Maximum Likelihood Estimators

Parameter estimation

Given independent samples $x_1, x_2, \ldots, x_n$ from a distribution $f(x|\theta)$ (either a PMF or PDF), where θ is a list of the distribution's parameters, estimate θ.

Ex: Given samples HHTHH from independent flips of a coin, estimate $\theta = P(\text{heads})$.

$P(x|\theta)$: Probability of event x given parameters θ.

Viewed as a function of x (θ fixed), it's a probability

Viewed as a function of θ (x fixed), it's called a likelihood and often written $L(x|\theta)$.

Maximum likelihood estimation:

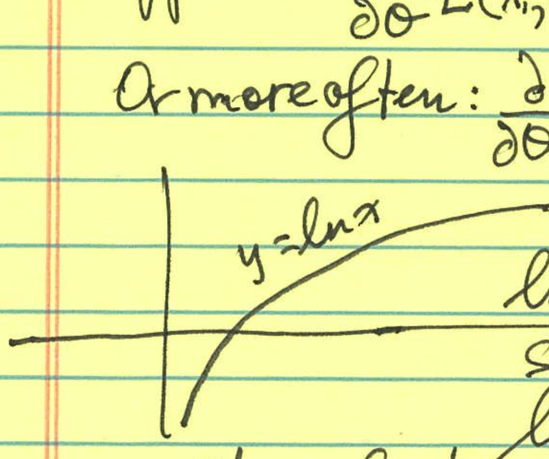
What θ maximizes $\cancel{L(x|\theta)}$ $L(x_1, x_2, \ldots, x_n|\theta)$

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

where $x_1, x_2, \ldots, x_n$ are ind. samples from $f(x|\theta)$?

Approach: $\frac{\partial}{\partial\theta} L(x_1, \ldots, x_n|\theta) = 0$

Or more often: $\frac{\partial}{\partial\theta} \ln L(x_1, \ldots, x_n|\theta) = 0$



ln monotonically increasing, so any θ that maximizes $\ln L$ also maximizes L.

Taking ln turns $\prod$ into $\sum$, and so taking the derivative is much simpler.

First application:

$\theta = P(\text{head})$, n ind. flips $x_1, x_2, \cdots, x_n$ yielding n_0 tails and n_1 heads, $n_0 + n_1 = n$.

$$L(x_1, x_2, \cdots, x_n | \theta) = (1-\theta)^{n_0} \theta^{n_1}$$

$$\ln L(x_1, x_2, \cdots, x_n | \theta) = n_0 \ln(1-\theta) + n_1 \ln(\theta)$$

$$\frac{\partial}{\partial \theta} \ln L(x_1, x_2, \cdots, x_n | \theta) = \frac{-n_0}{1-\theta} + \frac{n_1}{\theta} = 0$$

Let $\hat{\theta}$ be the solution to this.

$$\frac{n_1}{\hat{\theta}} = \frac{n_0}{1-\hat{\theta}}$$

$$n_1(1-\hat{\theta}) = n_0 \hat{\theta}$$

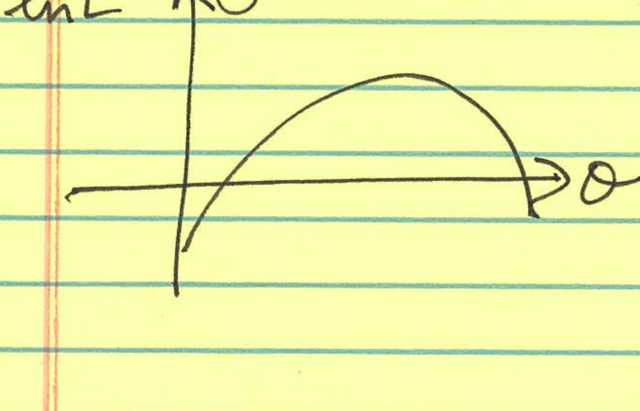
$$n_1 = n_0 \hat{\theta} + n_1 \hat{\theta}$$

$$\hat{\theta} = \frac{n_1}{n_0 + n_1} = \frac{n_1}{n}$$

Is $\theta = \hat{\theta}$ a maximum?

$$\frac{\partial^2}{\partial \theta^2} \ln L(x_1, x_2, \cdots, x_n | \theta) = \frac{-n_0}{(1-\theta)^2} - \frac{n_1}{\theta^2} < 0$$

so $\ln L(x_1, \cdots, x_n | \theta)$ is concave downward everywhere, and $\theta = \hat{\theta}$ is a maximum.



Second application:

$x_i \sim N(\mu, \sigma^2)$, where μ, σ^2 are unknown.

Let $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.

$$L(x_1, x_2, \ldots, x_n \mid \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} e^{-(x_i - \theta_1)^2/(2\theta_2)}$$

$$\ln L(x_1, x_2, \ldots x_n \mid \theta_1, \theta_2) = \sum_{i=1}^{n} \left(-\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \right)$$

$$\frac{\partial}{\partial\theta_1} \ln L(x_1, \ldots, x_n \mid \theta_1, \theta_2) = \sum_{i=1}^{n} + \frac{\cancel{2}(x_i - \theta_1)}{\cancel{2}\theta_2} = 0$$

$$\sum_{i=1}^{n} (x_i - \hat{\theta}_1) = 0$$

$$\sum_{i=1}^{n} x_i - n\hat{\theta}_1 = 0$$

$$\hat{\theta}_1 = \frac{1}{n} \sum_{i=1}^{n} x_i$$

So sample mean is MLE for population mean μ.

$$\frac{\partial^2}{\partial\theta_1^2} \ln L(x_1, \ldots, x_n \mid \theta_1, \theta_2) = \sum_{i=1}^{n} -\frac{1}{\theta_2} = -\frac{n}{\theta_2} < 0,$$

so $\ln L$ is concave downward everywhere as a function of θ_1, so $\theta_1 = \hat{\theta}_1$ is a max.

~~Next~~ Next up:

$$\frac{\partial}{\partial\theta_2} \ln L(x_1, \ldots, x_n \mid \theta_1, \theta_2)$$