Maximum likelihood estimation

Recap: maximum likelihood estimation

Definition: Let $\mathbf{Y} = (Y_1, \ldots, Y_n)$ 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

$$L(\theta|\mathbf{Y}) = f(\mathbf{Y}|\theta)$$

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

$$\hat{\theta} = \text{argmax}_\theta \, L(\theta|\mathbf{Y})$$

To maximize $L(\theta|Y)$, we often work with log likelihood

$$\ell(\theta|Y) = \log L(\theta|Y)$$

Continuing $N(\theta, 1)$ example

Note: The same argument shows that for any θ,

$$\sum_{i=1}^{n}(Y_i-\theta)^2 \geq \sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

(regardless of distribution of Y)

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

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

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

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

second derivative test: $\dfrac{d^2}{d\theta^2}\ell(\theta|Y) = \dfrac{d}{d\theta}\sum_{i=1}^{n}(Y_i-\theta)$

$$= -n < 0 \quad \checkmark$$

Boundaries: check $\theta = \pm\infty$

$$L(\pm\infty|Y) = (2\pi)^{-\frac{n}{2}} \exp\{-\infty\} = 0 \quad \checkmark$$

$$\Rightarrow \boxed{\hat{\theta} = \bar{Y}}$$

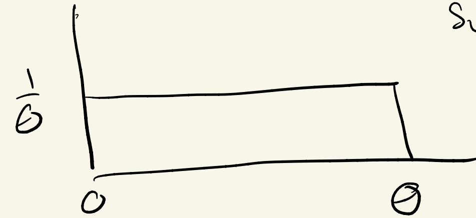
Example: $Uniform(0, \theta)$

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

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

Uniform(0, θ):

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



Suppose $Y_1, \dots, Y_n$

$Y_{(n)} = 0.8$

θ cannot be 0.5

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

$$= \frac{1}{\theta} \mathbb{1}\{\theta \ge y \ge 0\}$$

indicator function

$= 1$ if $0 \le y \le \theta$

0 otherwise

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

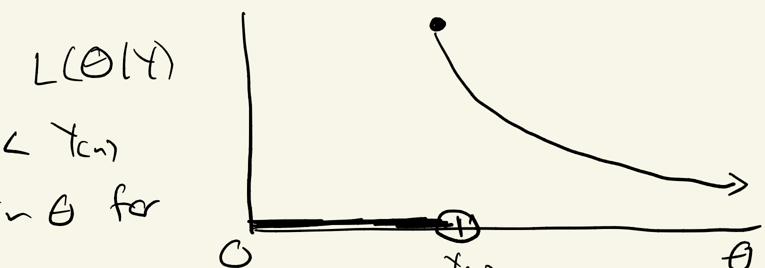
$$= \frac{1}{\theta^n} \underbrace{\prod_{i=1}^{n} \mathbb{1}\{\theta \ge Y_i \ge 0\}}$$

$= 1$ if $0 \le Y_i \le \theta$ for all $Y_1, \dots, Y_n$

$= 0$ otherwise

$$= \frac{1}{\theta^n} \mathbb{1}\{Y_1 \le \theta, Y_2 \le \theta, Y_3 \le \theta, \dots, Y_n \le \theta\}$$

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



- $L(\theta|Y) = 0$ if $\theta < Y_{(n)}$
- $L(\theta|Y)$ decreases in θ for $\theta \ge Y_{(n)}$

$\Rightarrow$ $\boxed{\hat{\theta} = Y_{(n)}}$

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

$\theta = (\mu, \sigma^2)$

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

$\sigma^2 > 0$

$Y_1, \dots, Y_n \overset{iid}{\sim} N(\mu, \sigma^2)$

$$f(Y_i \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{1}{2\sigma^2} (Y_i - \mu)^2 \right\}$$

$$\Rightarrow \quad L(\theta \mid Y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2 \right\}$$

$$\Rightarrow \quad \ell(\theta \mid Y) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \mu)^2$$

Option 1: Solve the system $\frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell}{\partial \mu} \\ \frac{\partial \ell}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

Option 2: Maximize wrt μ, then maximize wrt σ^2 (usually makes life easier)

Option 2:

1) Start w/ μ:

$$\sum_{i=1}^{n} (Y_i - \mu)^2 \geq \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

for any μ
(see $N(\theta, 1)$ example)

$\Rightarrow$ For any σ^2

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

$$\leq (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \right\} = \underbrace{\sup_{\mu} L(\theta|Y)}_{\text{profile likelihood for } \sigma^2}$$

"supremum" (basically max)

$\Rightarrow \boxed{\hat{\mu} = \bar{Y}}$, regardless of σ^2

2) Now σ^2: we want to maximize

$$L^*(\sigma^2|Y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \right\}$$

which is a univariate function of σ^2

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

$$\frac{d\ell^*}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \overset{\text{set}}{=} 0$$

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

N.B. compare to

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

why $\frac{1}{n-1}$ instead of $\frac{1}{n}$?

we'll see later....

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

Invariance of the MLE

Maximum likelihood estimation for logistic regression

$$Y_i \sim Bernoulli(p_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k}$$

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

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

for the logistic regression problem.