Motivation for Fisher Information, Continued Again

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

- Consider just the subset with 56 observations. The MLE is:

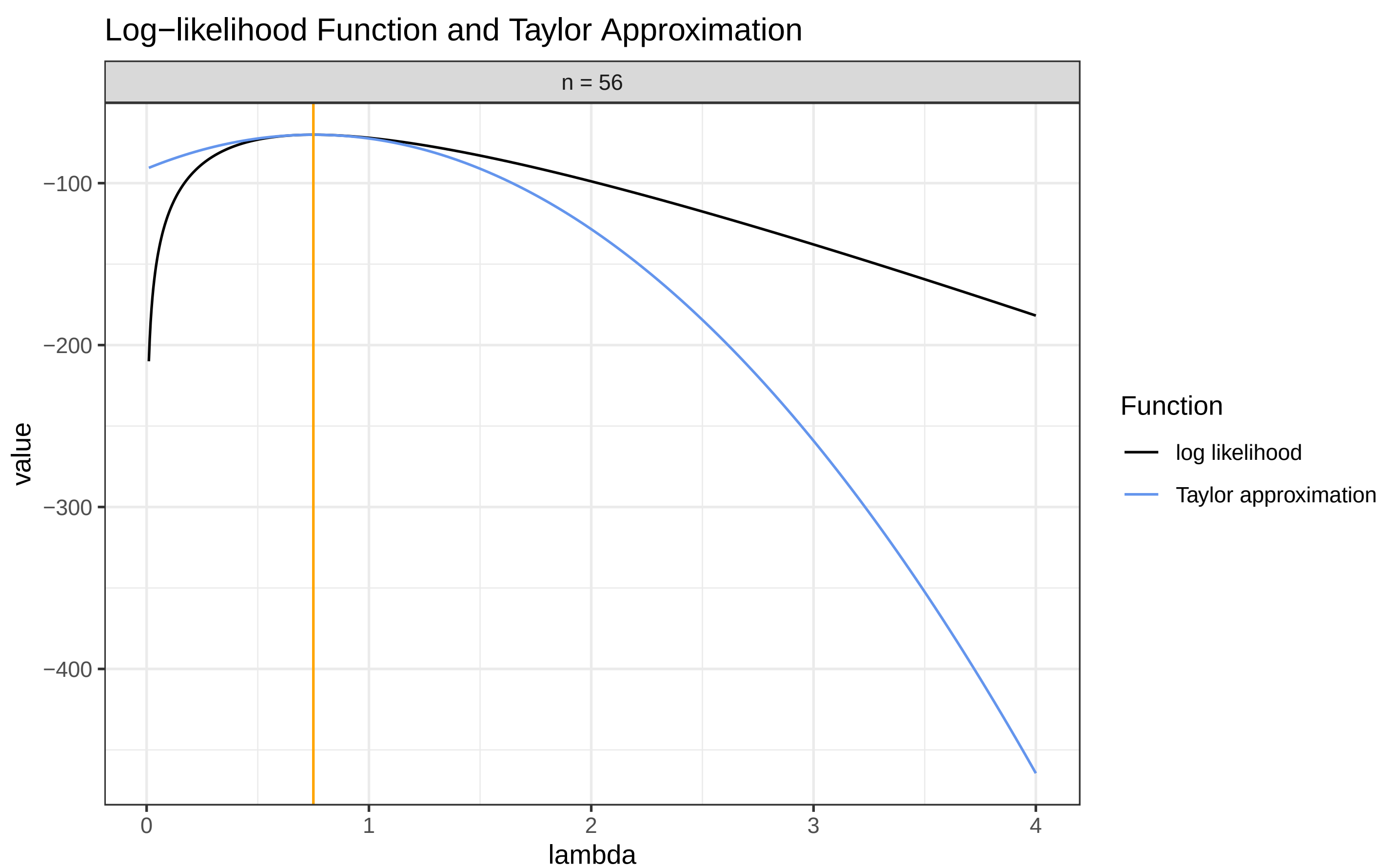
```
seedlings$new_1993[subset_inds]
```

```
##  [1] 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 0 1 0 3 3 0 1 4 2
## [36] 2 1 0 1 0 3 1 0 0 0 1 1 1 0 1 3 1 2 3 0 1
```

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

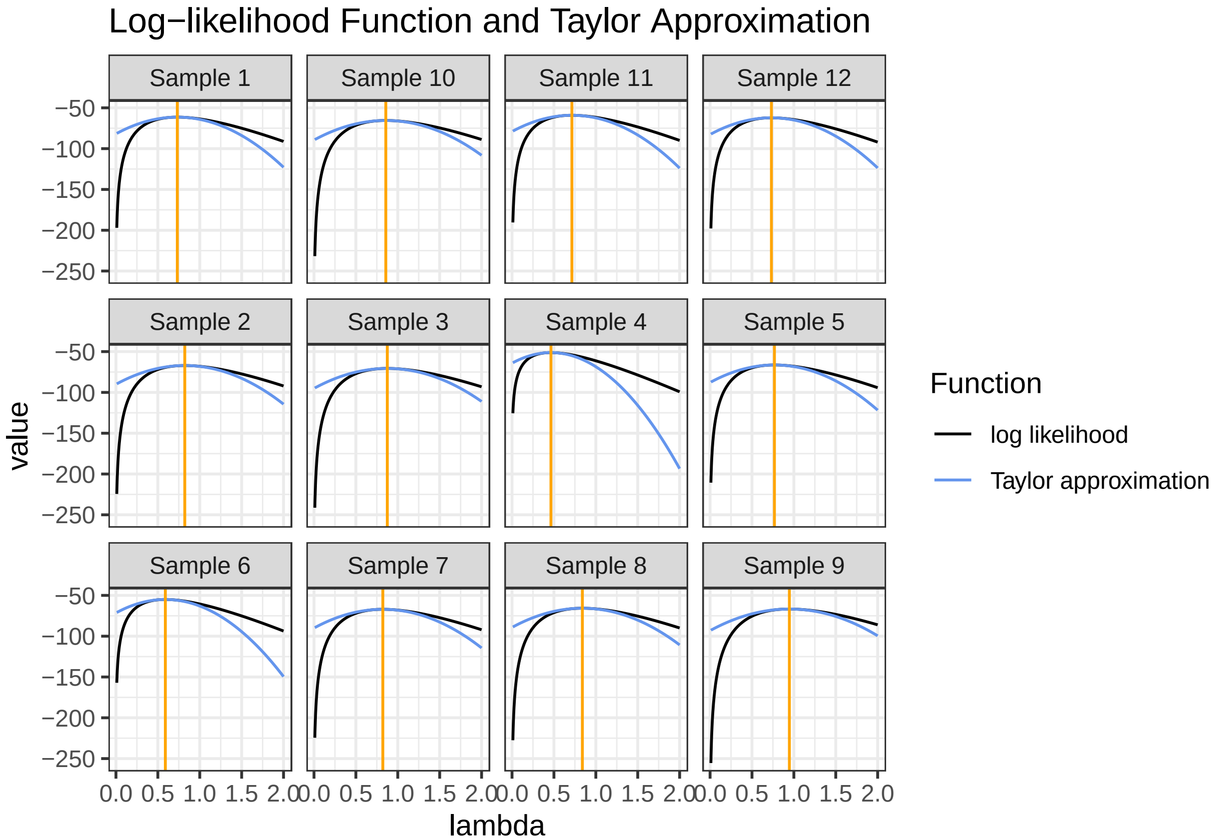
```
## [1] 0.75
```

- There are $n = 56$ observations
- The observed Fisher information is $J(\theta^*) = \frac{n}{\bar{x}} = \frac{56}{0.75} = 74.667$.
  - This is the negative second derivative of the log likelihood function.
- The Taylor series approximation about the maximum likelihood estimate $\hat{\lambda}^{MLE}$ is:
  - $\ell(\lambda | x_1, \ldots, x_n) \approx \ell(0.75 | x_1, \ldots, x_{56}) - \frac{1}{2} 74.667 (\lambda - 0.75)^2$
- Here is the log-likelihood function with an orange line at the MLE and the Taylor approximation in blue:



What if we took other samples?

- Suppose for the sake of the example that the true parameter value is $\lambda = 0.75$.
- Each random sample of size $n = 56$ has:
  - Different observed values
  - A different log-likelihood function
  - A different second derivative of the log-likelihood function at the maximum
  - A different observed Fisher information



Suppose now I simulate 1000 different samples and calculate the observed Fisher information from each:

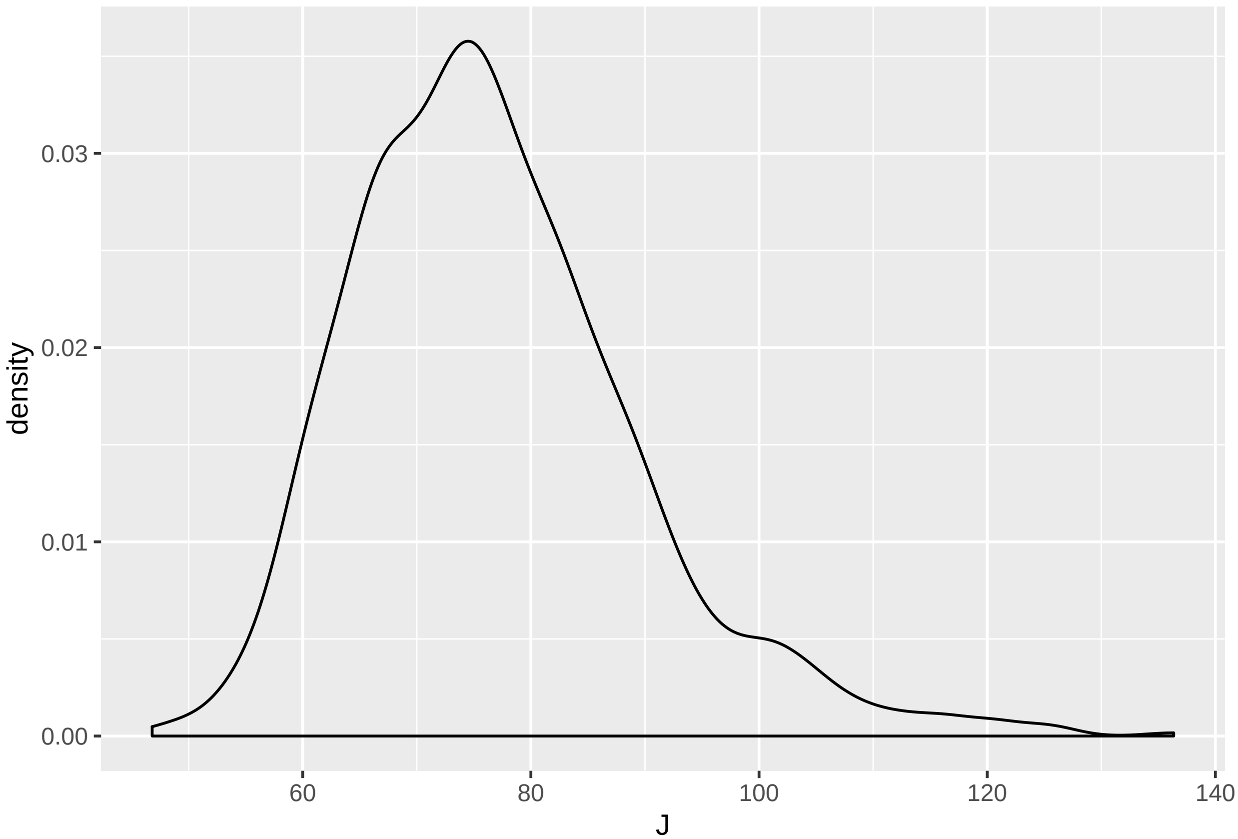
```
Fisher_informations <- data.frame(
  J = rep(NA, 1000)
)

for(i in 1:1000) {
  x <- rpois(n = 56, lambda = 0.75)
  Fisher_informations$J[i] <- 56 / mean(x)
}
head(Fisher_informations$J)
```

```
## [1]  65.33333  71.27273  78.40000 104.53333  69.68889  82.52632
```

```
length(Fisher_informations$J)
```

```
## [1] 1000
```

```
ggplot(data = Fisher_informations, mapping = aes(x = J)) +
  geom_density()
```



The Fisher information is the expected value (average) of the observed Fisher information across different samples:

```
mean(Fisher_informations$J)
```

```
## [1] 76.76499
```