Estimation Theory

What if X is a R.V. and we know (or assume) ~~th~~ the distribution of X belongs to a certain family (e.g., normal, exponential, ...) but one or more parameters is unknown:

- $X \sim N(\mu, 3)$, μ unknown
- $X \sim \text{Binomial}(20, p)$, p unknown
- ~~Neg Binom~~

  $X \sim N(\mu, \sigma^2)$, μ, σ^2 unknown.

How can we estimate the unknown parameter?

Idea: Collect a random sample and form a statistic that estimates the parameter.

Example: $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is an estimate of μ when $X \sim N(\mu, \sigma^2)$.

<u>Notation</u>: $X \sim f(x ; \theta)$

f = pmf or pdf

θ = unknown parameter

Ω = all possible θ = <u>parameter space</u>

$X_1, \ldots, X_n$: random sample ~~[illegible]~~ from f

<u>Example</u>:

① Exponential:

$$f(x ; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0$$

$$\theta \in \Omega = \{\theta : \theta > 0\}$$

② Poisson

$$f(x ; \theta) = e^{-\theta} \cdot \frac{\theta^x}{x!}, \quad x = 1, 2, \ldots$$

$$\theta \in \Omega = \{\theta : \theta > 0\}.$$

Two general kinds of estimation

1. point estimation : estimate a specific value of θ.

2. interval estimation : confidence intervals.

Point Estimation

An <u>estimator</u> is a statistic, i.e. a function of $X_1, \dots, X_n$. The output of this function on a realization $x_1, \dots, x_n$ is called an estimate of θ.

<u>Notation</u> $\hat{\theta}(X_1, \dots, X_n)$ (estimator)

$\hat{\theta}(x_1, \dots, x_n)$ (estimate)

Often we will abbreviate both with $\hat{\theta}$

Measuring Performance

- ~~How Big~~ What is $E\left[\left(\hat{\theta}(X_1, \dots, X_n) - \theta\right)^2\right]$ (mean squared error) where θ is the true value.

- Is $\hat{\theta}$ <u>unbiased</u>? An estimator is unbiased if $E\left[\hat{\theta}(X_1, \dots, X_n)\right] = \theta$.

Maximum Likelihood Estimation

Since $X_1, \dots, X_n$ are independent,

the joint pmf/pdf of $X_1, \dots, X_n$ is

$$f(x_1 ; \theta) f(x_2 ; \theta) \cdots f(x_n ; \theta).$$

Define $L(\theta ; x_1, \dots, x_n)$

$$= f(x_1 ; \theta) f(x_2 ; \theta) \cdots f(x_n ; \theta),$$

viewed as a function of θ, with $x_1, \dots, x_n$ fixed.

Shorthand: $L(\theta) = L(\theta ; x_1, \dots, x_n)$

Idea: After observing $X_1 = x_1$, $X_2 = x_2$, $\dots$, $X_n = x_n$,

use $L(\theta)$ as a basis for estimating θ.

Def The <u>maximum likelihood estimate</u> of θ is

the value of θ that maximizes $L(\theta)$:

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Omega} L(\theta).$$

Example Assume $X_1, \ldots, X_n \sim \exp(\theta)$, θ unknown, and we measure $x_1, \ldots, x_n$.

$$L(\theta) = f(x_1 ; \theta) \cdots f(x_n ; \theta)$$

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

$$= \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta},$$

$$= \frac{1}{\theta^n} e^{-(x_1 + x_2 + \cdots + x_n)/\theta}$$

How can I maximize? First take the ln of $L(\theta)$. Since ln is increasing, this will not change the maximizer.

$$\ln L(\theta) = -n \ln(\theta) - \frac{x_1 + x_2 + \cdots x_n}{\theta}$$

$$\Rightarrow \frac{\partial \ln L(\theta)}{\partial \theta} = \frac{-n}{\theta} + \frac{x_1 + \cdots + x_n}{\theta^2} = 0$$

$$\Longrightarrow \hat{\theta}_{ML} = \frac{x_1 + \cdots + x_n}{n} = \overline{X}.$$

The quantity $\ln L(\theta)$ is called the <u>log likelihood</u> of θ. Since many common pmfs / pdfs involve exponentials ~~expres~~, it is ~~oft~~ often more convenient to work with the log-likelihood. Since ln is monotonically increasing, the maximizer of $L(\theta)$ is the maximizer of $\ln L(\theta)$.

<u>Example</u> $X \sim \text{binom}(1, p)$, p unknown
(Bernoulli trial)

$$L(p) = \prod_{i=1}^{n} f(x_i \mid p)$$

$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$= p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$$

$$\ln L(p) = (\Sigma x_i)\ln p + (n-\Sigma x_i)\ln(1-p)$$

$$\frac{\partial \ln L(p)}{\partial p} = \frac{\Sigma x_i}{p} - \frac{n-\Sigma x_i}{1-p} = 0$$

$$\Rightarrow (1-p)\sum x_i = p\left(n - \sum x_i\right) \Rightarrow \sum x_i = p \cdot n$$

$$\Rightarrow \boxed{\hat{p} = \frac{\sum x_i}{n}}$$

MLE

Are these estimators unbiased? i.e., is $E[\hat{\theta}] = \theta$?

Exponential

We had $\hat{\theta}_{ML} = \frac{\sum x_i}{n}$.

$$E\left[\hat{\theta}_{ML}\right] = E\left[\frac{\sum x_i}{n}\right] = \frac{1}{n} E\left[\sum x_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} \theta$$

mean of exponential

$$= \frac{1}{n} \cdot n\theta = \theta.$$

So yes, $\hat{\theta}_{ML}$ is unbiased in this case.

Technicality: Just because $\frac{\partial L}{\partial \theta} = 0$ at a particular θ does not mean θ is a maximizer of L. It could be a minimizer, or it could be a local max.

To eliminate the possibility of a minimizer, we can confirm $\frac{\partial^2 L}{\partial \theta^2} < 0$. However, local extrema are much more difficult to handle. We will not concern ourselves with these issue in this course.

MLE For the Normal

Assume $X \sim \cancel{N(\mu, \sigma^2)}$ $N(\theta_1, \theta_2)$

$$\cancel{L(\theta) = L(\theta; x_1, \ldots, x_n)}$$

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

$$L(\theta_1, \theta_2) = L(\theta_1, \theta_2 ; x_1, \ldots, x_n)$$

$$= \prod_{i=1}^{n} f(x_i ; \theta_1, \theta_2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} \exp\left(-\frac{(x_i - \theta_1)^2}{2\theta_2}\right)$$

$$= (2\pi\theta_2)^{-\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^{n}(x_i - \theta_1)^2}{2\theta_2}\right)$$

$$\ln L(\theta_1, \theta_2) = -\frac{n}{2} \ln(2\pi\theta_2) - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

We need to maximize this w.r.t (θ_1, θ_2) jointly.

$$\frac{\partial \ln L(\theta_1, \theta_2)}{\partial \theta_1} = -\frac{1}{2\theta_2} \cdot \sum_{i=1}^{n} (-2) \cdot (x_i - \theta_1)$$

$$= \frac{1}{\theta_2} \sum_{i=1}^{n} (x_i - \theta_1) = 0$$

$$\Longrightarrow \sum_{i=1}^{n} (x_i - \theta_1) = 0$$

$$\parallel$$

$$\sum x_i - n\theta_1$$

$$\Longrightarrow \hat{\theta}_1 = \frac{\sum x_i}{n} = \bar{x}$$

$$\frac{\partial \ln L(\theta_1, \theta_2)}{\partial \theta_2} = \frac{-n}{2} \cdot \frac{1}{\theta_2} + \frac{1}{2\theta_2^2} \sum (x_i - \theta_1)^2 = 0.$$

At optimal solution, we know $\theta_1 = \bar{x}$, so plug this in to get

$$\frac{n}{2} \cdot \frac{1}{\theta_2} = \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Or

$$\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}).$$

Are $\hat{\theta}_1$, $\hat{\theta}_2$ unbiased? Let's check:

$$E[\hat{\theta}_1] = E\left[\frac{1}{n}\sum X_i\right]$$

$$= \frac{1}{n}\sum E[X_i] = \frac{1}{n}\cdot\sum \mu$$

$$= \frac{1}{n}\cdot n\mu = \mu.$$

It is a fact (which we did not prove) that $\dfrac{\sum (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$.

Hence

$$E[\hat{\theta}_2] = E\left[\frac{1}{n}\sum (X_i - \bar{X})^2\right]$$

$$= E\left[\frac{\sigma^2}{n}\cdot\frac{\sum (X_i - \bar{X})^2}{\sigma^2}\right]$$

$$= \frac{\sigma^2}{n}\cdot E[\chi^2(n-1)] = \sigma^2\cdot\frac{n-1}{n} \neq \sigma^2$$

So $\hat{\theta}_2$ is <u>biased</u>!

However, note that the

However, note that the sample variance

$$S^2 = \frac{1}{n-1} \Sigma (X_i - \bar{X})^2$$

is unbiased. In particular,

$$E[S^2] = E\left[\frac{1}{n-1}\Sigma(X_i - \bar{X})^2\right]$$

$$= \frac{\sigma^2}{n-1} \cdot E\left[\frac{\Sigma(X_i - \bar{X})^2}{\sigma^2}\right]$$

$$= \frac{\sigma^2}{n-1} \cdot (n-1)$$

$$= \sigma^2$$