# Maximum likelihood estimation

Ciaran Evans

# Warmup

Work on the warmup activity (handout) with your neighbors, then we will discuss as a class.

# Warmup

$Y_1, \ldots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Let $\boldsymbol{\theta} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$ and $\mathbf{y} = (Y_1, \ldots, Y_n)$.

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^{n} f(Y_i|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(Y_i-\mu)^2\right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2\right\}$$

$$\ell(\theta|y) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2$$

# Warmup

$$\ell(\boldsymbol{\theta}|\mathbf{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i-\mu)^2$$

$$\frac{\partial}{\partial\mu}\ell(\boldsymbol{\theta}|\mathbf{y}) = -\frac{1}{2\sigma^2}\sum_i (Y_i-\mu)(2)(-1) = \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\mu) \overset{\text{set}}{=} 0$$

$$\Rightarrow \sum_{i=1}^{n}(Y_i-\mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n} Y_i = n\mu \quad \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$$

$$\frac{\partial}{\partial\sigma^2}\ell(\boldsymbol{\theta}|\mathbf{y}) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i-\mu)^2 \overset{\text{set}}{=} 0$$

$$\Rightarrow \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (Y_i-\hat{\mu})^2 = 0$$

$$\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(Y_i-\hat{\mu})^2 = n$$

$$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

# Invariance of the MLE

For $Y_1, ..., Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$, we saw that the MLE of the variance is $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$. What if we instead want to estimate the *standard deviation*, $\sigma$?

<u>Theorem</u> (invariance of MLEs): (Theorem 7.2.10 in CB)

Let $\hat{\theta}$ be the MLE of $\theta$. For any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$

(see proof in CB)

<u>Ex</u>: $\sigma = \tau(\sigma^2) = \sqrt{\sigma^2}$

$\hat{\sigma} = \tau(\hat{\sigma}^2) = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$

# Another example: *Uniform*$(0, \theta)$

Let $Y_1, \ldots, Y_n \overset{iid}{\sim}$ *Uniform*$(0, \theta)$, where $\theta > 0$. We want the maximum likelihood estimator of $\theta$.

$$f(y_i \mid \theta) = \begin{cases} \frac{1}{\theta} & 0 \le y_i \le \theta \\ 0 & \end{cases} = \frac{1}{\theta} \underbrace{\mathbb{1}\{0 \le y_i \le \theta\}}_{\substack{=1 \text{ if true} \\ =0 \text{ if not}}}$$

$$L(\theta \mid y) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}\{0 \le y_i \le \theta\} = \frac{1}{\theta^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}\{0 \le y_i \le \theta\}}_{\substack{=1 \text{ if all } y_i \in [0,\theta] \\ =0 \text{ else}}}$$

$$= \frac{1}{\theta^n} \mathbb{1}\{0 \le y_1, \ldots, y_n \le \theta\}$$

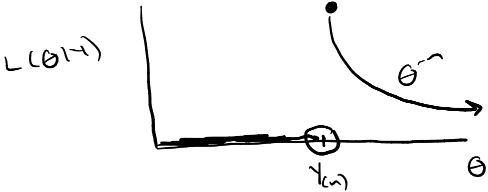
$$= \frac{1}{\theta^n} \mathbb{1}\{0 \le y_{(n)} \le \theta\}$$

$$y_{(n)} = \max\{y_1, \ldots, y_n\}$$

# Another example: *Uniform*$(0, \theta)$

Let $Y_1, \ldots, Y_n \overset{iid}{\sim} Uniform(0, \theta)$, where $\theta > 0$. We want the maximum likelihood estimator of $\theta$.

$$L(\theta|y) = \frac{1}{\theta^n} \mathbb{1}\{0 \leq Y_{(n)} \leq \theta\}$$




if $\theta < Y_{(n)}$:

$\to 0$

$\theta > Y_{(n)}$:

$\to \frac{1}{\theta^n}$

maximum likelihood estimator: $\hat{\theta} = Y_{(n)}$

often: if $\theta$ is related to the support/range of distribution, MLE involves order statistics

# Maximum likelihood estimation and linear regression

Let $(\mathbf{x}_1, Y_1), ..., (\mathbf{x}_n, Y_n)$ be iid samples from the model

$$Y_i | \mathbf{x}_i \sim N(\mu_i, \sigma^2)$$
$$\mu_i = \mathbf{x}_i^T \boldsymbol{\beta},$$

$$L(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) =$$