STSCI 5080
Probability Models and Inference
Lecture 19: Maximal Likelihood Estimation

November 6, 2018

Review of independence

Definition

Random variables $X_1, \ldots, X_n$ are independent if

$$P(X_1 \in A_1, \ldots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n)$$

for any subsets $A_1, \ldots, A_n \subset \mathbb{R}$.

Discrete case

If $X_1, \ldots, X_n$ are independent and discrete with pmfs

$$p_{X_1}(x_1), \ldots, p_{X_n}(x_n),$$

respectively, then the joint pmf of $(X_1, \ldots, X_n)$ is

$$\begin{aligned} p(x_1, \ldots, x_n) &= P(X_1 = x_1, \ldots, X_n = x_n) \\ &= P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= p_{X_1}(x_1) \cdots p_{X_n}(x_n) \\ &= \prod_{i=1}^{n} p_{X_i}(x_i). \end{aligned}$$

Continuous case

- A random vector $(X_1, \ldots, X_n)$ is continuous if there exists a joint pdf $f(x_1, \ldots, x_n)$ on $\mathbb{R}^n$ such that

$$P((X_1, \ldots, X_n) \in B) = \int \cdots \int_B f(x_1, \ldots, x_n) dx_1 \cdots dx_n$$

for any subset B of $\mathbb{R}^n$.

- If $X_1, \ldots, X_n$ are independent and continuous with pdfs

$$f_{X_1}(x_1), \ldots, f_{X_n}(x_n),$$

respectively, then $(X_1, \ldots, X_n)$ is continuous with joint pdf

$$\begin{aligned} f(x_1, \ldots .x_n) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \prod_{i=1}^{n} f_{X_i}(x_i). \end{aligned}$$

Setting

- Let $\{f_\theta \mid \theta \in \Theta\}$ be a class of pmfs/pdfs where $\Theta \subset \mathbb{R}^k$, and suppose that

$$X_1, \ldots, X_n \sim f_\theta \text{ i.i.d.}$$

  for some $\theta \in \Theta$.

- Find the joint pmf/pdf of $(X_1, \ldots, X_n)$.

Setting

- Let $\{f_\theta \mid \theta \in \Theta\}$ be a class of pmfs/pdfs where $\Theta \subset \mathbb{R}^k$, and suppose that

$$X_1, \ldots, X_n \sim f_\theta \text{ i.i.d.}$$

  for some $\theta \in \Theta$.

- Find the joint pmf/pdf of $(X_1, \ldots, X_n)$.

- The joint pmf/pdf is

$$\prod_{i=1}^{n} f_\theta(x_i).$$

Likelihood function

Plug in $X_1, \ldots, X_n$ and think of the joint pmf/pdf as a function of θ:

$$L_n(\theta) = L_n(\theta, X_1, \ldots, X_n) = \prod_{i=1}^{n} f_\theta(X_i).$$

This is called the likelihood function for θ.

Definition

An estimator $\widehat{\theta} = \widehat{\theta}(X_1, \dots, X_n)$ for θ is a function (statistic) of $X_1, \dots, X_n$ that takes values in $\mathbb{R}^k$. If the estimator $\widehat{\theta}$ is evaluated at some specific values of $X_1, \dots, X_n$, i.e., $X_1 = x_1, \dots, X_n = x_n$, then $\widehat{\theta}(x_1, \dots, x_n)$ is called an estimate.

An estimator is a random variable (vector), but an estimate is a non-random number.

MLE

Definition

The maximal likelihood estimator (MLE) $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ is defined by a point in Θ that maximizes $L_n(\theta)$:

$$L_n(\widehat{\theta}) = \max_{\theta \in \Theta} L_n(\theta).$$

If $\widehat{\theta}(X_1, \ldots, X_n)$ is evaluated at specific values of $X_1, \ldots, X_n$, i.e., $X_1 = x_1, \ldots, X_n = x_n$, then the value $\widehat{\theta}(x_1, \ldots, x_n)$ is called an maximum likelihood estimate.

Finding MLE

- In practice, it is easier to work with the log likelihood function

$$\begin{aligned}\ell_n(\theta) &= \ell_n(\theta, X_1, \ldots, X_n) \\ &= \log L_n(\theta) \\ &= \log \prod_{i=1}^{n} f_\theta(X_i) \\ &= \sum_{i=1}^{n} \log f_\theta(X_i).\end{aligned}$$

- We note that

maximizing $L_n(\theta) \Leftrightarrow$ maximizing $\ell_n(\theta)$

Theorem

The MLE $\widehat{\theta}$ can be defined as a point in Θ that maximizes $\ell_n(\theta)$:

$$\ell_n(\widehat{\theta}) = \max_{\theta \in \Theta} \ell_n(\theta).$$

Finding a maximizer of a smooth function

Theorem

Let $\Theta \subset \mathbb{R}^k$ *and let* $g : \Theta \to \mathbb{R}$ *be a smooth function. If* θ^* *maximizes* $g(\theta)$ *and is an interior point of* Θ*, then* θ^* *satisfies the first order condition (FOC):*

$$\frac{\partial g}{\partial \theta_1}(\theta) = 0,$$
$$\vdots$$
$$\frac{\partial g}{\partial \theta_k}(\theta) = 0.$$

Finding MLE

Rule of thumb

To find the MLE, find a point in Θ that satisfies the FOC

$$\begin{aligned} \frac{\partial \ell_n}{\partial \theta_1}(\theta) &= 0, \\ &\vdots \\ \frac{\partial \ell_n}{\partial \theta_k}(\theta) &= 0. \end{aligned} \tag{*}$$

If $k = 1$ (i.e., θ is one-dim.), then (*) simplifies to

$$\ell_n'(\theta) = 0.$$

Example 19.1

Example

Let

$$X_1, \ldots, X_n \sim Po(\lambda) \text{ i.i.d.}$$

for some $\lambda > 0$.

(a) Find the log likelihood function for λ.

(b) Find the FOC for the MLE of λ.

(c) Find the MLE.

The pmf of $Po(\lambda)$ is

$$f_\lambda(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

The joint pmf is

$$\begin{aligned}\prod_{i=1}^n f_\lambda(x_i) &= \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\ &= e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}. \quad \text{(why?)}\end{aligned}$$

The likelihood function is

$$L_n(\lambda) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!}.$$

The log likelihood function is

$$\ell_n(\lambda) = \log L_n(\lambda) = -n\lambda + (\sum_{i=1}^{n} X_i) \log \lambda - \log(\prod_{i=1}^{n} X_i!).$$

We note that

$$\ell_n'(\lambda) = -n + \frac{\sum_{i=1}^{n} X_i}{\lambda}.$$

So the FOC is

$$-n + \frac{\sum_{i=1}^{n} X_i}{\lambda} = 0.$$

Solving the FOC w.r.t. λ, we obtain the MLE

$$\widehat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

Example 19.2

Example

Let

$$X_1, \ldots, X_n \sim Ex(\lambda) \text{ i.i.d.}$$

for some $\lambda > 0$.

(a) Find the log likelihood function for λ.

(b) Find the FOC for the MLE of λ.

(c) Find the MLE.

The pdf of $Ex(\lambda)$ is

$$f_\lambda(x) = \lambda e^{-\lambda x}.$$

The joint pmf is

$$\begin{aligned}\prod_{i=1}^{n} f_\lambda(x_i) &= \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}.\end{aligned}$$

The likelihood function is

$$L_n(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} X_i}.$$

The log likelihood function is

$$\ell_n(\lambda) = \log L_n(\lambda) = n \log \lambda - \lambda \sum_{i=1}^{n} X_i.$$

We note that

$$\ell_n'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i.$$

So the FOC is

$$\frac{n}{\lambda} - \sum_{i=1}^{n} X_i = 0.$$

Solving the FOC w.r.t. λ, we obtain the MLE

$$\widehat{\lambda} = \frac{1}{n^{-1} \sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}.$$

Lifetimes of electronic components

- An exponential distribution is used for modeling lifetimes of electronic components (e.g. laptops).
- Suppose that we observe the lifetimes of three electronic components, and we fit an exponential distribution to them:

$$X_1, X_2, X_3 \sim Ex(\lambda) \text{ i.i.d.}$$

  for some λ.

- Now, the actual data are $X_1 = 3, X_2 = 1.5$, and $X_3 = 2.1$.
- The MLE for λ is

$$\widehat{\lambda} = \frac{1}{\overline{X}} = \frac{1}{2.2} \approx 0.45.$$

Example 19.3

Example

Suppose that σ_0^2 is known (e.g. $\sigma_0^2 = 9$). Let

$$X_1, \ldots, X_n \sim N(\mu, \sigma_0^2) \text{ i.i.d.}$$

for some $-\infty < \mu < \infty$.

(a) Find the log likelihood function for μ.
(b) Find the FOC for the MLE of μ.
(c) Find the MLE.

The pdf of $N(\mu, \sigma_0^2)$ is

$$f_\mu(x) = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma_0^2)}.$$

The joint pmf is

$$\begin{aligned}\prod_{i=1}^{n} f_\mu(x_i) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma_0^2)} \\ &= \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n}(x_i-\mu)^2/(2\sigma_0^2)}.\end{aligned}$$

The likelihood function is

$$L_n(\mu) = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^{n}(X_i-\mu)^2/(2\sigma_0^2)}.$$

The log likelihood function is

$$\ell_n(\mu) = \log L_n(\mu) = -\frac{n}{2}\log(2\pi) - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu)^2.$$

We note that

$$\ell_n'(\mu) = \frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu).$$

So the FOC is

$$\frac{1}{\sigma_0^2}\sum_{i=1}^{n}(X_i - \mu) = 0.$$

Solving the FOC w.r.t. μ, we obtain the MLE

$$\widehat{\mu} = \overline{X}.$$