Motivation for Fisher Information

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 $\hat{\lambda}^{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Results from 2 different samples

- Our full sample had $n = 60$ quadrats. I have selected two different subsets of these observations.
- For both subsets I have chosen, $\hat{\lambda}^{MLE}$ is the same:

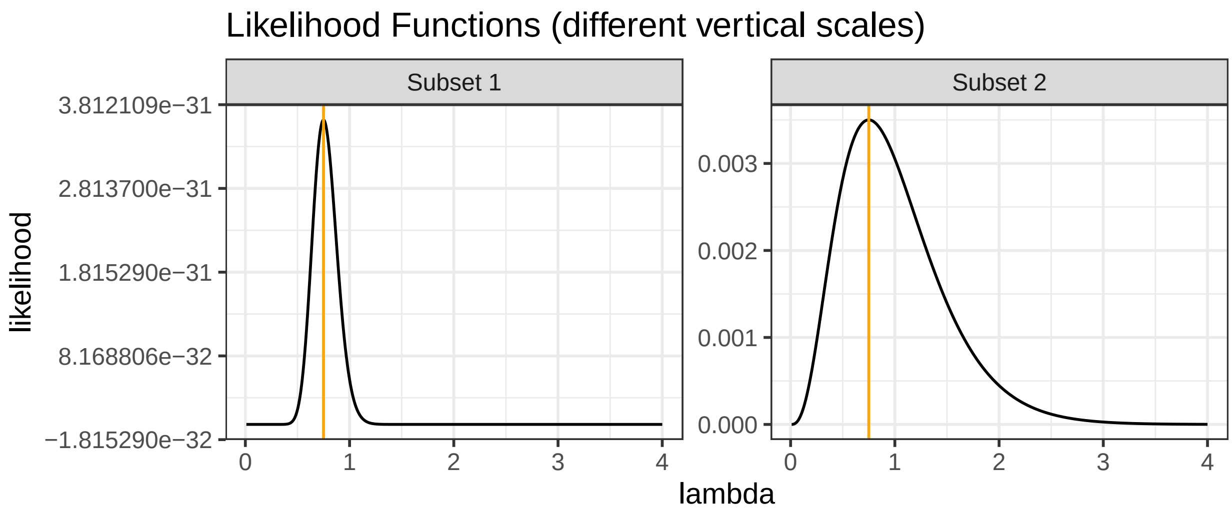
```
mean(seedlings$new_1993[subset_1_inds])
```

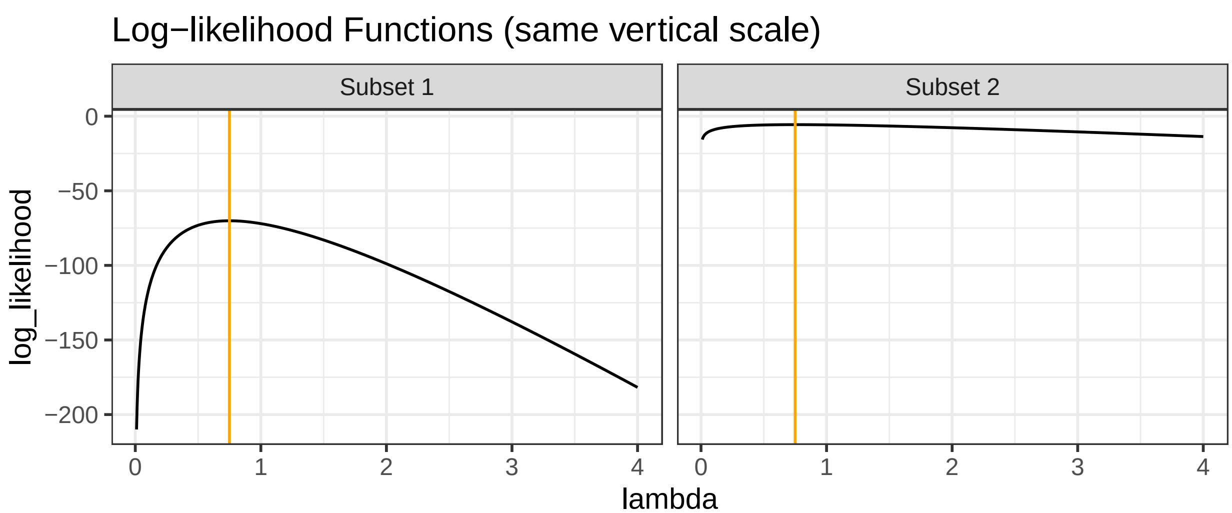
```
## [1] 0.75
```

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

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

- Here are the likelihood and log-likelihood functions based on the two different subsets, with orange lines at the MLE:





Some questions to consider:

1. Likelihood Ratios

The likelihood function measures the probability of the observed data, as a function of λ:

$$\mathcal{L}(\lambda|x_1, \ldots, x_n) = \text{Probability of observed data, if the parameter is } \lambda$$

Let's consider the *ratio* of the likelihood function at the MLE to the likelihood function at $\lambda = 2$.

For subset 1, this ratio is $\frac{\mathcal{L}(0.75|x_1,\ldots,x_{n_1})}{\mathcal{L}(2|x_1,\ldots,x_{n_1})} \approx \frac{3.630748e-31}{1.122197e-43} \approx 3.24 \times 10^{12}$

For subset 2, this ratio is $\frac{\mathcal{L}(0.75|x_1,\ldots,x_{n_2})}{\mathcal{L}(2|x_1,\ldots,x_{n_2})} \approx \frac{0.0035}{0.0004} \approx 7.826$

Based on a comparison of likelihood ratios, which data set provides more evidence that $\lambda = 0.75$ is better than $\lambda = 2$? Why?

2. Differences in Log-likelihoods

Taking the log of the likelihood ratio and using properties of logs, we have:

$$\begin{aligned} \log\left\{\frac{\mathcal{L}(\hat{\lambda}^{MLE}|x_1,\ldots,x_n)}{\mathcal{L}(\lambda^{alternative}|x_1,\ldots,x_n)}\right\} &= \log\left\{\mathcal{L}(\hat{\lambda}^{MLE}|x_1,\ldots,x_n)\right\} - \log\left\{\mathcal{L}(\lambda^{alternative}|x_1,\ldots,x_n)\right\} \\ &= \ell(\hat{\lambda}^{MLE}|x_1,\ldots,x_n) - \ell(\lambda^{alternative}|x_1,\ldots,x_n) \end{aligned}$$

So a comparison similar to the above can be done by looking at *differences* in the value of the log-likelihood function.

For subset 1, this difference is $\ell(0.75|x_1,\ldots,x_{n_1}) - \ell(2|x_1,\ldots,x_{n_1}) \approx -70.09 - (-98.90) = 28.81$

For subset 2, this difference is $\ell(0.75|x_1,\ldots,x_{n_2}) - \ell(2|x_1,\ldots,x_{n_2}) \approx -5.65 - (-7.71) = 2.06$

Based on a comparison of differences in log-likelihoods, which data set provides more evidence that $\lambda = 0.75$ is better than $\lambda = 2$? Why?

3. Looking at the plots of the log-likelihood functions, which subset provides more information about the value of λ, in the sense that it restricts the range of feasible values for λ to a smaller set? Why?

(Can you come up with a graphical rule based on the plots of the log-likelihood functions for determining which data set gives more information about λ?)

4. Looking at the plots of the log-likelihood functions, can you come up with a quantitative summary of the log-likelihood function that captures which data set gives more information about λ?

5. One of these subsets had a sample size of 4, and the other had a sample size of 56. Which is which?