Motivation for Fisher Information, Continued

Seedlings (Poisson Model)

Ecologists divided a region of the forest floor into n quadrats and counted the number of seedlings that sprouted in each quadrat as part of a study on climate change.

- Observe $X_1, \ldots, X_n$; X_i is the number of seedlings in quadrat number i.
- Data Model: $X_i | \Lambda = \lambda \overset{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$
- We have seen that the maximum likelihood estimate is $\hat{\lambda}^{MLE} = \frac{1}{n} \sum_{i=1}^n X_i$

Connection between Observed Fisher Information and Taylor series approximation to log-likelihood

- Second order Taylor approximation to the log-likelihood at the point λ^*:

$$\begin{aligned}
\ell(\lambda|x_1,\ldots,x_n) &\approx \ell(\lambda^*|x_1,\ldots,x_n) + \frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^*}(\lambda-\lambda^*) + \frac{1}{2}\frac{d^2}{d\lambda^2}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^*}(\lambda-\lambda^*)^2 \\
&= \ell(\lambda^*|x_1,\ldots,x_n) + \frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^*}(\lambda-\lambda^*) - \frac{1}{2}\left\{-\frac{d^2}{d\lambda^2}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^*}\right\}(\lambda-\lambda^*)^2 \\
&= \ell(\lambda^*|x_1,\ldots,x_n) + \frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^*}(\lambda-\lambda^*) - \frac{1}{2}\left\{J(\lambda^*)\right\}(\lambda-\lambda^*)^2
\end{aligned}$$

If we approximate at the maximum likelihood estimate then the second term goes away:

- At the MLE, the derivative of the log-likelihood is 0: $\frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^{MLE}} = 0$

$$\begin{aligned}
\ell(\lambda|x_1,\ldots,x_n) &\approx \ell(\lambda^{MLE}|x_1,\ldots,x_n) + \frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n)|_{\lambda=\lambda^{MLE}}(\lambda-\lambda^{MLE}) - \frac{1}{2}J(\lambda^{MLE})(\lambda-\lambda^{MLE})^2 \\
&\approx \ell(\lambda^{MLE}|x_1,\ldots,x_n) + 0(\lambda-\lambda^{MLE}) - \frac{1}{2}J(\lambda^{MLE})(\lambda-\lambda^{MLE})^2 \\
&\approx \ell(\lambda^{MLE}|x_1,\ldots,x_n) - \frac{1}{2}J(\lambda^{MLE})(\lambda-\lambda^{MLE})^2
\end{aligned}$$

Results from 2 different samples

- For both subsets I have chosen, $\hat{\lambda}^{MLE}$ is the same:

```
mean(seedlings$new_1993[subset_1_inds])
```

```
## [1] 0.75
```

```
mean(seedlings$new_1993[subset_2_inds])
```

```
## [1] 0.75
```

- The number of observations is different:
  - the first has $n = 56$ observations
  - the second has $n = 4$ observations
- The observed Fisher information is therefore different:
  - the first has observed Fisher information $J(\theta^*) = \frac{n}{\bar{x}} = \frac{56}{0.75} = 74.667$
  - the second has observed Fisher information $J(\theta^*) = \frac{n}{\bar{x}} = \frac{4}{0.75} = 5.333$
- Taylor series approximations about the maximum likelihood estimate $\hat{\lambda}^{MLE}$:
  - $\ell(\lambda|x_1, \dots, x_n) \approx \ell(0.75|x_1, \dots, x_{56}) - \frac{1}{2}74.667(\lambda - 0.75)^2$
  - $\ell(\lambda|x_1, \dots, x_n) \approx \ell(0.75|x_1, \dots, x_4) - \frac{1}{2}5.333(\lambda - 0.75)^2$
- Curvature of log-likelihood is greater with the sample size of 56 than with the sample size of 4.
- Here are the likelihood and log-likelihood functions based on the two different subsets, with orange lines at the MLE: