Recall: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \Rightarrow E[S_n^2] = var(X_i) = \sigma^2$ unbiased

$$\frac{n-1}{n}S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2 \Rightarrow E\left[\frac{n-1}{n}S_n^2\right] = \frac{n-1}{n}E[S_n^2] = \frac{n-1}{n}\sigma^2$$

$$\Rightarrow bias\left(\frac{n-1}{n}S_n^2\right) = E\left[\frac{n-1}{n}S_n^2\right] - \sigma^2$$

$$= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2 \xrightarrow{n\to\infty} 0$$

So $\frac{n-1}{n}S_n^2$ is biased but <u>asymptotically</u> unbiased.

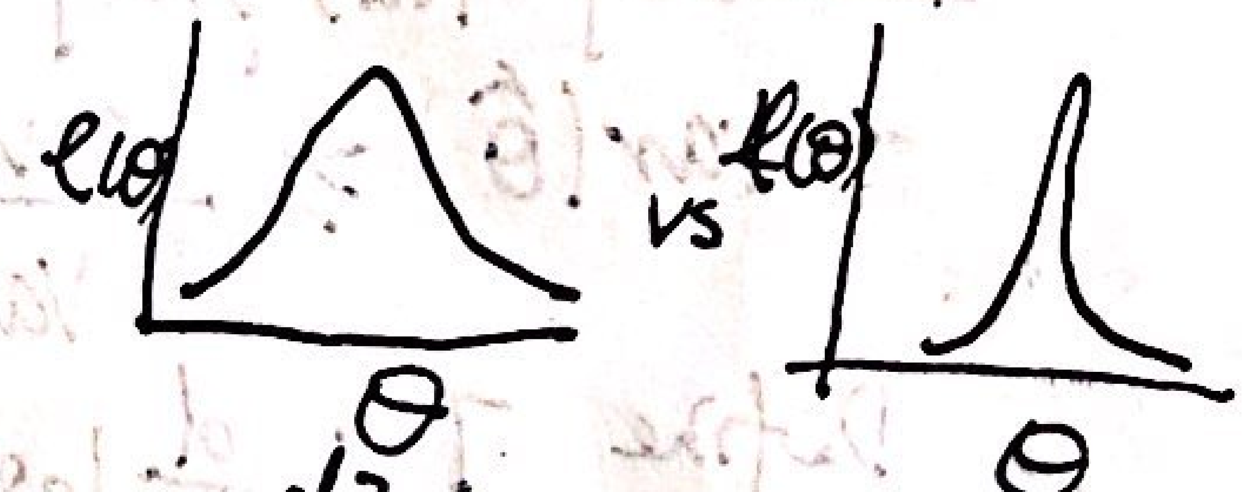
Fisher information: based on the log-likelihood of θ w/ 1 obs

$\ell(\theta) = \log f(X;\theta)$ $\quad$ f is the pdf or pmf of r.v. X

Fisher information: $I(\theta) = \underset{\text{w.r.t. } X}{E}\left[\left(\frac{d}{d\theta}\ell(\theta)\right)^2\right]$ $\quad$ θ is a constant

$$= E_X\left[\left(\frac{d}{d\theta}\log f(X;\theta)\right)^2\right]$$

Alternative def: $I(\theta) = -E\left[\frac{d^2}{d\theta^2}\ell(\theta)\right]$



$-\frac{d^2}{d\theta^2}\ell(\theta)$ is bigger on the right

Pf: Since $\int f(x;\theta)\,dx = 1$:

$$\frac{d}{d\theta}\int f(x;\theta)\,dx = \frac{d}{d\theta}1 = 0$$

Because of the smoothness conditions of $f(\cdot;\theta)$,

$$\frac{d}{d\theta}\int f(x;\theta)\,dx = \int \frac{d}{d\theta}f(x;\theta)\,dx = 0 \quad (1)$$

Now, $\frac{d}{d\theta}\log f(x;\theta) = \frac{\frac{d}{d\theta}f(x;\theta)}{f(x;\theta)}$

so $\frac{d}{d\theta}f(x;\theta) = \left(\frac{d}{d\theta}\log f(x;\theta)\right)\cdot f(x;\theta)$ $\quad$ (2)

Apply ~~Add~~ $\int \cdot\, dx$ to both

$$0 \overset{(1)}{=} \int \frac{d}{d\theta}f(x;\theta)\,dx \overset{(2)}{=} \int\left(\frac{d}{d\theta}\log f(x;\theta)\right)f(x;\theta)\,dx = E\left[\frac{d}{d\theta}\log f(X;\theta)\right]$$

Do $\frac{d}{d\theta}$ on both sides of (2):

$$\frac{d^2}{d\theta^2}f(x;\theta) = \left(\frac{d^2}{d\theta^2}\log f(x;\theta)\right)\cdot f(x;\theta) + \left(\frac{d}{d\theta}\log f(x;\theta)\right)\cdot\left(\frac{d}{d\theta}f(x;\theta)\right)$$

$\int dx$ on both sides:

$$\text{LHS} = \int \frac{d^2}{d\theta^2} f(x;\theta)\,dx = \frac{d^2}{d\theta^2} \underbrace{\int f(x;\theta)\,dx}_{=1} = 0$$

$$\| $$

$$\text{RHS} = \underbrace{\int \left(\frac{d^2}{d\theta^2}\log f(x;\theta)\right)\cdot f(x;\theta)\,dx}_{= E\left[\frac{d^2}{d\theta^2}\log f(X;\theta)\right]} + \underbrace{\int \left(\frac{d}{d\theta}\log f(x;\theta)\right)\cdot \boxed{\left(\frac{d}{d\theta}\log f(x;\theta)\right)\cdot f(x;\theta)}\,dx}_{}$$

by (2)

$$= \int \left(\frac{d}{d\theta}\log f(x;\theta)\right)^2 \cdot f(x;\theta)\,dx$$

$$= E\left[\left(\frac{d}{d\theta}\log f(X;\theta)\right)^2\right]$$

$$\Rightarrow E\left[\left(\frac{d}{d\theta}\log f(X;\theta)\right)^2\right] = -E\left[\frac{d^2}{d\theta^2}\log f(X;\theta)\right] \triangleq I(\theta).$$

□

Example: $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find $I(\mu)$. Assume σ^2 is known.

Ans: pdf: $f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$$\log f(x;\mu) = -\log(\sqrt{2\pi}\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{d}{d\mu}\log f(x;\mu) = -\frac{2(x-\mu)\cdot(-1)}{2\sigma^2} = \frac{x-\mu}{\sigma^2}$$

$$\frac{d^2}{d\mu^2}\log f(x;\mu) = -\frac{1}{\sigma^2} \quad \text{not involving } x$$

So $I(\mu) = -E\left[\frac{d^2}{d\mu^2}\log f(X;\mu)\right] = -E\left[\frac{-1}{\sigma^2}\right] = \frac{1}{\sigma^2}$.

Our definition of $I(\theta)$ is based on $X \sim f(\cdot;\theta)$

How about $X_1, \cdots, X_n \overset{iid}{\sim} f(\cdot;\theta)$? Define "joint Fisher information":

$$I_n(\theta) = E\Big[\Big(\frac{d}{d\theta}\log f(\underbrace{X_1, \cdots, X_n}_{(X_1, \cdots, X_n)};\theta)\Big)^2\Big] \qquad \boxed{f(X_1, \cdots, X_n;\theta) \overset{iid}{=} \prod_{i=1}^n f(X_i;\theta)}$$

$$\Updownarrow$$

$$= E\left[\left[\frac{d}{d\theta} \left(\sum_{i=1}^n \log f(X_i;\theta)\right)\right]^2\right] \qquad \boxed{\log f(X_1, \cdots, X_n;\theta) = \sum_{i=1}^n \log f(X_i;\theta)}$$

Alt def of $I_n(\theta)$:

$$I_n(\theta) = -E\left[\frac{d^2}{d\theta^2}\log f(X_1, \cdots, X_n;\theta)\right] = -E\left[\frac{d^2}{d\theta^2}\sum_{i=1}^n \log f(X_i;\theta)\right]$$

$$= -\sum_{i=1}^n \underbrace{E\left[\frac{d^2}{d\theta^2}\log f(X_i;\theta)\right]}_{\text{same for } i} \overset{iid}{=} n\cdot I(\theta)$$

Cramer-Rao lower bound:

$\hat{\theta} = g(X_1, \cdots, X_n)$ is unbiased ; $X_1, \cdots, X_n \overset{iid}{\sim} f(\cdot;\theta)$

then $var[\hat{\theta}] \geq \frac{1}{nI(\theta)} = \frac{1}{I_n(\theta)}$

Pf: A general fact: 2 r.v.s X, Y, we have

$$cov^2(X,Y) \leq var(X) \cdot var(Y) \quad \text{— a result of Cauchy-Schwartz inequality}$$

$$\underset{\left(E\left[(X-E[X])\cdot(Y-E[Y])\right]\right)^2}{\overset{\parallel}{cov^2(X,Y)}} \qquad \left(\sum_{i=1}^n a_i b_i\right)^2 \leq \left(\sum_{i=1}^n a_i^2\right) \cdot \left(\sum_{i=1}^n b_i^2\right)$$

$$var[X] = E\left[(X-E[X])^2\right] \qquad \cancel{xy \leq x^2+y^2}$$

$$cor(X,Y) = \frac{cov(X,Y)}{\sqrt{var(X)} \cdot \sqrt{var(Y)}} \in [-1,1]$$

$$cor^2(X,Y) = \frac{cov^2(X,Y)}{var(X) \cdot var(Y)} \in (0,1]$$

Idea of this proof: use some random variable T and

$$var(\hat{\theta}) \geq \frac{cov^2(\hat{\theta}, T)}{Var(T)} \overset{\text{for some } T}{=\!=} \frac{1}{nI(\theta)}.$$

Define $T = \frac{d}{d\theta} \log f(X_1, \cdots, X_n;\theta) = \frac{d}{d\theta} \sum_{i=1}^n \log f(X_i;\theta) = \sum_{i=1}^n \frac{d}{d\theta} \log f(X_i;\theta)$

$$= \sum_{i=1}^n \frac{\frac{d}{d\theta} f(X_i;\theta)}{f(X_i;\theta)}$$

$$var(T) = E[T^2] - (E[T])^2 = I_n(\theta) - 0 = n \cdot I(\theta)$$

$$E[T] = \int \cdots \int \left(\frac{d}{d\theta} \log f(x_1, \cdots, x_n;\theta)\right) \cdot f(x_1, \cdots, x_n;\theta)\, dx_1, \cdots, dx_n$$

$$= \int \cdots \int \left(\sum_{i=1}^n \frac{\frac{d}{d\theta} f(x_i;\theta)}{f(x_i;\theta)}\right) \cdot f(x_1;\theta) \cdots f(x_n;\theta)\, dx_1, \cdots, dx_n$$

$$\vdots$$

$$= 0$$

or can show $E[T] = \sum_{i=1}^n E\left[\frac{\frac{d}{d\theta} f(X_i;\theta)}{f(X_i;\theta)}\right] \overset{iid}{=} n \cdot E\left[\frac{\frac{d}{d\theta} f(X_1;\theta)}{f(X_1;\theta)}\right] = 0$

So we have $var(\hat{\theta}) \geq \frac{cov^2(\hat{\theta}, T)}{nI(\theta)}$, and we are left to show $cov(\hat{\theta}, T) = 1$.

$\mathrm{cov}(\hat\theta, T) = E[\hat\theta T] - (E[\hat\theta])\cdot(\underbrace{E[T]}_{=0}) = E[\hat\theta T]$

$\overset{\shortparallel}{E}[(\hat\theta - E[\hat\theta])\cdot(T-E[T])]$

$E[\hat\theta T] \overset{\text{by def}}{=} \int\cdots\int g(x_1,\cdots,x_n)\left(\sum_{i=1}^{n}\frac{\frac{d}{d\theta}f(x_i;\theta)}{f(x_i;\theta)}\right)\cdot\left(\prod_{i=1}^{n}f(x_i;\theta)\right)dx_1\cdots dx_n$

$i=1 \quad \frac{\frac{d}{d\theta}f(x_1;\theta)}{f(x_1;\theta)}\cdot\prod_{i=1}^{n}f(x_i;\theta) = \left(\frac{d}{d\theta}f(x_1;\theta)\right)\cdot\prod_{i=2}^{n}f(x_i;\theta)$

So $\sum_{i=1}^{n}\left(\frac{\frac{d}{d\theta}f(x_i;\theta)}{f(x_i;\theta)}\right)\cdot\left(\prod_{j=1}^{n}f(x_j;\theta)\right) = \frac{d}{d\theta}\prod_{j=1}^{n}f(x_j;\theta)$

$= \frac{d}{d\theta}\prod_{i=1}^{n}f(x_i;\theta)$

$E[\hat\theta T] = \int\cdots\int g(x_1,\cdots,x_n)\cdot\left(\frac{d}{d\theta}\prod_{i=1}^{n}f(x_i;\theta)\right)dx_1\cdots dx_n.$

$= \frac{d}{d\theta}\int\cdots\int g(x_1,\cdots x_n)\, f(x_1,\cdots,x_n;\theta)\, dx_1\cdots dx_n$

$\overset{\text{by def}}{=} \frac{d}{d\theta}\underset{(X_1,\cdots,X_n)}{E}[\hat\theta] \overset{\hat\theta \text{ is unbiased}}{=\!=} \frac{d}{d\theta}\cdot\theta = 1$

so $\mathrm{var}(\hat\theta) \geqslant \frac{1}{nI(\theta)}$ $\square$

back to $X_1,\cdots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$

$\hat\mu_{MOM} = \hat\mu_{MLE} = \bar X_n = \frac{1}{n}\sum_{i=1}^{n}X_i \qquad E[\bar X_n] = \mu$ "unbiased"

$\mathrm{var}(\bar X_n) = \frac{\sigma^2}{n} = \frac{1}{n\cdot\frac{1}{\sigma^2}} = \frac{1}{n\cdot I(\mu)}$ achieves the C-R lower bound

$\bar X_n$ is a minimum variance unbiased estimator (MVUE)

Relative efficiency (RE): compare two unbiased estimators $\hat\theta_1, \hat\theta_2$

$RE(\hat\theta_1,\hat\theta_2) = \frac{\mathrm{var}(\hat\theta_1)}{\mathrm{var}(\hat\theta_2)}$ If $RE(\hat\theta_1,\hat\theta_2) > 1 \Leftrightarrow \mathrm{var}(\hat\theta_1) > \mathrm{var}(\hat\theta_2)$

$\Leftrightarrow \hat\theta_2$ is more efficient

Example: $X_1,\cdots,X_n \overset{iid}{\sim}$ Poisson(λ). $n>2$. $\hat\lambda_1 = \bar X_n$. $\hat\lambda_2 = \frac{X_1+X_2}{2}$