1) Finish motivation for Fisher information worksheet

$$\ell(\lambda \mid x_1, \dots, x_n) = \log\{\mathcal{L}(\lambda \mid x_1, \dots, x_n)\}$$
$$= \log\{f(x_1, \dots, x_n \mid \lambda)\}$$
$$= \log\left\{\prod_{i=1}^{n} f(x_i \mid \lambda)\right\}$$
$$= \sum_{i=1}^{n} \log\{f_{X_i}(x_i \mid \lambda)\}$$

Last thing from last class:

$\frac{d^2}{d\lambda^2}\ell(\hat{\lambda}^{MLE} \mid x_1, \dots, x_n)$ large in magnitude

$\Rightarrow \ell(\hat{\lambda}^{MLE} \mid x_1, \dots, x_n)$ very curved at $\hat{\lambda}^{MLE}$

$\Rightarrow \ell(\lambda \mid x_1, \dots, x_n)$ changes quickly as we move away from $\hat{\lambda}^{MLE}$

$\Rightarrow$ a lot of "information" that $\hat{\lambda}^{MLE}$ is better than other alternatives.

There is an additive contribution to the log-likelihood from each of the n observations.

Consider the 2nd derivative:

$$\frac{d^2}{d\lambda^2}\ell(\lambda \mid x_1, \dots, x_n) = \cdots = \sum_{i=1}^{n} \frac{d^2}{d\lambda^2}\log\{f_{X_i}(x_i \mid \lambda)\}$$

Again, for each observation the information about λ grows.
(second derivative of log-likelihood larger in magnitude)

Subset 1 had a sample size of 56, more information about λ
Subset 2 had a sample size of 4, less information about λ.

<u>Note:</u> $\frac{d^2}{d\lambda^2}\ell(\hat{\lambda}^{MLE} \mid x_1, \dots, x_n) < 0$ (we're at a maximum of ℓ).

$-\frac{d^2}{d\lambda^2}\ell(\hat{\lambda}^{MLE} \mid x_1, \dots, x_n) > 0$ has a more intuitively meaningful sign.
(larger value $\Leftrightarrow$ more information about λ).

<u>Def.</u>: The <u>Observed Fisher Information</u> about a parameter θ is

$$J(\theta^*) = \frac{d^2}{d\theta^2}\ell(\theta \mid x_1, \dots, x_n)\Big|_{\theta=\theta^*}$$

<u>Note</u>: Usually evaluated at the MLE: $J(\hat{\theta}^{MLE}) = \frac{d^2}{d\theta^2}\ell(\theta \mid x_1, \dots, x_n)\Big|_{\theta=\hat{\theta}^{MLE}}$

The observed fisher information from one observation $X_i = x_i$ is

$$J_i(\theta^*) = \frac{d^2}{d\theta^2} \ell(\theta \mid x_i)\Big|_{\theta=\theta^*} = \frac{d^2}{d\theta^2} f_{X_i}(x_i \mid \theta)\Big|_{\theta=\theta^*}$$

The Fisher information is the expected value of the observed Fisher information.

$$I(\theta^*) = -E\left[\frac{d^2}{d\theta^2} \ell(\theta \mid X_1, \ldots, X_n)\Big|_{\theta=\theta^*}\right]$$

"On average, across all samples of size n, what is the curvature of the log-likelihood function at the parameter value θ^*?"

The Fisher information from one observation is

$$I_i(\theta^*) = -E\left[\frac{d^2}{d\theta^2} \ell(\theta \mid X_i)\Big|_{\theta=\theta^*}\right]$$

our book doesn't explicitly define Fisher information, but if it did, it would define it as the Fisher information from one observation. (p. 276)

Note: $$I(\theta^*) = -E\left[\frac{d^2}{d\theta^2} \ell(\theta \mid X_1, \ldots, X_n)\Big|_{\theta=\theta^*}\right]$$

$$= -E\left[\frac{d^2}{d\theta^2} \log\left\{ f_{X_1,\ldots,X_n}(X_1, \ldots, X_n \mid \theta)\Big|_{\theta=\theta^*}\right\}\right]$$

$$= -E\left[\frac{d^2}{d\theta^2} \sum_{i=1}^{n} \log\{f_{X_i}(X_i \mid \theta)\}\right]$$

$$= \sum_{i=1}^{n} -E\left[\frac{d^2}{d\theta^2} \log\{f_{X_i}(X_i \mid \theta)\}\Big|_{\theta=\theta^*}\right]$$

$$= \sum_{i=1}^{n} I_i(\theta^*)$$

$$= n \cdot I_i(\theta^*)$$

Theorem: If some conditions are satisfied (f is differentiable, the true parameter θ^* is not on the boundary of the parameter space...)

$$\sqrt{n I_1(\theta^*)}\,(\hat{\theta}^{MLE} - \theta^*) \longrightarrow \text{Normal}(0,1) \text{ in distribution as } n \to \infty$$

Intuitive statement: If n is large, it is approximately true that

$$\sqrt{n I_1(\theta^*)}\,(\hat{\theta}^{MLE} - \theta^*) \sim \text{Normal}(0,1)$$

... or ...

If n is large, it is approximately true that

$$\hat{\theta}^{MLE} \sim \text{Normal}\left(\theta^*, \frac{1}{n I_1(\theta^*)}\right)$$

... or ...

$$\hat{\theta}^{MLE} \sim \text{Normal}\left(\theta^*, \frac{1}{n \cdot -E\left[\frac{d^2}{d\theta^2}\ell(\theta|x_i)\big|_{\theta=\theta^*}\right]}\right)$$

... or ...

$$\hat{\theta}^{MLE} \sim \text{Normal}\left(\theta^*, \frac{1}{-E\left[\frac{d^2}{d\theta^2}\ell(\theta|x_1,\dots,x_n)\big|_{\theta=\theta^*}\right]}\right)$$

Fisher information, at the true θ^*

This result can also be stated in terms of the observed Fisher information at $\hat{\theta}^{MLE}$

Approximately, for large n,

$$\hat{\theta}^{MLE} \sim \text{Normal}\left(\theta^*, \frac{1}{-\frac{d^2}{d\theta^2}\ell(\theta|x_1,\dots,x_n)\big|_{\theta=\hat{\theta}^{MLE}}}\right)$$

For Poisson example:

$X_i \overset{iid}{\sim} \text{Poisson}(\lambda)$, $i=1,\ldots,n$, $E[X_i]=\lambda$, MLE is $\hat{\lambda}^{MLE} = \overline{X}$

$$f_{X_i}(x_i|\lambda) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}$$

$$\log\{f_{X_i}(x_i|\lambda)\} = x_i \log(\lambda) - \lambda - \log(x_i!)$$

$$\frac{d}{d\lambda}\log\{f_{X_i}(x_i|\lambda)\} = \frac{x_i}{\lambda} - 1$$

$$\frac{d^2}{d\lambda^2}\log\{f_{X_i}(x_i|\lambda)\} \quad \frac{-x_i}{\lambda^2}$$

Observed Fisher Information:

$$J(\lambda) = +\sum_{i=1}^{n} \frac{+x_i}{\lambda^2}$$

Evaluated at $\lambda = \hat{\lambda}^{MLE}$: $J(\hat{\lambda}^{MLE}) = \frac{1}{\overline{X}^2}\cdot\sum_{i=1}^{n} X_i$

$$= \frac{1}{\overline{X}}\cdot\frac{1}{\frac{1}{n}\sum_{i=1}^{n} X_i}\cdot\cancel{\sum_{i=1}^{n} X_i}$$

$$= \frac{n}{\overline{X}}$$

Fisher Information:

$$I(\lambda) = E\left[\sum_{i=1}^{n}\frac{+X_i}{\lambda^2}\right]$$

$$= \sum_{i=1}^{n} \frac{+1}{\lambda^2} E[X_i]$$

$$= \sum_{i=1}^{n} \frac{1}{\lambda^2}\cdot\lambda$$

$$= \frac{n}{\lambda}$$

Evaluated at $\lambda = \hat{\lambda}^{MLE}$:

$$I(\lambda) = \frac{n}{\overline{X}}$$

An approximate $(1-\alpha)*100\%$ CI is $\overline{X} \pm Z(\frac{\alpha}{2})\sqrt{\frac{\overline{X}}{n}}$