Mathematical statistics

October 18^{th}, 2018

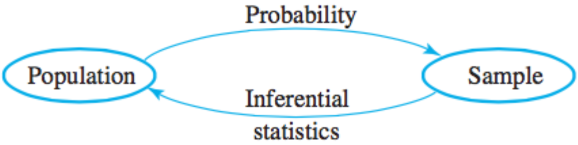
Lecture 16: Midterm review

Countdown to mid-term exam: 7 days

Week 1	**Chapter 1: Probability review**
Week 2	**Chapter 6: Statistics**
Week 4	**Chapter 7: Point Estimation**
Week 7	**Chapter 8: Confidence Intervals**
Week 10	Chapter 9: Test of Hypothesis
Week 14	Regression

Chapter 6: Summary

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.

Practice problems

Example 1*

Problem

Consider the distribution P

x	40	45	50
$p(x)$	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. *Derive the probability mass function of* T
2. *Compute the expected value and the standard deviation of* T

Question: If $T = X_1 - X_2$, can you still solve the problem?

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 $T = X_1 + X_2$.

1. *Compute the cumulative density function (cdf) of T*
2. *Compute the probability density function (pdf) of T*

Question: If $T = X_1 + 2X_2$, can you still solve the problem?

Example 3

Problem

Two airplanes are flying in the same direction in adjacent parallel corridors. At time $t = 0$*, the first airplane is 10 km ahead of the second one.*
Suppose the speed of the first plane (km/h) is normally distributed with mean 520 and standard deviation 10 and the second planes speed, independent of the first, is also normally distributed with mean and standard deviation 500 and 10, respectively.

What is the probability that after 2h of flying, the second plane has not caught up to the first plane?

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%?

Problem

The time that it takes a randomly selected rat of a certain subspecies to find its way through a maze is a normally distributed random variable with mean $\mu = 1.5$*(minutes) and standard deviation* $\sigma = .35$ *(minutes).*
Suppose five rats are selected. Let $X_1, X_2, \ldots, X_5$ *denote their times in the maze. Assuming the* X_i'*s to be a random sample from this normal distribution, what is the probability that the total time for the five*

$$T = X_1 + X_2 + X_3 + X_4 + X_5$$

is between 6 and 8 minutes.

Example

Problem

Let $X_1, X_2, \ldots, X_n$ *be random sample from a normally distribution with mean* 2.65 *and standard deviation* 0.85.

- *If* $n = 25$, *compute*
$$P[\bar{X} \leq 3]$$
- *Find* n *such that*
$$P[\bar{X} \leq 3] \geq 0.95$$

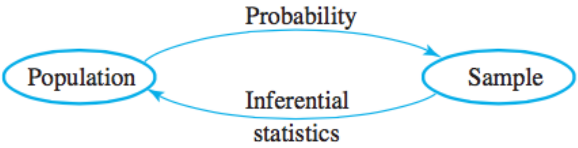
Chapter 7: Summary

Overview

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.

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

Mean Squared Error

- Measuring error of estimation

$$|\hat{\theta} - \theta| \qquad \text{or} \qquad (\hat{\theta} - \theta)^2$$

- The error of estimation is random

Definition

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

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

Bias-variance decomposition

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

Example 1

Problem

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

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

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 4

Problem

Suppose that for a parameter $0 \leq \theta \leq 1$, 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.

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: 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

Example 3

- Let $X_1, \ldots, X_{10}$ be a random sample of size $n = 10$ from a distribution with pdf

$$f(x) = \begin{cases} (\theta + 1)x^{\theta} & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- The observed x_i's are

$$0.92, 0.79, 0.90, 0.65, 0.86, 0.47, 0.73, 0.97, 0.94, 0.77$$

- Question: Use the method of maximum likelihood to obtain an estimator of θ.

Fisher-Neyman factorization theorem

Theorem

T is sufficient for if and only if nonnegative functions g and h can be found such that

$$f(x_1, x_2, \ldots, x_n; \theta) = g(t(x_1, x_2, \ldots, x_n), \theta) \cdot h(x_1, x_2, \ldots, x_n)$$

i.e. the joint density can be factored into a product such that one factor, h does not depend on θ; and the other factor, which does depend on θ, depends on x only through $t(x)$.

Fisher information

Definition

The Fisher information $I(\theta)$ in a single observation from a pmf or pdf $f(x;\theta)$ is the variance of the random variable $U = \frac{\partial \log f(X,\theta)}{\partial\theta}$, which is

$$I(\theta) = Var\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]$$

Note: We always have $E[U] = 0$.

The Cramer-Rao Inequality

Theorem

Assume a random sample $X_1, X_2, ..., X_n$ from the distribution with pmf or pdf $f(x, \theta)$ such that the set of possible values does not depend on θ. If the statistic $T = t(X_1, X_2, ..., X_n)$ is an unbiased estimator for the parameter θ, then

$$V(T) \geq \frac{1}{n \cdot I(\theta)}$$

Large Sample Properties of the MLE

Theorem

Given a random sample $X_1, X_2, ..., X_n$ from the distribution with pmf or pdf $f(x, \theta)$ such that the set of possible values does not depend on θ. Then for large n the maximum likelihood estimator $\hat{\theta}$ has approximately a normal distribution with mean θ and variance $\frac{1}{n \cdot I(\theta)}$.
More precisely, the limiting distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is normal with mean 0 and variance $1/I(\theta)$.

Chapter 8: Confidence intervals

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

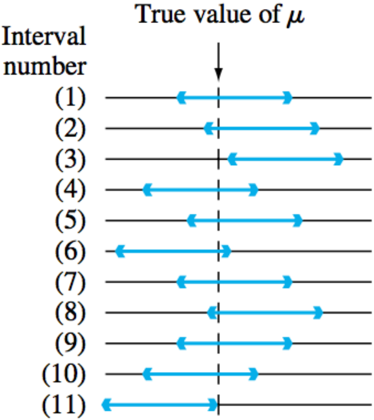
8.3 Intervals based on normal distribution
- Using Student's t-distribution

Overview

- Section 8.1
    - Normal distribution, σ is known
- Section 8.2
    - ~~Normal distribution~~ $\rightarrow$ Using Central Limit Theorem $\rightarrow$ needs $n > 30$
    - ~~σ is known~~ $\rightarrow$ needs $n > 40$
- Section 8.3
    - Normal distribution, ~~σ is known~~
    - n is small

$\rightarrow$ Introducing t-distribution

Interpreting confidence interval



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

Assumptions:

- Normal distribution
- σ is known

A 100(1 − α)% 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

Confidence intervals

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 − α)%.

One-sided CIs

CIs:

- $100(1-\alpha)\%$ confidence

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

- 95% confidence

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

One-sided CIs:

- $100(1-\alpha)\%$ confidence

$$\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

- 95% confidence

$$\left(-\infty, \bar{x} + 1.64\frac{\sigma}{\sqrt{n}}\right)$$

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 $100(1-\alpha)\%$.

Principles for deriving CIs

If $X_1, X_2, \ldots, X_n$ is a random sample from the normal distribution $\mathcal{N}(\mu, \sigma^2)$, then

- For μ

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

- For predicting X_{n+1}

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

- For σ

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

- For sample proportion (n large)

$$\frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \sim \mathcal{N}(0, 1)$$

Principles for deriving CIs

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

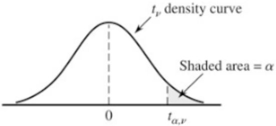
- Find a random variable $Y = h(X_1, X_2, \ldots, X_n; \theta)$ such that he probability distribution of Y does not depend on θ or on any other unknown parameters.
- Find constants a, b such that

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

- Manipulate these inequality to isolate θ

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

Table A.5 Critical Values for t Distributions



	α						
ν	.10	.05	.025	.01	.005	.001	.0005
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965