$\boxed{\textit{Example 4.2.6.}}$. $X_1, \ldots, X_n$ is a random sample from

$$f(x;\theta,\tau) = \begin{cases} \frac{1}{\theta}e^{-\frac{x-\tau}{\theta}}, & x \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

(θ, τ) is unknown, $\theta > 0, \tau \geq 0$. Find the ML estimates and ML estimators.

$$L(\theta,\tau)^{1} = \begin{cases} \prod_{i=1}^{n} \frac{1}{\theta}e^{-\frac{1}{\theta}(x_i-\tau)}, & x_i \geq \tau \ \forall i \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\tau)}, & \min\limits_{1\leq i\leq n} x_i \geq \tau \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{\theta^n}e^{-\frac{n}{\theta}(\overline{x}_n-\tau)}, & \min\limits_{1\leq i\leq n} x_i \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

where $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$. For fixed θ, $L(\theta,\tau)$ is maximal at $L(\theta, \min\limits_{1\leq i\leq n} x_i)$. Now maximize $L(\theta, \min\limits_{1\leq i\leq n} x_i)$ with θ.

$$\ln L(\theta, \min_{1\leq i\leq n} x_i) = -n\ln\theta - \frac{n}{\theta}(\overline{x}_n - \min_{1\leq i\leq n} x_i)$$

Take derivative and set it to 0

$$\frac{d\ln L(\theta, \min\limits_{1\leq i\leq n} x_i)}{d\theta} = -\frac{n}{\theta} + \frac{n}{\theta^2}(\overline{x}_n - \min_{1\leq i\leq n} x_i) = 0 \implies \theta = \overline{x}_n - \min_{1\leq i\leq n} x_i$$

Check second derivative

$$\left.\frac{d^2\ln L(\theta, \min\limits_{1\leq i\leq n} x_i)}{d\theta^2}\right|_{\theta=\overline{x}_n-\min\limits_{1\leq i\leq n} x_i} = \left.\frac{n}{\theta^2} - \frac{2n}{\theta^3}(\overline{x}_n - \min_{1\leq i\leq n} x_i)\right|_{\theta=\overline{x}_n-\min\limits_{1\leq i\leq n} x_i}$$

$$= \frac{-n}{\left(\overline{x}_n - \min\limits_{1\leq i\leq n} x_i\right)^2} < 0$$

[1] $L(\theta,\tau)$ is the abbreviation of $L(\theta,\tau;x_1,\ldots,x_n)$

Therefore the ML estimates are

$$\hat{\tau} = \min_{1\le i\le n} x_i \qquad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i - \min_{1\le i\le n} x_i$$

and the ML estimators are

$$\hat{\tau} = \min_{1\le i\le n} X_i \qquad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i - \min_{1\le i\le n} X_i$$

$\boxed{\textit{Example 4.2.7.}}$ **(Function of parameters)**. $X_1, \ldots, X_n$ is a random sample from $\mathcal{E}xp(\theta)$, θ is unknown, want to estimate $\tau(\theta) = e^{-t/\theta}$ $(= \mathbb{P}\{X_1 > t\})$, for some fixed $t \ge 0$.

Step 1: $f(x;\theta) = \begin{cases} \frac{1}{\theta}e^{-x/\theta}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$. Reparametrize in terms of $\tau = \tau(\theta) = e^{-t/\theta}$, i.e. plug in $\theta = -t/\ln\tau$.

$$f(x;\tau) = \begin{cases} \frac{\ln\tau}{-t}e^{\frac{x}{t}\ln\tau}, & x > 0 \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \frac{\ln\tau}{-t}\tau^{x/t}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Step 2: Now apply MLE method on $f(x;\tau)$

$$L(\tau) = \prod_{i=1}^{n} \frac{\ln\tau}{-t}\tau^{x_i/t} = \left(\frac{\ln\tau}{-t}\right)^n \tau^{\frac{1}{t}\sum_{i=1}^{n} x_i}, \quad x_i > 0\ \forall i$$

$$\ln L(\tau) = n\ln\left(\frac{\ln\tau}{-t}\right) + \left(\frac{1}{t}\sum_{i=1}^{n} x_i\right)\ln\tau, \quad x_i > 0\ \forall i$$

Take derivative and set it to 0

$$\frac{d\ln L(\tau)}{d\tau} = n\frac{-t}{\ln\tau}\frac{-1}{t\tau} + \left(\frac{1}{t}\sum_{i=1}^{n} x_i\right)\frac{1}{\tau} = \frac{n}{\tau\ln\tau} + \frac{n\overline{x}_n}{t\tau}[1] = 0, \quad x_i > 0\ \forall i$$

$$\implies \tau = e^{-t/\overline{x}_n}, \quad x_i > 0\ \forall i$$

[1] $\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

Check second derivative

$$\left.\frac{d^2 \ln L(\tau)}{d\tau^2}\right|_{\tau=e^{-t/\overline{x}_n}} = -\frac{n}{\tau^2 \ln \tau} - \frac{n}{\tau^2 \ln^2 \tau} - \left.\frac{n\overline{x}_n}{t\tau^2}\right|_{\tau=e^{-t/\overline{x}_n}}$$

$$= -\frac{n}{\tau^2}\left(\frac{1}{\ln \tau} + \frac{1}{\ln^2 \tau} + \frac{\overline{x}_n}{t}\right)\Bigg|_{\tau=e^{-t/\overline{x}_n}}$$

$$= -\frac{n}{\left(e^{-t/\overline{x}_n}\right)^2}\left(-\frac{\overline{x}_n}{t} + \frac{\overline{x}_n^2}{t^2} + \frac{\overline{x}_n}{t}\right) < 0, \quad x_i > 0 \;\forall i$$

So $\hat{\tau} = \widehat{\tau(\theta)} = e^{-t/\overline{X}_n}$ is the ML estimator of $\tau(\theta)$.

Recall last time, we have showed that $\hat{\theta} = \overline{X}_n$ is the ML estimator of θ in Example 4.2.3. So we can also express the ML estimator of $\tau(\theta)$ as $\hat{\tau} = e^{-t/\hat{\theta}}$. So $\widehat{\tau(\theta)} = \tau(\hat{\theta})$ for MLE method in this case. However, this is true in general, which is summarized in the following theorem.

Theorem 4.2.1. **(Invariance property of MLE)**. If $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ denotes the ML estimator of $\theta = (\theta_1, \ldots, \theta_k)$, then the ML estimator of $\tau = \tau(\theta_1, \ldots, \theta_k)$ is $\hat{\tau} = \tau(\hat{\theta}_1, \ldots, \hat{\theta}_k)$

Sketch Proof. Prove case $k = 1$, $\tau = \tau(\theta_1)$ is 1-1, differentiable everywhere, $0 < \frac{d\tau}{d\theta_1} < \infty$, $\forall \theta_1 \in \text{H}$.

$$L(\theta_1) = L(\tau^{-1}(\tau)) = L^*(\tau)$$

$$\frac{d \ln L^*(\tau)}{d\tau} = \frac{d \ln L(\theta_1)}{d\theta_1}\frac{d\theta_1}{d\tau} = 0$$

if and only if $\frac{d \ln L(\theta_1)}{d\theta_1} = 0$ since $\frac{d\theta_1}{d\tau} > 0$. So $\frac{d \ln L^*(\tau)}{d\tau} = 0$ for value of τ at which $\frac{d \ln L(\theta_1)}{d\theta_1} = 0$. I.e. $\hat{\tau} = \tau(\hat{\theta}_1)$. □

Example 4.2.8.. Continue on the previous example 4.2.6, where $X_1, \ldots, X_n$ is a random sample from

$$f(x;\theta,\tau) = \begin{cases} \frac{1}{\theta}e^{-\frac{x-\tau}{\theta}}, & x \geq \tau \\ 0, & \text{otherwise} \end{cases}$$

(θ, τ) is unknown, $\theta > 0, \tau \geq 0$. Find the MLE of $u(\theta, \tau) = \mathbb{P}\{X_1 > t\}$ for fixed t. Write this as a function of θ, τ

$$u(\theta, \tau) = \begin{cases} \int_t^\infty \frac{1}{\theta} e^{-\frac{x-\tau}{\theta}}\, dx = e^{-\frac{t-\tau}{\theta}}, & t \geq \tau \\ \int_\tau^\infty \frac{1}{\theta} e^{-\frac{x-\tau}{\theta}}\, dx = 1 & t < \tau \end{cases}$$

Know from the previous example 4.2.6 that the ML estimators of τ and θ are

$$\hat{\tau} = \min_{1 \leq i \leq n} X_i \qquad \hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i - \min_{1 \leq i \leq n} X_i$$

So, by invariance property, the estimator (MLE) of $u(\theta, \tau)$ is

$$\widehat{u(\theta, \tau)} = u(\hat{\theta}, \hat{\tau}) = \begin{cases} e^{-\frac{t-\hat{\tau}}{\hat{\theta}}}, & t \geq \hat{\tau} \\ 1 & t < \hat{\tau} \end{cases} = \begin{cases} e^{-\frac{t - \min_{1 \leq i \leq n} X_i}{\frac{1}{n}\sum_{i=1}^n X_i - \min_{1 \leq i \leq n} X_i}}, & t \geq \min_{1 \leq i \leq n} X_i \\ 1 & t < \min_{1 \leq i \leq n} X_i \end{cases}$$

4.3 Criteria for Evaluating Estimators

4.3.1 Bias and variance

Definition 4.3.1. **(Bias)**. If T is an estimator of $\tau(\theta)$, then bias is

$$b(T) = \mathbb{E}T - \tau(\theta)$$

Definition 4.3.2. **(Unbiased estimator)**. An estimator T is said to be an unbiased estimator of $\tau(\theta)$ if $\mathbb{E}T = \tau(\theta)$ for all $\theta \in \Theta$. Otherwise, T is said to be biased.

Example 4.3.1.. $X_1, \ldots, X_n$ is a random sample from $\mathcal{E}xp(\theta)$. Recall we have find the MLE of θ is $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n X_i$ in Example 4.2.3.

$$\mathbb{E}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i = \theta$$

so $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an unbiased estimator for θ.

By invariance property, the MLE of $\tau(\theta) = 1/\theta$ is $\widehat{\tau(\theta)} = \tau(\hat{\theta}) = 1/\hat{\theta} = 1/\overline{X}_n$. What is $\mathbb{E}(\widehat{\tau(\theta)})$? By Lemma 2.3.1 and Theorem 2.2.1, $\frac{2n\overline{X}_n}{\theta} \sim \chi^2(2n)$ (student verify it). So

$$\mathbb{E}\left(\widehat{\tau(\theta)}\right) = \mathbb{E}\left(1/\overline{X}_n\right) = \mathbb{E}\left(1/\left(\frac{\theta Y}{2n}\right)\right) = \frac{2n}{\theta}\mathbb{E}Y^{-1}$$

where $Y \sim \chi^2(2n)$. Recall the moment of chi-square (you proved it in your HW1, 1))[1],

$$\mathbb{E}Y^{-1} = \frac{\Gamma(n-1)}{2\Gamma(n)} = \frac{1}{2(n-1)}$$

So

$$\mathbb{E}\left(\widehat{\tau(\theta)}\right) = \frac{n}{n-1}\frac{1}{\theta}$$

which is biased for $\tau(\theta) = \frac{1}{\theta}$.

But we can convert for bias here. Let $T = \frac{n-1}{n}\frac{1}{\overline{X}_n}$,

$$\mathbb{E}T = \frac{n-1}{n}\frac{n}{n-1}\frac{1}{\theta} = \frac{1}{\theta}$$

So $T = \frac{n-1}{n}\frac{1}{\overline{X}_n}$ is unbiased for $\frac{1}{\theta}$.

However, it is often possible to derive several unbiased estimators of a parameter, so need additional evaluation criteria: variance.

Example 4.3.2.. $X_1, \ldots, X_n$ is a random sample from $\mathcal{U}nif(0, \theta)$. We showed in Example 4.2.5 that MM estimator of θ is $T_1 = 2\overline{X}_n$ of θ and T_1 is unbiased since $\mathbb{E}T_1 = 2\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i = \theta$. We also showed that the MLE is $T_2 = \max_{1\le i\le n} X_i$. Is T_2 unbiased?

$$\mathbb{P}\{T_2 \le t\} = \mathbb{P}\{\max_{1\le i\le n} X_i \le t\} = \mathbb{P}\{X_1 \le t, \ldots, X_n \le t\}$$

[1] $Y \sim \chi^2(v)$, $\mathbb{E}Y^r = \frac{2^r\Gamma(v/2+r)}{\Gamma(v/2)}$, $v/2 + r > 0$

$$\stackrel{i.i.d.}{=} (\mathbb{P}\{X_1 \le t\})^n = \begin{cases} (t/\theta)^n & 0 \le t \le \theta \\ 0 & t < 0 \\ 1 & t > \theta \end{cases}$$

So the PDF of T_2 is

$$f_{T_2}(t) = \begin{cases} \frac{n}{\theta}(t/\theta)^{n-1} = \frac{nt^{n-1}}{\theta^n} & 0 \le t \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}T_2 = \int_0^\theta t\frac{nt^{n-1}}{\theta^n}\,dt = \frac{n}{\theta^n}\frac{t^{n+1}}{n+1}\Bigg|_{t=0}^{\theta} = \frac{n\theta}{n+1}$$

So T_2 is biased. Consider $T_3 = \frac{n+1}{n}T_2 = \frac{n+1}{n}\max\limits_{1\le i\le n} X_i$, $\mathbb{E}T_3 = \frac{n+1}{n}\frac{n}{n+1}\theta$, so T_3 is unbiased. We have both of T_1 and T_3 unbiased, but which is better? Let's compute their variances.

$$\mathbb{V}arT_1 = \mathbb{V}ar\left(\frac{2}{n}\sum_{i=1}^n X_i\right) \stackrel{i.i.d.}{=} \frac{4}{n^2}\sum_{i=1}^n \mathbb{V}arX_i = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}$$

$$\mathbb{E}T_3^2 = \mathbb{E}\left(\frac{(n+1)^2}{n^2}T_2^2\right) = \frac{(n+1)^2}{n^2}\mathbb{E}T_2^2 = \frac{(n+1)^2}{n^2}\int_0^\theta t^2\frac{nt^{n-1}}{\theta^n}\,dt$$

$$= \frac{(n+1)^2}{n^2}\frac{n}{\theta^n}\frac{t^{n+2}}{n+2}\Bigg|_{t=0}^{\theta} = \frac{(n+1)^2}{n^2}\frac{n\theta^2}{n+2}$$

$$\mathbb{V}arT_3 = \frac{(n+1)^2}{n}\frac{\theta^2}{n+2} - \theta^2 = \frac{\theta^2}{n(n+2)}$$

For $n > 1$, $\mathbb{V}arT_3 < \mathbb{V}arT_1$ and for large n, $\mathbb{V}arT_3$ decreases like $\frac{1}{n^2}$ and $\mathbb{V}arT_1$ decreases like $\frac{1}{n}$.

Variance is related to how concentrated the distribution of the estimator is about the $\tau(\theta)$. It is because that we want to select the estimator that tends to be closest or most concentrated about the true value. So estimator T_1 is more concentrated than T_2 about $\tau(\theta)$ if

$$\mathbb{P}\{\tau(\theta) - \epsilon < T_1 < \tau(\theta) + \epsilon\} \ge \mathbb{P}\{\tau(\theta) - \epsilon < T_2 < \tau(\theta) + \epsilon\}$$

for all $\epsilon > 0$. Now assume T is an unbiased estimator of $\tau(\theta)$, it follows from the Chebychev inequality[1] that

$$\mathbb{P}\{\tau(\theta) - \epsilon < T < \tau(\theta) + \epsilon\} \geq 1 - \frac{\mathbb{V}arT}{\epsilon^2}$$

This suggests that for unbiased estimator, one with a smaller variance will tend to be more concentrated and thus be preferable.

Definition 4.3.3. **(Uniformly minimum variance unbiased estimator)**. An estimator T^* of $\tau(\theta)$ is called a uniformly minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$ if

1. $\mathbb{E}T^* = \tau(\theta),\ \forall\theta \in Ⓗ$
2. $\mathbb{V}arT^* \leq \mathbb{V}arT$ for any other unbiased estimator T of $\tau(\theta)$

Definition 4.3.4. **(Efficiency)**. Let T, T^* be unbiased estimator of $\tau(\theta)$. We say T^* is more efficient than T if $\mathbb{V}arT^* \leq \mathbb{V}arT$, $\forall\theta \in Ⓗ$ with strict $<$ for at least one $\theta \in Ⓗ$. We simply say T^* is efficient, without refer to any other estimators if T^* is a UMVUE

4.3.2 Cramer-Rao lower bound

It is useful to have lower bound on variance of unbiased estimators. If an unbiased estimator attains such a lower bound, then the estimator is UMVUE.

Theorem 4.3.1. **(Cramer-Rao lower bound)**. Let $X_1, \dots, X_n$ have PDF (or PMF) $f(x_1, \dots, x_n; \theta), \theta \in Ⓗ$ = open interval of $\mathbb{R}$. Suppose $\{(x_1, \dots, x_n) : f(x_1, \dots, x_n; \theta) = 0\}$ does not depend on θ. Let $T = t(X_1, \dots, X_n)$ be an unbiased estimator $\tau(\theta)$ such that $\mathbb{V}arT < \infty \quad \forall\theta \in Ⓗ$ where τ is differentiable function of θ, $\theta \in Ⓗ$.

Assume interchange the order of differentiation and integral:

1. $$\frac{\partial}{\partial\theta}\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1, \dots, x_n) f(x_1, \dots, x_n; \theta)\, dx_1 \dots dx_n$$

[1] $\mathbb{P}\{|X - \mathbb{E}X| \geq \epsilon\} \leq \frac{\mathbb{V}arX}{\epsilon^2}$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} f(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n$$

or

$$\frac{\partial}{\partial \theta} \sum_{x_1} \cdots \sum_{x_n} t(x_1, \ldots, x_n) f(x_1, \ldots, x_n; \theta)$$

$$= \sum_{x_1} \cdots \sum_{x_n} t(x_1, \ldots, x_n) \frac{\partial}{\partial \theta} f(x_1, \ldots, x_n; \theta)$$

2.

$$0 = \frac{\partial}{\partial \theta} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n$$

$$= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n$$

or

$$0 = \frac{\partial}{\partial \theta} \sum_{x_1} \cdots \sum_{x_n} f(x_1, \ldots, x_n; \theta) = \sum_{x_1} \cdots \sum_{x_n} \frac{\partial}{\partial \theta} f(x_1, \ldots, x_n; \theta)$$

Then

$$\mathbb{V}ar T \geq \underbrace{\frac{(\tau'(\theta))^2}{\mathbb{E}\left(\frac{\partial}{\partial \theta} \ln f(X_1, \ldots, X_n; \theta)\right)^2}}_{\text{this is the CRLB}}$$

If $X_1, \ldots, X_n$ is a random sample from $f(x; \theta)$, then

$$\mathbb{V}ar T \geq \underbrace{\frac{(\tau'(\theta))^2}{n\mathbb{E}\left(\frac{\partial}{\partial \theta} \ln f(X_1; \theta)\right)^2}}_{\text{this is the CRLB}}$$

Proof. Consider the random variable $U = \frac{\partial}{\partial \theta} \ln f(X_1, \ldots, X_n; \theta)$. Because $\mathbb{C}ov(U, T) = \mathbb{E}(UT) - \mathbb{E}(U)\mathbb{E}(T)$. Now by assumption 2

$$\mathbb{E}U = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \frac{\partial \ln f(x_1, \ldots, x_n; \theta)}{\partial \theta} f(x_1, \ldots, x_n; \theta)\, dx_1 \ldots dx_n$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \frac{1}{f(x_1,\ldots,x_n;\theta)} \frac{\partial f(x_1,\ldots,x_n;\theta)}{\partial\theta} f(x_1,\ldots,x_n;\theta)\, dx_1 \ldots dx_n \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \frac{\partial f(x_1,\ldots,x_n;\theta)}{\partial\theta}\, dx_1 \ldots dx_n \\
&= \frac{\partial}{\partial\theta} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1,\ldots,x_n;\theta)\, dx_1 \ldots dx_n = 0
\end{aligned}
$$

So $\mathbb{C}ov(U,T) = \mathbb{E}(UT)$. Then we have $(\mathbb{E}(UT))^2 = (\mathbb{C}ov(U,T))^2 \leq$ [1] $\mathbb{V}ar U \mathbb{V}ar T$.Now because

$$
\mathbb{V}ar U = \mathbb{E}U^2 = \mathbb{E}\left(\frac{\partial}{\partial\theta} \ln f(X_1,\ldots,X_n;\theta)\right)^2
$$

$$
\begin{aligned}
\mathbb{E}(UT) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1,\ldots,x_n;\theta) \frac{\partial \ln f(x_1,\ldots,x_n;\theta)}{\partial\theta} \\
&\times f(x_1,\ldots,x_n;\theta)\, dx_1 \ldots dx_n \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1,\ldots,x_n;\theta) \frac{\partial f(x_1,\ldots,x_n;\theta)}{\partial\theta}\, dx_1 \ldots dx_n \\
&= \frac{\partial}{\partial\theta} \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} t(x_1,\ldots,x_n;\theta) f(x_1,\ldots,x_n;\theta)\, dx_1 \ldots dx_n}_{=\mathbb{E}(T)=\tau(\theta),\text{ since } T \text{ is unbiased estimator of } \tau(\theta)} = \frac{\partial}{\partial\theta}\tau(\theta) = \tau'(\theta)
\end{aligned}
$$

Plugging back into $(\mathbb{E}(UT))^2 \leq \mathbb{V}ar U \mathbb{V}ar T$ yields

$$
\mathbb{V}ar T \geq \frac{(\mathbb{E}(UT))^2}{\mathbb{V}ar U} = \frac{(\tau'(\theta))^2}{\mathbb{E}\left(\frac{\partial}{\partial\theta} \ln f(X_1,\ldots,X_n;\theta)\right)^2}
$$

Now if $X_1,\ldots,X_n$ is a random sample from $f(x;\theta)$,

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

So $U = \frac{\partial}{\partial\theta} \ln f(X_1,\ldots,X_n;\theta) = \frac{\partial}{\partial\theta} \sum_{i=1}^{n} \ln f(X_i;\theta) = \sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta}$

$$
\mathbb{E}\left(\frac{\partial \ln f(X_i;\theta)}{\partial\theta}\right) = \int_{\mathbb{R}} \frac{\partial \ln f(x_i;\theta)}{\partial\theta} f(x_i;\theta)\, dx_i
$$

[1] Recall the correlation inequality. $-1 \leq \rho = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar X \mathbb{V}ar Y}} \leq 1$

$$= \int_{\mathbb{R}} \frac{1}{f(x_i;\theta)} \frac{\partial f(x_i;\theta)}{\partial \theta} f(x_i;\theta)\, dx_i = \int_{\mathbb{R}} \frac{\partial f(x_i;\theta)}{\partial \theta}\, dx_i$$

$$= \frac{\partial}{\partial \theta} \int_{\mathbb{R}} f(x_i;\theta)\, dx_i = \frac{\partial}{\partial \theta} 1 = 0$$

So

$$\mathbb{V}ar U = \mathbb{V}ar \left(\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial \theta} \right) \overset{i.i.d.}{=} n \mathbb{V}ar \left(\frac{\partial \ln f(X_1;\theta)}{\partial \theta} \right)$$

$$= n\mathbb{E} \left(\frac{\partial \ln f(X_1;\theta)}{\partial \theta} \right)^2 - 0$$

Therefore

$$\mathbb{V}ar T \geq \frac{(\mathbb{E}(UT))^2}{\mathbb{V}ar U} = \frac{(\tau'(\theta))^2}{n\mathbb{E} \left(\frac{\partial \ln f(X_1;\theta)}{\partial \theta} \right)^2}$$

□

Remark 4.3.1.. If T is unbiased estimator of $\tau(\theta)$ and $\mathbb{V}ar T =$CRLB, then T is UMVUE.

Example 4.3.3.. $X_1, \dots, X_n$ is a random sample from $\mathbb{E}xp(\theta)$. Find CRLB for $\tau(\theta) = \theta$

$$f(x;\theta) = \frac{1}{\theta} e^{-x/\theta}, \qquad \ln f(x;\theta) = \ln \frac{1}{\theta} - \frac{x}{\theta} = -\ln \theta - \frac{x}{\theta}$$

$$\frac{\partial \ln f(x;\theta)}{\partial \theta} = -\frac{1}{\theta} + \frac{x}{\theta^2}$$

$$\mathbb{E} \left(\frac{\partial \ln f(X_1;\theta)}{\partial \theta} \right)^2 = \mathbb{E} \left(-\frac{1}{\theta} + \frac{X_1}{\theta^2} \right)^2 = \mathbb{E} \frac{(X_1 - \theta)^2}{\theta^4} = \frac{\mathbb{V}ar X_1}{\theta^4} = \frac{1}{\theta^2}$$

So CRLB $= \frac{1}{n/\theta^2} = \theta^2/n$. Note that we showed that (in Example 4.2.3) the MLE is $\overline{X}_n$, and $\mathbb{V}ar\overline{X}_n = \frac{\theta^2}{n}$, $\mathbb{E}\overline{X}_n = \theta$, so $\overline{X}_n$ is UMVUE.