Regression and Simulation Methods

Week 4: Likelihood

Chris Oldnall, 31st October 2023

How is likelihood the next step in our journey?

- So far we have seen that we can fit a ‘normal linear model’ and estimate the parameters of the model through the least squares procedure.
- Well maybe the underlying distribution of data isn’t normal?
- That is ok! That is why GLMs will come in useful.
- However how do we know we are estimating the parameters properly…the answer: likelihood.

What is likelihood?

- The probability of observing the data with the parameters in the system set to the vector theta.

$$L(\underline{\theta};\underline{x}) = \prod_{i=1}^{n} p(x_i;\underline{\theta}) \iff l(\underline{\theta};\underline{x}) = \log\left(\prod_{i=1}^{n} p(x_i;\underline{\theta})\right)$$

- The maximum likelihood estimator is then the value of theta which maximises the likelihood function.

Maximum Likelihood Estimation

- This is the procedure in which we can find the value of the parameter of interest which maximises the likelihood function.
- The value which maximises the likelihood will also maximise the log-likehood which turns out to be easier to use.
- We achieve this value by setting the first derivative of the log-likelihood to 0 - this equation is the ‘score’ function, U.
- However we must check it is a maximum by seeing that the second derivative is < 0 for all values of the parameter.

What is the sample information?

$$k(\underline{x}) = -\, l''(\hat{\theta})$$

- This tells us the information about the maximum point.

Theoretical Question 1...

Cauchy distribution

Consider a random sample $y_1, y_2, \ldots, y_n$ from a Cauchy distribution with probability density function

$$f(y;\theta) \;=\; \frac{1}{\pi\{1+(y-\theta)^2\}} \qquad (-\infty < y < \infty),$$

where θ is an unknown parameter. θ is a location parameter.

ML equation

Find an equation satisfied by the maximum likelihood estimator of θ (i.e. $U(\widehat{\theta}) = 0$).

Cauchy distribution

Likelihood:

$$L(\theta) = \prod_{i=1}^{n} f(y_i; \theta) = \prod_{i=1}^{n} \frac{1}{\pi\{1 + (y_i - \theta)^2\}}$$

Log likelihood:

$$l(\theta) = -\sum_{i=1}^{n} \log[\pi\{1 + (y_i - \theta)^2\}] = -\sum_{i=1}^{n} \log\{1 + (y_i - \theta)^2\} + \text{constant}$$

ML equation

The score function is:

$$U(\theta) = \sum_{i=1}^{n} \frac{2(y_i - \theta)}{1 + (y_i - \theta)^2}$$

Therefore an equation satisfied by the maximum likelihood estimator of θ is

$$\sum_{i=1}^{n} \frac{2(y_i - \theta)}{1 + (y_i - \theta)^2} = 0$$

Numerical MLE...

Sometimes we sadly cannot find a closed form solution of the maximum…

…but Newton’s method says we can always approximate anything by a quadratic.

The procedure explored...

1 Start with an initial parameter guess $\hat{\theta}_0$ and set index $k = 0$.

2 Approximate $l(\theta)$ with a quadratic, by making a Taylor expansion around $\hat{\theta}_k$.

3 Find $\hat{\theta}_{k+1}$ to maximize this quadratic.

4 If $\partial l / \partial \theta |_{\hat{\theta}_{k+1}} \approx 0$ then stop, returning $\hat{\theta}_{k+1}$ as the m.l.e. Otherwise increase k by one and return to step 2.

The procedure explored...

In the **single parameter** case step 2 is as follows. First write $\theta = \hat{\theta}_k + \Delta$ and then

$$l(\theta) \simeq l(\hat{\theta}_k) + \Delta \left.\frac{\partial l}{\partial \theta}\right|_{\hat{\theta}_k} + \frac{1}{2}\Delta^2 \left.\frac{\partial^2 l}{\partial \theta^2}\right|_{\hat{\theta}_k}.$$

Differentiating w.r.t. Δ gives

$$\frac{\partial l}{\partial \Delta} \simeq \left.\frac{\partial l}{\partial \theta}\right|_{\hat{\theta}_k} + \Delta \left.\frac{\partial^2 l}{\partial \theta^2}\right|_{\hat{\theta}_k}$$

while setting the differential to zero and solving yields

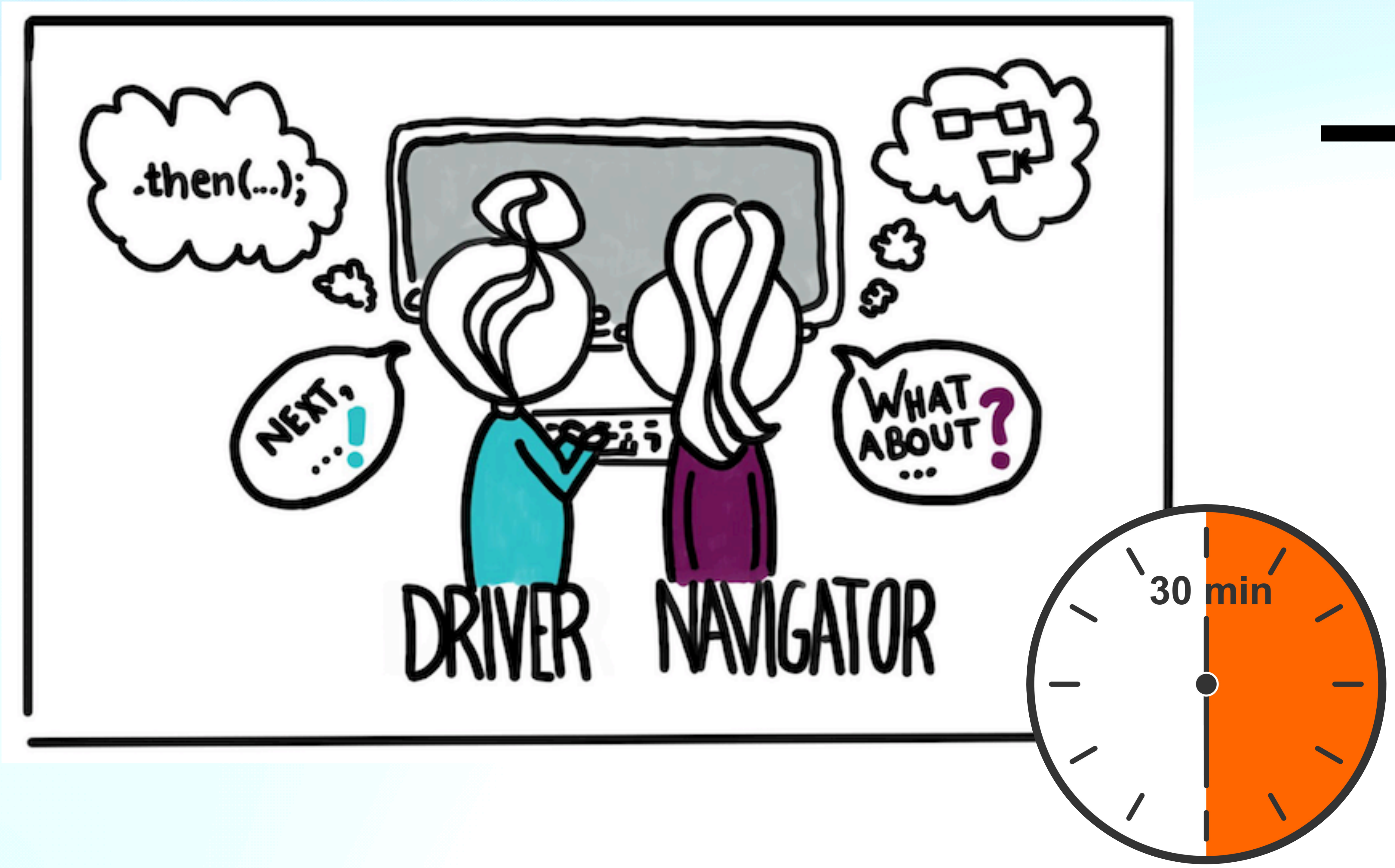
$$\hat{\Delta} = -\left(\left.\frac{\partial^2 l}{\partial \theta^2}\right|_{\hat{\theta}_k} \right)^{-1} \left.\frac{\partial l}{\partial \theta}\right|_{\hat{\theta}_k}$$

which maximizes the quadratic approximation to l. Hence step 3 is

$$\hat{\theta}_{k+1} = \hat{\theta}_k + \hat{\Delta}.$$

Rest of the tutorial...

- In pairs work on the third notebook. Paired programming will continue!



Rest of the week...

From the end of the slides attempt the question.

Following this attempt Question 4-1, 4-2 and 4-3.

- If you would like a challenge then try and implement a numerical equivalent solution for some of the models that we use.

An extra theory question...

Theoretical Question 2...

Let $x_1, x_2, \ldots, x_n$ be a random i.i.d. sample from the exponential distribution $Exp(\lambda)$ with density $f(x;\lambda) = \lambda e^{-\lambda x}$, $x > 0$, and
$y_1, y_2, \ldots, y_n$ be a random i.i.d. sample from the exponential distribution $Exp(\lambda - \mu)$, with density $f(y;\lambda,\mu) = (\lambda - \mu)e^{-(\lambda-\mu)y}$, $y > 0$.
Denote the unknown parameter $\theta = (\lambda, \mu)$, with $\lambda > 0$ and $\mu \in (-\infty, \lambda)$.

Write down the likelihood and hence the log likelihood function for the unknown parameter $\theta = (\lambda, \mu)$ given observed samples $\underline{x} = (x_1, \ldots, x_n)$ and $\underline{y} = (y_1, \ldots, y_n)$. Explain your reasoning.

Obtain expressions for the score vector and for the MLE for the vector of the unknown parameters $\theta = (\lambda, \mu)$.

1. The likelihood for θ based on both samples is the joint probability density of the two samples. The two samples are independent, and within each sample the observations are iid, hence the likelihood function for θ using both samples is

$$L(\theta; x, y) = \prod_{i=1}^{n} f(x_i; \lambda) \prod_{i=1}^{n} f(y_i; \lambda, \mu) = \lambda^n (\lambda - \mu)^n e^{-\lambda \sum_i x_i - (\lambda - \mu) \sum_i y_i}.$$

The corresponding log likelihood is

$$\ell(\theta; x, y) = \log L(\theta; x, y) = n \log \lambda + n \log(\lambda - \mu) - \lambda \sum_i x_i - (\lambda - \mu) \sum_i y_i.$$

2. The derivatives of the log likelihood are

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial \lambda} &= n/\lambda + n/(\lambda - \mu) - \sum_i x_i - \sum_i y_i, \\ \frac{\partial \ell(\theta)}{\partial \mu} &= -n/(\lambda - \mu) + \sum_i y_i.\end{aligned}$$

Hence, the score function is

$$U(\theta) = \begin{pmatrix} \frac{\partial \ell(\theta)}{\partial \mu} \\ \frac{\partial \ell(\theta)}{\partial \mu} \end{pmatrix} = \begin{pmatrix} n/\lambda + n/(\lambda - \mu) - \sum_i x_i - \sum_i y_i \\ -n/(\lambda - \mu) + \sum_i y_i) \end{pmatrix}.$$

Solving the MLE equations $U(\hat{\theta}) = 0$, we have

$$\begin{aligned} n/\hat{\lambda} + n/(\hat{\lambda} - \hat{\mu}) - \sum_i x_i - \sum_i y_i = 0, \\ -n/(\hat{\lambda} - \hat{\mu}) + \sum_i y_i = 0 \end{aligned}$$

which can be rewritten as (by adding the two equations)

$$\begin{aligned} n/\hat{\lambda} - \sum_i x_i = 0, \\ n/(\hat{\lambda} - \hat{\mu}) = \sum_i y_i. \end{aligned}$$

Denote $\bar{x} = n^{-1} \sum_i x_i$ and $\bar{y} = n^{-1} \sum_i y_i$. Therefore, the solutions are

$$\hat{\lambda} = 1/\bar{x}, \hat{\mu} = \hat{\lambda} - 1/\bar{y} = 1/\bar{x} - 1/\bar{y}.$$

We'll talk about verification of this when we look at Fisher Information next time.