Lecture 4: Maximum likelihood estimation

Ciaran Evans

Recap: maximum likelihood estimation

Definition: Let $\mathbf{Y} = (Y_1, ..., Y_n)$ be a sample of n observations, and let $f(\mathbf{y}|\theta)$ denote the joint pdf or pmf of $\mathbf{Y}$, with parameter(s) θ. The *likelihood function* is

function of θ $\rightsquigarrow$ $L(\theta|\mathbf{Y}) = f(\mathbf{Y}|\theta)$

Definition: Let $\mathbf{Y} = (Y_1, ..., Y_n)$ be a sample of n observations. The *maximum likelihood estimator* (MLE) is

$$\widehat{\theta} = \text{argmax}_{\theta} \; L(\theta|\mathbf{Y})$$

Common approach:

- write down likelihood, take the log
- differentiate wrt θ, $\stackrel{\text{set}}{=} 0$
- solve

Example: $N(\theta, 1)$

$\theta \in (-\infty, \infty)$

$Y_1, \ldots, Y_n \overset{iid}{\sim} N(\theta, 1)$

$$L(\theta | Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(Y_i - \theta)^2\right\}$$

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

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

$$\frac{\partial}{\partial \theta} \ell(\theta | Y) = -\frac{1}{2} \sum_{i=1}^{n} 2(Y_i - \theta)(-1) = \sum_{i=1}^{n} (Y_i - \theta) \overset{set}{=} 0$$

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

$$\frac{\partial^2}{\partial \theta^2} \ell(\theta | Y) = \frac{\partial}{\partial \theta} \sum_{i=1}^{n} (Y_i - \theta) = -n < 0 \checkmark$$

unique maximum

MLE: $\hat{\theta} = \bar{Y}$

$$\mathbb{E}[T] = \mathbb{E}\left[\frac{1}{n} \sum_i Y_i\right] = \frac{1}{n} \sum_i \mathbb{E}[Y_i]$$

$$= \frac{1}{n} \sum_i \theta$$

$$= \frac{1}{n} \cdot n \cdot \theta = \theta$$

Sidebar: Let $\hat{\theta}$ is an estimator of θ

$\hat{\theta} \leftarrow$ r.v.

ideally: $\mathbb{E}[\hat{\theta}] = \theta$

Bias: $\mathbb{E}[\hat{\theta}] - \theta$

unbiased estimator: $\text{Bias}(\hat{\theta}) = 0$

i.e. $\mathbb{E}[\hat{\theta}] = \theta$

Consistency "asymptotic unbiasedness" $(n \to \infty)$

Suppose estimate Variance σ^2 of $N(\mu, \sigma^2)$

$$s^2 = \frac{1}{n-1} \sum_i (Y_i - \bar{Y})^2 \qquad \mathbb{E}[s^2] = \sigma^2$$

Also: $$\hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2 \qquad \mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{n-1}{n} s^2\right] = \frac{n-1}{n}\sigma^2$$

$\text{Bias}(\hat{\sigma}^2) \to 0$ as $n \to \infty$

Example: *Uniform*$(0, \theta)$

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

Discuss with your neighbors what the MLE of θ might be. *Hint: focus on finding and sketching the likelihood function* $L(\mathbf{Y}|\theta)$

$Y_1, \dots, Y_n \overset{iid}{\sim} \text{Uniform}(0, \theta)$

$$f(y_i|\theta) = \begin{cases} \frac{1}{\theta} & 0 \le y_i \le \theta \\ 0 & \text{else} \end{cases}$$

$$= \frac{1}{\theta} \underbrace{\mathbb{1}\{0 \le y_i \le \theta\}}_{\substack{=1 \text{ if true} \\ =0 \text{ if not}}}$$

$$L(\theta|Y) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}\{0 \le Y_i \le \theta\}$$

$$= \frac{1}{\theta^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}\{0 \le Y_i \le \theta\}}$$

$Y_{(1)} = \min\{Y_1, \dots, Y_n\}$

$Y_{(n)} = \max\{Y_1, \dots, Y_n\}$

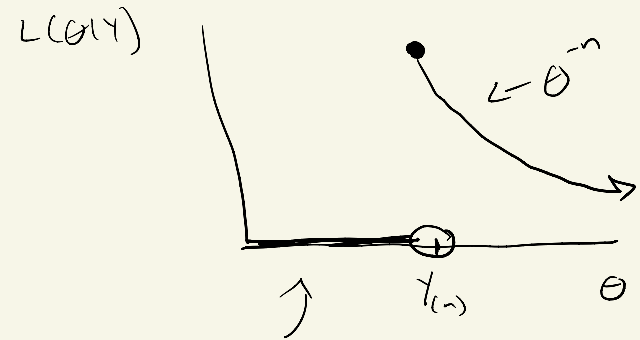
$= 1$ if $0 \le Y_i \le \theta \ \forall i$

$= 0$ else

$= \mathbb{1}\{0 \le Y_1, \dots, Y_n \le \theta\}$

$= \mathbb{1}\{0 \le Y_{(n)} \le \theta\}$

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



$L(\theta|Y) = 0$ for $\theta < Y_{(n)}$

maximum likelihood estimator:

$$\hat{\theta} = Y_{(n)} \quad //$$

often: MLE involves order statistics if θ is related to the support / range of the distribution

$Y_1, \ldots, Y_n \overset{iid}{\sim} U(\theta, 1)$ $f(y_i$

$$L(\theta \mid Y) = \left(\frac{1}{1-\theta}\right)^n \mathbb{1}\{\theta \le Y_{(1)} \le 1\}$$

$$\hat{\theta} = Y_{(1)}$$

Theorem (invariance of the MLE): (Thm 7.2.10 in CB)

Let $\hat{\theta}$ be the MLE of θ

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

Ex: $N(\mu, \sigma^2)$ MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i (Y_i - \bar{Y})^2$$

if we want MLE for σ:

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n} \sum_i (Y_i - \bar{Y})^2}$$

Example: $N(\mu, \sigma^2)$

Linear regression with normal errors

$$Y_i \sim N(\mu_i, \sigma^2)$$

$$\mu_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k}$$

Suppose we observe independent samples $(X_1, Y_1), ..., (X_n, Y_n)$.
Write down the likelihood function

$$L(\beta|\mathbf{X}, \mathbf{Y}) \propto \prod_{i=1}^{n} f(Y_i|\beta, X_i)$$

for the linear regression problem.