**
  `pf(x, df1, df2, lower.tail=TRUE)` $\implies$ P(X $\leq x$).
  `df(x, df1, df2)` gives $P(X = x)$.
  `qf(quantile, df1, df2)` $\implies$ $x_{1-\alpha}$.