STA 2002, Summer 2023
Probability and Statistics II

Point Estimators

Outline

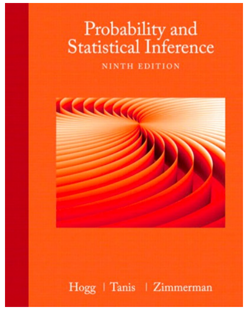
- **Estimator: Definition**
- **Basic properties**
- **Methods for finding point estimators**
  - **Maximum likelihood estimate (MLE)**
  - **Method of Moments (MOM)**

Questions we aim to address

- **What is a good estimator?**
- **How to find estimators?**

Data $\longrightarrow$ **Statistics** $\longleftarrow$ **Sampling distribution**

Estimator

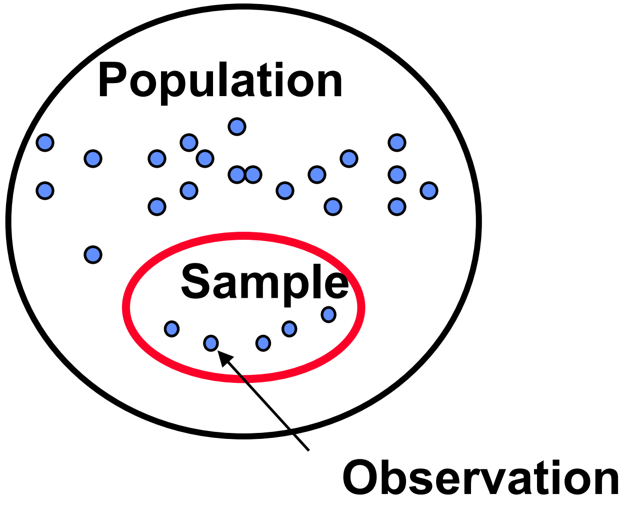


Reading: Chapter 6.4 in textbook

Population Vs. Sample

- **Population:** a finite well-defined group of ALL objects which, although possibly large, can be enumerated in theory
  (e.g. investigating ALL the bearings manufactured today).

- **Sample:** A sample is a SUBSET of a population



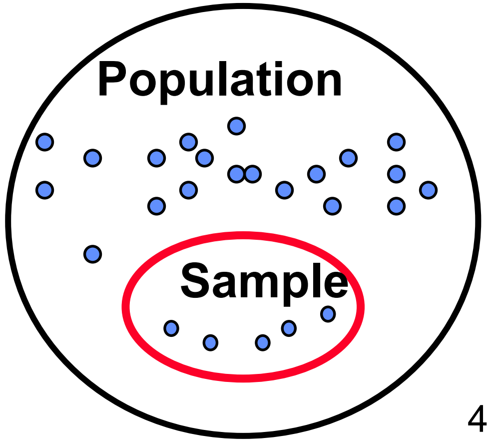
Population Vs. Sample

Suppose you are trying to compare two brands of ice cream: Haagen-Dazs and Wall's, in terms of the preference among all CUHK(SZ) students. You randomly select 500 students in CUHK(SZ) and ask them their preference (either Haagen-Dazs or Wall's) using an online survey.

vs

1) All students in CUHK(SZ) are:
2) The proportion of students preferring Haagen-Dazs among all CUHK(SZ) students is:
3) The 500 selected students that were asked are:
4) The sample proportion of students preferring Haagen-Dazs among 500 students who took the survey is

(A) population (B) sample (C) parameter (D) statistic



<u>Note: Sometimes by “population” we mean an underlying distribution (e.g., Guassian).</u>

Estimator

Suppose X is a random variable with $f(x;\theta)$ as the pdf. If $X_1, X_2, \ldots X_n$ is a random sample of size n from X, the statistic

$$\hat{\Theta} = h(X_1, X_2, \ldots, X_n)$$

Is called a **point estimator** of θ.

After the sample has been selected, $\hat{\Theta}$ takes on a particular numerical value called the **point estimate** of θ.

if $x_1 = 25, x_2 = 30, x_3 = 29$, and $x_4 = 31$

Parameter: μ **Estimator:** $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **Estimate:** $\bar{x} = \frac{25+30+29+31}{4} = 28.75$

Note that $\hat{\Theta}$ is a random variable because it is a statistic (function of random variables)

Internet service provider

- **Two Internet providers**

- **Observe download rate is as follows (mbp)**

Provider 1	5.34	5.16	5.043	4.661	4.521	5.25	5.245
Provider 2	5.363	4.797	5.28	4.666	4.927	5.286	5.37
Provider 1	5.276	4.508	4.558	5.478	4.919	4.708	
Provider 2	5.109	5.113	5.157	5.145	4.801	4.948	

- What's the difference of their rate?

中国电信 CHINA TELECOM

China Unicom 中国联通

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \dots, n_1$
  - Second service provider $Y_i, i = 1, 2, \dots, n_2$
- Assumption
  - $X_i \sim N(\mu_1, \sigma_1^2)$
  - $Y_i \sim N(\mu_2, \sigma_2^2)$
- **Parameters of interest:** $\mu_1 - \mu_2$
- **Estimator:** $\bar{X} - \bar{Y}$
- **Estimate: 4.9744- 5.0740 = -0.0996 (mbp)**

- **How accurate is the estimate?**
- **Is the estimator (method) unbiased?**

Basic properties of estimators

Standard error of estimator

ties to sampling distribution

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error,** denoted by $\hat{\sigma}_{\hat{\Theta}}$.

$$\begin{tikzcd} \text{Data} \arrow[r] & \text{Statistics} \arrow[dr] & \text{Sampling distribution} \arrow[l] \arrow[d] \\ & & \text{Estimator} \end{tikzcd}$$

Internet service provider

- **Two Internet providers**

- **Observe download rate is as follows (mbp)**

Provider 1	5.34	5.16	5.043	4.661	4.521	5.25	5.245
Provider 2	5.363	4.797	5.28	4.666	4.927	5.286	5.37
Provider 1	5.276	4.508	4.558	5.478	4.919	4.708	
Provider 2	5.109	5.113	5.157	5.145	4.801	4.948	

- What's the standard error of the estimator for the difference of their rate?

- What's the difference of their rate?
- Samples
  - First service provider $X_i, i = 1, 2, \ldots, n_1$
  - Second service provider $Y_i, i = 1, 2, \ldots, n_2$
- Assumptions
  - $X_i \sim N(\mu_1, \sigma_1^2)$
  - $Y_i \sim N(\mu_2, \sigma_2^2)$
- **Parameters of interest:** $\mu_1 - \mu_2$
- **Estimator:** $\bar{X} - \bar{Y}$

- **Standard error of the estimator**

$$\sigma_{\bar{X}-\bar{Y}} = \sqrt{Var(\bar{X} - \bar{Y})} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Exercise

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

41.60, 41.48, 42.34, 41.95, 41.86,
42.18, 41.72, 42.26, 41.81, 42.04

What is the estimator for the conductivity?

What is the standard error of the estimator?

Unbiased Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

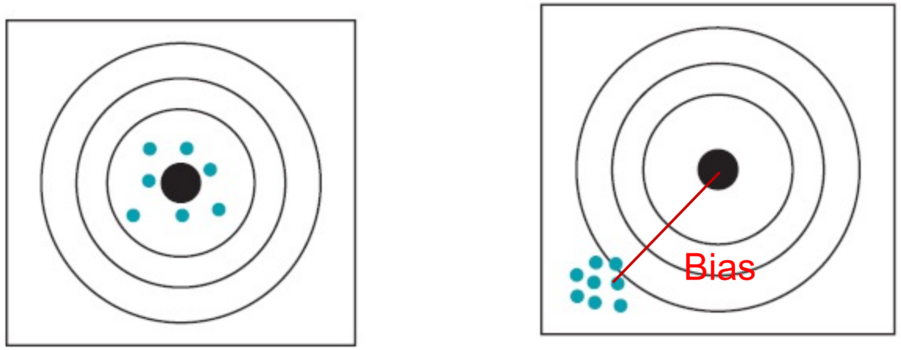
$$\boxed{E(\hat{\Theta})} = \theta$$

ties to sampling distribution

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta$$

is called the **bias** of the estimator $\hat{\Theta}$.



Sample mean is unbiased estimator

- **Assume $x_1, \dots, x_n \sim N(\mu, \sigma^2)$**
- **Then $\overline{x}$ is an unbiased estimator of μ**

Sample variance is unbiased estimator

- Assume $x_1, \ldots, x_n \sim N(\mu, \sigma^2)$
- Then S^2 is an unbiased estimator of σ^2

Variance of a Point Estimator

If two estimators are unbiased, the one with smaller variance is preferred.

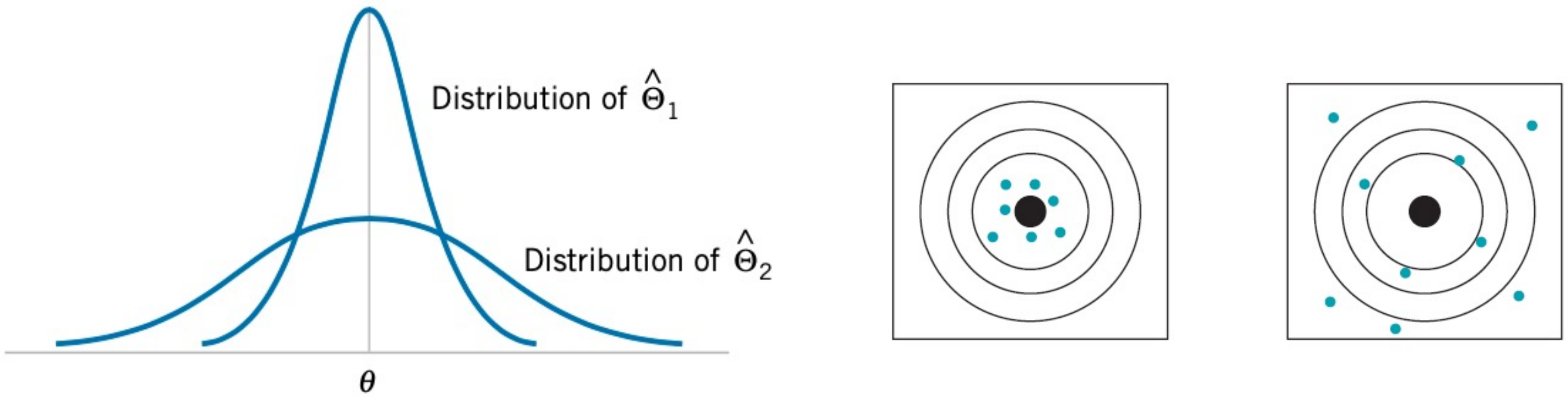


Figure The sampling distributions of two unbiased estimators

$$\boxed{\mathrm{var}(\hat{\Theta}_1)} < \mathrm{var}(\hat{\Theta}_2)$$

ties to sampling distribution

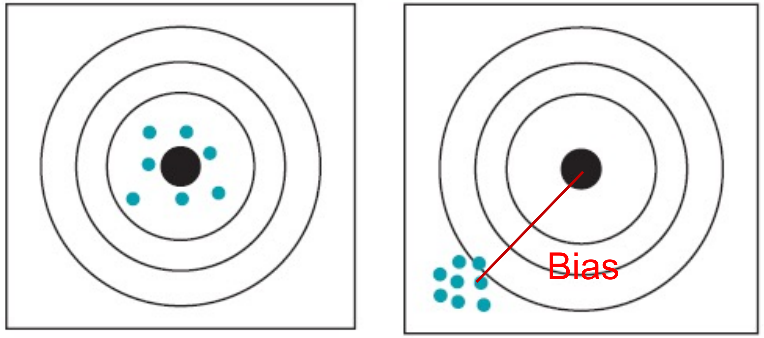
Mean Square Error (MSE)

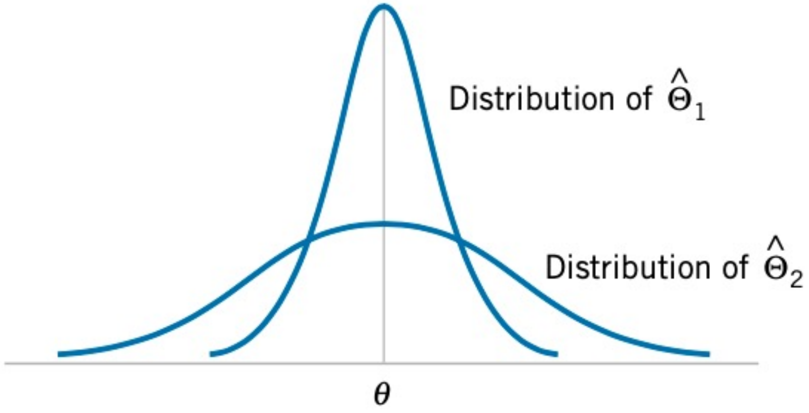
The **mean square error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2$$

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \Theta)^2 = \left[E(\hat{\Theta} - \Theta)\right]^2 + \text{var}(\hat{\Theta} - \Theta)$$

$$MSE(\hat{\Theta}) = \left[\text{Bias}(\hat{\Theta})\right]^2 + \text{var}(\hat{\Theta})$$





Example: find bias and variance of estimator

Let X_1, X_2 be independent random variables with mean μ and variance σ^2.
Suppose that we have two estimators of μ:

$$\widehat{\Theta}_1 = \frac{X_1 + X_2}{2}$$

$$\widehat{\Theta}_2 = \frac{X_1 + 3X_2}{4}$$

(a) Are both estimators unbiased estimators of μ?

There is not a unique unbiased estimator!

(b) What is the variance of each estimator?

(c) What's the MSE of two estimators?

Compare the MSE of estimators

Let $X_1, X_2, \ldots X_7$ denote a random sample from a population with mean μ and variance σ^2. Calculate the MSE of the following estimators of μ.

$$\hat{\Theta}_1 = \frac{\sum_{i=1}^{7} X_i}{7}$$

$$\hat{\Theta}_2 = \frac{2X_1 - X_6 + X_4}{2}$$

$$\hat{\Theta}_3 = \frac{4X_2 + 2X_3 - 2X_5}{2}$$

- **Is either estimator unbiased?**
- **Which estimator is best? In what sense is it best?**

Example

Suppose $X \sim Uniform(\theta, 3\theta)$, $\theta > 0$

- Show that $\frac{\overline{X}}{2}$ is an unbiased estimator of θ

- Calculate the MSE of $\frac{\overline{X}}{2}$ and $\overline{X}$

Methods for Finding Estimators

- **Assume a distribution for the samples**
- **Estimate the parameter of the distribution**
- **Several methods**
  - **Maximum likelihood**
  - **Method of moment**

Method of Maximum Likelihood

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \qquad (7\text{-}5)$$

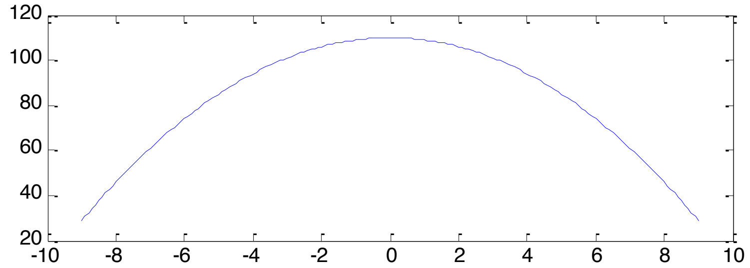
Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

$$L(\theta;x) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta)\ldots f(x_n;\theta)$$

$$\boxed{l(\theta;x) = \sum_{i=1}^{n} \log[f(x_i;\theta)]}$$

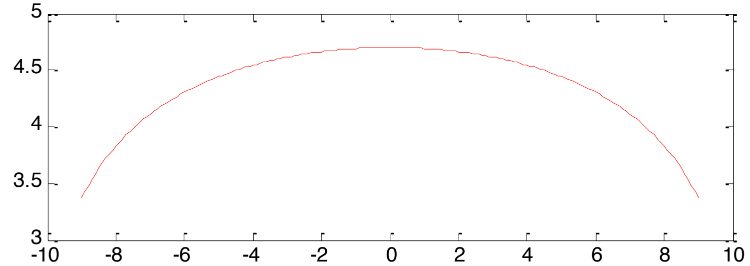
$$\hat{\Theta}(x) = \arg\max_{\theta} L(\theta;x) = \arg\max_{\theta} l(\theta;x)$$

$L(\theta;x)$

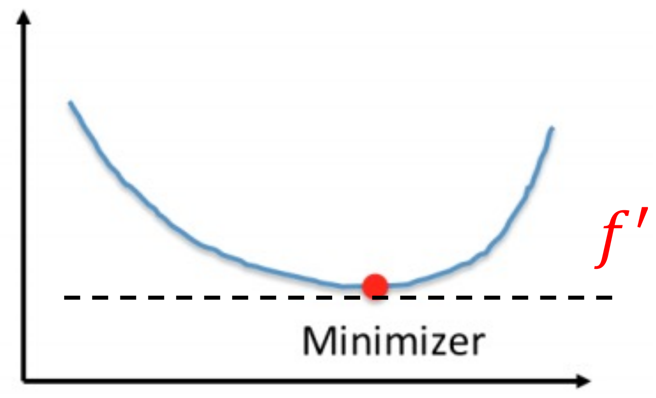


θ

$l(\theta;x)$



First order optimality condition



$f'(x^*) = 0$

First order optimality condition

- Still need to verify $f''(x^*) \geq 0$ so that x^* is a (local) minimum

- Similarly, when we are maximizing a function, we need to verify $f''(x^*) \leq 0$ so that x^* is a (local) maximum

A random variable x has probability density function

$$f(x) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty$$

Given samples $x_1, \ldots x_n$,
find the maximum likelihood estimator for θ

Example: Bernoulli

Let X be a Bernoulli random variable. The probability mass function is

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$\begin{aligned} L(p) &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

$$\longrightarrow \quad l(p) = \ln L(p) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

$$\longrightarrow \quad \frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} \quad \longrightarrow \quad \hat{P} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Example: Normal

Let X be normally distributed with unknown μ and known variance σ^2. The likelihood function of a random sample of size n, say $X_1, X_2, \ldots, X_n$, is

$$L(\mu) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i - \mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i - \mu)^2}$$

Now

$$\ln L(\mu) = -(n/2)\ln(2\pi\sigma^2) - (2\sigma^2)^{-1} \sum_{i=1}^{n} (x_i - \mu)^2$$

and

$$\frac{d \ln L(\mu)}{d\mu} = (\sigma^2)^{-1} \sum_{i=1}^{n} (x_i - \mu)$$

$\longrightarrow$ What is the MLE for μ?

Example (Continued, unknown variance)

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

MLE: Exponential

Let X be an exponential random variable with parameter θ.
The likelihood function of a random sample of size n is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \frac{1}{\theta^n} e^{-\sum_{i=1}^{n} x_i/\theta}$$

$$l(\theta) = -n \ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i, \quad 0 < \theta < \infty$$

$$l'(\theta) = \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0$$

$$\theta = \frac{\sum_{i=1}^{n} x_i}{n} = \overline{x}$$

MLE: Geometric

Let X be a Geometric random variable with parameter p.
The likelihood function of a random sample of size n is:

$$L(p) = \prod_{i=1}^{n} f(x_i; p) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p = p^n (1-p)^{\sum_{i=1}^{n} x_i - n}$$

$$l(p) = n \ln p + \left(\sum_{i=1}^{n} x_i - n \right) \ln(1-p)$$

$$l'(p) = \frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p} = 0$$

$$\hat{p} = \frac{1}{\overline{X}}$$

MLE: Graphical Illustration

The time to failure is exponentially distributed. Eight units are randomly selected and tested, resulting in the following failure time (in hours): $x_1 = 11.96$, $x_2 = 5.03$, $x_3 = 67.40$, $x_4 = 16.07$, $x_5 = 31.50$, $x_6 = 7.73$, $x_7 = 11.10$, and $x_8 = 22.38$.

"peaked" vs **"flat"**

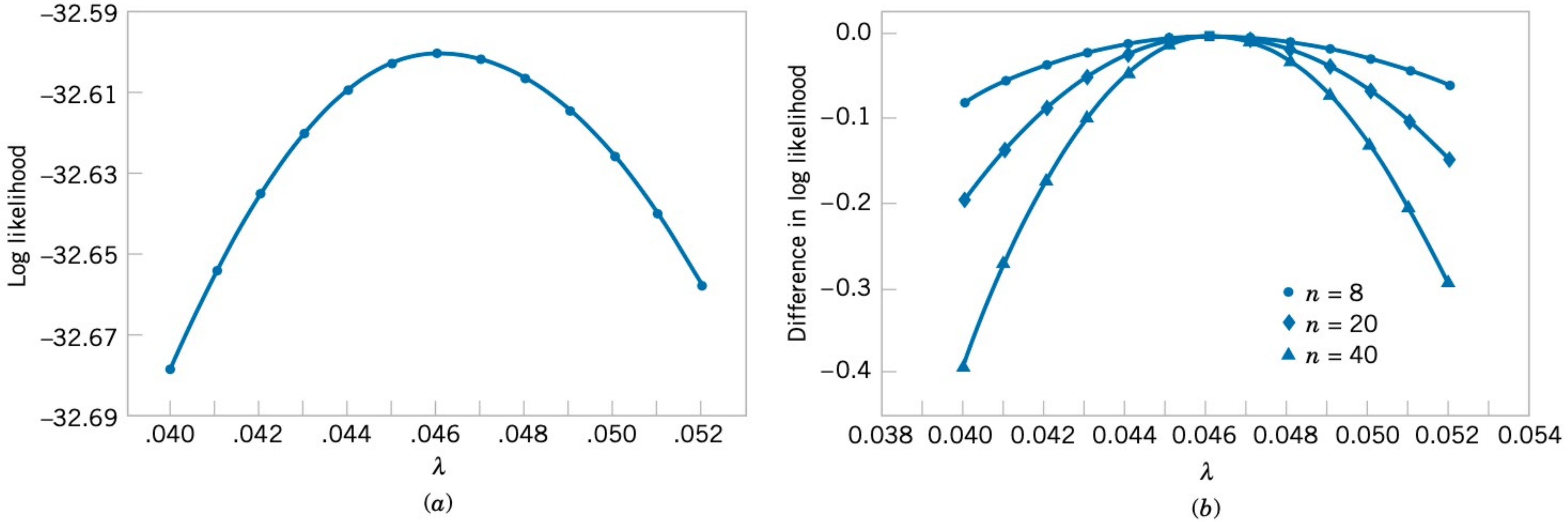


Figure -- Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

Why use maximum likelihood estimator?

It enjoys the following good properties:

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ $[E(\hat{\Theta}) \simeq \theta]$,

(2) the variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\Theta}$ has an approximate normal distribution.

Chapter 6.6*

Complications in Using MLE

- It is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\Theta)/d\Theta = 0$ may be difficult to solve.
- It may not always be possible to use calculus methods directly to determine the maximum of $L(\Theta)$.
- At times it can be hard to give an explicit formula for the maximizer, and so numerical optimization methods are required. (MAT3007 Optimization)

Exercise: baseball team

- **The weight for a baseball team players are**

  {150, 143, 132, 160, 175, 190, 123, 154}

- **Assume their weights are uniformly distributed over an interval [a, b]**

- **What are good estimators for a? for b?**

MLE: Uniform

Let X be a Uniform random variable on the interval $[0, \theta]$

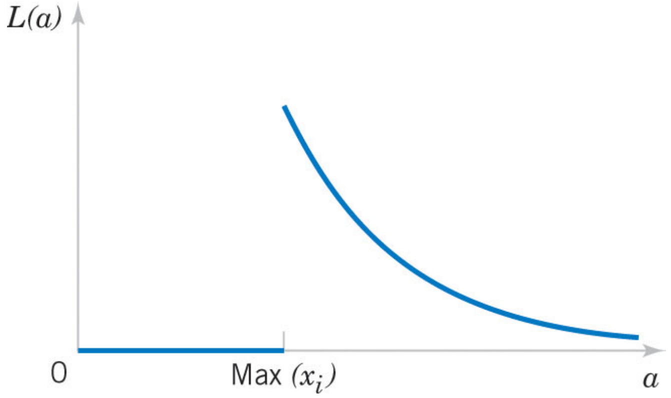
$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & \text{for } 0 \le x \le \theta, \\ 0, & \text{otherwise,} \end{cases} = \frac{1}{\theta}\mathbf{1}_{\{0 \le x \le \theta\}}$$

indicator function $\mathbf{1}_A(x)$

$$\mathbf{1}_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise} \end{cases}$$

The likelihood function of a random sample of size n is:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbf{1}_{\{0 \le x_i \le \theta\}} = \begin{cases} \frac{1}{\theta^n}, & \text{if } \theta \ge x_{(n)}, \\ 0, & \text{if } \theta < x_{(n)}. \end{cases}$$



$$\hat{\theta} = X_{(n)} = max\{X_1, X_2, \dots, X_n\}$$

Calculus methods don't work here because $L(\theta)$ is maximized at the discontinuity.
Clearly, θ cannot be smaller than max(x_i), thus the MLE is max(x_i)= $X_{(n)}$.

Methods of Moments

Population and samples moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Population moments

$$\mu'_k = \begin{cases} \int_x x^k f(x)dx & \textbf{If } \boldsymbol{x} \textbf{ is continuous} \\ \sum_x x^k f(x) & \textbf{If } \boldsymbol{x} \textbf{ is discrete} \end{cases}$$

Sample moments

$$m'_k = \frac{\sum_{i=1}^{n} X_i^k}{n}$$

Method of Moments

- **Equating empirical moments to theoretical moments**

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

***m* equations for *m* parameters**

$$\begin{cases} m_1' = \mu_1' \\ m_2' = \mu_2' \\ \vdots \\ m_m' = \mu_m' \end{cases}$$

Example

MoM estimator for exponential parameter?

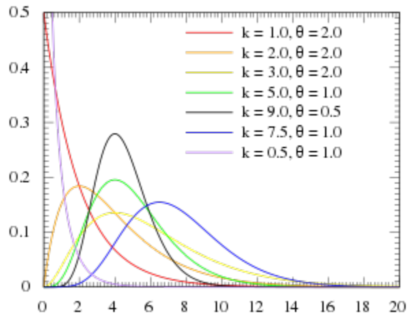
MoM estimator for normal distribution?

MoM: Gamma

Method of moment estimator for Gamma distribution?

$$f(x_i) = \frac{1}{\Gamma(\alpha)\theta^\alpha} x^{\alpha-1} e^{-x/\theta}$$

The likelihood function is difficult to differentiate because of the Gamma function $\Gamma(\alpha)$.



$$L(\alpha, \theta) = \left(\frac{1}{\Gamma(\alpha)\theta^\alpha}\right)^n (x_1 x_2 \cdots x_n)^{\alpha-1} \exp\left[-\frac{1}{\theta}\sum x_i\right]$$

We will use method of moment estimator

$$E(X) = \alpha\theta = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

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

$\Rightarrow$

$$\alpha = \frac{\bar{X}}{\theta}$$

$$\hat{\theta}_{MM} = \frac{1}{n\bar{X}}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

MoM: Gamma (known α)

A random variable x has probability density function

$$f(x) = \frac{1}{2\theta^3} x^2 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty$$

Given samples $x_1, \ldots x_n$,
find the MoM estimator for θ

Gamma distribution with $\alpha = 3$

MoM: Uniform

Let X be a Uniform random variable on the interval [0, θ]

$$f(x;\theta) = \begin{cases} \frac{1}{\theta}, & \text{for } 0 \le x \le \theta, \\ 0, & \text{otherwise,} \end{cases} = \frac{1}{\theta}\mathbf{1}_{\{0 \le x \le \theta\}}$$

The mean of X is $\dfrac{\theta}{2}$

$$\frac{\theta}{2} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \qquad \Longrightarrow \qquad \widetilde{\theta} = 2\overline{X}$$