6.434/16.391 Statistics for Engineers and Scientists

Lecture 12 10/24/2012

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

Lecture 11 10/24/2011

SCORE FUNCTION AND FISHER INFORMATION

Score function

- Definition: Let $X_1, X_2, \ldots, X_n$ be random variables with joint PDF (or PMF) $f_{\mathbf{X}}(\mathbf{x}|\theta)$. The score function is defined as

$$V(\mathbf{x},\theta) = \frac{f'(\mathbf{x},\theta)}{f(\mathbf{x},\theta)}, \quad \text{where } f'(\mathbf{x},\theta) = \frac{\partial}{\partial\theta} f(\mathbf{x},\theta)$$

- Note:
  - $V(\mathbf{X},\theta)$ is random
  - The score function can be expressed as

$$\begin{aligned} V(\mathbf{x},\theta) &= \frac{\partial}{\partial\theta} \ln f(\mathbf{x},\theta) \\ &= \frac{\partial}{\partial\theta} L(\theta|\mathbf{x}) \end{aligned}$$

  where $L(\theta|\mathbf{x})$ is the log likelihood function

Fisher information

- Definition: The Fisher information in the observation $\mathbf{X}$ about the parameter θ is given by

$$I_{\mathbf{X}}(\theta) = \mathbb{E}_{\mathbf{X}} \left\{ [V(\mathbf{X}, \theta)]^2 \right\}$$
$$= \mathbb{E}_{\mathbf{X}} \left\{ \left[\frac{\partial}{\partial \theta} L(\theta | \mathbf{X}) \right]^2 \right\}$$

- One should start to suspect that there is a strong connection between MLE and the Fisher information
- In fact, under so called "regularity conditions", one can show that MLE $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ is a CAN estimator with

$$J^2(\theta) = \frac{1}{I_X(\theta)}$$

Example 1

- Let X be distributed as $\mathcal{N}(\mu, \sigma^2)$. Find $I(\mu)$ and $I(\sigma^2)$
- First of all, the likelihood function is

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

Then we have

$$\ln f(x|\mu, \sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial}{\partial\mu}\ln f(x|\mu, \sigma^2) = +\frac{1}{\sigma^2}(x-\mu)$$

$$\frac{\partial}{\partial\sigma^2}\ln f(x|\mu, \sigma^2) = -\frac{1}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}(x-\mu)^2$$

$$\frac{\partial}{\partial\sigma}\ln f(x|\mu, \sigma^2) = -\frac{1}{2}\frac{2\sigma}{\sigma^2} + \frac{2}{2\sigma^3}(x-\mu)^2$$

$$= -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2$$

Example 1

- For parameter μ,

$$V(x,\mu) = \frac{1}{\sigma^2}(x-\mu)$$

$$\begin{aligned} I_X(\mu) &= \mathbb{E}\left\{V^2(X,\mu)\right\} \\ &= \frac{1}{\sigma^4}\mathbb{E}\left\{(X-\mu)^2\right\} \\ &= \frac{1}{\sigma^2} \end{aligned}$$

Example 1

- For parameter σ^2,

$$V(x,\sigma^2) = -\frac{1}{2}\frac{1}{\sigma^2} + \frac{1}{2}\frac{1}{\sigma^4}(x-\mu)^2$$

$$\begin{aligned} I_X(\sigma^2) &= \mathbb{E}\left\{\frac{1}{4}\frac{1}{\sigma^4} - \frac{2}{4}\frac{1}{\sigma^6}(X-\mu)^2 + \frac{1}{4}\frac{1}{\sigma^8}(X-\mu)^4\right\} \\ &= \frac{1}{4}\frac{1}{\sigma^4} - \frac{1}{2}\frac{1}{\sigma^6}\mathbb{E}\left\{(X-\mu)^2\right\} + \frac{1}{4\sigma^4}\mathbb{E}\left\{\frac{(X-\mu)^4}{\sigma^4}\right\} \\ &= -\frac{1}{4\sigma^4} + \frac{1}{4\sigma_4}\alpha^4 \\ &= \frac{1}{4\sigma^4}(\alpha_4 - 1) \end{aligned}$$

where $\alpha_4 = \mathrm{E}\left\{\frac{(X-\mu)^4}{\sigma^4}\right\}$

Example 1

- For parameter σ,

$$
\begin{aligned}
V(x,\sigma) &= -\frac{1}{\sigma} + \frac{1}{\sigma^3}(x-\mu)^2 \\
I_X(\sigma^2) &= \mathbb{E}\left\{\frac{1}{\sigma^2} - \frac{2}{\sigma^4}(X-\mu)^2 + \frac{1}{\sigma^6}(X-\mu)^4\right\} \\
&= \frac{1}{\sigma^2} - \frac{2}{\sigma^4}\mathbb{E}\left\{(X-\mu)^2\right\} + \frac{1}{\sigma^2}\mathbb{E}\left\{\frac{(X-\mu)^4}{\sigma^4}\right\} \\
&= -\frac{1}{\sigma^2} + \frac{1}{\sigma^2}\alpha_4 \\
&= \frac{1}{\sigma^2}(\alpha_4 - 1)
\end{aligned}
$$

Example 2

- Let X be a Bernoulli random variable with p the probability of success, i.e.,
$$\mathbb{P}\{X = x|p\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$
Find the Fisher information for p

- First,
$$\ln \mathbb{P}\{X = x|p\} = x \ln p + (1-x)\ln(1-p)$$
The score function is
$$V(x,p) = \frac{\partial}{\partial p} \ln \mathbb{P}\{X = x|p\} = \frac{x}{p} - \frac{1-x}{1-p}$$

Example 2

- Finally,

$$\begin{aligned} I_X(p) &= \mathbb{E}\left\{V^2(X,p)\right\} \\ &= \mathbb{E}\left\{\frac{X^2}{p^2} - \frac{2X(1-X)}{p(1-p)} + \frac{(1-X)^2}{(1-p)^2}\right\} \\ &= \frac{1}{p(1-p)} \end{aligned}$$

- Note: Since

$$V(x,p) = \frac{(1-p)x - p(1-x)}{p(1-p)} = \frac{x-p}{p(1-p)}$$

we have $\mathbb{E}\{V(X,p)\} = 0$. Therefore,

$$I_X(p) = \mathbb{V}\{V(X,p)\} = \frac{1}{p^2(1-p)^2}\mathbb{V}\{X-p\} = \frac{1}{p(1-p)}$$

Fisher information in independent observations

- Lemma: Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent observation for θ. Then

$$I_{\mathbf{XY}}(\theta) = I_{\mathbf{X}}(\theta) + I_{\mathbf{Y}}(\theta)$$

- Proof: The score function

$$\begin{aligned}
V(\mathbf{x}, \mathbf{y}, \theta) &= \frac{\partial}{\partial \theta} L(\theta | \mathbf{x}, \mathbf{y}) \\
&= \frac{\partial}{\partial \theta} \ln f(\mathbf{xy}|\theta) \\
&= \frac{\partial}{\partial \theta} \ln \left[f(\mathbf{x}|\theta) f(\mathbf{y}|\theta) \right] \\
&= \frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta) + \frac{\partial}{\partial \theta} \ln f(\mathbf{y}|\theta) \\
&= V(\mathbf{x}, \theta) + V(\mathbf{y}, \theta)
\end{aligned}$$

Fisher information in independent observations

- Therefore,

$$\begin{aligned} I_{\mathbf{XY}}(\theta) &= \mathbb{V}\left\{V(\mathbf{X}, \mathbf{Y}, \theta)\right\} \\ &= \mathbb{V}\left\{V(\mathbf{X}, \theta) + V(\mathbf{Y}, \theta)\right\} \\ &= \mathbb{V}\left\{V(\mathbf{X}, \theta)\right\} + \mathbb{V}\left\{V(\mathbf{Y}, \theta)\right\} \\ &= I_{\mathbf{X}}(\theta) + I_{\mathbf{Y}}(\theta) \end{aligned}$$

  where the third equality is due to the independence between $\mathbf{X}$ and $\mathbf{Y}$

- In particular, if $X_1, X_2, \ldots, X_n$ are i.i.d., then

$$I_{\mathbf{X}}(\theta) = n I_{X_1}(\theta)$$

Lecture 12 10/24/2012

INFORMATION INEQUALITY

Regularity conditions

- Satisfy certain technical conditions
  - The set of possible values of the random variable X do not depend on the unknown parameter θ
  - Integration and differentiation can be interchanged
- Mathematically
  - The set $\mathcal{A} = \{\mathbf{x} : \ f_{\mathbf{X}}(\mathbf{x}|\theta) > 0\}$ does not depend on θ for all $\mathbf{x} \in \mathcal{A}$ and $\theta \in \Theta$, $\frac{\partial}{\partial\theta} \ln f(\mathbf{x}, \theta)$ exists and is finite
  - If T is any statistic such that

$$\mathbb{E}_X \left\{ |T(\mathbf{X})| \right\} < \infty, \ \ \forall \, \theta \in \Theta$$

Then we require that

$$\frac{\partial}{\partial\theta} \int \cdots \int T(\mathbf{x}) f(\mathbf{x}, \theta) d\mathbf{x} = \int \cdots \int T(\mathbf{x}) \frac{\partial}{\partial\theta} f(\mathbf{x}, \theta) d\mathbf{x}$$

Regularity conditions

- Under regularity conditions,

$$
\begin{aligned}
\mathbb{E}\left\{V(\mathbf{X},\theta)\right\} &= \mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f(\mathbf{X}|\theta)\right\} \\
&= \int\cdots\int \frac{f'(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} f(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\int\cdots\int f(\mathbf{x}|\theta)d\mathbf{x} \\
&= 0
\end{aligned}
$$

- Fisher information

$$
\begin{aligned}
I_{\mathbf{X}}(\theta) &= \mathbb{E}\left\{V^2(\mathbf{X},\theta)\right\} \\
&= \mathbb{E}\left\{V^2(\mathbf{X},\theta)\right\} - \underbrace{\mathbb{E}^2\left\{V(\mathbf{X},\theta)\right\}}_{=0} \\
&= \mathbb{V}\left\{V(\mathbf{X},\theta)\right\}
\end{aligned}
$$

Information inequality

- Theorem (Information inequality): Let $T(\mathbf{X})$ be any estimator of some function of θ such that $\mathbb{V}\{T(\mathbf{X})\} < \infty$ for all $\theta \in \Theta$. Denote $\psi(\theta) = \mathbb{E}\{T(\mathbf{X})\}$. Suppose regularity conditions holds and the condition $0 < I_{\mathbf{X}}(\theta) < \infty$ is satisfied, then for all $\theta \in \Theta$,

$$\mathbb{E}\left\{[T(\mathbf{X}) - g(\theta)]^2\right\} \geq \frac{[\psi'(\theta)]^2}{I_{\mathbf{X}}(\theta)}$$

- Note
  - $\psi(\theta) = \mathbb{E}\{T(\mathbf{X})\} - g(\theta) + g(\theta) = b(\theta) + g(\theta)$
  - Information inequality gives lower bound on MSE of the estimate of $g(\theta)$

Information inequality

- Proof: We show the following relationships step by step

(A) $$\mathbb{E}\left\{T(\mathbf{X})V(\mathbf{X},\theta)\right\} = \psi'(\theta)$$

(B) $$\mathbb{E}\left\{g(\theta)V(\mathbf{X},\theta)\right\} = 0$$

(C) $$\mathbb{E}\left\{\left[T(\mathbf{X}) - g(\theta)\right]V(\mathbf{X},\theta)\right\} = \psi'(\theta)$$

(D) $$\mathbb{E}\left\{\left[T(\mathbf{X}) - g(\theta)\right]^2\right\}\mathbb{E}\left\{V^2(\mathbf{X},\theta)\right\} \geq \left[\psi'(\theta)\right]^2$$

$$\mathbb{E}\left\{\left[T(\mathbf{X}) - g(\theta)\right]^2\right\} \geq \frac{\left[\psi'(\theta)\right]^2}{I_{\mathbf{X}}(\theta)}$$

Information inequality

- Proof of (A)

$$\mathbb{E}\left\{T(X)V(X,\theta)\right\} = \int \cdots \int T(\mathbf{x})V(\mathbf{x},\theta)f(\mathbf{x}|\theta)d\mathbf{x}$$

Using the relationship

$$V(\mathbf{x},\theta) = \frac{\partial}{\partial\theta}\ln f(\mathbf{x}|\theta) = \frac{f'(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)}$$

we have

$$\begin{aligned}
\mathbb{E}\left\{T(\mathbf{X})V(\mathbf{X},\theta)\right\} &= \int \cdots \int T(\mathbf{x})\frac{\partial}{\partial\theta}f(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\int \cdots \int T(\mathbf{x})f(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\mathbb{E}\left\{T(\mathbf{X})\right\}
\end{aligned}$$

i.e., $\mathbb{E}\left\{T(\mathbf{X})V(\mathbf{X},\theta)\right\} = \psi'(\theta)$

Information inequality

- Proof of (B)

$$\mathbb{E}\left\{g(\theta)V(\mathbf{X},\theta)\right\} = g(\theta)\mathbb{E}\left\{V(\mathbf{X},\theta)\right\} = 0$$

where the second equality is due to

$$\begin{aligned}
\mathbb{E}\left\{V(\mathbf{X},\theta)\right\} &= \mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f(\mathbf{X}|\theta)\right\} \\
&= \int\cdots\int \frac{f'(\mathbf{x}|\theta)}{f(\mathbf{x}|\theta)} f(\mathbf{x}|\theta)d\mathbf{x} \\
&= \frac{\partial}{\partial\theta}\underbrace{\int\cdots\int f(\mathbf{x}|\theta)d\mathbf{x}}_{1} \\
&= 0
\end{aligned}$$

Information inequality

- Proof of (C)

  (A) – (B) gives $\mathbb{E}\{[T(\mathbf{X}) - g(\theta)]\, V(\mathbf{X}, \theta)\} = \psi'(\theta)$

Information inequality

- Proof of (D)
  We use the Cauchy-Schwarz inequality

$$\left[\int a(z)b(z)dz\right]^2 \leq \left[\int a(z)^2 dz\right]\left[\int b(z)^2 dz\right]$$

where equality holds if and only if $b(z) = c\,a(z)$. Following this inequality, we have

$$\mathbb{E}^2\{XY\} \leq \mathbb{E}\{X^2\}\mathbb{E}\{Y^2\}$$

where equality holds if and only if $Y = c\,X$. Therefore,

$$\mathbb{E}\left\{[T(\mathbf{X}) - g(\theta)]^2\right\}\mathbb{E}\left\{V^2(\mathbf{X},\theta)\right\} \geq \mathbb{E}^2\left\{[T(\mathbf{X}) - g(\theta)]\,V(\mathbf{X},\theta)\right\} = [\psi'(\theta)]^2$$

i.e.,

$$\mathbb{E}\left\{[T(\mathbf{X}) - g(\theta)]^2\right\} \geq \frac{[\psi'(\theta)]^2}{I_{\mathbf{X}}(\theta)}$$

Equality in information inequality

- Corollary: equality in the information equality holds if and only if

$$V(\mathbf{X}, \theta) = k(\theta) \left[T(\mathbf{X}) - g(\theta) \right], \quad \forall\, \theta \in \Theta$$

  - Note: $k(\theta)$ is a constant with respect to $\mathbf{x}$, but can be a function of θ

Information inequality for unbiased estimator

- Corollary: If $T(\mathbf{X})$ is unbiased estimator of $g(\theta)$, then

$$\mathbb{V}\{T(\mathbf{X})\} \geq \frac{[g'(\theta)]^2}{I_{\mathbf{X}}(\theta)}$$

- Proof: The result follows directly from information inequality

- Corollary: If $T(\mathbf{X})$ is unbiased estimator of θ, then

$$\mathbb{V}\{T(\mathbf{X})\} \geq \frac{1}{I_{\mathbf{X}}(\theta)}$$

- Proof: The result follows directly from information inequality

Independent samples

- Corollary: For n stationary independent samples, the lower bound goes to zero as $1/n$
- Proof: Stationary X_i are identically distributed. Recall that for $X_1, X_2, \ldots, X_n$ that are i.i.d.,

$$I_{\mathbf{X}}(\theta) = nI_X(\theta)$$

Therefore,

$$\mathbb{E}\left[T(\mathbf{X}) - g(\theta)\right]^2 \geq \frac{1}{n} \frac{\left[\psi'(\theta)\right]^2}{I_X(\theta)}$$

Other forms for Fisher information

- Fisher information can be expressed as

$$
\begin{aligned}
I_{\mathbf{X}}(\theta) &= \mathbb{E}\left\{V^2(\mathbf{X},\theta)\right\} \\
&= \mathbb{V}\left\{V(\mathbf{X},\theta)\right\} \\
&= -\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\ln f(\mathbf{X}|\theta)\right\}
\end{aligned}
$$

- Proof: First

$$
\begin{aligned}
\underbrace{\mathbb{E}\left\{V(X,\theta)\right\}}_{=0} &= \mathbb{E}\left\{\frac{\partial}{\partial\theta}\ln f(\mathbf{x}|\theta)\right\} \\
&= \int\cdots\int\frac{\partial}{\partial\theta}\ln f(\mathbf{x}|\theta) f(\mathbf{x}|\theta)d\mathbf{x}
\end{aligned}
$$

Other forms for Fisher information

- Take $\frac{\partial}{\partial\theta}$ for both sides, we have

$$0 = \int \cdots \int \frac{\partial^2}{\partial\theta^2} \ln f(\mathbf{x}|\theta) f(\mathbf{x}|\theta) d\mathbf{x} + \int \cdots \int \frac{\partial}{\partial\theta} \ln f(\mathbf{x}|\theta) \frac{\partial}{\partial\theta} f(\mathbf{x}|\theta) d\mathbf{x}$$

$$= \mathbb{E}\left\{ \frac{\partial^2}{\partial\theta^2} \ln f(\mathbf{x}|\theta) \right\} + \underbrace{\mathbb{E}\left\{ V^2(\mathbf{X}, \theta) \right\}}_{=I_{\mathbf{X}}(\theta)}$$

Therefore,

$$I_{\mathbf{X}}(\theta) = -\mathbb{E}\left\{ \frac{\partial^2}{\partial\theta^2} \ln f(\mathbf{x}|\theta) \right\}$$

Lecture 13 11/05/2012

EXAMPLES ON INFORMATION INEQUALITY

Example 1

- The iid Gaussian case: Let $X_1, X_2, \ldots, X_n$ be iid $N(0, \sigma^2)$
  Recall that MLE of σ^2 is given by $T_{\mathrm{ML}}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$\mathbb{E}\left\{T_{\mathrm{ML}}(\mathbf{X})\right\} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{X_i^2\} = \sigma^2$$

and

$$\begin{aligned}\mathbb{V}\left\{T_{\mathrm{ML}}(\mathbf{X})\right\} &= \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{V}\{X_i^2\} \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^4\mathbb{V}\left\{Y_i\right\} \\ &= \frac{2\sigma^4}{n}\end{aligned}$$

where $Y_i = X_i^2/\sigma^2$ has gamma distribution with one degree of freedom

Example 1

- Let's look at the information inequality. We saw that

$$I_X(\sigma^2) = \frac{1}{4\sigma^4}(\alpha^4 - 1) = \frac{1}{2\sigma^4}$$

where

$$\alpha^4 = \frac{1}{\sigma^4}\underbrace{\mathbb{E}\left\{(x-\mu)^4\right\}}_{=3\sigma^4} = 3$$

- For any unbiased estimator $T(\mathbf{X})$ for σ^2, we know from CRLB:

$$\begin{aligned}\mathbb{E}\left\{T(\mathbf{X}) - \sigma^2\right\} &\geq \frac{1}{I_{\mathbf{X}}(\sigma^2)} \\ &= \frac{1}{nI_X(\theta)} \\ &= \frac{2\sigma^4}{n}\end{aligned}$$

Therefore, MLE achieves the lower bound in this case

Example 1

- Conversely: Let $T(\mathbf{X})$ be any estimator that achieves information inequality
- Let's analyze the equality condition

$$k(\sigma^2)\left[T(\mathbf{x})-\sigma^2\right]=V(\mathbf{x},\theta)$$

$$V(\mathbf{x},\sigma^2)=\frac{\partial}{\partial\sigma^2}\ln f(\mathbf{x}|\sigma^2)=\frac{1}{2\sigma^4}\sum_{i=1}^{n}x_i-\frac{n}{2\sigma^2}$$

Therefore,

$$k(\sigma^2)\left[T(\mathbf{x})-\sigma^2\right]=\frac{n}{2\sigma^4}\left[\frac{1}{n}\sum_{i=1}^{n}x_i^2-\sigma^2\right]$$

Thus, $T(\mathbf{X})=\frac{1}{n}\sum_{i=1}^{n}X_i^2$ and $k(\sigma^2)=\frac{n}{2\sigma^4}$

- Therefore, $T(\mathbf{X})$ that satisfies the equality condition is MLE of σ^2

Example 2

- The iid exponential case: Let $X_1, X_2, \ldots, X_n$ be iid random variables with

$$f(x|\lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  We also have $\mathbb{E}\{X\} = \lambda$ and $\mathbb{V}\{X\} = \lambda^2$

- Recall that MLE of λ is given by $T_{\mathrm{ML}}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{n} X_i$
  $\mathbb{E}\left\{T_{\mathrm{ML}}(\mathbf{X})\right\} = \lambda$ and $\mathbb{V}\left\{T_{\mathrm{ML}}(\mathbf{X})\right\} = \frac{\lambda^2}{n}$

Example 2

- Let's look at the information inequality. The score function is

$$\begin{aligned} V(x,\theta) &= \frac{\partial}{\partial\theta}\ln f(x|\lambda) \\ &= \frac{\partial}{\partial\lambda}\left(-\ln\lambda - \frac{x}{\lambda}\right) \\ &= -\frac{1}{\lambda} + \frac{x}{\lambda^2} \\ &= \frac{1}{\lambda^2}(x-\lambda) \end{aligned}$$

and the Fisher information is

$$I_X(\theta) = \frac{1}{\lambda^4}\mathbb{E}\left\{(x-\lambda)^2\right\} = \frac{1}{\lambda^2}$$

Therefore, information inequality is $\mathbb{V}\left\{T(\mathbf{X})\right\} \geq \frac{\lambda^2}{n}$
We see that MLE achieves the lower bound in this example

Example 2

- Conversely, let $T(\mathbf{X})$ be any estimator that achieves information inequality. What can we say about $T(\mathbf{X})$?
- The equality condition is

$$k(\lambda)\left[T(\mathbf{x})-\lambda\right]=V(\mathbf{x},\theta)$$

  where $V(\mathbf{x},\theta)=\frac{1}{\lambda^2}\left[\sum_{i=1}^n x_i-n\lambda\right]$.
  Therefore,

$$k(\lambda)\left[T(\mathbf{x})-\lambda\right]=\frac{n}{\lambda^2}\left[\frac{1}{n}\sum_{i=1}^n x_i-\lambda\right]$$

  Thus $T(\mathbf{X})=\frac{1}{n}\sum_{i=1}^n X_i$ and $k(\lambda)=\frac{n}{\lambda^2}$
- Therefore, $T(\mathbf{X})$ that satisfies equality condition is MLE of λ

MLE and information inequality

- Corollary: If an estimator $T(\mathbf{X})$ for θ satisfies the condition for equality in the information inequality, then $T(\mathbf{X})$ is a MLE of θ i.e., $\widehat{\theta}_{\mathrm{ML}} = T(\mathbf{X})$
- Proof: For the maximum likelihood estimate $\widehat{\theta}_{\mathrm{ML}}$, we have

$$V(\mathbf{x}, \theta)|_{\theta=\widehat{\theta}_{\mathrm{ML}}} = \frac{\partial}{\partial \theta} \ln f(\mathbf{x}|\theta)\bigg|_{\theta=\widehat{\theta}_{\mathrm{ML}}} = 0$$

  Since $T(\mathbf{X})$ achieves equality,

$$k(\theta)\left[T(\mathbf{X}) - \theta\right]|_{\theta=\widehat{\theta}_{\mathrm{ML}}} = V(\mathbf{X}, \theta)|_{\theta=\widehat{\theta}_{\mathrm{ML}}} = 0$$

  Therefore, $\widehat{\theta}_{\mathrm{ML}} = T(\mathbf{X})$

Example 4

- The iid uniform case: Let $X_1, X_2, \ldots, X_n$ be iid random variables with pdf

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

Recall that $Y_n = \max\{X_1, \ldots, X_n\}$ is a MLE of θ, and

$$\mathbb{E}\{Y_n\} = \frac{n}{n+1}\theta$$

i.e., Y_n is biased. An unbiased estimator for θ is given by

$$T_1(\mathbf{X}) = \frac{n+1}{n}\max\{X_1, \ldots, X_n\}$$

Another unbiased estimator for θ is

$$T_2(\mathbf{X}) = \frac{2}{n}\sum_{i=1}^{n} X_i$$

Example 4

- One can verify that

$$\mathbb{V}\left\{T_1(\mathbf{X})\right\} = \frac{1}{n(n+2)}\theta^2$$

$$\begin{aligned}\mathbb{V}\left\{T_2(\mathbf{X})\right\} &= \frac{4}{n}\mathbb{V}\{X_i\} \\ &= \frac{4}{n}\frac{\theta^2}{12} \\ &= \frac{\theta^2}{3n}\end{aligned}$$

Example 4

- The score function is

$$V(x,\theta) = \frac{\partial}{\partial\theta} \ln f(x|\theta)$$
$$= \frac{\partial}{\partial\theta}(-\ln\theta)$$
$$= \frac{1}{\theta}$$

and Fisher information is

$$I_X(\theta) = \mathbb{V}\left\{\left(\frac{1}{\theta}\right)^2\right\} = \frac{1}{\theta^2}$$

We find that $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$ both beats "CRLB". The reason is regularity conditions are not satisfied, i.e., the support of $f(x|\theta)$ depends on θ