Chapter 3

Point Estimation

STAT2602A Probability and statistics II
(2024-2025 1st Semester)

Contents

Introduction

A random variable X from a random experiment is assumed to have a distribution with the p.d.f $f(x; \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \mathbb{R}^s$ is a unknown parameter taking a value in the parameter space Ω.

Aim: To estimate the unknown parameter $\boldsymbol{\theta}$ based on a random sample $\boldsymbol{X}$.

3.1 Maximum likelihood estimation

(**Example 3.1**)

Bernoulli Distribution

Suppose X follows a Bernoulli distribution, the p.d.f. of X is

$$f(x;p) = p^x(1-p)^{1-x}, x = 0, 1.$$

The unknown parameter $p \in \Omega$ with $\Omega = \{p : p \in (0,1)\}$.
A random sample $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ with the observed values $\boldsymbol{x} = \{x_1, \ldots, x_n\}$.

- The probability that $\boldsymbol{X} = \boldsymbol{x}$:
  $L(x_1, \cdots, x_n; p) = P(X_1 = x_1, \ldots, X_n = x_n)$
  $= \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i}$.
  (Joint p.d.f. of $X_1, X_2, \ldots, X_n$ evaluated at the observed values.)
  Aim: find the value of p that maximizes the joint p.d.f., i.e.,

$$p_* = \text{argmax}_{p \in \Omega} L(x_1, \ldots, x_n; p)$$

Bernoulli Distribution

$$p_* = \text{argmax}_{p \in \Omega} L(x_1, \ldots, x_n; p)$$

p_* most likely has produced the observed values $x_1, \ldots, x_n$.
p_* is called the maximum likelihood estimate.

$$\ell(p) = \log L(p) \quad = \quad \log p \cdot \sum_{i=1}^{n} x_i + \log(1-p) \cdot \sum_{i=1}^{n} (1 - x_i)$$

$$\begin{aligned}
\frac{\mathrm{d}\ell(p)}{\mathrm{d}p} &= \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i \right) \\
&= \frac{(1-p)\sum_{i=1}^{n} x_i - np + p \sum_{i=1}^{n} x_i}{p(1-p)} \\
&= \frac{n(\overline{x} - p)}{p(1-p)}
\end{aligned}$$

$\ell(p)$ attains its maximum at $p_* = \overline{x}$

3.1 Maximum likelihood estimation

(Def 3.1, 3.2)

Likelihood Function

Let $\boldsymbol{X}$ be a random sample with a joint p.d.f. $f(x_1, \ldots, x_n; \theta)$, where the parameter θ is within a certain parameter space Ω. Then the likelihood function of this random sample is defined as $L(\theta) = \boldsymbol{f}(\boldsymbol{X};\, \theta) = \boldsymbol{f}(X_1, X_2, \cdots, X_n; \theta)$ for $\theta \in \Omega$.

Log Likelihood Function

$\ell(\theta) = \log L(\theta)$

Maximum Likelihood Estimator

MLE of θ: $\hat{\theta} = \arg\max_{\theta \in \Omega} L(\theta) = \arg\max_{\theta \in \Omega} \ell(\theta)$

Maximum likelihood Estimate

The observed value of $\hat{\theta}$

3.1 Maximum likelihood estimation

(**Example 3.2**)

Bernoulli Distribution

Let $\boldsymbol{X}$ be an independent random sample from a Bernoulli distribution with parameter p with $0 < p < 1$. Find the MLE of p.
$f(x; p) = p^x(1-p)^{1-x}, \quad x = 0, 1$ and $\Omega = \{p : p \in (0, 1)\}$.

- Likelihood function:

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

Log-likelihood function:

$$\ell(p) = \log L(p) = \log p \cdot \sum_{i=1}^{n} X_i + \log(1-p) \cdot \sum_{i=1}^{n}(1 - X_i)$$

$\ell(p)$ attains its maximum at $p = \overline{X}$.
The maximum likelihood estimator of p is $\overline{X}$.

3.1 Maximum likelihood estimation

(**Example 3.3**)

Uniform Distribution

Let $\boldsymbol{X}$ be an independent random sample from a uniformly distribution over $[0, \beta]$. Find the maximum likelihood estimator of β.

- The p.d.f. is

  $$f(x; \beta) = \frac{1}{\beta} \mathrm{I}(0 \leq x \leq \beta),$$

  The likelihood function is

  $$L(\beta) = \prod_{i=1}^{n} \frac{1}{\beta} \mathrm{I}(0 \leq X_i \leq \beta) = \frac{1}{\beta^n} \prod_{i=1}^{n} \mathrm{I}(0 \leq X_i \leq \beta).$$

3.1 Maximum likelihood estimation

(**Example 3.3 con't**)

Uniform Distribution

In order that $L(\beta)$ attains its maximum, β must satisfies

$$0 \leq X_i \leq \beta, \ \ i = 1, 2, \ldots, n. \text{ Otherwise,} L(\beta) = 0.$$

i.e.,

$$0 \leq X_{(1)} \leq \cdots \leq X_{(n)} \leq \beta,$$

Meanwhile, $\dfrac{1}{\beta^n}$ increases as β decreases.

$\hat{\beta} = X_{(n)} = \max_{1 \leq i \leq n} X_i$

3.1 Maximum likelihood estimation

(**Example 3.4**)

Normal Distribution

Let $\boldsymbol{X}$ be an independent random sample from a normal distribution $N(\theta_1, \theta_2)$, where $(\theta_1, \theta_2) \in \Omega$ and $\Omega = \{(\theta_1, \theta_2) : \theta_1 \in \mathbb{R}, \theta_2 > 0\}$. Find the maximum likelihood estimator of θ_1 and θ_2.

- Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$.

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp\left[-\frac{(X_i - \theta_1)^2}{2\theta_2}\right]$$

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi\theta_2) - \frac{\sum_{i=1}^{n}(X_i - \theta_1)^2}{2\theta_2}$$

3.1 Maximum likelihood estimation

(**Example 3.4 con't**)

Normal Distribution

Since $\hat{\boldsymbol{\theta}} = arg\ max\, \ell(\boldsymbol{\theta})$

$$0 = \frac{\partial \ell(\theta)}{\partial \theta_1} = \frac{1}{\theta_2} \sum_{i=1}^{n} (X_i - \theta_1),$$

$$0 = \frac{\partial \ell(\theta)}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (X_i - \theta_1)^2$$

$\hat{\theta}_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\hat{\theta}_2 = S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2$

(**Example 3.4 con't**)

Normal Distribution

$$\boxed{\textbf{Check:}\ \text{Second derivatives}\,(<0)}$$

$$\hat{\theta}_1 = \overline{X} \quad \text{and} \quad \hat{\theta}_2 = S^2$$

3.2 Method of moments estimator

(Def 3.1)

The r-th moment about the origin of X

$$\mu_r = \mathrm{E}X^r$$

The r-th sample moment of a random sample $\boldsymbol{X}$

is defined as

$$m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \ldots$$

μ_r may contain information about the unknown parameter.

3.2 Method of moments estimator

Motivation Example:
$X \sim N(\mu, \sigma^2)$.

$$\mu_1 = EX = \mu, \;\; \mu_2 = E[X^2] = Var(X) + [E(X)]^2 = \sigma^2 + \mu^2$$

or

$$\mu = \mu_1, \;\; \sigma^2 = \mu_2 - \mu_1^2$$

The unknown parameter μ and σ^2 can be estimated if we find "good" estimators for μ_1 and μ_2.
By WLLN,

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p E(X) = \mu_1, \;\; m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_p E(X^2) = \mu_2$$

So $\hat{\mu} = m_1$ and $\hat{\sigma}^2 = m_2 - m_1^2$.

3.2 Method of moments estimator

Unknown parameter $\boldsymbol{\theta} \in \mathbb{R}^s$ can be expressed by

$$\boldsymbol{\theta} = h(\mu_1, \mu_2, \ldots, \mu_k), \;\; h : \mathbb{R}^k \to \mathbb{R}^s$$

In the **Motivation Example**:

$$s = 2, \;\; k = 2, \;\; \boldsymbol{\theta} = (\mu, \sigma^2), \;\; h = (h_1, h_2)$$

$$\mu_1 = h_1(\mu_1, \mu_2) = \mu_1, \;\; \sigma^2 = h_2(\mu_1, \mu_2) = \mu_2 - \mu_1^2$$

Method of Moments Estimator (MME)

$$\tilde{\boldsymbol{\theta}} = h(m_1, m_2, \cdots, m_k)$$

The observed value of $\tilde{\boldsymbol{\theta}}$ is called the method of moments estimate.

3.2 Method of Moments Estimator

(**Example 3.4 con't**)

Normal Distribution

$\boldsymbol{X} \sim \mathrm{N}(\theta_1, \theta_2)$

$\mu = \mu_1$ and $\sigma^2 = \mu_2 - \mu_1^2$

By the weak law of large numbers,

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \to_p \mathrm{E}(X) = \mu_1 \quad \text{and} \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_p \mathrm{E}(X^2) = \mu_2$$

MME of μ: m_1
MME of σ^2: $m_2 - m_1^2$

3.2 Method of Moments Estimator

(**Example 3.5**)

Gamma Distribution
Let $\boldsymbol{X}$ be an independent random sample from a Gamma distribution $Gamma(\alpha, \lambda)$, where $(\alpha, \lambda) \in \Omega$ and $\Omega = \{(\alpha, \lambda) : \alpha > 0, \lambda > 0\}$. Find the MME of α and λ.

- $$f(x) = \begin{cases} \dfrac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, & \text{for } x > 0; \\ 0, & \text{otherwise.} \end{cases}$$

  $$\mu_1 = \frac{\alpha}{\lambda} \text{ and } \mu_2 = \frac{\alpha^2 + \alpha}{\lambda^2}$$

3.2 Method of Moments Estimator

(**Example 3.5 con't**)

Gamma Distribution

$$\alpha = \lambda\mu_1 = \frac{(\mu_1)^2}{\mu_2 - (\mu_1)^2} \text{ and } \lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

<u>One</u> MME of (α, λ) is $(\tilde{\alpha}, \tilde{\lambda})$, where

$$\tilde{\alpha} = \frac{m_1^2}{m_2 - m_1^2} = \frac{\overline{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2} \quad \text{and} \quad \tilde{\lambda} = \frac{m_1}{m_2 - m_1^2} = \frac{\overline{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2}$$

3.2 Method of Moments Estimator

Note: The way to construct h is not unique. Usually, we use the lowest possible order moments to construct h, although this may not be the optimal way.

3.3 Estimator Properties

Criteria of goodness to compare different estimators:

- Unbiasedness
- Efficiency
- Consistency

3.3.1 Unbiasedness

(Def 3.4, 3.5) Suppose that $\hat{\theta}$ is an estimator of θ.

Unbiased estimator

The bias of an estimator $\hat{\theta}$ is defined as

$$\mathrm{Bias}(\hat{\theta}) = \mathrm{E}(\hat{\theta}) - \theta.$$

If $\mathrm{Bias}(\hat{\theta}) = 0$, $\hat{\theta}$ is called an unbiased estimator of θ. Otherwise, it is said to be biased.

Asymptotically unbiased estimator

$\hat{\theta}$ is an asymptotically unbiased estimator if

$$\lim_{n\to\infty} \mathrm{Bias}(\hat{\theta}) = \lim_{n\to\infty} [\mathrm{E}(\hat{\theta}) - \theta] = 0,$$

where n is the sample size.

3.3.1 Unbiasedness

(**Example 3.3 con't**)

Uniform Distribution

(i)Show that $\hat{\beta} = X_{(n)}$ is an asymptotically unbiased estimator of β.

- Let $Y = X_{(n)}$

  For $0 \le y \le \beta$,

$$P(Y \le y) = \prod_{i=1}^{n} P(X_i \le y) = \left(\frac{y}{\beta}\right)^n .$$

By Property 1.10,

$$\begin{aligned} \mathrm{E}(Y) &= \int_0^\infty P(Y > y)\mathrm{d}y = \int_0^\beta \left[1 - \left(\frac{y}{\beta}\right)^n\right] \mathrm{d}y = \beta - \frac{\beta^{n+1}}{(n+1)\beta^n} \\ &= \frac{n\beta}{n+1} \to \beta \text{ as } n \to \infty \end{aligned}$$

3.3.1 Unbiasedness

(**Example 3.3 con't**)

Uniform Distribution

$\hat{\beta} = X_{(n)}$

(ii)Modify this estimator of β to make it unbiased.

- From (i), $\mathrm{E}(Y) = \frac{n\beta}{n+1}$

$$\mathrm{E}\left(\frac{n+1}{n}Y\right) = \beta$$

$\dfrac{n+1}{n}Y$ is an unbiased estimator of β

3.3.1 Unbiasedness

(**Example 3.4 con't**)

Normal Distribution

$X_i \sim \mathrm{N}(\theta_1, \theta_2)$

$\hat{\theta}_1 = \overline{X}$ and $\hat{\theta}_2 = S^2$

Show that $\overline{X}$ is an unbiased estimator of θ_1, and S^2 is an asymptotically unbiased estimator of θ_2.

- $$\mathrm{E}(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} \theta_1 = \theta_1$$

$$\mathrm{E}(S^2) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]$$

For each i, we have

$$\begin{aligned}
\mathrm{E}\left[(X_i - \overline{X})^2\right] &= \mathrm{E}\left[(X_1 - \overline{X})^2\right] = \mathrm{Var}(X_1 - \overline{X}) \\
&= \mathrm{Var}(X_1 - \frac{X_1 + X_2 \cdots + X_n}{n}) \\
&= \mathrm{Var}(\frac{n-1}{n} X_1 - \sum_{i=2}^{n} \frac{X_i}{n}) \\
&= \frac{(n-1)^2}{n^2}\theta_2 + \frac{(n-1)}{n^2}\theta_2 = \frac{n-1}{n}\theta_2
\end{aligned}$$

So

$$\mathrm{E}(S^2) = \mathrm{E}(S^2) = \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right] = \frac{n-1}{n}\theta_2$$

$$\lim_{n\to\infty} \mathrm{E}(S^2) = \theta_2$$

3.3.2 Efficiency

(Def 3.6)

Mean squared error

Suppose that $\hat{\theta}$ is an estimator of θ. The mean squared error of $\hat{\theta}$ is

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

Use MSE to compare the goodness of two **unbiased** estimators

3.3.2 Efficiency

(Property 3.1)

If $\mathrm{Var}(\hat{\theta})$ exists, then the mean squared error of $\hat{\theta}$ is

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \left[\mathrm{Bias}(\hat{\theta})\right]^2.$$

(Remark 3.1)

$$\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}) = 0 \Longleftrightarrow \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}) = 0 \text{ and } \lim_{n\to\infty} \mathrm{Bias}(\hat{\theta}) = 0$$

For the unbiased estimators,

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta})$$

3.3.2 Efficiency

(Def 3.7)

Suppose that $\hat{\theta}$ and $\tilde{\theta}$ are two unbiased estimators of θ. The efficiency of $\hat{\theta}$ relative to $\tilde{\theta}$ is defined by

$$\text{Eff}\,(\hat{\theta}, \tilde{\theta}) = \frac{\text{Var}(\tilde{\theta})}{\text{Var}(\hat{\theta})}.$$

If $\text{Eff}\,(\hat{\theta}, \tilde{\theta}) > 1$, then we say that $\hat{\theta}$ is relatively more efficient than $\tilde{\theta}$.

3.3.2 Efficiency

(**Example 3.6**)

Relationship between sample size and efficiency

Let $(X_n : n \geq 1)$ be a sequence of independent random variables having the same finite mean and variance,

$\mu = \mathrm{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1)$

Then $\overline{X}$ is an unbiased estimator of μ, and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$.
Suppose that we now take two samples one of size n_1 and one of size n_2, and denote the sample means as $\overline{X}^{(1)}$ and $\overline{X}^{(2)}$.

$$\mathrm{Eff}\,(\overline{X}^{(1)}, \overline{X}^{(2)}) = \frac{\mathrm{Var}(\overline{X}^{(2)})}{\mathrm{Var}(\overline{X}^{(1)})} = \frac{n_1}{n_2}$$

Therefore, the larger is the sample size, the more efficient is the sample mean for estimating μ.

(**Example 3.3 con't**)

Uniform Distribution

Note that $\frac{n+1}{n}X_{(n)}$ is an unbiased estimator of β

Show that

(i) $2\overline{X}$ is also an unbiased estimator of β.

- $\mathrm{E}(\overline{X}) = \beta/2$

  $\mathrm{E}(2\overline{X}) = \beta$

3.3.2 Efficiency

(**Example 3.3 con't**)

Uniform Distribution

(ii) Compare the efficiency of these two estimators of β

- Variance of the estimator $X_{(n)}$

  Let $Y = X_{(n)}$

  For $0 \le y \le \beta$, $\mathrm{P}(Y \le y) = \prod_{i=1}^{n} \mathrm{P}(X_i \le y) = \left(\frac{y}{\beta}\right)^n$

  $\mathrm{Var}\left(\frac{n+1}{n} Y\right) = \mathrm{E}\left[\left(\frac{n+1}{n} Y\right)^2\right] - \left[\mathrm{E}\left(\frac{n+1}{n} Y\right)\right]^2$

  We have obtained that $\mathrm{E}(Y) = \frac{n\beta}{n+1}$ (**Example 3.3**)

  Next step is to calculate $\mathrm{E}(Y^2)$.

3.3.2 Efficiency

To calculate $\mathrm{E}(Y^2)$, let $Z = Y^2$

$\mathrm{P}(Z \leq z) = \mathrm{P}(Y \leq \sqrt{z}) = \left(\frac{\sqrt{z}}{\beta}\right)^n$ for $0 \leq z \leq \beta^2$

$$\begin{aligned}
\mathrm{E}(Z) &= \int_0^{\beta^2} 1 - \left(\frac{\sqrt{z}}{\beta}\right)^n dz \\
&= 2\int_0^{\beta} t\left(1 - \frac{t^n}{\beta^n}\right) \mathrm{d}t \quad (\text{ by setting } t = \sqrt{z}) \\
&= 2\left[\frac{t^2}{2} - \frac{t^{n+2}}{(n+2)\beta^n}\right]\Bigg|_0^{\beta} = 2\left[\frac{\beta^2}{2} - \frac{\beta^{n+2}}{(n+2)\beta^n}\right] \\
&= \frac{n}{n+2}\beta^2
\end{aligned}$$

Hence, $\mathrm{Var}\left(\frac{n+1}{n}Y\right) = \frac{\beta^2}{n(n+2)}$

3.3.2 Efficiency

(**Example 3.3 con't**)

Uniform Distribution

(ii) Compare the efficiency of these two estimators of β

- Variance of the estimator $2\overline{X}$

  $\mathrm{Var}(2\overline{X}) = 4\mathrm{Var}(\overline{X}) = 4 \cdot \frac{\beta^2}{12n} = \frac{\beta^2}{3n}$

- Compare the efficiency

  $\mathrm{Eff}\left(\frac{n+1}{n}Y, 2\overline{X}\right) = \frac{\mathrm{Var}(2\overline{X})}{\mathrm{Var}\left(\frac{n+1}{n}Y\right)} = \frac{n+2}{3}$

  For $n = 1$, $\frac{n+1}{n}Y$ and $2\overline{X}$ has the same efficiency.
  For $n > 1$, $\frac{n+1}{n}Y$ is more efficient than $2\overline{X}$.

3.3.2 Efficiency

- **Question 1**

  For a given unbiased estimator $\tilde{\theta}$, could we find another unbiased estimator $\tilde{\theta}_*$, which has a smaller variance than $\tilde{\theta}$?

3.3.2 Efficiency

- **Question 2**

  Among all unbiased estimators, could we find the **uniformly minimum variance unbiased estimator** (UMVUE)

$$\text{UMVUE} = \underset{\tilde{\theta} \text{ is unbiased}}{\arg\min} \ \text{Var}(\tilde{\theta}).$$

3.3.2 Efficiency

(Def 3.8)

Sufficient Statistic / Factorization Theorem

Suppose that the random sample $\boldsymbol{X}$ has a joint p.d.f. $\boldsymbol{f}(x_1, \ldots, x_n; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown parameter. The statistic $T := T(\boldsymbol{X})$ is sufficient for $\boldsymbol{\theta}$ if and only if

$$f(x_1, \cdots, x_n; \boldsymbol{\theta}) = g\big(T(x_1, \cdots, x_n); \boldsymbol{\theta}\big)\, h(x_1, \cdots, x_n),$$

where g depends on $x_1, \cdots, x_n$ only through $T(x_1, \cdots, x_n)$, and h does not depend on $\boldsymbol{\theta}$.

3.3.2 Efficiency

Sufficient statistic is not unique

$$
\begin{aligned}
&T = T(\boldsymbol{X}) \text{ is a sufficent statistic for } \theta \\
\Longleftrightarrow\; &v(T) = v\big(T(\boldsymbol{X})\big) \text{ is also a sufficient statistic for } \theta,
\end{aligned}
$$

where $v(\cdot)$ is an **invertible** function.

Example: If T is a sufficient statistic for $\boldsymbol{\theta}$, then T^3 is also a sufficient statistic for $\boldsymbol{\theta}$.
While, T^2 is NOT a sufficient statistic for $\boldsymbol{\theta}$.

3.3.2 Efficiency

(**Example 3.7**)

Uniform Distribution

Suppose that $\boldsymbol{X}$ is an independent random sample from a uniform distribution $U(\alpha, \beta)$. Find a sufficient statistic for (α, β).

- Let $\boldsymbol{\theta} = (\alpha, \beta)$

  The joint p.d.f. of $\boldsymbol{X}$ is

$$\begin{aligned} & \boldsymbol{f}(x_1, \cdots, x_n; \boldsymbol{\theta}) \\ = & \prod_{i=1}^{n} \left(\frac{1}{\beta - \alpha} \right) \mathrm{I}(\alpha \leq x_i \leq \beta) \\ = & \left(\frac{1}{\beta - \alpha} \right)^n \mathrm{I}(\alpha \leq \min_{1 \leq i \leq n} X_i) I(\max_{1 \leq i \leq n} X_i \leq \beta) \end{aligned}$$

$\left(\min_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} X_i \right)$ is a sufficient statistic for (α, β).

3.3.2 Efficiency

(**Example 3.8**)

Normal Distribution

Suppose that $\boldsymbol{X}$ is an independent random sample from a normal distribution $N(\mu, \sigma^2)$. Find a sufficient statistic for (μ, σ^2).

- **Method 1**
  Let $\theta = (\mu, \sigma^2)$
  The joint p.d.f. of $\boldsymbol{X}$ is

$$
\begin{aligned}
&\boldsymbol{f}(x_1, \cdots, x_n; \theta) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{n}{2\sigma^2} s^2\right) \exp\left(-\frac{n}{2\sigma^2}(\mu - \overline{x})^2\right).
\end{aligned}
$$

$(\overline{X}, S^2)$ is a sufficient statistic for (μ, σ^2)

3.3.2 Efficiency

(Property 3.2)

Sufficient Statistics for Exponential family

Let $\boldsymbol{X}$ be an i.i.d. random sample from a p.d.f. having the form

$$f(x;\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{s} p_i(\boldsymbol{\theta})t_i(x)\right) \quad \text{(expontial family)},$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_s) \in \Theta \subset \mathbb{R}^s$.

Then,

$$T(\boldsymbol{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \cdots, \sum_{j=1}^{n} t_s(X_j)\right)$$

is a sufficient statistic for $\boldsymbol{\theta}$.

3.3.2 Efficiency

(Remark 3.3)

Exponential Family

includes many of the most common distributions, eg

Normal; Exponential; Gamma; Chi Squared; Beta; Bernoulli; Poisson; Geometric; ...

3.3.2 Efficiency

(**Example 3.8 con't**)

Normal Distribution

- Find a sufficient statistic for (μ, σ^2)

Method 2

Let $\boldsymbol{\theta} = (\mu, \sigma^2)$

$$
\begin{aligned}
& f(x; \boldsymbol{\theta}) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2+\mu^2-2\mu x}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\mu^2}{2\sigma^2}\right) \exp\left(\frac{n\mu}{\sigma^2}\frac{x}{n} - \frac{n}{2\sigma^2}\frac{x^2}{n}\right)
\end{aligned}
$$

$(\overline{X}, \overline{X^2})$ is *ALSO* a sufficient statistic for (μ, σ^2)

3.3.2 Efficiency

Note that both $(\overline{X}, S^2)$ and $(\overline{X}, \overline{X^2})$ are sufficient statistics for $\boldsymbol{\theta}$, and

$$v(\overline{X}, \overline{X^2}) = (\overline{X}, S^2)$$

where $(z_1, z_2) \rightarrow_v (z_1, z_2 - z_1^2)$ is invetible.

Fact: $S^2 = \overline{X^2} - (\overline{X})^2$

3.3.2 Efficiency

- **Question 1**

  For a given unbiased estimator $\tilde{\theta}$, could we find another unbiased estimator $\tilde{\theta}_*$, which has a <u>smaller variance</u> than $\tilde{\theta}$?

3.3.2 Efficiency

(Theorem 3.1)

Rao-Blackwell Theorem

Let $\tilde{\theta}$ be an unbiased estimator of θ with $\mathrm{E}(\tilde{\theta}^2) < \infty$,

and $T := T(\boldsymbol{X})$ be a **sufficient statistic** for θ.

Let $w(t) = \mathrm{E}(\tilde{\theta}|T = t)$. Then, $\tilde{\theta}_* = w(T)$ is an unbiased

estimator of θ and $\mathrm{Var}(\tilde{\theta}_*) \leq \mathrm{Var}(\tilde{\theta})$.

3.3.2 Efficiency

- **Answer 1**

  By using the sufficient statistic T, we can always get a better unbiased estimator (in terms of efficiency) from an initial unbiased estimator.

3.3.2 Efficiency

From Rao-Blackwell Theorem,

the UMVUE is a function of the sufficient statistic T.

(Property 3.3)

If $\hat{\theta}$ is the UMVUE of θ, then $\hat{\theta}$ is unique.

Next step: How to find the UMVUE?

- Two approaches:
  1. Complete and sufficient statistic method
  2. CRLR method

3.3.2.1 UMVUE: Complete and sufficient statistic method

(Def 3.9)

Complete Statistic

For a given random sample $\boldsymbol{X}$, $T := T(\boldsymbol{X})$ is a complete statistic of θ if

$$\mathrm{E}\left[z(T)\right] = 0 \text{ for all } \theta \text{ implies } \mathrm{P}\left(z(T) = 0\right) = 1 \text{ for all } \theta.$$

3.3.2.1 UMVUE: Complete and sufficient statistic method

(Property 3.4)

Complete statistic for exponential family

Let $\boldsymbol{X}$ be an i.i.d. random sample from a p.d.f. having the form

$$f(x;\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{i=1}^{s} p_i(\boldsymbol{\theta})t_i(x)\right) \quad \text{(expontial family)},$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_s) \in \Theta \subset \mathbb{R}^s$. Then,

$$T(\boldsymbol{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \cdots, \sum_{j=1}^{n} t_s(X_j)\right)$$

is a complete statistic for θ as long as the parameter Θ contains an open set in $\mathbb{R}^s$.

3.3.2.1 UMVUE: Complete and sufficient statistic method

(Property 3.4)
From Properties 3.2 and 3.4,

if $\boldsymbol{X}$ is from an exponential family and the related parameter Θ contains an open set in $\mathbb{R}^s$, then

$$T(\boldsymbol{X}) = \left(\sum_{j=1}^{n} t_1(X_j), \sum_{j=1}^{n} t_2(X_j), \cdots, \sum_{j=1}^{n} t_s(X_j) \right)$$

is a <u>complete and sufficient statistic</u> for θ

(**Example 3.8 con't**)

$(\overline{X}, \overline{X^2})$ is also a complete statistic for $\boldsymbol{\theta}$.

3.3.2.1 UMVUE: Complete and sufficient statistic method

(Theorem 3.2)

Let $T := T(\boldsymbol{X})$ be a complete and sufficient statistic for θ,

and $\phi(T)$ be any estimator based only on T.

Then, $\phi(T)$ is the unique UMVUE of its expected value $\mathrm{E}[\phi(T)]$.

3.3.2.1 UMVUE: Complete and sufficient statistic method

- **First approach to find the UMVUE**

1. Find a complete and sufficient statistic T for θ
2. Find a functional $\phi_0(\cdot)$ such that $\mathrm{E}[\phi_0(T)] = \theta$

3.3.2.1 UMVUE: Complete and sufficient statistic method

(**Example 3.8 con't**)

Normal Distribution

$\boldsymbol{X} \sim \mathrm{N}(\mu, \sigma^2)$

- Find the UMVUE of $\boldsymbol{\theta} = (\mu, \sigma^2)$

  $(\overline{X}, \overline{X^2})$ is a complete and sufficient statistic for θ

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{E}\left(\frac{n}{n-1} S^2\right) = \sigma^2$$

  $\overline{X}$ is the UMVUE of μ,
  $\frac{n}{n-1} S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ is the UMVUE of σ^2.

3.3.2.1 UMVUE: Complete and sufficient statistic method

(**Example 3.9**)

Poisson Distribution

Let $\boldsymbol{X}$ be i.i.d. random sample from Poisson distribution with parameter λ. Find the UMVUE of λ.

- $f(x;\theta) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{e^{-\lambda}}{x!}e^{\log(\lambda^x)} = \frac{e^{-\lambda}}{x!}e^{x\log(\lambda)} = \frac{e^{-\lambda}}{x!}e^{\frac{x}{n}n\log(\lambda)}$

  $T(\boldsymbol{X}) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a complete and sufficient statistic for λ.

  $\mathrm{E}(\overline{X}) = \lambda$

  Therefore, $\overline{X}$ is the UMVUE of λ

3.3.2.2 UMVUE: CRLB method

- **Second approach to find the UMVUE**

1. Find the lower bound of $\text{Var}(\tilde{\theta})$ for all unbiased estimators;
2. Find an unbiased estimator $\hat{\theta}$ whose variance achieves this lower bound.

Note: Sufficient but not necessary condition for the UMVUE

3.3.2.2 UMVUE: CRLB method

(Def 3.10) To consider the lower bound of $\mathrm{Var}(\tilde{\theta})$, we introduce

Fisher information about θ

$$I_n(\theta) = \mathrm{E}\left[\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)^2\right]$$

where $\ell(\theta) = \log L(\theta)$ is the log-likelihood function of the random sample.

3.3.2.2 UMVUE: CRLB method

(Theorem 3.3)

1. $I_n(\theta) = nI(\theta)$, where

$$I(\theta) = \mathrm{E}\left[\left(\frac{\partial \log f(X;\theta)}{\partial \theta}\right)^2\right],$$

and X has the same distribution as the population

2. $I(\theta) = -\mathrm{E}\left[\frac{\partial^2 \log f(X;\theta)}{\partial \theta^2}\right]$;

3. Cramer-Rao inequality:

$$\mathrm{Var}(\hat{\theta}) \geq \frac{1}{I_n(\theta)},$$

where $\hat{\theta}$ is an unbiased estimator of θ, and $\frac{1}{I_n(\theta)}$ is called the $\boxed{Cramer-Rao\ lower\ bound(CRLB)}$

3.3.2.2 UMVUE: CRLB method

(Corollary 3.1)

If $\hat{\theta}$ is an unbiased estimator of θ and $\mathrm{Var}(\hat{\theta}) = \dfrac{1}{I_n(\theta)}$, then $\hat{\theta}$ is the UMVUE of θ.

3.3.2.1 UMVUE: Complete and sufficient statistic method

(**Example 3.10**)

Normal Distribution

$\boldsymbol{X} \sim \mathrm{N}(\mu, \sigma^2)$

- Find the UMVUE of $\boldsymbol{\theta} = (\mu, \sigma^2)$

  For $-\infty < x < \infty$,

$$f(x; \mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

  which implies that

$$\log f(x; \mu) = -\log(\sigma\sqrt{2\pi}) - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2,$$

$$\text{and } \frac{\partial \log f(x; \mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}.$$

3.3.2.1 UMVUE: Complete and sufficient statistic method

(**Example 3.8 con't**)

Normal Distribution

$\boldsymbol{X} \sim \mathrm{N}(\theta_1, \theta_2)$

- Find the UMVUE of θ

$$I(\mu) = \mathrm{E}\left[\left(\frac{\partial \log f(X;\mu)}{\partial \mu}\right)^2\right] = \mathrm{E}\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}$$

$$CRLB = \frac{1}{I_n(\mu)} = \frac{1}{nI(\mu)} = \frac{\sigma^2}{n}.$$

Recall that $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$.

Thus, $\overline{X}$ is the UMVUE of μ.

3.3.2.1 UMVUE: CRLB method

(**Example 3.10**)

Bernoulli Distribution

- Show that $\overline{X}$ is the UMVUE of the parameter θ of a Bernoulli population.

  For $x = 0$ or 1,

$$f(x;\theta) = \theta^x(1-\theta)^{1-x},$$

  which implies that

$$\begin{aligned}
\frac{\partial \log f(x;\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta}\left[x\log\theta + (1-x)\log(1-\theta)\right] \\
&= \frac{x}{\theta} - \frac{1-x}{1-\theta} \\
&= \frac{x}{\theta(1-\theta)} - \frac{1}{1-\theta}.
\end{aligned}$$

3.3.2.1 UMVUE: CRLB method

(**Example** 3.10)

Bernoulli Distribution

- Show that $\overline{X}$ is the UMVUE of the parameter θ of a Bernoulli population.

  Noting that

  $$\mathrm{E}\left[\frac{X}{\theta(1-\theta)}\right] = \frac{1}{1-\theta},$$

  we have

  $$I(\theta) = \mathrm{E}\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right] = \mathrm{Var}\left[\frac{X}{\theta(1-\theta)}\right] = \frac{1}{\theta(1-\theta)}.$$

  Hence,

  $$CRLB = \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}.$$

  Since $\mathrm{E}(\overline{X}) = \theta$ and $\mathrm{Var}(\overline{X}) = \frac{\theta(1-\theta)}{n}$, $\overline{X}$ is the UMVUE of θ.

3.3.2.1 UMVUE: CRLB method

(Theorem 3.4)

Suppose that $\hat{\theta}$ is the MLE of a parameter θ of a population distribution. Then, under certain regular conditions, as $n \to \infty$,

$$\frac{\hat{\theta} - \theta}{\sqrt{1/I_n(\theta)}} \to_d \mathrm{N}(0, 1).$$

If $\hat{\theta}$ is MLE of θ, the above theorem implies that $\mathrm{Var}(\hat{\theta}) \approx \dfrac{1}{I_n(\theta)}$ when n is large. That is, the MLE $\hat{\theta}$ can achieve the CRLB asymptotically.

3.3.3 Consistency

(Def 3.11)

Consistent estimator

$\hat{\theta}$ is a consistent estimator of θ, if

$$\hat{\theta} \rightarrow_p \theta,$$

that is, for any $\epsilon > 0$,

$$\mathrm{P}\left(|\hat{\theta} - \theta| > \epsilon\right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

3.3.3 Consistency

Note: The unbiasedness alone does not imply the consistency.
Example:

$$\hat{\theta} = I(0 < X_1 < 1/2) - I(1/2 < X_1 < 1), \;\; X \sim U(0,1).$$

$$E(\hat{\theta}) = 0.$$

$\hat{\theta}$ is an unbiased estimator of $\theta = 0$.
But as $\hat{\theta}$ takes value of either 1 or -1, $\hat{\theta}$ is not consistent.

3.3.3 Consistency

(Property 3.5, 3.6)

If $\hat{\theta}$ is an (asymptotically) unbiased estimator of a parameter θ and

$\mathrm{Var}(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta}$ is a consistent estimator of θ.

3.3.3 Consistency

(Property 3.7)

If $\hat{\theta} \to_p \theta$ and $\tilde{\theta} \to_p \theta'$, then

(i) $\hat{\theta} \pm \tilde{\theta} \to_p \theta \pm \theta'$;

(ii) $\hat{\theta} \cdot \tilde{\theta} \to_p \theta \cdot \theta'$;

(iii) $\hat{\theta}/\tilde{\theta} \to_p \theta/\theta'$ assuming that $\tilde{\theta} \neq 0$ and $\theta' \neq 0$;

(iv) if g is any real-valued function that is continuous at θ, $g(\hat{\theta}) \to_p g(\theta)$.

3.3.3 Consistency

(**Example 3.12**)

Suppose that $\boldsymbol{X}$ is an independent random sample from a population with the finite mean $\mu = \mathrm{E}(X_1)$, finite variance $\sigma^2 = \mathrm{Var}(X_1)$, and finite fourth moment $\mu_4 = \mathrm{E}(X_1^4)$.

- Show that $\overline{X}$ is a consistent estimator of μ, and S^2 is a consistent estimator of σ^2.

  - For μ:

    $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

  - For S^2:

$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \to_p \mu_2 = \mathrm{E}(X_1^2).$$

$$(\overline{X})^2 \to_p \mu^2.$$

$$S^2 \to_p \mu_2 - \mu^2 = \sigma^2$$

3.3.3 Consistency

Comprehensive Question

(**Example 3.13**)

Let $\boldsymbol{X}$ be an independent random sample from a population with a p.d.f.

$$f(x;\theta) = \frac{2x}{\theta^2} \mathrm{I}(0 < x \leq \theta).$$

(i) Find the UMVUE of θ;
(ii) Show that this UMVUE is a consistent estimator of θ;
(iii) Find the MLE of θ;
(iv) Find a MME of θ;
(v) Will the MLE be better than the MME in terms of efficiency?

3.3.3 Consistency

(**Example** 3.13)

- Find the UMVUE of θ

  Complete and sufficient statistic

  - Sufficient statistic

$$
\begin{aligned}
\boldsymbol{f}(x_1, x_2, \cdots, x_n; \theta) &= f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) \\
&= \prod_{i=1}^{n} \frac{2x_i}{\theta^2} \mathrm{I}(0 < x_i < \theta) \\
&= \left[\prod_{i=1}^{n} x_i \mathrm{I}(x_i > 0) \right] \left[\left(\frac{2}{\theta^2} \right)^n \mathrm{I}\left(\max_{1 \le i \le n} x_i \le \theta \right) \right]
\end{aligned}
$$

By Def 3.8, $T := \max_{1 \le i \le n} X_i$ is a sufficient statistic for θ.

3.3.3 Consistency

(**Example** 3.13)

- Find the UMVUE of θ

  Complete and sufficient statistic

  - Let $z(\cdot)$ be a functional such that $\mathrm{E}(z(T)) = 0$ for all $\theta > 0$, i.e.,

    $$0 = \mathrm{E}(z(T)) = \int_0^\theta z(t)\frac{2nt^{2n-1}}{\theta^{2n}}dt = \frac{2n}{\theta^{2n}}\int_0^\theta z(t)t^{2n-1}dt$$

    for all $\theta > 0$.

    $\int_0^\theta z(t)t^{2n-1}dt = 0$ for all $\theta > 0$, which

    implies that $z(\theta)\theta^{2n-1} = 0$ and hence $z(\theta) = 0$.

    By Def 3.9, T is a complete statistic for θ

3.3.3 Consistency

(**Example** 3.13)

- Find the UMVUE of θ

$$E(T) = \int_0^\theta t\frac{2nt^{2n-1}}{\theta^{2n}}dt = \frac{2n}{\theta^{2n}}\int_0^\theta t^{2n}dt = \frac{2n}{2n+1}\theta,$$

By Theorem 3.2, $Y := \dfrac{2n+1}{2n}T$ is the UMVUE of θ.

3.3.3 Consistency

(**Example 3.13**)

- Show that this UMVUE is a consistent estimator of θ

$$E(Y^2) = \frac{(2n+1)^2}{(2n)\theta^{2n}} \int_0^\theta t^{2n+1} dt = \frac{(2n+1)^2}{(2n)(2n+2)}\theta^2,$$

$$\mathrm{Var}(Y) = E(Y^2) - [E(Y)]^2 = \frac{(2n+1)^2}{(2n)(2n+2)}\theta^2 - \theta^2 \to 0$$

as $n \to \infty$.

By Property 3.5, Y is a consistent estimator of θ.

3.3.3 Consistency

(**Example** 3.13)

- Find the MLE of θ

  The MLE of θ is T.

3.3.3 Consistency

(**Example** 3.13)

- Find a MME of θ

$$\mathrm{E}(X) = \int_0^{\theta} x \frac{2x}{\theta^2} dx = \frac{2}{3}\theta$$

Hence, $\frac{3}{2}\overline{X}$ is a MME of θ.