MATH 450: Mathematical statistics

Oct 1st, 2019

Lecture 11: Sufficient statistic

Topics

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

Chapter 7: Overview

7.1 Point estimate

- unbiased estimator
- mean squared error

7.2 Methods of point estimation

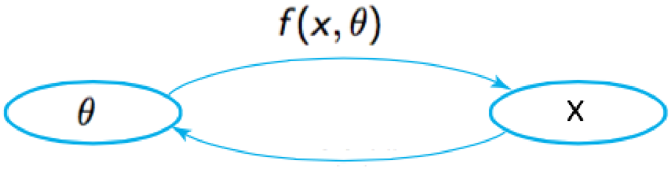
- method of moments
- method of maximum likelihood.

7.3 Sufficient statistic

7.4 Information and Efficiency

- Large sample properties of the maximum likelihood estimator

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

- Measuring error of estimation

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

- The error of estimation is random

Definition

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

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

Theorem

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

Bias-variance decomposition

Mean squared error $=$ 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

Moments

- Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with pmf or pdf $f(x)$.
- For $k = 1, 2, 3, \ldots$, the k^{th} population moment, or k^{th} moment of the distribution $f(x)$, is

$$E(X^k)$$

  - First moment: the mean
  - Second moment: $E(X^2)$

Sample moments

- Let $X_1, \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f(x)$.
- For $k = 1, 2, 3, \ldots,$ the k^{th} sample moment is

$$\frac{X_1^k + X_2^k + \ldots + X_n^k}{n}$$

The law of large numbers provides that when $n \to \infty$

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

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 $k = 1, \ldots, m$

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

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

Method of maximum likelihood

Random sample

Let $X_1, X_2, ..., X_n$ be a random sample of size n from a distribution with density function $f_X(x)$.

Then the density of the joint distribution of $(X_1, X_2, ..., X_n)$ is

$$f_{joint}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_X(x_i)$$

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

Connecting everything

Example 3

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*
- *Given that*

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

  construct an estimator of θ based on the method of moments.

Example 4

Problem

Let $0 < \theta < \infty$ *and* $X_1, X_2, \ldots, X_n$ *sample from a distribution with density function*

$$f(x;\theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta.$$

- *Construct an estimator of* θ *by the method of moments.*
- *Compute the mean squared error (MSE) of this estimator.*

Example 5

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.

Sufficient statistic

Example

- Your professor stores a dataset $x_1, x_2, ..., x_n$ in his computer. He says it is a random sample from the exponential distribution

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

  where λ is an unknown parameter. He wants you to work on the dataset and give him a good estimate of λ
- Assume that the sample size is very large, $n = 10^{20}$, and you could not copy the whole dataset
- You can compute any summary statistics of the dataset using the computer, but the lab is closing in 5 minutes
- What will you do?

Example

- If you are using the method of moments

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

- If you are using the method of maximum likelihood

$$L(\lambda) = \lambda^n e^{-\lambda(x_1+x_2+\ldots+x_n)}$$

- In both case, it seems that you need to only save n and $t = x_1 + x_2 + \ldots + x_n$

Conditional probability

- For discrete random variables, the conditional probability mass function of Y given the occurrence of the value x of X can be written according to its definition as:

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

- For continuous random variables, the conditional probability of Y given the occurrence of the value x of X has density function

$$f_Y(y|X = x) = \frac{f_{joint}(y, x)}{f(x)}$$

Some observations

- Basic estimation problem:
  - Given a density function $f(x, \theta)$ and a sample $X_1, X_2, \ldots, X_n$
  - Construct a statistic $\hat{\theta} = T(X_1, X_2, \ldots, X_n)$
  - Different methods lead to different estimates with different accuracies
- If, however, the distribution of $t(X_1, X_2, \ldots, X_n)$ does not depend on θ, then it is no good
- Similarly, if the conditional probability

$$P(X_1, X_2, \ldots, X_n | T)$$

  does not depend on θ, then this means that $T(X_1, X_2, \ldots, X_n)$ contained all the information to estimate θ

Sufficient statistic

Definition

A statistic $T = t(X_1, \ldots, X_n)$ is said to be sufficient for making inferences about a parameter θ if the joint distribution of $X_1, X_2, \ldots, X_n$ given that $T = t$ does not depend upon θ for every possible value t of the statistic T.

Fisher-Neyman factorization theorem

Theorem

T is sufficient for θ 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 $t(x)$.

Example 1

Problem

Let $X_1, X_2, ..., X_n$ be a random sample of from a Poisson distribution with parameter λ

$$f(x, \lambda) = \frac{1}{x!} e^{-\lambda x} \qquad x = 0, 1, 2, \ldots,$$

where λ is unknown.
Find a sufficient statistic of λ.

Example 2

Problem

Let $X_1, X_2, ..., X_n$ be a random sample of from a Poisson distribution with parameter λ

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

where β is unknown.
Find a sufficient statistic of β.

Jointly sufficient statistic

Definition

The m statistics $T_1 = t_1(X_1, \dots, X_n)$, $T_2 = t_2(X_1, \dots, X_n)$, ..., $T_m = t_m(X_1, \dots, X_n)$ are said to be jointly sufficient for the parameters $\theta_1, \theta_2, \dots, \theta_k$ if the joint distribution of $X_1, \dots, X_n$ given that

$$T_1 = t_1, T_2 = t = 2, \dots, T_m = t_m$$

does not depend upon $\theta_1, \theta_2, \dots, \theta_k$ for every possible value $t_1, t_2, \dots, t_m$ of the statistics.

Fisher-Neyman factorization theorem

Theorem

$T_1, T_2, \ldots, T_m$ are sufficient for $\theta_1, \theta_2, \ldots, \theta_k$ if and only if nonnegative functions g and h can be found such that

$$f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \ldots, \theta_k) = g(t_1, t_2, \ldots, t_m, \theta_1, \theta_2, \ldots, \theta_k) \cdot h(x_1, x_2, \ldots, x_n)$$

Example 3

- Let $X_1, X_2, ..., X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$

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

- Prove that

$$T_1 = X_1 + \ldots + X_n, \qquad T_2 = X_1^2 + X_2^2 + \ldots + X_n^2$$

are jointly sufficient for the two parameters μ and σ.

Example 4

- Let $X_1, X_2, ..., X_n$ be a random sample from a Gamma distribution
$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}$$
where α, β is unknown.
- Prove that
$$T_1 = X_1 + \ldots + X_n, \qquad T_2 = \prod_{i=1}^{n} X_i$$
are jointly sufficient for the two parameters α and β.