Packet 3: Point Estimation

Maximum Likelihood Estimator

Likelihood function (R. A. Fisher, 1922) of a model $f(x \mid \theta)$ is the joint probability density or mass function of the observed data $x = \{x_1, x_2, \ldots, x_n\}$, viewed as a function of θ. For example, if $X = \{X_1, X_2, \ldots, X_n\}$ are continuous r.v.s,

$$L(\theta) = f(x \mid \theta) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta), \text{ if independent.}$$

If the data are discrete r.v.s,

$$L(\theta) = P(X = x \mid \theta) = P(X_1 = x_1, \ldots, X_n = x_n \mid \theta) = \prod_{i=1}^{n} P(X_i = x_i \mid \theta), \text{ if independent.}$$

In this discrete case, the likelihood function is the "probability" that we observe the data $\{X = x\}$ under θ. For example, let's say $L(0.8) \gg L(0.2)$. It means that the probability of observing the current data $P(X = x \mid \theta)$ is much higher when $\theta = 0.8$. So, the data seem to support $\theta = 0.8$ much more than $\theta = 0.2$; the data themselves speak about θ! In general, $L(\theta)$ indicates how likely the observed data are as a function of θ, and maximizing the likelihood function determines the parameters that are most likely to produce the observed data.

Example: We want to know the number of ducks living at Penn State Duck Pond (Hintz Alumni Garden) in this summer, and we count the number of ducks in 3 consecutive days $x = (12, 13, 17)$. Assume the number of observed ducks follows a uniform distribution, Uniform$[0, \theta]$, where θ is the total number of ducks. The p.d.f. of Uniform$[0, \theta]$ is given by

$$f(x \mid \theta) = \frac{1}{\theta} I_{\{0 \le x \le \theta\}}.$$

Which θ most likely generate those three observations?

A: $\theta = 30$, B: $\theta = 20$, C: $\theta = 10$.

Maximum likelihood estimator: A widely used method of obtaining a point estimate for a parameter θ is to find the maximum likelihood estimate (MLE). As the name suggests, the MLE is defined as some value maximizing $L(\theta)$ in the parameter space Ω.

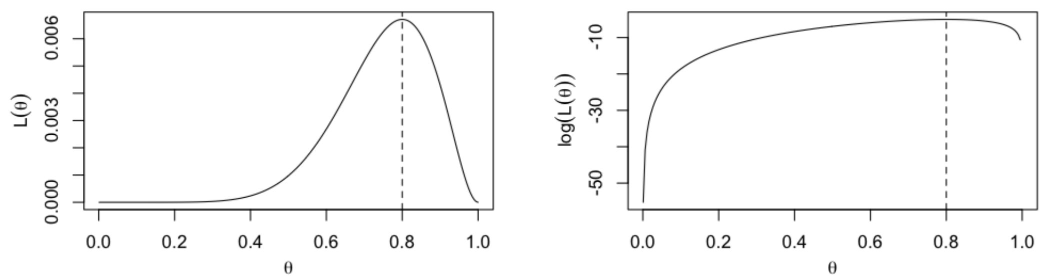
In practice, we obtain the MLE by maximizing $\ell(\theta) = \log(L(\theta))$ instead of maximizing $L(\theta)$ for a few reasons.

1. Since $L(\theta)$ involves a product when the data are independent, it is mathematically more convenient to work with the (natural) logarithm of the likelihood function.
2. The logarithmic function is strictly increasing, preserving the maximizing value, i.e., the value of θ that maximizes $\ell(\theta)$ also maximizes $L(\theta)$.
3. When an analytic solution is not available, we need to find a numerical solution and it is computationally more stable to find the value of θ that maximizes $\ell(\theta)$.

Example: If we knew there were 10 ducks and observed 8 of them on a random day. We assume that $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} \text{Bernoulli}(\theta)$ for some $\theta \in [0, 1]$, where X_i is 1 if we observe duck i and 0 otherwise. We want to find the most likely value of θ that maximizes the probability of observing these data.

Write down the likelihood function and log-likelihood function.

What is the MLE of θ?



Example: The lifetime of a particular type of light bulb can be modeled by an exponential distribution, and its p.d.f. is

$$f(x \mid \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \text{ for } x > 0.$$

Suppose the average lifetime θ is unknown, and we want to estimate it. We independently observe the lifetime of n such light bulbs, $x_1, x_2, \ldots, x_n$. What is the MLE of the expected lifetime θ?