Likelihood : recap

Definition 8.2 :

For any $\theta \in \Theta$, $L(\theta | x)$ is called the likelihood of θ given we have observed the data x.

Example: 8.1.4

Suppose our observed data are $x = \{x_1, x_2, \dots, x_n\}$.
Assume independently sampled from $N(\theta, 1)$.
Then

$$
\begin{aligned}
L(\theta | x) &= f(x | \theta) \\
&= \prod_{i=1}^{n} f(x_i | \theta) \\
&= \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_i - \theta)^2\right) \right] \\
&\;\vdots \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2} \sum (x_i - \theta)^2\right)
\end{aligned}
$$

8.2 Interpreting the likelihood

$$L(\theta|x) = f(x|\theta)$$

DISCRETE CASE

$$L(\theta|x) = f(x|\theta) = P_\theta(X=x)$$

$L(\theta|x)$ is the probability of observing the data x given that the true parameter of our model is θ.

Two values of θ, say θ_1 and θ_2

$$P_{\theta_1}(X=x) = L(\theta_1|x) > L(\theta_2|x) = P_{\theta_2}(X=x)$$

θ_1 is a more plausible value for θ.

Example 8.2.1. Suppose that one has a (possibly unfair) coin and wishes to determine the probability θ of obtaining a head when the coin is tossed, with $\theta \in \Theta = [0, 1]$. The coin is tossed $n = 10$ times and exactly $x = 3$ heads are observed. An appropriate statistical model for the data is the $\text{Bin}(10, \theta)$ model, with likelihood function given by

$$L(\theta|x) = f(x|\theta) = \binom{10}{x} \theta^x (1-\theta)^{10-x}$$

we observe $x = 3$

$$L(\theta|3) = \binom{10}{3} \theta^3 (1-\theta)^7$$

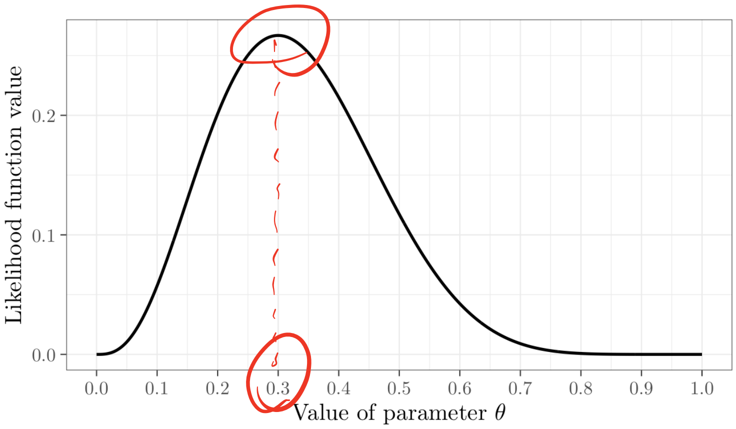

Figure 8.1: Plot showing likelihood of $L(\theta|3) = \binom{10}{3}\theta^3(1-\theta)^7$ for $\theta \in [0,1]$.

Continuous data

$$L(\theta|x) = \underset{\uparrow \text{ density}}{f(x|\theta)} = \underbrace{P_\theta(X = x)}_{=0}$$

Discrete data (again)

Suppose $\Omega = \{1, 2, 3, \ldots\}$

and our model is $\{P_\theta : \theta \in \{1, 2\}\}$

P_1: discrete uniform dist. on $\{1, 2, \ldots, 10^5\}$

P_2: discrete uniform dist. on $\{1, 2, \ldots, 10^8\}$

Suppose we observe $x = 100$

$$L(\theta = 1 | 100) = \frac{1}{10^5}$$

$$L(\theta = 2 | 100) = \frac{1}{10^8}$$

Ratio: $$\frac{L(\theta = 1 | 100)}{L(\theta = 2 | 100)} = \frac{\left(\frac{1}{10^5}\right)}{\left(\frac{1}{10^8}\right)} = 1000$$

Likelihood ratios

Discrete case : $\dfrac{L(\theta_1|x)}{L(\theta_2|x)} = \dfrac{P_{\theta_1}(X=x)}{P_{\theta_2}(X=x)}$

Continuous case:

$$P_\theta(X=x) = 0$$

look in small interval $(x-\delta, x+\delta)$

Exercise

$$L(\theta|x) \approx P_\theta(x-\delta < X < x+\delta) \simeq 2\delta f(x|\theta)$$

$$\frac{L(\theta_1|x)}{L(\theta_2|x)} \approx \frac{P_{\theta_1}(x-\delta < X < x+\delta)}{P_{\theta_2}(x-\delta < X < x+\delta)} \approx \frac{2\delta f(x|\theta_1)}{2\delta f(x|\theta_2)}$$

$$\approx \frac{f(x|\theta_1)}{f(x|\theta_2)}$$

Equivalent likelihood functions

Suppose $L(\theta|x) = cL'(\theta|x)$, $c>0$

c is a constant w.r.t. θ

$$\frac{L(\theta_1|x)}{L(\theta_2|x)} = \frac{cL'(\theta_1|x)}{cL'(\theta_2|x)} = \frac{L'(\theta_1|x)}{L'(\theta_2|x)}$$

Exercise 8.4.1

Show that $L_1 \sim L_2$

$\iff L_1(\cdot|x)$ and $L_2(\cdot|x)$ are equivalent likelihood functions

ie $L_1 = cL_2$

is an equivalence relation.

$x \sim x$

$x \sim y \Rightarrow y \sim x$

$x \sim y,\ y \sim z \Rightarrow x \sim z$

Example 8.4.2

Suppose that $X = (x_1, x_2, \ldots, x_n)$ is a sample of iid observations following a $N(\theta, \sigma^2)$ distribution; σ^2 is known but $\theta \in \mathbb{R}$ is unknown

$$L(\theta|x) = f(x|\theta)$$

$$= \prod_{i=1}^{n} f(x_i|\theta) \quad \text{(independence)}$$

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

Exercise 8.4.3

$$= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\theta)^2\right]\right)$$

$$L(\theta|x) = (2\pi\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}(n-1)s^2\right] \cdot$$

$$\cdot \exp\left[-\frac{1}{2\sigma^2}n(\bar{x}-\theta)^2\right]$$

Then an equivalent likelihood is

$$L'(\theta|x) = \exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)$$

MAXIMUM LIKELIHOOD ESTIMATION

Def. 8.4

Given likelihood $L(\theta|x)$ for our unknown parameter θ and observed data x

Then the parameter value $\hat{\theta}(x) = \hat{\theta}$ at which $L(\theta|x)$ attains its maximum as a function of θ, is called the maximum likelihood estimate of θ (MLE)

Return to example above

$$L'(\theta|x) = \exp\left(\frac{-n}{2\sigma^2}(\bar{x}-\theta)^2\right)$$

$$\frac{d}{d\theta}L'(\theta|x)$$

$$= \exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right) \cdot$$
$$\cdot \left(-\frac{n}{2\sigma^2} \cdot 2(\bar{x}-\theta) \cdot (-1)\right)$$

$$\frac{d}{d\theta} L'(\theta|x) = 0$$

$$\Rightarrow -\frac{n}{2\sigma^2}\left(2(\bar{x}-\theta)(-1)\right) = 0$$

$$\Rightarrow \theta = \bar{x}$$

Maximum?

so: • need to check 2nd derivative < 0
• need to check boundary values for θ on its range

This derivative was complicated ...

Use log-likelihood

$$L(\theta|x) \longrightarrow \log L(\theta|x)$$

log is a monotonic tranformation

$$\max L(\theta|x) \Leftrightarrow \max \log L(\theta|x)$$

$$L'(\theta|x) = \exp\left(-\frac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)$$

$$\log L'(\theta|x) = -\frac{n}{2\sigma^2}(\bar{x}-\theta)^2$$

$X_1, X_2, \ldots, X_n$ - random variables

$$\begin{array}{ccc} X_1 & X_2 & X_n \\ \downarrow \text{pdf} & \downarrow & \downarrow \\ f(x_1) & f(x_2) & f(x_n) \end{array}$$

independent $f(x|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$