$\boxed{\text{Lemma 4.3.1.}}$ **(Second derivative method of CRLB).** Let $X_1, \ldots, X_n$ be a random sample from $f(x;\theta)$. Under regular conditons of CRLB. Assume that

$$0 = \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}} f(x;\theta)\, dx = \int_{\mathbb{R}} \frac{\partial^2}{\partial \theta^2} f(x;\theta)\, dx$$

or

$$0 = \frac{\partial^2}{\partial \theta^2} \sum_x f(x;\theta) = \sum_x \frac{\partial^2}{\partial \theta^2} f(x;\theta)$$

Then

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

Proof.

$$\frac{\partial}{\partial \theta} \ln f(X_1;\theta) = \frac{1}{f(X_1;\theta)} \frac{\partial}{\partial \theta} f(X_1;\theta)$$

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

So

$$\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \ln f(X_1;\theta)\right) = \underbrace{\int_{\mathbb{R}} \frac{\partial^2}{\partial \theta^2} f(x;\theta)\, dx}_{\substack{\text{by assumption, can} \\ \text{interchange the order} \\ \text{of the second derivative} \\ \text{and the integral}}} - \int_{\mathbb{R}} \left(\underbrace{\frac{\frac{\partial}{\partial \theta} f(x;\theta)}{f(x;\theta)}}_{= \frac{\partial}{\partial \theta} \ln f(x;\theta)}\right)^2 f(x;\theta)\, dx$$

$$= \frac{\partial^2}{\partial \theta^2} \underbrace{\int_{\mathbb{R}} f(x;\theta)\, dx}_{\substack{\text{integral of PDF} \\ \text{over entire plane is 1}}} - \int_{\mathbb{R}} \left(\frac{\partial}{\partial \theta} \ln f(x;\theta)\right)^2 f(x;\theta)\, dx$$

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

$\square$

Corollary 4.3.1. let $X_1,\ldots,X_n$ be a random sample from $f(x;\theta)$, assume the conditions holding for CRLB, and let $T = t(X_1,\ldots,X_n)$ be unbiased for $\tau(\theta)$. Then $\mathbb{V}ar(T)$ =CRLB if and only if $T = a\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta} + b$.

Proof. Recall the proof of CRLB Theorem, $U = \sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta}$ and $(\mathbb{C}ov(U,T))^2 \le \mathbb{V}arU\mathbb{V}arT = n\mathbb{V}arT\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2$, where according to the correlation inequality , the equality holds if and only if $T = aU + b$ for some $a \ne 0$. □

Remark 4.3.2. *a and b can depend on θ, but only in a way that T does not.*

Example 4.3.4. let $X_1,\ldots,X_n$ be a random sample from $\mathcal{P}oi(\theta)$,

$$f(x;\theta) = \mathbb{P}\{X_1 = x\} = \frac{\theta^x e^{-\theta}}{x!}, \quad x = 0,1,2,\ldots$$

Find the UMVUE of $\tau(\theta) = \theta$.

$$\ln f(X_1;\theta) = \ln \frac{\theta^{X_1} e^{-\theta}}{X_1!} = X_1 \ln\theta - \theta - \ln(X_1!)$$

$$\frac{\partial \ln f(X_1;\theta)}{\partial\theta} = \frac{X_1 - \theta}{\theta}$$

So

$$\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2 = \mathbb{E}\frac{(X_1-\theta)^2}{\theta^2} = \frac{\mathbb{V}arX_1}{\theta^2} = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Or apply the second derivative method of CRLB (i.e. Lemma 4.3.1)

$$\frac{\partial^2 \ln f(X_1;\theta)}{\partial\theta^2} = \frac{-X_1}{\theta^2}$$

So

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

Now by CRLB Theorem, CRLB= $\frac{1}{n/\theta} = \frac{\theta}{n}$. One can verify that the MLE of θ is $T = \overline{X}_n$ and it is unbiased. To check the variance

$$\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{X_i - \theta}{\theta} = \frac{1}{\theta}\sum_{i=1}^{n} X_i - n = \frac{n\overline{X}_n}{\theta} - n$$

So

$$\overline{X}_n = \theta + \frac{\theta}{n}\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta}$$

By Corollary 4.3.1, $\mathbb{V}ar\overline{X}$ =CRLB= $\frac{\theta}{n}$. We can check this directly by

$$\mathbb{V}ar\overline{X} \overset{i.i.d.}{=} \frac{1}{n^2} n\mathbb{V}ar X_1 = \frac{\theta}{n}$$

Corollary 4.3.2. If T is unbiased for $\tau(\theta)$ and $\mathbb{V}ar T$ =CRLB, then the only other parameters that admit unbiased estimators that attain CRLB are linear functions of $\tau(\theta)$.

Proof. Suppose T^* is unbiased for $u(\theta)$ and $\mathbb{V}ar T^*$ =CRLB. Then by Corollary 4.3.1

$$T^* = a\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta} + b$$

Also, since T is unbiased for $\tau(\theta)$ and $\mathbb{V}ar T$ =CRLB

$$T = c\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta} + d$$

So

$$\sum_{i=1}^{n} \frac{\partial \ln f(X_i;\theta)}{\partial\theta} = \frac{T-d}{c}, \quad T^* = a\frac{T-d}{c} + b = \frac{a}{c}T + b - \frac{ad}{c}$$

but

$$\mathbb{E}T^* = u(\theta) = \frac{a}{c}\tau(\theta) + b - \frac{ad}{c}$$

□

4.3.3 Mean squared error

Definition 4.3.5. **(Mean squared error)**. Let T be an estimator of $\tau(\theta)$ (not necessary unbiased), the mean squared error (MSE) of T is

$$MSE(T) = \mathbb{E}\left(T - \tau(\theta)\right)^2$$

Theorem 4.3.2. Let T be an estimator of $\tau(\theta)$

$$MSE(T) = \mathbb{V}ar T + b^2(T)$$

Proof. Recall bias $b(T) = \mathbb{E}T - \tau(\theta)$

$$MSE(T) = \mathbb{E}\left(T - \tau(\theta)\right)^2 = \mathbb{E}\left(T - \mathbb{E}T + \mathbb{E}T - \tau(\theta)\right)^2$$

$$= \mathbb{E}\left(\left(T - \mathbb{E}T\right)^2 + \left(\mathbb{E}T - \tau(\theta)\right)^2 + 2\left(T - \mathbb{E}T\right)\left(\mathbb{E}T - \tau(\theta)\right)\right)$$

$$= \mathbb{E}\left(T - \mathbb{E}T\right)^2 + \left(\mathbb{E}T - \tau(\theta)\right)^2 + 2\left(\mathbb{E}T - \tau(\theta)\right)\underbrace{\mathbb{E}\left(T - \mathbb{E}T\right)}_{=\mathbb{E}T - \mathbb{E}T = 0}$$

$$= \mathbb{V}ar T + b^2(T)$$

□

MSE takes into account both the variance and the bias, and it agrees with the variance criterion if is restricted to unbiased estimators. It is possible to have $MSE(T) < MSE(T^*)$ if T^* is UMVUE. Here is an example:

Example 4.3.5. $X_1, \ldots, X_n$ is a random sample from $\mathcal{E}xp(\theta)$. We have showed in Examples 4.3.3 and 4.2.3 that the MLE of $\tau(\theta) = \theta$ is $\overline{X}_n$, which is a UMVUE. So

$$MSE(\overline{X}_n) = \mathbb{V}ar\overline{X}_n = \frac{\theta^2}{n}$$

Now consider a biased estimator $T = c\overline{X}_n$, where c is to be determined.

$$MSE(T) = \mathbb{E}\left(c\overline{X}_n - \theta\right)^2 = \mathbb{E}\left(c^2\overline{X}_n^2 - 2c\theta\overline{X}_n + \theta^2\right)$$

$$= c^2\mathbb{E}\left(\overline{X}_n^2\right) - 2c\theta\mathbb{E}\overline{X}_n + \theta^2 = c^2\left(\frac{\theta^2}{n} + \theta^2\right) - 2c\theta^2 + \theta^2$$
$$= \theta^2\left(c^2\left(\frac{n+1}{n}\right) - 2c + 1\right)$$

We want to find c such that $MSE(T) < MSE(\overline{X}_n)$. Take derivative and set it to 0

$$\frac{dMSE(T)}{dc} = \theta^2\left(2c\left(\frac{n+1}{n}\right) - 2\right) = 0 \implies c = \frac{n}{n+1}$$

$c = \frac{n}{n+1}$ minimizes MSE since $\frac{d^2MSE(T)}{dc^2} = 2\theta^2\frac{n+1}{n} > 0$.

$$MSE(\frac{n}{n+1}\overline{X}_n) = \theta^2\left(\left(\frac{n}{n+1}\right)^2\left(\frac{n+1}{n}\right) - 2\frac{n}{n+1} + 1\right) = \frac{\theta^2}{n+1}$$

Therefore $T = \frac{n}{n+1}\overline{X}_n$ has smaller MSE than $\overline{X}_n$.

MSE is useful for comparing estimators, but not for selecting one estimator because there is no estimator with minimum MSE for all $\theta \in Ⓗ$. Consider a constant estimator $T = c$ of θ has $MSE(T) = (c-\theta)^2$, which is 0 if $\theta = c$. This means that for a minimum MSE estimator T^* of θ, T^* is a constant and $MSE(T^*) = 0$ for $\theta = T^*$. But $MSE(T^*) = (T^* - \theta)^2 > 0$ for $\theta \neq T^*$, in which case T^* is not uniformly minimum MSE.

4.4 Large Sample Properties

We have discussed fixed sample size and/or small sample properties of estimators. It is possible that an estimator has undesirable properties with small n, but becomes more reasonable as the sample size increases.

Definition 4.4.1. **(Simple consistency)**. Let T_n be a sequence of estimators for $\tau(\theta)$, $n = 1, 2, \ldots$. T_n is said to be a consistent sequence of estimators of $\tau(\theta)$ if $T_n \xrightarrow{P} \tau(\theta)$ as $n \to \infty$ $\forall\theta \in Ⓗ$. I.e. $\forall\epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\{|T_n - \tau(\theta)| < \epsilon\} = 1$$

Remark 4.4.1. $T_n \xrightarrow{P} \tau(\theta)$ as $n \to \infty$ $\forall \theta \in Ⓗ$ if and only if

$$\mathbb{P}\{T_n \le t\} \to \begin{cases} 1, & t > \tau(\theta) \\ 0, & t < \tau(\theta) \end{cases} \quad \text{as } n \to \infty$$

Definition 4.4.2. **(MSE consistency).** Let T_n be a sequence of estimators for $\tau(\theta)$, $n = 1, 2, \ldots$. T_n is MSE consistent if

$$\lim_{n\to\infty} MSE(T_n) = 0 \quad \forall \theta \in Ⓗ$$

Definition 4.4.3. **(Asymptotic unbiased).** Let T_n be a sequence of estimators for $\tau(\theta)$, $n = 1, 2, \ldots$. T_n is asymptotic unbiased for $\tau(\theta)$ if

$$\lim_{n\to\infty} \mathbb{E}T_n = \tau(\theta) \quad \forall \theta \in Ⓗ$$

i.e.

$$\lim_{n\to\infty} b(T_n) = 0 \quad \forall \theta \in Ⓗ$$

Theorem 4.4.1.

1. T_n is MSE consistent if and only if $\mathbb{V}arT_n \to 0$ and T_n asymptotic unbiased
2. MSE consistency $\implies$ simple consistency
3. If T_n is simple consistent for $\tau(\theta)$ and g is continuous at $\tau(\theta)$ $\forall \theta \in Ⓗ$, then $g(T_n)$ is simple consistent for $g(\tau(\theta))$
4. Asymptotic unbiased $\implies$ asymptotically MSE and variance are same.

Proof.

1. Follows from the relation $MSE(T_n) = \mathbb{V}arT_n + b^2(T_n)$
2. By Markov's inequality[1]

$$\mathbb{P}\{|T_n - \tau(\theta)| \ge \epsilon\} \le \frac{\mathbb{E}|T_n - \tau(\theta)|^2}{\epsilon^2} = \frac{MSE(T_n)}{\epsilon^2}$$

[1] Markov's inequality: $\mathbb{P}\{|X| > c\} \le \frac{\mathbb{E}|X|^r}{c^r}$

3. Follows from $T_n \xrightarrow{P} \tau(\theta)$ and g is continuous at $\tau(\theta) \implies g(T_n) \xrightarrow{P} g(\tau(\theta))$ by Theorem 3.3.3.

$\square$

Definition 4.4.4. **(Asymptotic efficiency)**. Let T_n, T_n^* be sequences of asymptotically unbiased estimators of $\tau(\theta)$. The asymptotic relative efficiency (ARE) of T_n relative to T_n^* is

$$ARE(T_n, T_n^*) = \lim_{n\to\infty} \frac{\mathbb{V}ar T_n^*}{\mathbb{V}ar T_n}$$

T_n^* is asymptotically efficient sequence if $ARE(T_n, T_n^*) \leq 1 \ \forall \theta \in \Theta$ for all asymptotically unbiased sequence T_n. The asymptotic efficiency (AE) of T_n is

$$AE(T_n) = ARE(T_n, T_n^*)$$

if T_n^* is asymptotically efficient.

However, $\mathbb{V}ar T_n$, $\mathbb{V}ar T_n^*$ may not exit, yet $\sqrt{n}(T_n - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, k(\theta))$ and $\sqrt{n}(T_n^* - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, k^*(\theta))$. Then we could alternatively define

$$ARE(T_n, T_n^*) = \frac{k^*(\theta)}{k(\theta)}$$

4.5 Asymptotic Properties of MLE's

Theorem 4.5.1. let $X_1, \ldots, X_n$ be a random sample from $f(x;\theta)$, and assume the conditions of CRLB theorem are met. Suppose MLE of θ, $\hat{\theta}_n$, exists, is unique, is consistent, and satisfies

$$\left. \frac{\partial \ln\left(\prod_{i=1}^n f(X_i;\theta)\right)}{\partial \theta} \right|_{\theta=\hat{\theta}_n} = 0$$

Then

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{as } n \to \infty$$

where $\sigma^2 = \frac{1}{\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2}$. I.e. for large n

$$\hat{\theta}_n \overset{d}{\approx} \mathcal{N}\left(\theta, \frac{1}{n\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2}\right) = \mathcal{N}(\theta, \text{CRLB})$$

I.e. $\hat{\theta}_n$ is asymptotically efficient.

Sketch of proof. By assumption and definition of MLE and consistent estimator[1]

$$0 = \frac{\partial \ln\left(\prod_{i=1}^n f(X_i;\hat{\theta}_n)\right)}{\partial\hat{\theta}_n} = \sum_{i=1}^n \frac{\partial \ln f(X_i;\hat{\theta}_n)}{\partial\hat{\theta}_n}$$

Expand in Taylor series near θ

$$0 = \sum_{i=1}^n \frac{\partial \ln f(X_i;\hat{\theta}_n)}{\partial\hat{\theta}_n} = \sum_{i=1}^n \frac{\partial \ln f(X_i;\theta)}{\partial\theta} + \left(\sum_{i=1}^n \frac{\partial^2 \ln f(X_i;\theta)}{\partial\theta^2}\right)$$

$$\times\left(\hat{\theta}_n - \theta\right) + \text{higher order (i.e. } \left(\hat{\theta}_n - \theta\right)^2 \text{ and higher order)}$$

Solve for

$$\hat{\theta}_n - \theta = \frac{-\sum_{i=1}^n \frac{\partial \ln f(X_i;\theta)}{\partial\theta}}{\sum_{i=1}^n \frac{\partial^2 \ln f(X_i;\theta)}{\partial\theta^2}} + \text{higher order}$$

which implies

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \frac{\partial \ln f(X_i;\theta)}{\partial\theta} - 0\right)}{\frac{1}{n}\sum_{i=1}^n\left(-\frac{\partial^2 \ln f(X_i;\theta)}{\partial\theta^2}\right)} + \underbrace{\sqrt{n}\times\text{higher order}}_{\xrightarrow{P} 0,\text{ by consistency}}$$

By the law of large numbers (Theorem 3.3.2)

$$\frac{1}{n}\sum_{i=1}^n\left(\frac{-\partial^2 \ln f(X_i;\theta)}{\partial\theta^2}\right) \xrightarrow{P} \mathbb{E}\left(\frac{-\partial^2 \ln f(X_1;\theta)}{\partial\theta^2}\right) = \mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2$$

[1] Recall $\hat{\theta}_n$ is a consistent estimator of θ if $\hat{\theta}_n \xrightarrow{P} \theta$ as $n \to \infty$

By central limit theorem

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\theta)}{\partial\theta}-0\right)\xrightarrow{d}\mathcal{N}\left(0,\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2\right)$$

By Corollary 3.4.1 of the continuous mapping theorem

$$\sqrt{n}\left(\hat{\theta}_n-\theta\right)\xrightarrow{d}\frac{\mathcal{N}\left(0,\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2\right)}{\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2}=\mathcal{N}\left(0,\frac{1}{\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2}\right)$$

□

Corollary 4.5.1.. Under the conditions of Theorem 4.5.1

$$\sqrt{n}\left(\tau\left(\hat{\theta}_n\right)-\tau(\theta)\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\left(\tau'(\theta)\right)^2}{\mathbb{E}\left(\frac{\partial \ln f(X_1;\theta)}{\partial\theta}\right)^2}\right)$$

i.e. for large n

$$\tau\left(\hat{\theta}_n\right)\overset{d}{\approx}\mathcal{N}\left(\tau(\theta),\mathrm{CRLB}\right)$$

Proof. Applying delta rule yields the result. □

Example 4.5.1.. $X_1,\ldots,X_n$ is a random sample from a Pareto distribution[1] $\mathcal{P}ar(1,\kappa)$, $f(x;\kappa)=\kappa(1+x)^{-\kappa-1}, x>0, \kappa>0$. One have verified in HW 3 Prob 5h that the MLE of κ is $\hat{\kappa}=\frac{n}{\sum_{i=1}^{n}\ln(1+X_i)}$. What is the asymptotic behavior of the MLE?

$$\frac{\partial \ln f(X_1;\kappa)}{\partial\kappa}=\frac{\partial}{\partial\kappa}\left(\ln\kappa-(\kappa+1)\ln(1+X_1)\right)=\frac{1}{\kappa}-\ln(1+X_1)$$

$$\frac{\partial^2 \ln f(X_1;\kappa)}{\partial\kappa^2}=-\frac{1}{\kappa^2}$$

$$\mathrm{CRLB}=\frac{1}{n\mathbb{E}\left(\frac{\partial \ln f(X_1;\kappa)}{\partial\kappa}\right)^2}=\frac{1}{-n\mathbb{E}\left(\frac{\partial^2 \ln f(X_1;\kappa)}{\partial\kappa^2}\right)}=\frac{\kappa^2}{n}$$

[1] refer to the special distribution table

Apply Theorem 4.5.1

$$\sqrt{n}\,(\hat{\kappa} - \kappa) \xrightarrow{d} \mathcal{N}(0, \kappa^2) \quad \text{as } n \to \infty$$

I.e. for large n

$$\hat{\kappa} \overset{d}{\approx} \mathcal{N}\left(\kappa, \frac{\kappa^2}{n}\right)$$