MATH 450: Mathematical statistics

Oct 22nd, 2019

Lecture 17: Midterm review

Att.

- Midterm exam:

  Thursday, 10/24/2019, 9:30 am –10:45 am

- Closed-book. Books/notes/computers are not allowed
- Calculators allowed
- One-sided hand-written A4-size note
- z and t tables are provided

Midterm

Week 2	**Chapter 6: Statistics and Sampling Distributions**
Week 4	**Chapter 7: Point Estimation**
Week 7	**Chapter 8: Confidence Intervals**
Week 10	Chapter 9: Test of Hypothesis
Week 11	Chapter 10: Two-sample inference
Week 13	Regression

6.1 Statistics and their distributions
6.2 The distribution of the sample mean
6.3 The distribution of a linear combination

Given a random sample $X_1, X_2, \ldots, X_n$, and

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

- If we **know** the distribution of X_i's, can we obtain the distribution of T?
  - Simple cases
  - If $X_i's$ follow normal distribution, then so does T.
- If we **don't know** the distribution of X_i's, can we still obtain/approximate the distribution of T?
  - Can we at least compute the mean and the variance?
  - When T is the sample mean, i.e. $a_1 = a_2 = \ldots = \frac{1}{n}$

Example 1

Problem

Consider the distribution P

x	10	15	20
$p(x)$	0.2	0.3	0.5

Let $\{X_1, X_2\}$ *be a random sample of size 2 from* P*, and* $T = X_1 + X_2$.

1. *Compute* $P[T = 40]$
2. *Derive the probability mass function of* T
3. *Compute the expected value and the standard deviation of* T

Example 2

Problem

Let $\{X_1, X_2\}$ *be a random sample of size 2 from the exponential distribution with parameter* λ

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

and $T = X_1 + X_2$.
What is the distribution of T*?*

Linear combination of random variables

Theorem

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

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

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

- $E(T) = a_1E(X_1) + a_2E(X_2) + \ldots + a_nE(X_n)$
- $\sigma_T^2 = a_1^2\sigma_{X_1}^2 + a_2^2\sigma_{X_2}^2 + \ldots + a_n^2\sigma_{X_n}^2$

Mean and variance of the sample mean

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$

2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

The Central Limit Theorem

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.

Example

Problem

The tip percentage at a restaurant has a mean value of 18% and a standard deviation of 6%.

What is the approximate probability that the sample mean tip percentage for a random sample of 40 bills is between 16% and 19%?

Chapter 7: Overview

7.1 Point estimate

- unbiased estimator
- mean squared error

7.2 Methods of point estimation

- method of moments
- method of maximum likelihood.

Bias-variance decomposition

Definition

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

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

Theorem

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

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

Method of moments

- 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 $k = 1, \ldots, m$

$$\hat{u}_k = \frac{X_1^k + X_2^k + \ldots + X_n^k}{n} = E(X^k)$$

- Solve the system of equations for $\theta_1, \theta_2, \ldots, \theta_m$

Maximum likelihood estimator

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

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

  where θ is unknown.

- When $x_1, \ldots, x_n$ 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**:

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

How to find the MLE?

- Step 1: Write down the likelihood function.
- Step 2: Try taking the logarithm of this function.
- Step 3: Find the maximum of this new function.
  - compute the derivative of the function with respect to θ
  - set this expression of the derivative to 0
  - solve the equation

8.1 Basic properties of confidence intervals (CIs)
- Interpreting CIs
- General principles to derive CI

8.2 Large-sample confidence intervals for a population mean
- Using the Central Limit Theorem to derive CIs

8.3 Intervals based on normal distribution
- Using Student's t-distribution

Framework

- Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution $f(x, \theta)$
- In Chapter 7, we learnt methods to construct an estimate $\hat{\theta}$ of θ
- Goal: we want to indicate the degree of uncertainty associated with this random prediction
- One way to do so is to construct a *confidence interval* $[\hat{\theta} - a, \hat{\theta} + b]$ such that

$$P[\theta \in [\hat{\theta} - a, \hat{\theta} + b]] = 95\%$$

Principles for deriving CIs

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

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

$$P\left[a < h(X_1, X_2, \dots, X_n; \theta) < b\right] = 0.95$$

- Manipulate these inequalities to isolate θ

$$P\left[\ell(X_1, X_2, \dots, X_n) < \theta < u(X_1, X_2, \dots, X_n)\right] = 0.95$$

Confidence intervals for a population mean

- Section 8.1: Normal distribution with known σ
  - Normal distribution
  - σ is known
- Section 8.2: Large-sample confidence intervals
  - ~~Normal distribution~~
    $\rightarrow$ use Central Limit Theorem $\rightarrow$ needs $n > 30$
  - ~~σ is known~~
    $\rightarrow$ replace σ by s $\rightarrow$ needs $n > 40$
- Section 8.3: Intervals based on normal distributions
  - Normal distribution
  - ~~σ is known~~

$\rightarrow$ Introducing t-distribution

z-critical value

NOTATION z_α will denote the value on the measurement axis for which α of the area under the z curve lies to the right of z_α. (See Figure 4.19.)

For example, $z_{.10}$ captures upper-tail area .10 and $z_{.01}$ captures upper-tail area .01.

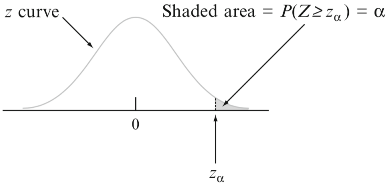


Figure 4.19 z_α notation illustrated

Since α of the area under the standard normal curve lies to the right of z_α, $1-\alpha$ of the area lies to the left of z_α. Thus z_α *is the* $100(1-\alpha)$*th percentile of the standard normal distribution*. By symmetry the area under the standard normal curve to the left of $-z_\alpha$ is also α. The z_α's are usually referred to as ***z* critical values**. Table 4.1 lists the most useful standard normal percentiles and z_α values.

t distributions

Let $t_{\alpha,\nu}$ = the number on the measurement axis for which the area under the t curve with ν df to the right of $t_{\alpha,\nu}$, is α; $t_{\alpha,\nu}$ is called a ***t* critical value.**

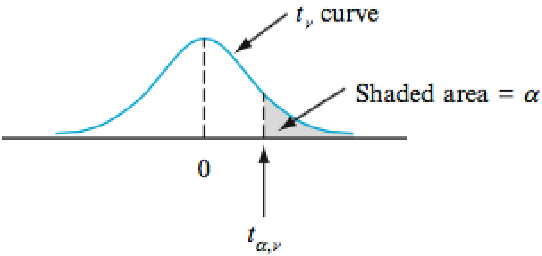


Figure 8.7 A pictorial definition of $t_{\alpha,\nu}$

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

One-sided CIs

A **large-sample upper confidence bound for μ** is

$$\mu < \bar{x} + z_\alpha \cdot \frac{s}{\sqrt{n}}$$

and a **large-sample lower confidence bound for μ** is

$$\mu > \bar{x} - z_\alpha \cdot \frac{s}{\sqrt{n}}$$

Let $\bar{x}$ 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 **$100(1-\alpha)\%$ confidence interval for μ**, the **one-sample t CI**, is

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

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

An **upper confidence bound for μ** is

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

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

Prediction intervals

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

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

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

Example

Problem

Here are the alcohol percentages for a sample of 16 beers:

4.68	4.13	4.80	4.63	5.08	5.79	6.29	6.79
4.93	4.25	5.70	4.74	5.88	6.77	6.04	4.95

(a) *Assume the distribution is normal, construct the 95% confidence interval for the population mean.*

(b) *Assume the distribution is normal, construct the 95% lower confidence bound for the population mean.*

(c) *Assume that another beer is sampled from the same distribution, construct the 95% prediction interval for the alcohol percentages of that beer.*

Example

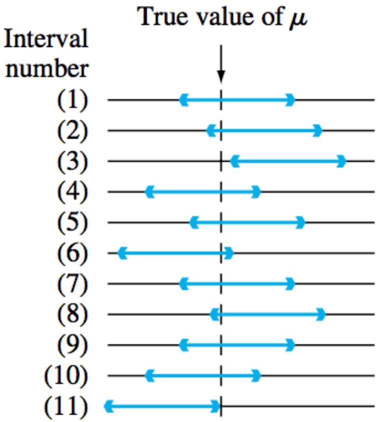
Problem

Suppose that against a certain opponent, the number of points a basketball team scores is normally distributed with unknown mean μ *and unknown variance* σ^2*. Suppose that over the course of the last 10 games, the team scored the following points:*

$$59, \quad 62, \quad 59, \quad 74, \quad 70, \quad 61, \quad 62, \quad 66, \quad 62, \quad 75$$

- *Construct a 95% confidence interval for* μ*.*
- *Now suppose that you learn that* $\sigma^2 = 25$*. Construct a 95% confidence interval for* μ*.*

Interpreting confidence intervals



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