STATS 2107
Statistical Modelling and Inference II
Solutions
Workshop 12:
MLE for SLR

Matt Ryan

Semester 2 2022

Contents

The set up

Simple linear regression

Consider data $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ and the simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$ independently for each $i = 1, 2, \ldots, n$.

Likelihood estimation

How does SLR fit into likelihood estimation? For likelihood estimation we need:

1. Independent data $y_1, y_2, \ldots, y_n$.
2. A pdf for each y_i, $f_{Y_i}(y_i)$.
3. Some parameters $\boldsymbol{\theta}$ to estimate

What is the pdf for the SLR?

pdf for SLR

We may write $Y_i \sim N(\mu_i, \sigma^2)$ where $\mu_i = \beta_0 + \beta_1 x_i$ for each $i = 1, 2, \ldots, n$. Hence $\boldsymbol{\theta} = \left(\beta_0, \beta_1, \sigma^2\right)$, and

$$\begin{aligned} f_{Y_i}(y_i; \boldsymbol{\theta}) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu_i)^2}{2\sigma^2}} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2} \end{aligned}$$

Calculating the likelihood

By definition,

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{n} f_{Y_i}(y_i; \boldsymbol{\theta}).$$

Calculating the likelihood

$$\begin{aligned} L(\boldsymbol{\theta}; \boldsymbol{y}) &= \prod_{i=1}^{n} f_{Y_i}(y_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - (\beta_0 + \beta_1 x_i))^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2} \end{aligned}$$

Your turn

What to do

1. Calculate the log-likelihood $\ell(\boldsymbol{\theta}; \boldsymbol{y})$.

Solutions:
The log-likelihood is:

$$
\begin{aligned}
\ell(\boldsymbol{\theta}; \boldsymbol{y}) &= \log(L(\boldsymbol{\theta}; \boldsymbol{y})) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2}\right) \\
&= \log\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\right) + \log\left(e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2}\right) \\
&= -\log\left((2\pi\sigma^2)^{\frac{n}{2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2 \\
&= -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2 \\
&= -\frac{n}{2}\log\left(2\pi\right) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2
\end{aligned}
$$

The log-likelihood

You should get:

$$\ell(\boldsymbol{\theta}; \boldsymbol{y}) = -\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\beta_0+\beta_1 x_i))^2 + C$$

for a constant C.

The score

The score vector

We define the score vector for a parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$ by

$$[S(\boldsymbol{\theta}; \boldsymbol{y})]_i = \left[\frac{\partial \ell}{\partial \theta_i}\right]$$

For SLR?

In our case, we have

- $\theta_1 = \beta_0$
- $\theta_2 = \beta_1$
- $\theta_3 = \sigma^2$

The first element

$$
\begin{aligned}
\frac{\partial \ell}{\partial \beta_0} &= \frac{\partial}{\partial \beta_0}\left[-\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 + C\right] \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial \beta_0}\left[\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2)\left(y_i - (\beta_0 + \beta_1 x_i)\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)
\end{aligned}
$$

Your turn

What to do

1. Show that

$$
S(\boldsymbol{\theta}; \boldsymbol{y}) = \begin{bmatrix} \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) \\ \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) x_i \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \end{bmatrix}
$$

Solutions:
This amounts to calculating

$$
\frac{\partial \ell}{\partial \beta_1} \quad \text{and} \quad \frac{\partial \ell}{\partial \sigma^2},
$$

Notice how our parameter is σ^2, not σ!

We find

$$
\begin{aligned}
\frac{\partial \ell}{\partial \beta_1} &= \frac{\partial}{\partial \beta_1}\left[-\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 + C\right] \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{\partial}{\partial \beta_1}\left[\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \\
&= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(-2x_i)\left(y_i - (\beta_0 + \beta_1 x_i)\right) \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) x_i\,,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial \ell}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 + C\right] \\
&= \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2}\log\left(\sigma^2\right)\right] + \frac{\partial}{\partial \sigma^2}\left[-\frac{1}{2\sigma^2}\right]\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \,.
\end{aligned}
$$

Maximum Likelihood estimates

How do we get the MLE?

To find the MLE, we solve the equation

$$S(\boldsymbol{\theta}; \boldsymbol{y}) = \mathbf{0}\,.$$

For SLR

This gives the following three equations:

$$
\begin{aligned}
\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) &= 0\,, \\
\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) x_i &= 0\,, \\
-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 &= 0\,.
\end{aligned}
$$

Solving for $\hat{\beta}_0$

The first equation gives:

$$
\begin{aligned}
0 &= \sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right) \\
&= \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\beta_0 - \sum_{i=1}^{n}\beta_1 x_i \\
&= n\bar{y} - n\beta_0 - n\beta_1\bar{x}\,,
\end{aligned}
$$

hence

$$\hat{\beta}_0 = \bar{y} - \beta_1\bar{x}$$

Solving for $\hat{\beta}_1$

$$\begin{aligned}
0 &= \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right) x_i \\
&= \sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} y_i x_i - \beta_0 n\bar{x} - \beta_1 \sum_{i=1}^{n} x_i^2
\end{aligned}$$

Evaluate at $\hat{\beta}_0$

$$\begin{aligned}
0 &= \sum_{i=1}^{n} y_i x_i - \hat{\beta}_0 n\bar{x} - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} y_i x_i - (\bar{y} - \beta_1 \bar{x}) n\bar{x} - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x} - \beta_1 \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2 \right) \\
&= S_{XY} - \beta_1 S_{XX} \,,
\end{aligned}$$

hence

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} \,.$$

Your turn

What to do

1. Calculate $\widehat{\sigma^2}$, the MLE for σ^2.

Solutions:
We have

$$\begin{aligned}
0 &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \,, \\
\frac{n}{2\sigma^2} &= \frac{1}{2(\sigma^2)^2} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \,, \\
n\sigma^2 &= \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2 \,,
\end{aligned}$$

hence

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

---

2. Compare the MLEs to the least squares estimates for simple linear regression.

Solutions:
You clearly see that the MLE and LSE for β_1 and β_0 are identical. However, the MLE for σ^2 is

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$$

where as the LSE for σ^2 is

$$s_e^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - (\beta_0 + \beta_1 x_i))^2$$

We have previously shown that s_e^2 is unbiased for σ^2, hence this shows that the MLE is a biased estimate for σ^2.

Fisher Information

The Fisher information matrix

Under some regularity conditions, the Fisher information matrix is given by

$$[I_{\boldsymbol{\theta}}]_{ij} = \left[\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial\theta_i \partial\theta_j}\right]\right]$$

For SLR

This will be a 3×3 matrix, so first we need to calculate the following partials:

$$\begin{array}{ccc} \dfrac{\partial^2 \ell}{\partial\beta_0^2} & \dfrac{\partial^2 \ell}{\partial\beta_1^2} & \dfrac{\partial^2 \ell}{\partial(\sigma^2)^2} \\ \dfrac{\partial^2 \ell}{\partial\beta_0\partial\beta_1} & \dfrac{\partial^2 \ell}{\partial\beta_0\partial\sigma^2} & \dfrac{\partial^2 \ell}{\partial\beta_1\partial\sigma^2} \end{array}$$

Your turn

What to do

1. Show that

$$I_{\boldsymbol{\theta}} = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n\bar{x}}{\sigma^2} & 0 \\ \frac{n\bar{x}}{\sigma^2} & \frac{1}{\sigma^2}\sum\limits_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & \frac{n-4}{2(\sigma^2)^2} \end{bmatrix}$$

Solutions:
First up, the partials:

$$
\begin{aligned}
\frac{\partial^2 \ell}{\partial \beta_0^2} &= \frac{\partial}{\partial \beta_0}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(-1) \\
&= -\frac{n}{\sigma^2}\,. \\
\frac{\partial^2 \ell}{\partial \beta_1^2} &= \frac{\partial}{\partial \beta_1}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)x_i\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(-x_i^2\right) \\
&= -\frac{1}{\sigma^2}\sum_{i=1}^{n}x_i^2\,. \\
\frac{\partial^2 \ell}{\partial (\sigma^2)^2} &= \frac{\partial}{\partial \sigma^2}\left[-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right] \\
&= \frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\,. \\
\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1} &= \frac{\partial}{\partial \beta_1}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n}(-x_i) \\
&= -\frac{n\bar{x}}{\sigma^2}\,. \\
\frac{\partial^2 \ell}{\partial \beta_0 \partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)\right] \\
&= -\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)\,. \\
\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[\frac{1}{\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)x_i\right] \\
&= -\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)x_i\,.
\end{aligned}
$$

Now, expected values of the negative:

$$
\begin{aligned}
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \beta_0^2}\right] &= \mathrm{E}\left[-\left(-\frac{n}{\sigma^2}\right)\right] \\
&= \frac{n}{\sigma^2}\,. \\
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \beta_1^2}\right] &= \mathrm{E}\left[-\left(-\frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2\right)\right] \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n} x_i^2\,. \\
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial (\sigma^2)^2}\right] &= \mathrm{E}\left[-\left(\frac{n}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)^2\right)\right] \\
&= -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}\,\mathrm{E}\left[(n-2)S_e^2\right] \\
&= -\frac{n}{2(\sigma^2)^2} + \frac{1}{(\sigma^2)^3}(n-2)\sigma^2 \\
&= \frac{-n + 2n - 4}{2(\sigma^2)^2} \\
&= \frac{n-4}{2(\sigma^2)^2}\,. \\
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \beta_0 \partial \beta_1}\right] &= \mathrm{E}\left[-\left(-\frac{n\bar{x}}{\sigma^2}\right)\right] \\
&= \frac{n\bar{x}}{\sigma^2}\,. \\
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \beta_0 \sigma^2}\right] &= \mathrm{E}\left[-\left(-\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)\right)\right] \\
&= \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(\mathrm{E}[y_i] - (\beta_0 + \beta_1 x_i)\right) \\
&= \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i)\right) \\
&= 0\,. \\
\mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \beta_1 \partial \sigma^2}\right] &= \mathrm{E}\left[-\left(-\frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_i)\right)x_i\right)\right] \\
&= \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(\mathrm{E}[y_i] - (\beta_0 + \beta_1 x_i)\right)x_i \\
&= \frac{1}{(\sigma^2)^2}\sum_{i=1}^{n}\left(\beta_0 + \beta_1 x_i - (\beta_0 + \beta_1 x_i)\right)x_i \\
&= 0\,.
\end{aligned}
$$

Hence, we have:

$$
I_{\boldsymbol{\theta}} = \begin{bmatrix} \frac{n}{\sigma^2} & \frac{n\bar{x}}{\sigma^2} & 0 \\ \frac{n\bar{x}}{\sigma^2} & \frac{1}{\sigma^2}\sum\limits_{i=1}^{n} x_i^2 & 0 \\ 0 & 0 & \frac{n-4}{2(\sigma^2)^2} \end{bmatrix}
$$
