Definition 4

Let $X_1, \dots, X_n$ be i.i.d with the pdf $f(x, \theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \dots, X_n)$ is an estimator of θ_0 such that $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{\hat{\theta}_1})$.

(i) The number

$$e(\hat{\theta}_1) = \frac{\frac{1}{I(\theta_0)}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic efficiency of $\hat{\theta}_{1n}$.

(ii) If $e(\hat{\theta}_1) = 1$, it is said that $\hat{\theta}_{1n}$ is asymptotically efficient.

(iii) Suppose $\hat{\theta}_{2n} = \hat{\theta}_{2n}(X_1, \dots, X_n)$ is an estimate of θ_0 such that $\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{\hat{\theta}_2})$.

The number

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma^2_{\hat{\theta}_2}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic relative efficiency of $\hat{\theta}_{1n}$ with respect to $\hat{\theta}_{2n}$.

Example 6

$X_i = \theta + e_i$, $i = 1, \dots, n$, $e_1, \dots, e_n$ i.i.d

(i) $\frac{1}{2}e^{-|x|}$

$e_1 \sim$ Laplace distribution

MLE of θ is $\hat{\theta}_{1n} = Me\{X_1, \dots, X_n\} = Q_2$, $I(\theta_0) = 1$.

Also $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1)$.

Let $\hat{\theta}_{2n} = \bar{X}_n$

CLT implies

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = \text{var} X_1 = \text{var}(e_1 + \theta) = \text{var}\, e_1 = E e_1^2 = \int_{-\infty}^{+\infty} z^2 \frac{1}{2} e^{-|z|} dz$

$$= \int_0^{\infty} z^2 e^{-z} dz = \Gamma(3) = 2$$

Thus, $e(Q_2, \bar{X}) = \frac{2}{1} = 2$.

The sample median is twice as efficient as the sample mean (asymptotically).

(ii) $e_i \sim N(0,1)$.

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \tfrac{\pi}{2}), \qquad \frac{\pi}{2} = \frac{1}{[2f(0)]^2}$$

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \sim N(0,1),$$

$$e(Me, \bar{X}) = \frac{1}{\frac{\pi}{2}} = \frac{2}{\pi} \approx 0.636 \approx \frac{1}{1.57}$$

$\bar{X}$ is 1.57 times more efficient than Q_2.

Corollary 3

Under the assumptions of Theorem 2, suppose $g(x)$ is a continuous function of x which is differentiable at θ_0 such that $g'(\theta_0) \neq 0$. Then

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{\mathcal{D}} N\left(0, \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right).$$

3. Numerical finding of MLEs (Newton's method)

$\hat{\theta}^{(0)}$ - initial guess, $\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \dfrac{l'(\hat{\theta}^{(0)})}{l''(\hat{\theta}^{(0)})}$, and so on ~~etc~~

Example 1

$X_1, \dots, X_n$ i.i.d. $f(x,\theta) = \dfrac{\exp\{-(x-\theta)\}}{[1+\exp\{-(x-\theta)\}]^2}$ $\quad x \in \mathbb{R}$, $\theta \in \mathbb{R}$

No explicit form

II Sufficiency

1. Measures of Quality of Estimators

Suppose $f(x,\theta)$, $\theta \in \Theta$ is the "density" of a variable X. Consider a point estimator $Y_n = u(X_1, .., X_n)$ based on a sample $X_1, .., X_n$.

Definition 1

For a given integer n, $Y = u(X_1, .., X_n)$ is called a minimum variance unbiased estimator, (MVUE), of the parameter θ, if Y is unbiased and its variance is smaller than or equal to the variance of every other unbiased estimate of θ.

Example 1

$X_1, .., X_9$ i.i.d. $X_1 \sim N(\theta, \sigma^2)$.

X_1 - unbiased estimator of θ, $Var\, X_1 = \sigma^2$

$\bar{X}$ - —"— —"—, $Var\, \bar{X} = \frac{\sigma^2}{n}$

Problem: Is there ~~any~~ ~~another~~ an unbiased estimate of θ with variance smaller than $\frac{\sigma^2}{n}$?

2. Sufficient Statistics

Suppose that $X_1, .., X_n$ are i.i.d. random variables with the density $f(x,\theta)$, $\theta \in \Theta$, ~~and~~ Let $Y_1 = u_1(X_1, .., X_n)$ be a statistic.

Example 1

$X_1, \dots, X_n$ i.i.d.

$$X_i \sim f(x, \theta) = \theta^x (1-\theta)^{1-x}, \quad x = 0, 1, \quad \theta \in (0,1).$$

Then,

$$Y_1 = \sum_{i=1}^n X_i \sim f_{Y_1}(y_1, \theta) = \binom{n}{y_1} \theta^{y_1} (1-\theta)^{n-y_1}, \quad y_1 = 0, 1, \dots, n.$$

We will find the conditional probability

$$P(X_1 = x_1, \dots, X_n = x_n \mid Y_1 = y_1) = P(A \mid B),$$

say, where $y_1 = 0, 1, \dots, n$.

(i) If $\sum_{i=1}^n x_i \neq y_1$, then $P(A|B) = 0$ because $A \cap B = \emptyset$.

(ii) If $\sum_{i=1}^n x_i = y_1$, then $A \subset B$, ~~and~~ Thus $A \cap B = A$ & $P(A|B) = \frac{P(A)}{P(B)} =$

$$\frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}}{\binom{n}{y_1} \theta^{y_1} (1-\theta)^{n-y_1}} = \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}}{\binom{n}{\sum_i x_i} \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n - \sum_{i=1}^n x_i}} = \frac{1}{\binom{n}{\sum_i x_i}} \text{ and}$$

it does not depend on θ.

In general, let $f_{Y_1}(y_1, \theta)$ be the "density" of the statistic $Y_1 = u_1(X_1, \dots, X_n)$. The conditional probability

$$X_1 = x_1, \dots, X_n = x_n \mid Y_1 = y_1$$

equals

$$\frac{f(x_1, \theta) \cdot \ldots \cdot f(x_n, \theta)}{f_{Y_1}(u_1(X_1, \dots, X_n), \theta)}.$$

Definition 1

The statistic Y_1 is called a sufficient statistic for the parameter θ if and only if

$$\frac{f(x_1,\theta)\cdot\ldots\cdot f(x_n,\theta)}{f_{Y_1}\left(u_1(x_1,\ldots,x_n),\theta\right)} = H(x_1,\ldots,x_n),$$

where $H(x_1,\ldots,x_n)$ does not depend upon θ.

Example 2

$X_1,\ldots,X_n$ i.i.d $X_1 \sim \Gamma(2,\theta)$ $\quad f(x,\theta) = \frac{x e^{-\frac{x}{\theta}}}{\Gamma(2)\,\theta^2} \mathbb{1}_{(0,+\infty)}(x)$

$Y_1 = \sum_{i=1}^n X_i \sim \Gamma(2n,\theta)$ $\quad f_{Y_1}(y_1,\theta) = \frac{1}{\Gamma(2n)\,\theta^{2n}} y_1^{2n-1} e^{-y_1/\theta} \mathbb{1}_{(0,+\infty)}(y_1)$

We have

$$\frac{\prod_{i=1}^n f(x_i,\theta)}{f_{Y_1}(y_1,\theta)} = \frac{\frac{x_1 e^{-x_1/\theta}}{\Gamma(2)\,\theta^2}\cdot\ldots\cdot\frac{x_n e^{-x_n/\theta}}{\Gamma(2)\,\theta^2}}{\frac{1}{\Gamma(2n)\,\theta^{2n}} y_1^{2n-1} e^{-y_1/\theta}} = \frac{\frac{1}{\theta^{2n}}\left(\prod_{i=1}^n x_i\right) e^{-\sum_{i=1}^n x_i/\theta}}{\frac{1}{\Gamma(2n)\,\theta^{2n}}\left(\sum_{i=1}^n x_i\right)^{2n-1} e^{-\sum_{i=1}^n x_i/\theta}} =$$

$$\frac{\Gamma(2n)\left(\prod_{i=1}^n x_i\right)}{\left(\sum_{i=1}^n x_i\right)^{2n-1}}$$ does not depend on θ.

Y_1 - sufficient statistic for θ.

Example 3

$Y_1 \leqslant \ldots \leqslant Y_n$ - the order statistics of a random sample of size n from the distribution with pdf

$$f(x,\theta) = e^{-(x-\theta)} \mathbb{1}_{(\theta,+\infty)}(x).$$