Mathematical statistics

May 13th, 2019

Review

Att.

- Final exam:

  Wednesday, 5/29/2019, 10:30am –12:30pm

  Ewing Hall Room 101

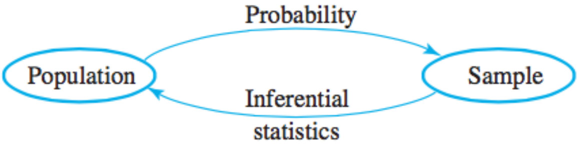
- Course evaluation

Overview

Week 1	Probability reviews
Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	Chapter 7: Point Estimation
Week 7	Chapter 8: Confidence Intervals
Week 10	Chapter 9, 10: Test of Hypothesis

Chapter 6: Statistics and Sampling Distributions

Random sample



Definition

The random variables $X_1, X_2, ..., X_n$ are said to form a (simple) random sample of size n if

1. the X_i's are independent random variables
2. every X_i has the same probability distribution

1. If the distribution and the statistic T is simple, try to construct the pmf of the statistic
2. If the probability density function $f_X(x)$ of X's is known, the
   - try to represent/compute the cumulative distribution (cdf) of T
     $$\mathbb{P}[T \leq t]$$
   - take the derivative of the function (with respect to t)

Section 6.3: Linear combination of normal random variables

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent normal random variables (with possibly different means and/or variances). Then

$$T = a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

also follows the normal distribution.

If X has a normal distribution with mean μ and standard deviation σ, then

$$Z = \frac{X-\mu}{\sigma}$$

has a standard normal distribution. Thus

$$\begin{aligned} P(a \le X \le b) &= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \end{aligned}$$

$$P(X \le a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \quad P(X \ge b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Theorem

Let $X_1, X_2, \ldots, X_n$ be independent random variables (with possibly different means and/or variances). Define

$$T = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n,$$

then the mean and the standard deviation of T can be computed by

- $E(T) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$
- $\sigma_T^2 = a_1^2 \sigma_{X_1}^2 + a_2^2 \sigma_{X_2}^2 + \ldots + a_n^2 \sigma_{X_n}^2$

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then, in the limit when $n \to \infty$, the standardized version of $\bar{X}$ have the standard normal distribution

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right) = \mathbb{P}[Z \leq z] = \Phi(z)$$

Rule of Thumb:

If $n > 30$, the Central Limit Theorem can be used for computation.

Chapter 7: Point Estimation

Chapter 7: Point estimates

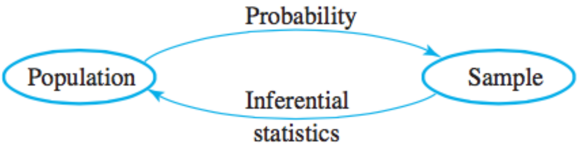
7.1 Point estimate

- unbiased estimator
- mean squared error

7.2 Methods of point estimation

- method of moments
- method of maximum likelihood.

Point estimate



Definition

A point estimate $\hat{\theta}$ of a parameter θ is a single number that can be regarded as a sensible value for θ.

$$\begin{array}{ccccc} \text{population parameter} & \Longrightarrow & \textit{sample} & \Longrightarrow & \textit{estimate} \\ \theta & \Longrightarrow & X_1, X_2, \ldots, X_n & \Longrightarrow & \hat{\theta} \end{array}$$

Mean Squared Error & Bias-variance decomposition

Definition

The mean squared error of an estimator $\hat{\theta}$ is

$$E[(\hat{\theta} - \theta)^2]$$

Theorem

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + \Big(E(\hat{\theta}) - \theta\Big)^2$$

Bias-variance decomposition

$$\text{Mean squared error} = \text{variance of estimator} + (bias)^2$$

Unbiased estimators

Definition

A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if

$$E(\hat{\theta}) = \theta$$

for every possible value of θ.

Unbiased estimator

$\Leftrightarrow$ Bias $= 0$

$\Leftrightarrow$ Mean squared error $=$ variance of estimator

Example

Problem

Consider a random sample $X_1, \ldots, X_n$ from the pdf

$$f(x) = \frac{1 + \theta x}{2} \qquad -1 \le x \le 1$$

Show that $\hat{\theta} = 3\bar{X}$ is an unbiased estimator of θ.

Method of moments: ideas

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with pmf or pdf
$$f(x; \theta_1, \theta_2, \ldots, \theta_m)$$
- Assume that for $k = 1, \ldots, m$
$$\frac{X_1^k + X_2^k + \ldots + X_n^k}{n} = E(X^k)$$
- Solve the system of equations for $\theta_1, \theta_2, \ldots, \theta_m$

Method of moments: example

Problem

Suppose that for a parameter $0 \le \theta \le 1$, X is the outcome of the roll of a four-sided tetrahedral die

x	1	2	3	4
$p(x)$	$\frac{3\theta}{4}$	$\frac{\theta}{4}$	$\frac{3(1-\theta)}{4}$	$\frac{(1-\theta)}{4}$

Suppose the die is rolled 10 times with outcomes

$$4, 1, 2, 3, 1, 2, 3, 4, 2, 3$$

Use the method of moments to obtain an estimator of θ.

Midterm: Problem 2a

Problem

Let $X_1, X_2, \ldots, X_n$ represent a random sample from a distribution with pdf

$$f(x, \theta) = \frac{2x}{\theta + 1} e^{-x^2/(\theta+1)}, \quad x > 0$$

It can be shown that

$$E(X^2 - 1) = \theta$$

Use this fact to construct an estimator of θ based on the method of moments.

Sketch:

$$E(X^2) = \theta - 1$$

MoM:

$$E(X^2) = \frac{X_1^2 + X_2^2 + \ldots + X_n^2}{n}$$

Maximum likelihood estimator

- Let $X_1, X_2, ..., X_n$ have joint pmf or pdf

$$f_{joint}(x_1, x_2, \dots, x_n; \theta)$$

  where θ is unknown.
- When $x_1, \dots, x_n$ are the observed sample values and this expression is regarded as a function of θ, it is called the likelihood function.
- The maximum likelihood estimates θ_{ML} are the value for θ that maximize the likelihood function:

$$f_{joint}(x_1, x_2, \dots, x_n; \theta_{ML}) \geq f_{joint}(x_1, x_2, \dots, x_n; \theta) \qquad \forall \theta$$

How to find the MLE?

- Step 1: Write down the likelihood function.
- Step 2: Can you find the maximum of this function?
- Step 3: Try taking the logarithm of this function.
- Step 4: Find the maximum of this new function.

To find the maximum of a function of θ:

- compute the derivative of the function with respect to θ
- set this expression of the derivative to 0
- solve the equation

Problem

Let $X_1, X_2, \ldots, X_n$ represent a random sample from a distribution with pdf

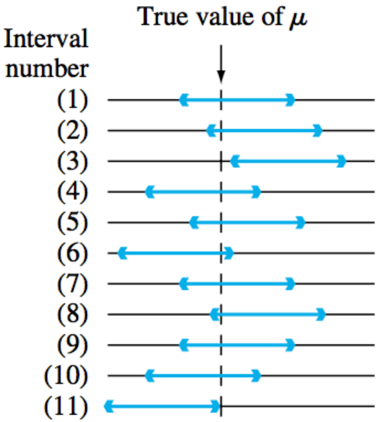
$$f(x, \theta) = \frac{2x}{\theta + 1} e^{-x^2/(\theta+1)}, \quad x > 0$$

Derive the maximum-likelihood estimator for parameter θ based on the following dataset with $n = 10$

17.85, 11.23, 14.43, 19.27, 5.59, 6.36, 9.41, 6.31, 13.78, 11.95

Chapters 8 and 10: Confidence intervals

Interpreting confidence interval



95% confidence interval: If we repeat the experiment many times, the interval contains μ about 95% of the time

Confidence intervals

- By target
  - Chapter 8: Confidence intervals for population means
  - Chapter 8: Prediction intervals for an additional sample
  - Chapter 10: Confidence intervals for difference between two population means
    - independent samples
    - paired samples
- By types
  - (Standard) two-sided confidence intervals
  - One-sided confidence intervals (confidence bounds)
- By distributions of the statistics
  - z-statistic
  - t-statistic

- Section 8.1
  - Normal distribution, σ is known
- Section 8.2
  - ~~Normal distribution~~, ~~σ is known~~
  - $n > 40$
- Section 8.3
  - Normal distribution, ~~σ is known~~
  - n is small

$\rightarrow$ t-distribution

Assumptions:

- Normal distribution
- σ is known

A **$100(1-\alpha)\%$ confidence interval** for the mean μ of a normal population when the value of σ is known is given by

$$\left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) \tag{8.5}$$

or, equivalently, by $\bar{x} \pm z_{\alpha/2} \cdot \sigma/\sqrt{n}$.

If after observing $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$ $(n > 40)$, we compute the observed sample mean $\bar{x}$ and sample standard deviation s. Then

$$\left(\bar{x} - z_{\alpha/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right)$$

is a 95% confidence interval of μ

Let $\overline{x}$ and s be the sample mean and sample standard deviation computed from the results of a random sample from a normal population with mean μ. Then a **100(1 − α)% confidence interval for μ**, the **one-sample t CI**, is

$$\left(\overline{x} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}\right) \tag{8.15}$$

or, more compactly, $\overline{x} \pm t_{\alpha/2,n-1} \cdot s/\sqrt{n}$.

An **upper confidence bound for μ** is

$$\overline{x} + t_{\alpha,n-1} \cdot \frac{s}{\sqrt{n}}$$

and replacing + by − in this latter expression gives a **lower confidence bound for μ**; both have confidence level $100(1 - \alpha)\%$.

Prediction intervals

- We have available a random sample $X_1, X_2, \ldots, X_n$ from a normal population distribution
- We wish to predict the value of X_{n+1}, a single future observation.

A prediction interval (PI) for a single observation to be selected from a normal population distribution is

$$\bar{x} \pm t_{\alpha/2,n-1} \cdot s\sqrt{1 + \frac{1}{n}} \tag{8.16}$$

The *prediction level* is $100(1 - \alpha)\%$.

Confidence intervals: difference between two means

- Independent samples
  1. $X_1, X_2, \ldots, X_m$ is a random sample from a population with mean μ_1 and variance σ_1^2.
  2. $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with mean μ_2 and variance σ_2^2.
  3. The X and Y samples are independent of each other.
- Paired samples
  1. There is only one set of n individuals or experimental objects
  2. Two observations are made on each individual or object

Difference between population means: independent samples

The **two-sample t confidence interval for $\mu_1 - \mu_2$** with confidence level $100(1-\alpha)\%$ is then

$$\bar{x} - \bar{y} \pm t_{\alpha/2,\nu}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

A one-sided confidence bound can be calculated as described earlier.

Difference between population means: paired samples

- The paired t CI for μ_D is

$$\bar{d} \pm t_{\alpha/2,n-1}\frac{s_D}{\sqrt{n}}$$

- A one-sided confidence bound results from retaining the relevant sign and replacing $t_{\alpha/2,n-1}$ by $t_{\alpha,n-1}$.

Principles for deriving CIs

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution $f(x, \theta)$, then

- Find a random variable $Y = h(X_1, X_2, \ldots, X_n; \theta)$ such that he probability distribution of Y does not depend on θ or on any other unknown parameters.
- Find constants a, b such that

$$P\left[a < h(X_1, X_2, \ldots, X_n; \theta) < b\right] = 1 - \alpha$$

- Manipulate these inequality to isolate θ

$$P\left[\ell(X_1, X_2, \ldots, X_n) < \theta < u(X_1, X_2, \ldots, X_n)\right] = 1 - \alpha$$

Examples

- For μ and X_{n+1}

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}, \quad \frac{\bar{X}-X_{n+1}}{S\sqrt{1+1/n}} \sim t_{n-1}$$

- Difference between two means [independent samples]

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{m}+\frac{S_2^2}{n}}} \sim t_\nu$$

- Difference between two means [paired samples]

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

Chapters 9 and 10: Tests of hypotheses

Test of hypotheses

- By target
  - Chapter 9: population mean
  - Chapter 10: difference between two population means
    - independent samples
    - paired samples
- By the alternative hypothesis
  - $>$
  - $<$
  - $\neq$
- By the type of test
  - z-test
  - t-test
- By method of testing
  - Rejection region
  - p-value

Hypothesis testing

In any hypothesis-testing problem, there are two contradictory hypotheses under consideration

- The null hypothesis, denoted by H_0, is the claim that is initially assumed to be true
- The alternative hypothesis, denoted by H_a, is the assertion that is contradictory to H_0.

Implicit rules

- H_0 will always be stated as an equality claim.
- If θ denotes the parameter of interest, the null hypothesis will have the form
$$H_0 : \theta = \theta_0$$
- θ_0 is a specified number called the *null value*
- The alternative hypothesis will be either:
  - $H_a : \theta > \theta_0$
  - $H_a : \theta < \theta_0$
  - $H_a : \theta \neq \theta_0$

Test procedures

A test procedure is specified by the following:

- A test statistic T: a function of the sample data on which the decision (reject H_0 or do not reject H_0) is to be based
- A rejection region $\mathcal{R}$: the set of all test statistic values for which H_0 will be rejected
- A type I error consists of rejecting the null hypothesis H_0 when it is true
- A type II error involves not rejecting H_0 when H_0 is false.

Hypothesis testing for one parameter: rejection region method

1. Identify the parameter of interest
2. Determine the null value and state the null hypothesis
3. State the appropriate alternative hypothesis
4. Give the formula for the test statistic
5. State the rejection region for the selected significance level α
6. Compute statistic value from data
7. Decide whether H_0 should be rejected and state this conclusion in the problem context

Normal population with known σ

Null hypothesis: $\mu = \mu_0$
Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

Alternative Hypothesis	**Rejection Region for Level α Test**
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
H_a: $\mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

Large-sample tests

Null hypothesis: $\mu = \mu_0$
Test statistic:

$$Z = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

Alternative Hypothesis	**Rejection Region for Level α Test**
H_a: $\mu > \mu_0$	$z \geq z_\alpha$ (upper-tailed test)
H_a: $\mu < \mu_0$	$z \leq -z_\alpha$ (lower-tailed test)
H_a: $\mu \neq \mu_0$	either $z \geq z_{\alpha/2}$ or $z \leq -z_{\alpha/2}$ (two-tailed test)

[Does not need the normal assumption]

t-test

Null hypothesis: H_0: $\mu = \mu_0$

Test statistic value: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Alternative Hypothesis	**Rejection Region for a Level α Test**
H_a: $\mu > \mu_0$	$t \geq t_{\alpha,n-1}$ (upper-tailed)
H_a: $\mu < \mu_0$	$t \leq -t_{\alpha,n-1}$ (lower-tailed)
H_a: $\mu \neq \mu_0$	either $t \geq t_{\alpha/2,n-1}$ or $t \leq -t_{\alpha/2,n-1}$ (two-tailed)

[Require normal assumption]

P-value

DEFINITION The ***P*-value** (or *observed significance level*) is the smallest level of significance at which H_0 would be rejected when a specified test procedure is used on a given data set. Once the P-value has been determined, the conclusion at any particular level α results from comparing the P-value to α:

1. P-value $\leq \alpha \Rightarrow$ reject H_0 at level α.
2. P-value $> \alpha \Rightarrow$ do not reject H_0 at level α.

Testing by P-value method

DECISION RULE BASED ON THE *P*-VALUE	Select a significance level α (as before, the desired type I error probability). Then reject H_0 if P-value $\leq \alpha$; do not reject H_0 if P-value $> \alpha$

Remark: the smaller the P-value, the more evidence there is in the sample data against the null hypothesis and for the alternative hypothesis.

P-values for z-tests

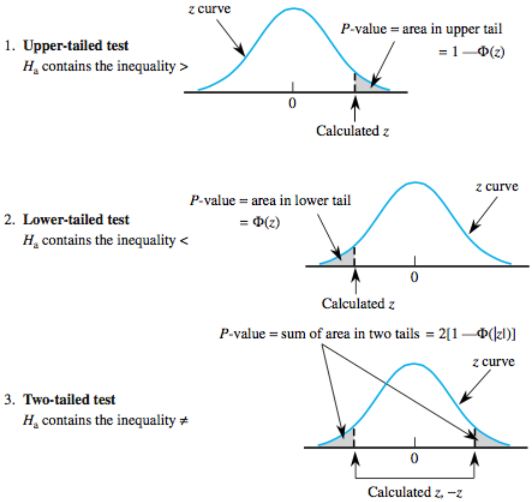


Figure 9.7 Determination of the P-value for a z test

P-values for t-tests

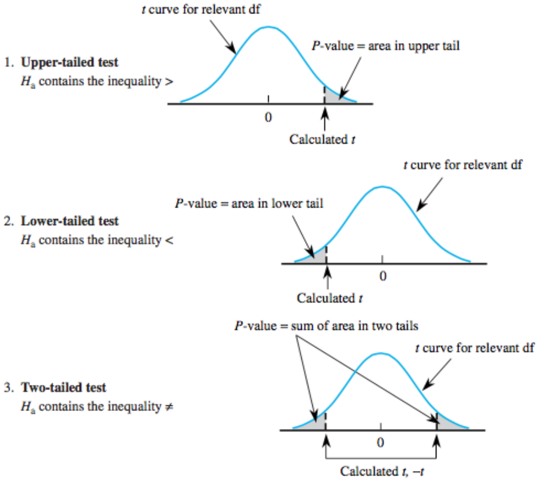


Figure 9.8 P-values for t tests

Testing by rejection region method

- Parameter of interest: μ = true average activation temperature
- Hypotheses

$$H_0 : \mu = 130$$
$$H_a : \mu \neq 130$$

- Test statistic:

$$z = \frac{\bar{x} - 130}{1.5/\sqrt{n}}$$

- Rejection region: either $z \leq -z_{0.005}$ or $z \geq z_{0.005} = 2.58$
- Substituting $\bar{x} = 131.08$, $n = 25 \to z = 2.16$.
- Note that $-2.58 < 2.16 < 2.58$. We fail to reject H_0 at significance level 0.01.
- The data does not give strong support to the claim that the true average differs from the design value.

1. Parameter of interest: μ = true average wafer thickness
2. Null hypothesis: H_0: $\mu = 245$
3. Alternative hypothesis: H_a: $\mu \neq 245$
4. Formula for test statistic value: $z = \dfrac{\bar{x} - 245}{s/\sqrt{n}}$
5. Calculation of test statistic value: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$
6. Determination of P-value: Because the test is two-tailed,

$$P\text{-value} = 2[1 - \Phi(2.32)] = .0204$$

7. Conclusion: Using a significance level of .01, H_0 would not be rejected since $.0204 > .01$. At this significance level, there is insufficient evidence to conclude that true average thickness differs from the target value.

A P-value:

- is not the probability that H_0 is true
- is not the probability of rejecting H_0
- is the probability, calculated assuming that H_0 is true, of obtaining a test statistic value at least as contradictory to the null hypothesis as the value that actually resulted

Testing the difference between two population means

- Setting: independent normal random samples $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ with known values of σ_1 and σ_2. Constant Δ_0.
- Null hypothesis:

$$H_0 : \mu_1 - \mu_2 = \Delta_0$$

- Alternative hypothesis:
  (a) $H_a : \mu_1 - \mu_2 > \Delta_0$
  (b) $H_a : \mu_1 - \mu_2 < \Delta_0$
  (c) $H_a : \mu_1 - \mu_2 \neq \Delta_0$
- When $\Delta = 0$, the test (c) becomes

$$H_0 : \mu_1 = \mu_2$$
$$H_a : \mu_1 \neq \mu_2$$

Difference between 2 means (independent samples)

Proposition

The **two-sample t test** for testing H_0: $\mu_1 - \mu_2 = \Delta_0$ is as follows:

$$\text{Test statistic value: } t = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}}$$

Alternative Hypothesis	**Rejection Region for Approximate Level α Test**
H_a: $\mu_1 - \mu_2 > \Delta_0$	$t \geq t_{\alpha,\nu}$ (upper-tailed test)
H_a: $\mu_1 - \mu_2 < \Delta_0$	$t \leq -t_{\alpha,\nu}$ (lower-tailed test)
H_a: $\mu_1 - \mu_2 \neq \Delta_0$	either $t \geq t_{\alpha/2,\nu}$ or $t \leq -t_{\alpha/2,\nu}$ (two-tailed test)

A P-value can be computed as described in Section 9.4 for the one-sample t test.