Newton Raphson for Optimization

Evan L. Ray

February 3, 2020

Poisson Model Example

Example A in Section 8.4 of Rice

The National Institute of Science and Technology did a study where they wanted to develop measurement standards for asbestos concentration. Asbestos dissolved in water was spread on a filter, and 3-mm diameter punches were taken from the filter and mounted on a transmission electron microscope. An operator counted the number of fibers in each of 23 grid squares. For the sake of illustration I am using just the first 5 observations here.

Let X_i be the number of fibers of asbestos found in square number i.

Model: $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$

Log-likelihood and its derivatives

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

$$\ell(\lambda|x_1,\ldots,x_n) = \cdots = -n\lambda + \log(\lambda)\sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\log(x_i)$$

$$\frac{d}{d\lambda}\ell(\lambda|x_1,\ldots,x_n) = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n$$

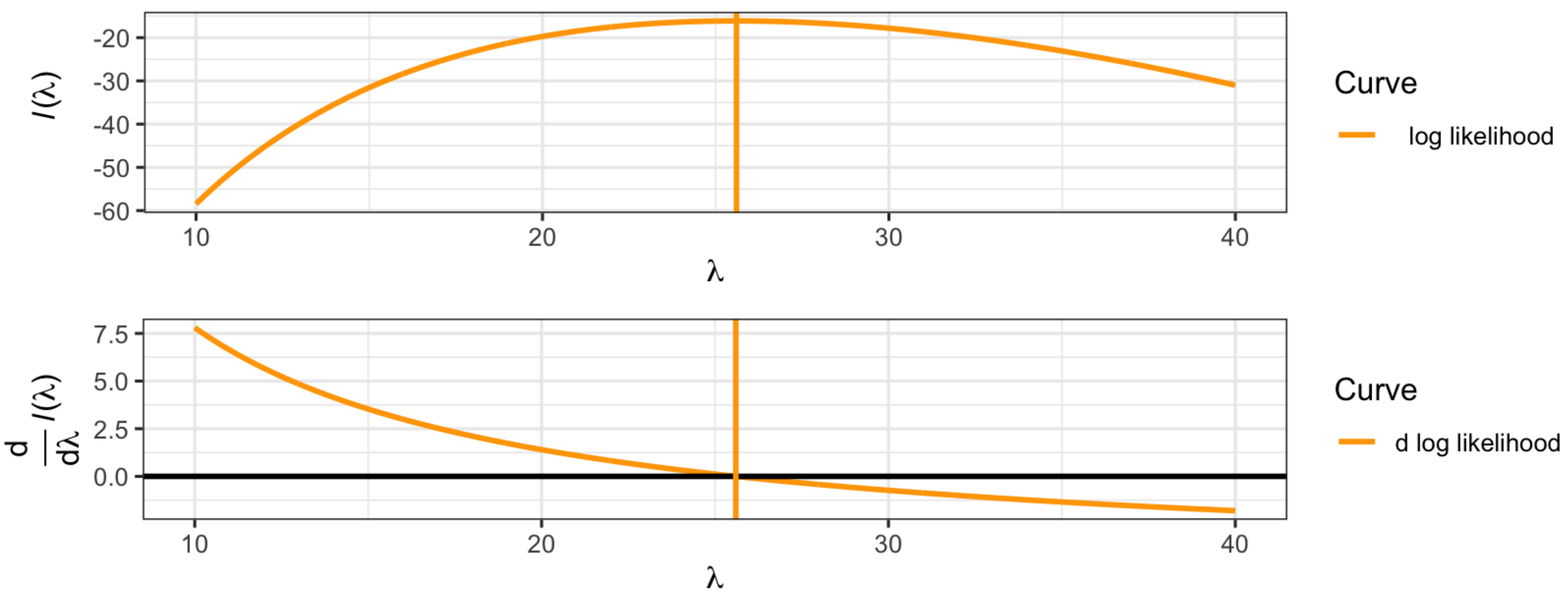
$$\frac{d^2}{d\lambda^2}\ell(\lambda|x_1,\ldots,x_n) = \frac{-1}{\lambda^2}\sum_{i=1}^{n} x_i$$

We saw previously that $\hat{\lambda}_{MLE} = \frac{1}{n}\sum_{i=1}^{n} x_i$

In our example, $\hat{\lambda}_{MLE} = 25.6$

(This means we really don't need numerical maximization methods for this example – I'm using this example so that we can check that the estimation worked by comparing to what we know to be the correct answer.)

Log-likelihood function, MLE



The vertical orange line is at the MLE:

- maximizes log-likelihood function
- at a root (0) of the first derivative of the log-likelihood function

…But what if we couldn't solve for the MLE directly?

Maximize Taylor Approx. to $\ell(\lambda)$

Pick a value λ_0. The second-order Taylor Series approximation to $\ell(\lambda|x_1, \dots, x_n)$ around λ_0 is

$$P_2(\lambda) = \ell(\lambda_0|x_1, \dots, x_n) + \frac{d}{d\lambda}\ell(\lambda_0|x_1, \dots, x_n)(\lambda - \lambda_0) + \frac{1}{2}\frac{d^2}{d\lambda^2}\ell(\lambda_0|x_1, \dots, x_n)(\lambda - \lambda_0)^2$$

If $\frac{d^2}{d\lambda^2}\ell(\lambda_0|x_1, \dots, x_n) < 0$, P_2 is maximized when its first derivative is 0:

$$0 = \frac{d}{d\lambda}P_2(\lambda) = \frac{d}{d\lambda}\ell(\lambda_0|x_1, \dots, x_n) + \frac{d^2}{d\lambda^2}\ell(\lambda_0|x_1, \dots, x_n)(\lambda - \lambda_0)$$

$$\Rightarrow \lambda = \lambda_0 - \frac{\frac{d}{d\lambda}\ell(\lambda_0|x_1,\dots,x_n)}{\frac{d^2}{d\lambda^2}\ell(\lambda_0|x_1,\dots,x_n)}$$

This will be our updated estimate, λ_1

Now repeat, but centering the Taylor Series approximation at λ_1.

Find Root of Taylor Approx. to $\frac{d}{d\lambda}\ell(\lambda)$

Pick a value λ_0. The first-order Taylor Series approximation to $\frac{d}{d\lambda}\ell(\lambda|x_1, \dots, x_n)$ around λ_0 is

$$P_1(\lambda) = \frac{d}{d\lambda}L(\lambda_0|x_1, \dots, x_n) + \frac{d^2}{d\lambda^2}\ell(\lambda_0|x_1, \dots, x_n)(\lambda - \lambda_0)$$

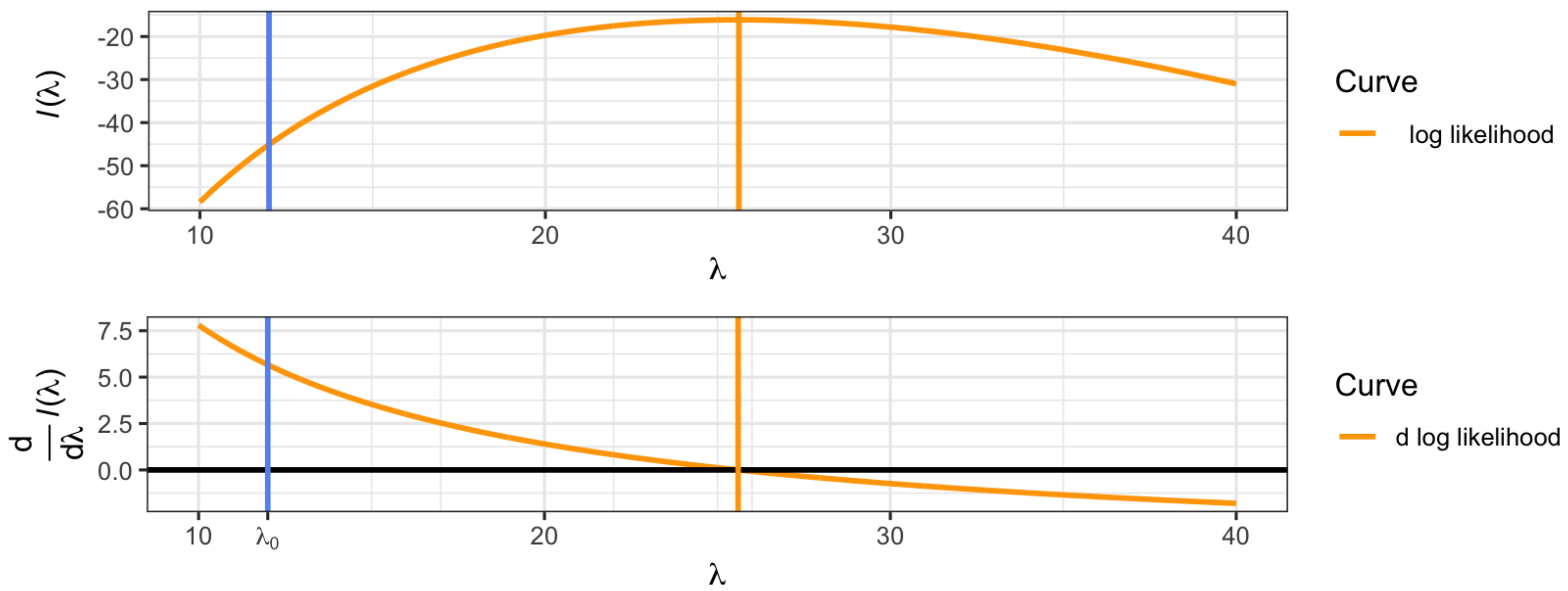
The root of $P_1(\lambda)$ is at $\lambda = \lambda_0 - \frac{\frac{d}{d\lambda}L(\lambda_0|x_1,\dots,x_n)}{\frac{d^2}{d\lambda^2}L(\lambda_0|x_1,\dots,x_n)}$

This will be our updated estimate, λ_1

Now repeat, but centering the Taylor Series approximation at λ_1.

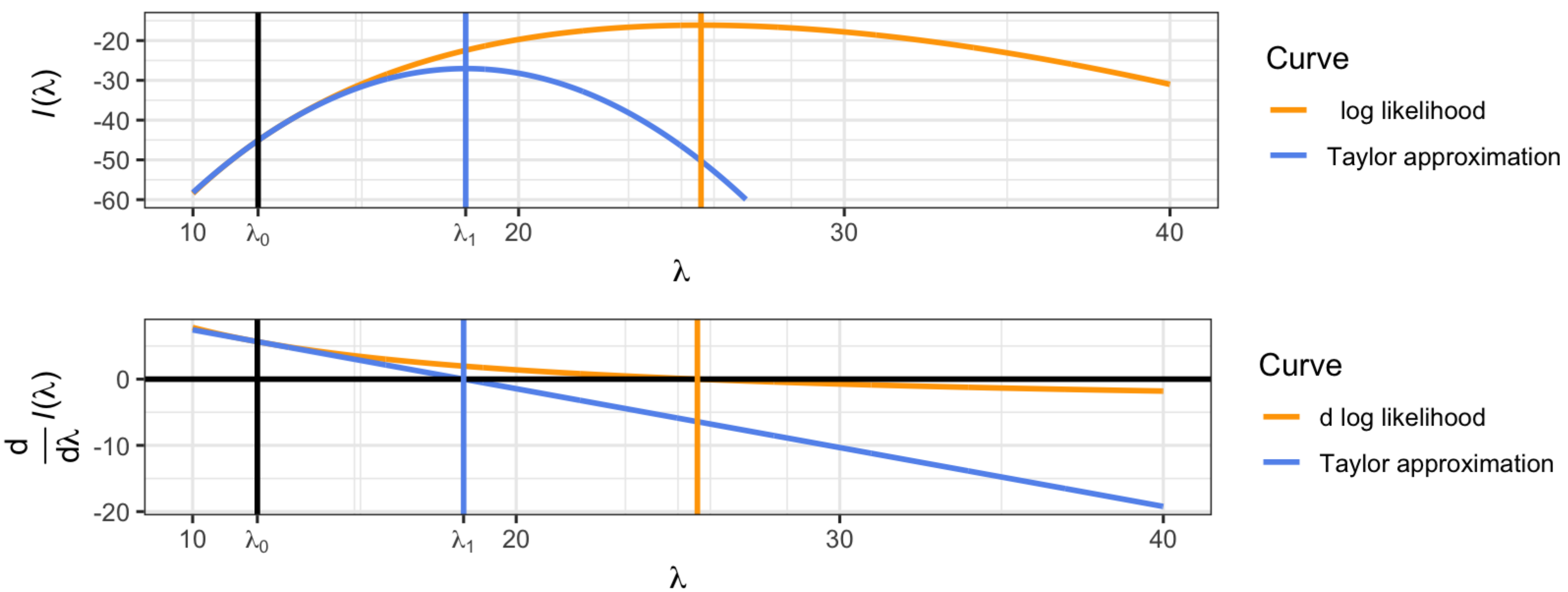
Pick λ_0

- Often the initial value of λ_0 is selected by the method of moments
- For this example, I picked a number arbitrarily



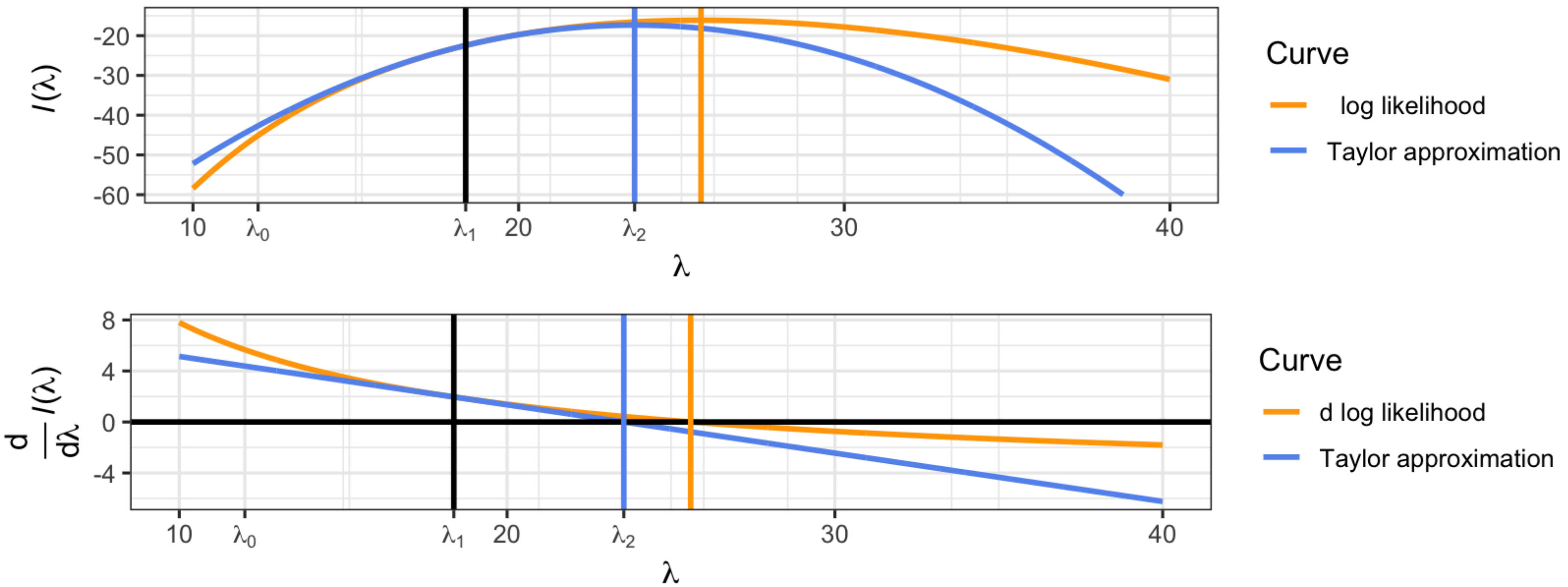
$$\lambda_0 = 12$$

Approximate $\ell(\lambda)$ around λ_0, get λ_1



$$\lambda_0 = 12, \lambda_1 = \lambda_0 - \frac{\frac{d}{d\lambda} L(\lambda_0 | x_1, \ldots, x_n)}{\frac{d^2}{d\lambda^2} L(\lambda_0 | x_1, \ldots, x_n)} = 18.375$$

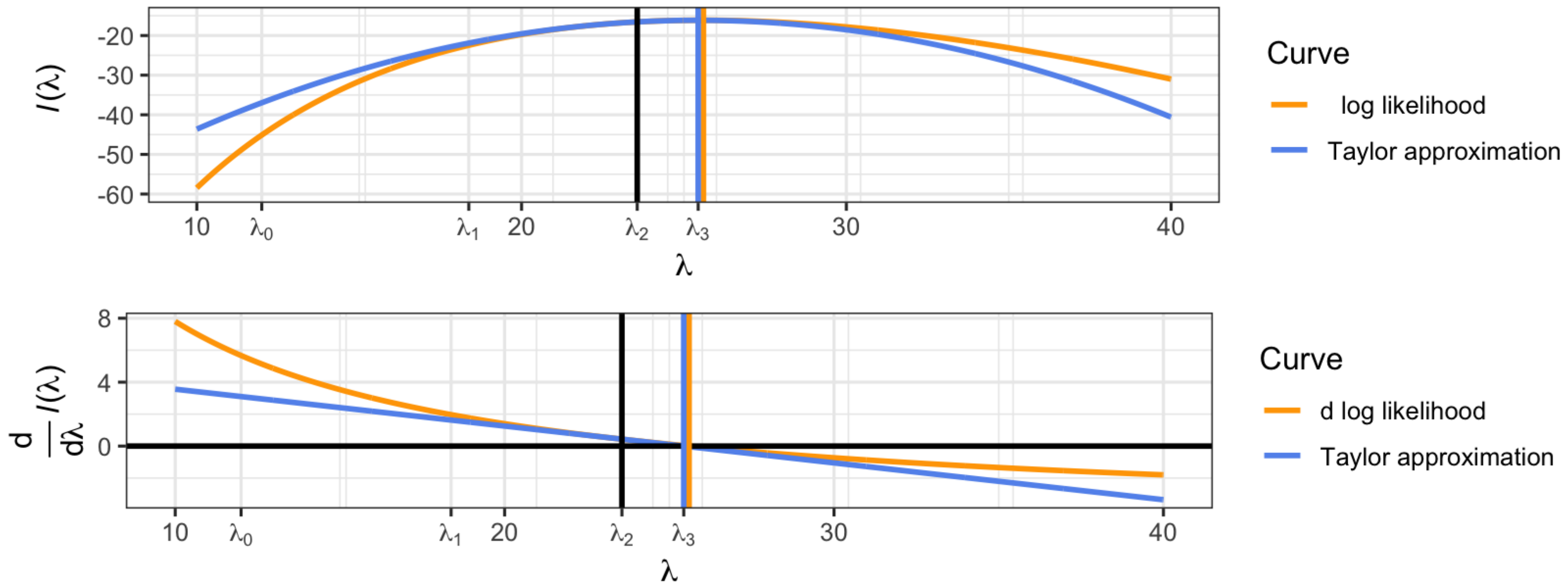
Approximate $\ell(\lambda)$ around λ_1, get λ_2



$$\lambda_0 = 12, \ \lambda_1 = 18.375,$$

$$\lambda_2 = \lambda_1 - \frac{\frac{d}{d\lambda} L(\lambda_1 | x_1, \ldots, x_n)}{\frac{d^2}{d\lambda^2} L(\lambda_1 | x_1, \ldots, x_n)} = 23.561$$

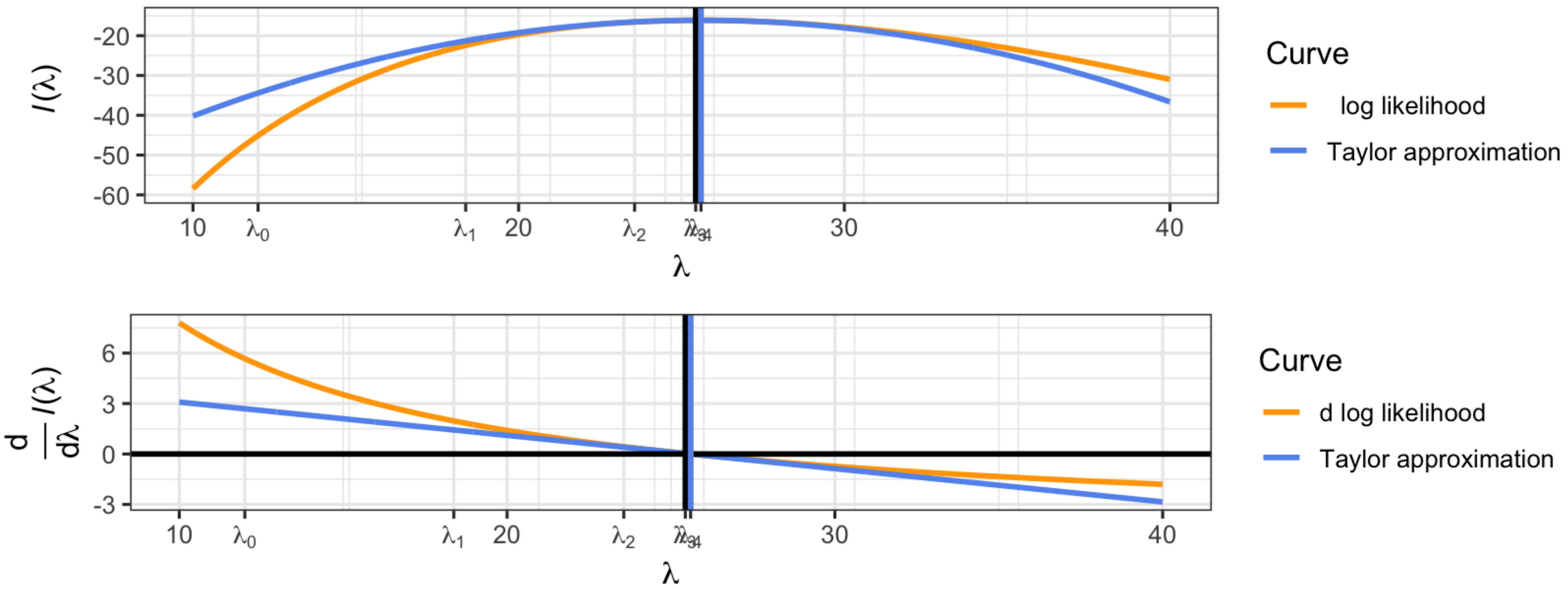
Approximate $\ell(\lambda)$ around λ_2, get λ_3



$$\lambda_0 = 12,\ \lambda_1 = 18.375, \lambda_2 = 23.561$$

$$\lambda_3 = \lambda_2 - \frac{\frac{d}{d\lambda}L(\lambda_2 | x_1, \ldots, x_n)}{\frac{d^2}{d\lambda^2}L(\lambda_2 | x_1, \ldots, x_n)} = 25.438$$

Approximate $\ell(\lambda)$ around λ_3, get λ_4



$$\lambda_0 = 12,\ \lambda_1 = 18.375, \lambda_2 = 23.561, \lambda_3 = 25.438$$

$$\lambda_4 = \lambda_3 - \frac{\frac{d}{d\lambda} L(\lambda_3 | x_1, \ldots, x_n)}{\frac{d^2}{d\lambda^2} L(\lambda_3 | x_1, \ldots, x_n)} = 25.599$$