Fisher Information

Assume $X \sim f(x|\theta)$ with $\theta \in \Theta \subset \mathbb{R}$.

pdf or pmf

Define

$$I_X(\theta) = E_\theta\left(\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right)$$

the derivative of the log-likelihood function evaluated at the true value of θ.

Fisher information is meaningful for families of distributions which are <u>regular</u>:

1. Fixed support: $\{x : f(x|\theta) > 0\}$ is the same for all θ.
2. $\frac{\partial}{\partial \theta} \log f(x|\theta)$ must exist and be finite for all x and θ.
3. If $E_\theta |W(X)| < \infty \ \forall \theta$, then

$$\left(\frac{\partial}{\partial \theta}\right)^k E_\theta W(X) = \left(\frac{\partial}{\partial \theta}\right)^k \int W(x) f(x|\theta)\, dx$$

$\rightarrow$ or Σ

$$= \int W(x) \left(\frac{\partial}{\partial \theta}\right)^k f(x|\theta)\, dx.$$

Regular Families

One-parameter exponential families,
Cauchy location or scale family:

$$f(x|\theta) = \frac{1}{\pi(1+(x-\theta)^2)}$$

$$f(x|\theta) = \frac{1}{\pi\theta\left(1+\left(\frac{x}{\theta}\right)^2\right)}$$

and lots more.

(Most families of distns used in applications are regular.)

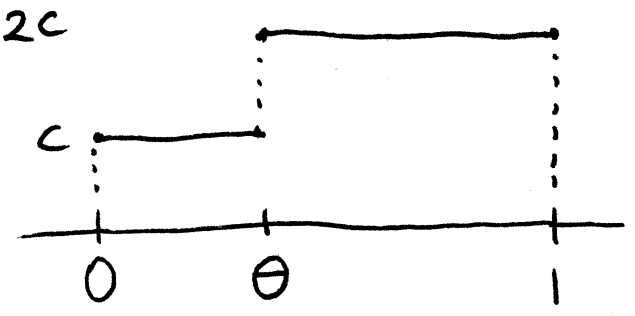
Non-regular Families

Uniform$(0,\theta)$

Uniform$(\theta,\theta+1)$

$$f(x|\theta) = c(\theta)\left[1+I(x>\theta)\right]$$

for $0<x<1$,
$0<\theta<1$.



Facts about Fisher Information

(assume a regular family)

① $E_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = 0$

$\frac{\partial}{\partial\theta}\log f(X|\theta)$ → The "score" function $S(\theta)$.

Proof:

$$E_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = \int\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right] f(x|\theta)\,dx$$

$$= \int \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} \cdot f(x|\theta)\,dx$$

$$= \int \frac{\partial}{\partial\theta} f(x|\theta)\,dx = \frac{\partial}{\partial\theta}\underbrace{\int f(x|\theta)\,dx}_{=1 \text{ for all } \theta}$$

$$= 0$$

② $I_X(\theta) = \mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)$

Proof: Since $E\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = 0$,

$$\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right) = E_\theta\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2$$

$$= I_X(\theta).$$

③ If $X = (X_1, X_2, \ldots, X_n)$ and $X_1, X_2, \ldots, X_n$ are independent rv's, then $I_X(\theta) = I_{X_1}(\theta) + I_{X_2}(\theta) + \cdots + I_{X_n}(\theta)$.

Proof:

$$f(x|\theta) = \prod_{i=1}^{n} f_i(x_i|\theta)$$

where $f_i(\cdot|\theta)$ is the pdf (pmf) of X_i.

$$\frac{\partial}{\partial\theta} \log f(x|\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log f_i(X_i|\theta)$$

and the rv's in the sum are independent.

Thus

$$\mathrm{Var}\left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right] = \sum_{i=1}^{n} \mathrm{Var}\left[\frac{\partial}{\partial\theta} \log f_i(X_i|\theta)\right]$$

so that $I_X(\theta) = \sum_{i=1}^{n} I_{X_i}(\theta)$ by ②.

④ If $X_1, X_2, \ldots, X_n$ are iid and $X = (X_1, \ldots, X_n)$, then $I_{X_i}(\theta) = I_{X_1}(\theta)$ for all i so that $I_X(\theta) = n I_{X_1}(\theta)$.

⑤ $I_X(\theta) = E_\theta\left(-\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right)$

(Alternate formula for Fisher information)

Proof:

Abbreviate $\int f(x|\theta)\,dx$ as $\int f$, etc.

$$1 = \int f.$$

Apply $\frac{\partial}{\partial\theta}$ to both sides.

$$0 = \frac{\partial}{\partial\theta}\int f = \int \frac{\partial f}{\partial\theta} = \int \frac{\frac{\partial f}{\partial\theta}}{f}\cdot f$$

$$= \int \left(\frac{\partial}{\partial\theta}\log f\right)\cdot f\ .$$

Apply $\frac{\partial}{\partial\theta}$ again.

$$0 = \frac{\partial}{\partial\theta}\int \left(\frac{\partial}{\partial\theta}\log f\right)\cdot f$$

$$= \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f\right)\cdot f\right]$$

$$= \int \left(\frac{\partial^2}{\partial\theta^2}\log f\right)\cdot f + \int \left(\frac{\partial}{\partial\theta}\log f\right)\cdot\frac{\partial f}{\partial\theta}$$

Noting that $\frac{\partial f}{\partial \theta} = \frac{\frac{\partial f}{\partial \theta}}{f} \cdot f$

$$= \left(\frac{\partial}{\partial \theta} \log f\right) \cdot f$$

this becomes

$$0 = \int \left(\frac{\partial^2}{\partial \theta^2} \log f\right) \cdot f + \int \left(\frac{\partial}{\partial \theta} \log f\right)^2 \cdot f$$

or

$$0 = E\left(\frac{\partial^2}{\partial \theta^2} \log f(X|\theta)\right) + I_X(\theta)$$

QED

Example: Fisher information for a Poisson sample.
Observe $\underset{\sim}{\boldsymbol{X}} = (X_1, \ldots, X_n)$ iid Poisson(λ). Find $I_{\underset{\sim}{\boldsymbol{X}}}(\lambda)$.

We know $I_{\underset{\sim}{\boldsymbol{X}}}(\lambda) = nI_{X_1}(\lambda)$.

We shall calculate $I_{X_1}(\lambda)$ in three ways. Let $X = X_1$.
Preliminaries:

$$\begin{aligned} f(x \mid \lambda) &= \frac{\lambda^x e^{-\lambda}}{x!} \\ \log f(x \mid \lambda) &= x \log \lambda - \lambda - \log x! \\ \frac{\partial}{\partial \lambda} \log f(x \mid \lambda) &= \frac{x}{\lambda} - 1 \\ -\frac{\partial^2}{\partial \lambda^2} \log f(x \mid \lambda) &= \frac{x}{\lambda^2} \end{aligned}$$

Method # 1:

$$\begin{aligned} I_X(\lambda) &= E_\lambda \left[\left(\frac{\partial}{\partial \lambda} \log f(X \mid \lambda) \right)^2 \right] = E_\lambda \left[\left(\frac{X}{\lambda} - 1 \right)^2 \right] \\ &= \mathrm{Var}_\lambda \left(\frac{X}{\lambda} \right) \qquad \text{since } E\left(\frac{X}{\lambda} \right) = \frac{EX}{\lambda} = \frac{\lambda}{\lambda} = 1 \\ &= \frac{\mathrm{Var}(X)}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} \end{aligned}$$

Method # 2 (almost the same):

$$\begin{aligned} I_X(\lambda) &= \mathrm{Var}_\lambda \left(\frac{\partial}{\partial \lambda} \log f(X \mid \lambda) \right) = \mathrm{Var} \left(\frac{X}{\lambda} - 1 \right) \\ &= \mathrm{Var} \left(\frac{X}{\lambda} \right) = \frac{1}{\lambda} \qquad \text{as in Method \#1} \end{aligned}$$

Method #3:

$$I_X(\lambda) = E_\lambda \left(-\frac{\partial^2}{\partial \lambda^2} \log f(X \mid \lambda) \right) = E_\lambda \left(\frac{X}{\lambda^2} \right) = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

- Thus $I_{\underset{\sim}{\boldsymbol{X}}}(\lambda) = nI_{X_1}(\lambda) = \dfrac{n}{\lambda}$.

Example: Fisher information for Cauchy location family.

Suppose $X_1, X_2, \ldots, X_n$ iid with pdf $f(x|\theta) = \dfrac{1}{\pi(1+(x-\theta)^2)}$.

Let $\underset{\sim}{X} = (X_1, \ldots, X_n)$, $X \sim f(x|\theta)$.

Find $I_{\underset{\sim}{X}}(\theta)$.

$$I_{\underset{\sim}{X}}(\theta) = n\, I_{X_1}(\theta) = n\, I_X(\theta).$$

$$\frac{\partial}{\partial\theta} \log f(x|\theta) = \frac{\frac{\partial f}{\partial\theta}}{f}$$

$$= \frac{\dfrac{-1}{\pi(1+(x-\theta)^2)^2} \cdot 2(x-\theta)(-1)}{\dfrac{1}{\pi(1+(x-\theta)^2)}}$$

$$= \frac{2(x-\theta)}{1+(x-\theta)^2}$$

$$I_X(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log f(X|\theta)\right)^2\right]$$

$$= E\left(\frac{2(X-\theta)}{1+(X-\theta)^2}\right)^2$$

$$= \int_{-\infty}^{\infty}\left(\frac{2(x-\theta)}{1+(x-\theta)^2}\right)^2 \frac{1}{\pi(1+(x-\theta)^2)}\,dx$$

$$= \frac{4}{\pi}\int_{-\infty}^{\infty}\frac{(x-\theta)^2}{(1+(x-\theta)^2)^3}\,dx$$

Let $u = x-\theta$, $du = dx$.

$$= \frac{4}{\pi}\int_{-\infty}^{\infty}\frac{u^2}{(1+u^2)^3}\,du$$

$$= \frac{8}{\pi}\int_{0}^{\infty}\frac{u^2}{(1+u^2)^3}\,du$$

Substitute

$$x = \frac{1}{1+u^2}, \quad 1-x = \frac{u^2}{1+u^2}$$

$$dx = \frac{-2u}{(1+u^2)^2}\,du \quad \text{(don't use)}$$

$$u = \left(\frac{1}{x} - 1\right)^{1/2}$$

$$du = \frac{1}{2}\left(\frac{1}{x} - 1\right)^{-1/2}\left(-\frac{1}{x^2}\right)dx$$

Then

$$\frac{8}{\pi}\int_0^\infty \frac{u^2}{(1+u^2)^3}\,du$$

$$= \frac{8}{\pi}\int_0^\infty \frac{u^2}{1+u^2}\left(\frac{1}{1+u^2}\right)^2 du$$

$$= \frac{8}{\pi}\int_1^0 (1-x)\,x^2 \cdot \frac{1}{2}\left(\frac{1}{x} - 1\right)^{-1/2}\left(-\frac{1}{x^2}\right)dx$$

$$= \frac{4}{\pi}\int_0^1 x^{+1/2}(1-x)^{1/2}\,dx$$

$$= \frac{4}{\pi}\int_0^1 x^{3/2-1}(1-x)^{3/2-1}\,dx \quad \left(\begin{matrix}\text{Beta}\\ \text{integral}\end{matrix}\right)$$

$$= \frac{4}{\pi}\frac{\Gamma(3/2)\,\Gamma(3/2)}{\Gamma(3/2+3/2)} = \frac{4}{\pi}\frac{\left(\frac{1}{2}\sqrt{\pi}\right)^2}{2!}$$

$$= \frac{1}{2}. \quad \text{Therefore } I_{\underset{\sim}{X}}(\theta) = \frac{n}{2}.$$

Uses of Fisher Information:

- Asymptotic distribution of MLE's
- Cramér-Rao Inequality (Information Inequality)

Asymptotic distribution of MLE's

• IID Case:

If $f(x \mid \theta)$ is a regular one-parameter family of pdf's (or pmf's) and $\widehat{\theta}_n = \widehat{\theta}_n(\boldsymbol{X}_n)$ is the MLE based on $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ where n is large and $X_1, \ldots, X_n$ are iid from $f(x \mid \theta)$, then

$$\widehat{\theta}_n \sim \text{approx}\, N\left(\theta\,,\, \frac{1}{nI(\theta)}\right) \qquad \text{where } I(\theta) \equiv I_{X_1}(\theta)$$

and θ is the true value. Note that $nI(\theta) = I_{\boldsymbol{X}_n}(\theta)$.

More formally,

$$\frac{\widehat{\theta}_n - \theta}{\sqrt{\frac{1}{nI(\theta)}}} = \sqrt{nI(\theta)}(\widehat{\theta}_n - \theta) \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty.$$

• More general case:

(Assuming various regularity conditions)
If $f(\underset{\sim}{\mathbf{x}} \mid \theta)$ is a one-parameter family of joint pdf's (or joint pmf's) for data $\boldsymbol{X}_n = (X_1, \ldots, X_n)$ where n is large (think of a large data set arising from a regression or time series model) and $\widehat{\theta}_n = \widehat{\theta}_n(\boldsymbol{X}_n)$ is the MLE, then

$$\widehat{\theta}_n \sim \text{approx}\, N\left(\theta\,,\, \frac{1}{I_{\boldsymbol{X}_n}(\theta)}\right) \qquad (\theta \equiv \text{true value}).$$

Estimation of the Fisher Information

If θ is unknown, then so is $I_{\boldsymbol{X}}(\theta)$.

Two estimates $\hat{I}$ of the Fisher information $I_{\boldsymbol{X}}(\theta)$ are

$$\hat{I}_1 = I_{\boldsymbol{X}}(\hat{\theta}) \quad \text{and} \quad \hat{I}_2 = -\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X} \mid \theta)\bigg|_{\theta=\hat{\theta}}$$

where $\hat{\theta}$ is the MLE of θ based on the data $\boldsymbol{X}$.

$\hat{I}_1$ is the obvious plug-in estimator. It can be difficult to compute when $I_{\boldsymbol{X}}(\theta)$ does not have a known closed form.

The estimator $\hat{I}_2$ is suggested by the formula

$$I_{\boldsymbol{X}}(\theta) = E\left(-\frac{\partial^2}{\partial\theta^2} \log f(\boldsymbol{X} \mid \theta)\right).$$

It is often easy to compute, and is required in many Newton-Raphson style algorithms for finding the MLE (so that it is already available without extra computation).

The two estimates $\hat{I}_1$ and $\hat{I}_2$ are often referred to as the "expected" and "observed" Fisher information, respectively.

As $n \to \infty$, both estimators are consistent (after normalization) for $I_{\boldsymbol{X}_n}(\theta)$ under various regularity conditions.

For example: in the iid case: $\hat{I}_1/n$, $\hat{I}_2/n$, and $I_{\boldsymbol{X}_n}(\theta)/n$ all converge to $I(\theta) \equiv I_{X_1}(\theta)$.

Approximate Confidence Intervals for θ

Choose $0 < \alpha < 1$ (say, $\alpha = 0.05$). Let z^* be such that

$$P(-z^* < Z < z^*) = 1 - \alpha \quad \text{where } Z \sim N(0,1).$$

When n is large,

$$\sqrt{I_{\boldsymbol{X}}(\theta)}(\widehat{\theta} - \theta) \sim \text{approx } N(0,1) \quad \text{so that}$$

$$P\left\{-z^* < \sqrt{I_{\boldsymbol{X}}(\theta)}(\widehat{\theta} - \theta) < z^*\right\} \approx 1 - \alpha \quad \text{or equivalently}$$

$$P\left\{\widehat{\theta} - z^*\sqrt{\frac{1}{I_{\boldsymbol{X}}(\theta)}} < \theta < \widehat{\theta} + z^*\sqrt{\frac{1}{I_{\boldsymbol{X}}(\theta)}}\right\} \approx 1 - \alpha\,.$$

This approximation continues to hold when $I_{\boldsymbol{X}}(\theta)$ is replaced by an estimate $\widehat{I}$ (either $\widehat{I}_1$ or $\widehat{I}_2$):

$$P\left\{\widehat{\theta} - z^*\sqrt{\frac{1}{\widehat{I}}} < \theta < \widehat{\theta} + z^*\sqrt{\frac{1}{\widehat{I}}}\right\} \approx 1 - \alpha\,.$$

Thus

$$\left(\widehat{\theta} - z^*\sqrt{\frac{1}{\widehat{I}}}\,,\, \widehat{\theta} + z^*\sqrt{\frac{1}{\widehat{I}}}\right)$$

is an approximate $1 - \alpha$ confidence interval for θ. (Here $\widehat{\theta}$ is the MLE and $\widehat{I}$ an estimate of the Fisher information.)

Cramer-Rao Inequality ($\underset{\sim}{X} \sim P_\theta, \theta \in \Theta \subset \mathbb{R}$)

If $f(\underset{\sim}{x}|\theta)$ is a regular one-parameter family,

$E_\theta W(\underset{\sim}{X}) = \tau(\theta)$ for all θ, and

$\tau(\theta)$ is differentiable,

then $\mathrm{Var}_\theta(W(\underset{\sim}{X})) \geq \dfrac{(\tau'(\theta))^2}{I_{\underset{\sim}{X}}(\theta)}$.

Proof:

Preliminary Facts

(A) $[\mathrm{Cov}(X,Y)]^2 \leq (\mathrm{Var}\,X)(\mathrm{Var}\,Y)$.

This is a special case of the Cauchy-Schwarz inequality. It is better known to statisticians as $\rho^2 \leq 1$ where $\rho = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}X \cdot \mathrm{Var}Y}}$ is the correlation between X and Y.

(B) $\mathrm{Cov}(X,Y) = EXY$ if either $EX=0$ or $EY=0$.

This follows from the well known formula $\mathrm{Cov}(X,Y) = EXY - (EX)(EY)$.

Body of Proof: From (A) we have

$$\left[\mathrm{Cov}_\theta\left(W(\underset{\sim}{X}), \frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta)\right)\right]^2 \le \left(\mathrm{Var}\, W(\underset{\sim}{X})\right)\underbrace{\left(\mathrm{Var}\left(\frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta)\right)\right)}_{I_{\underset{\sim}{X}}(\theta)}.$$

From (B) we have

$$\mathrm{Cov}_\theta\left(W(\underset{\sim}{X}), \frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta)\right) = E\left[W(\underset{\sim}{X})\frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta)\right]$$

(since $E_\theta \frac{\partial}{\partial\theta}\log f(\underset{\sim}{X}|\theta) = 0$)

$$= \int W(\underset{\sim}{x})\underbrace{\left(\frac{\partial}{\partial\theta}\log f(\underset{\sim}{x}|\theta)\right)}_{= \frac{\frac{\partial f}{\partial\theta}}{f}} f(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x}$$

$$= \int W(\underset{\sim}{x})\frac{\partial f(\underset{\sim}{x}|\theta)}{\partial\theta}\, d\underset{\sim}{x}$$

$$= \frac{\partial}{\partial\theta}\int W(\underset{\sim}{x}) f(\underset{\sim}{x}|\theta)\, d\underset{\sim}{x}$$

(since $f(\underset{\sim}{x}|\theta)$ is a regular family)

$$= \frac{\partial}{\partial\theta} E_\theta W(\underset{\sim}{X}) = \tau'(\theta).$$

Thus $[\tau'(\theta)]^2 \le \left(\mathrm{Var}_\theta W(\underset{\sim}{X})\right) I_{\underset{\sim}{X}}(\theta)$.

QED

Addendum:

Equality in (A) is achieved iff

$$Y = aX + b \text{ for some constants } a, b.$$

Moreover, if $EY = 0$, then $E(aX+b) = 0$ forces $b = -aEX$ so that

$$Y = a(X - EX) \text{ for some constant } a.$$

Applying this to the proof of CRLB with $X = W(\underset{\sim}{X})$, $Y = \frac{\partial}{\partial\theta} \log f(\underset{\sim}{X}|\theta)$ tells us that

$$\text{Var}_\theta\, W(\underset{\sim}{X}) = \frac{(\tau'(\theta))^2}{I_{\underset{\sim}{X}}(\theta)}$$

iff $\frac{\partial}{\partial\theta} \log f(\underset{\sim}{x}|\theta) = a(\theta)\left[W(\underset{\sim}{x}) - \tau(\theta)\right]$ (*)

for some function $a(\theta)$.

Note: (*) is true only when

$f(\underset{\sim}{x}|\theta)$ is a 1pef and

$W(\underset{\sim}{X}) = cT(\underset{\sim}{X}) + d$ for some c, d

where $T(\underset{\sim}{X})$ is the natural suff. stat of the 1pef.

Asymptotic Efficiency

Given: A sequence of estimators

$$W_n = W_n(\underbrace{X_1, X_2, \ldots, X_n}_{\underset{\sim}{X}_n}).$$

If $E W_n = \tau(\theta)$ for all n,
then $\{W_n\}$ is <u>asymptotically efficient</u> if

$$\lim_{n\to\infty} \underbrace{\frac{\mathrm{Var}_\theta W_n}{(\tau'(\theta))^2 / I_{\underset{\sim}{X}_n}(\theta)}}_{\text{call this } V_n(\theta)} = 1 .$$

What if $\mathrm{Var}_\theta W_n = \infty$ or if W_n is biased?

<u>An alternative definition</u>:

$\{W_n\}$ is asymptotically efficient for estimating $\tau(\theta)$ if

$$W_n \sim \underbrace{AN}_{\text{asymptotically normal}}(\tau(\theta), V_n(\theta)).$$

Meaning of AN: $\dfrac{W_n - \tau(\theta)}{\sqrt{V_n(\theta)}} \xrightarrow{d} N(0,1)$ as $n \to \infty$.

Example: Observe $X_1, \ldots, X_n$ iid Poisson(λ).

- **Estimation of $\tau(\lambda) = \lambda$**

$E\bar{X} = \lambda$

Does $\bar{X}$ achieve the CRLB? Yes!

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X_1)}{n} = \frac{\lambda}{n}$$

$$\text{CRLB} = \frac{(\tau'(\lambda))^2}{I_{\boldsymbol{X}}(\lambda)} = \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

Alternative: Check condition for exact attainment of CRLB.

$$\begin{aligned} \frac{\partial}{\partial\lambda} \log f(\boldsymbol{X} \,|\, \lambda) &= \sum_{i=1}^{n} \frac{\partial}{\partial\lambda} \log f(X_i \,|\, \lambda) = \sum_i \left(\frac{X_i}{\lambda} - 1 \right) \\ &= \frac{n}{\lambda} \left(\bar{X} - \lambda \right) \end{aligned}$$

Note: Since $\bar{X}$ attains the CRLB (for all λ), it must be the best unbiased estimator of λ.

Showing that an estimator attains the CRLB is one way to show it is best unbiased. (But see later remark.)

- **Estimation of $\tau(\lambda) = \lambda^2$**

Define $W = \dfrac{T(T-1)}{n^2}$ where $T = \sum_{i=1}^{n} X_i$.

$EW = \lambda^2$ (see calculations below) and W is a function of the CSS T. Thus W is best unbiased for λ^2.

Does W achieve the CRLB? No!!!

$$\text{CRLB} = \frac{(\tau'(\lambda))^2}{I_{\boldsymbol{X}}(\lambda)} = \frac{(2\lambda)^2}{n/\lambda} = \frac{4\lambda^3}{n}$$

$$\text{Var}(W) = \frac{4\lambda^3}{n} + \frac{2\lambda^2}{n^2} \quad \text{(see calculations below)}$$

Alternative: Show condition for achievement of CRLB **fails**.

As shown earlier:

$$\frac{\partial}{\partial\lambda}\log f(\boldsymbol{X}\,|\,\lambda) = \sum_i \left(\frac{X_i}{\lambda} - 1\right) = \frac{T}{\lambda} - n$$

The CRLB is attained **iff** there exists $a(\lambda)$ such that

$$\frac{T}{\lambda} - n = a(\lambda)\left(\frac{T(T-1)}{n^2} - \lambda^2\right).$$

But the left side is linear in T, and the right is quadratic in T, so that no multiplier $a(\lambda)$ can make them equal for all possible values of $T = 0, 1, 2, \ldots$

Remark: This situation is not unusual. The best unbiased estimator often fails to achieve the CRLB.

But W is asymptotically efficient:

$$\lim_{n\to\infty}\frac{\mathrm{Var}(W)}{CRLB} = \lim_{n\to\infty}\frac{\frac{4\lambda^3}{n} + \frac{2\lambda^2}{n^2}}{\frac{4\lambda^3}{n}} = \lim_{n\to\infty}\left(1 + \frac{1}{2n\lambda}\right) = 1$$

- **Calculations:**

Suppose $Y \sim$ Poisson(ξ).

The factorial moments of the Poisson follow a simple pattern:

$$\begin{aligned} EY &= &&= \xi \\ EY(Y-1) &= &&= \xi^2 \\ EY(Y-1)(Y-2) &= &&= \xi^3 \\ EY(Y-1)(Y-2)(Y-3) &= &&= \xi^4 \quad \text{etc.} \end{aligned}$$

Proof of one case:

$$\begin{aligned} EY(Y-1)(Y-2) &= \sum_{i=0}^{\infty} i(i-1)(i-2)\frac{\xi^i e^{-\xi}}{i!} \\ &= \xi^3 \sum_{i=3}^{\infty} \frac{\xi^{i-3} e^{-\xi}}{(i-3)!} = \xi^3 \sum_{j=0}^{\infty} \frac{\xi^j e^{-\xi}}{j!} = \xi^3 \end{aligned}$$

From the factorial moments, we can calculate everything else. For example:

$$\begin{aligned} \text{Var}(Y(Y-1)) &= E\left[\{Y(Y-1)\}^2\right] - [EY(Y-1)]^2 \\ &= E\left[Y^2(Y-1)^2\right] - \left[\xi^2\right]^2 \\ &= E\left[\langle Y\rangle_4 + 4\langle Y\rangle_3 + 2\langle Y\rangle_2\right] - \xi^4 \\ &= \left[\xi^4 + 4\xi^3 + 2\xi^2\right] - \xi^4 = 4\xi^3 + 2\xi^2 \\ &\quad \text{where } \langle Y\rangle_k \equiv Y(Y-1)\cdots(Y-k+1)\,. \end{aligned}$$

In our case $T \sim \text{Poisson}(n\lambda)$ so that substituting $\xi = n\lambda$ in the above results leads to

$$\begin{aligned} ET(T-1) &= (n\lambda)^2 = n^2\lambda^2 \\ \text{Var}[T(T-1)] &= 4(n\lambda)^3 + 2(n\lambda)^2 = 4n^3\lambda^3 + 2n^2\lambda^2 \end{aligned}$$

so that $W = T(T-1)/n^2$ satisfies:

$$\begin{aligned} EW &= \lambda^2 \\ \text{Var}(W) &= \frac{4\lambda^3}{n} + \frac{2\lambda^2}{n^2} \end{aligned}$$

An asymptotically inefficient estimator

Example: Let $X_1, X_2, \ldots, X_n$ be iid with pdf $f(x|\alpha) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$ for $x > 0$.

For this pdf, $EX = \text{Var}\, X = \alpha$.

Clearly $E\bar{X} = \alpha$. Thus $\bar{X}$ is MOM estimator of α.

Is it asymptotically efficient? No. (Verified below.)

Note: This is lpef with natural sufficient statistic $T = \sum_{i=1}^{n} \log X_i$. Since T is complete, $E(\bar{X}|T)$ is the UMVUE of α. Since $\bar{X}$ is <u>not</u> a function of T, we know

$\text{Var}(\bar{X}) > \text{Var}[E(\bar{X}|T)]$.

But $\text{Var}[E(\bar{X}|T)] \geq$ CRLB. Thus, without calculation, we know that $\bar{X}$ cannot achieve the CRLB for any value of n. We now show it does <u>not</u> achieve it asymptotically either.

$$\text{Var}\,\bar{X} = \frac{\text{Var}\,X_1}{n} = \frac{\alpha}{n}.$$

$$I_{\underset{\sim}{X}_n}(\alpha) = n\, I_{X_1}(\alpha)$$

$$= n\left[\frac{\Gamma''(\alpha)\Gamma(\alpha) - (\Gamma'(\alpha))^2}{(\Gamma(\alpha))^2}\right]$$

by a routine calculation

$$\text{CRLB} = \frac{1}{n\, I_{X_1}(\alpha)}.$$

Thus

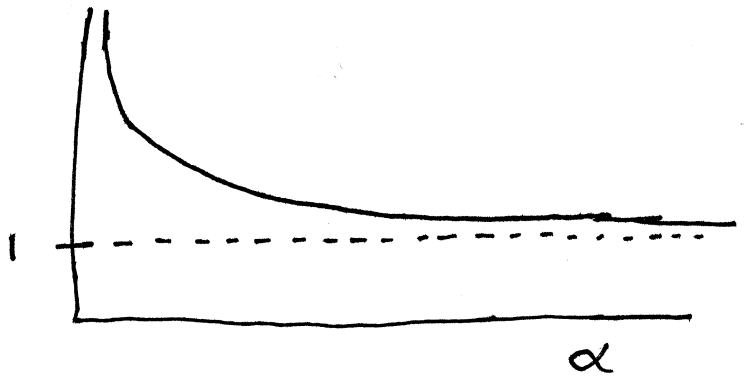
$$\frac{\text{Var}\,\bar{X}}{\text{CRLB}} = \alpha\, I_{X_1}(\alpha)$$ which does <u>not</u> depend on n.

since $\bar{X}$ does <u>not</u> achieve CRLB for any n, we know $\alpha\, I_{X_1}(\alpha) > 1$. Thus

$$\lim_{n\to\infty} \frac{\text{Var}\,\bar{X}}{\text{CRLB}} = \alpha\, I_{X_1}(\alpha) > 1$$

so that $\bar{X}$ is <u>not</u> asymp. eff.

Note: The function $\alpha I_{X_1}(\alpha)$ looks like this:



$$\lim_{\alpha \to 0} \alpha I_{X_1}(\alpha) = \infty$$

$$\lim_{\alpha \to \infty} \alpha I_{X_1}(\alpha) = 1$$

When α is small, $\bar{X}$ is horrible.
When α is large, $\bar{X}$ is pretty good.

MOM is inefficient (in general).

General Comment: For regular families, the MLE is asymptotically efficient.

Thus

$\lim_{n \to \infty} \frac{\text{Var } W_n}{\text{CRLB}(n)}$ essentially compares the variance of W_n with that of the MLE in large samples.