## §5.2 Parameter Estimation

Suppose a random variable X has a pdf with unknown parameter $\theta$. We need to estimate the unknown parameter $\theta$ using a random sample of $n$ observations, $X_1, X_2, \cdots, X_n$.

**Example 1.** Suppose we have a biased coin with unknown probability $\theta = P(head)$. We know that it satisfies the Bernoulli distribution $Bernoulli(\theta)$. We toss it 10 times, and get $HTHHTHHTHT$ (6 heads, 4 tails). It is natural to estimate $\theta = 6/10$.

What is the theory behind the estimation? The method is called **Maximum likelihood**. The best estimation for $\theta$ from the sample data(observations) is to maximize the likelihood of getting the sample.

### Maximum Likelihood Estimate (MLE)

Step 1: Find the **likelihood function**, Continuous discrete

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i; \theta) \qquad \text{or} \qquad L(\theta) = \prod_{i=1}^{n} P_X(x_i; \theta)$$

Step 2: To determine the value of $\theta$ that maximize the likelihood, let

$$l(\theta) = \ln L(\theta).$$

Step 3: Set $\frac{dl(\theta)}{d\theta}|_{\theta=\hat{\theta}} = 0$. Solve for the estimator $\hat{\theta}$.

### Definition.

An **estimator** is a RV (or a function) used to generate the estimate.

An **estimate** is the particular value of the estimator for a sample that is actually taken.

**Remark:** $\ln(L(\theta))$ and $L(\theta)$ have the same critical points. Sometimes, it is easy to use $\ln(L(\theta))$ to find critical points of $L(\theta)$.

**Example 1.** (Cont'd).

$$L(\theta) = P(H)P(T)P(H)P(H)P(T)P(H)P(H)P(T)P(H)P(T) = \theta^6(1-\theta)^4$$

We want to maximize $L(\theta)$ by finding a critical point.

$\ell(\theta)$

$\ln(L(\theta)) = \ln(\theta^6(1-\theta)^4) = 6\ln(\theta) + 4\ln(1-\theta)$

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\hat{\theta}} = \frac{6}{\hat{\theta}} + \frac{-4}{1-\hat{\theta}} \overset{set}{=} 0$$

$$\frac{6}{\hat{\theta}} = \frac{4}{1-\hat{\theta}} \Rightarrow 6 - 6\hat{\theta} = 4\hat{\theta}$$

$$6 = 10\hat{\theta} \Rightarrow \boxed{\hat{\theta} = \frac{6}{10}}$$

**Example 2.** Use the sample $Y_1 = 8.2$, $Y_2 = 9.1$, $Y_3 = 10.6$, and $Y_4 = 4.9$ to calculate the maximum likelihood estimate for $\theta$ in the exponential pdf

$$f_Y(y{:}\,\theta) = \theta\, e^{-\theta y}, \; y \geq 0.$$

$$L(\theta) = \prod_{i=1}^{4} f_Y(Y_i\,;\theta)$$

$$= f_Y(8.2)\cdot f_Y(9.1)\cdot f_Y(10.6)\cdot f_Y(4.9)$$

$$= \theta e^{-8.2\theta}\cdot \theta e^{-9.1\theta}\cdot \theta e^{-10.6\theta}\cdot \theta e^{-4.9\theta}$$

$$= \theta^4 e^{-32.8\theta}$$

$$\ell(\theta) = \ln\left(\theta^4 e^{-32.8\theta}\right) = 4\ln(\theta) - 32.8\theta$$

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\hat{\theta}} = \frac{4}{\hat{\theta}} - 32.8 \overset{set}{=} 0$$

$$\boxed{\text{in general: } \hat{\theta} = \frac{n}{\sum_i n}}$$

$$\frac{4}{\hat{\theta}} = 32.8 \Rightarrow \boxed{\hat{\theta} = \frac{4}{32.8} \approx 0.12195}$$

**Example 3.** A random sample of size 8, $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $X_4 = 1$, $X_5 = 0$, $X_6 = 1$, $X_7 = 1$, and $X_8 = 0$, is taken from the probability function

$$p_X(k;\theta) = \theta^k(1-\theta)^{1-k}, \quad k = 0, 1; \quad 0 < \theta < 1$$

Find the maximum likelihood estimate for $\theta$.

$X \sim \text{Bernoulli}(\theta)$

$$L(\theta) = \prod_{i=1}^{8} p_X(x_i;\theta)$$

$$= \theta^5(1-\theta)^3$$

$$\ell(\theta) = 5\ln(\theta) + 3\ln(1-\theta)$$

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\hat\theta} = \frac{5}{\hat\theta} + \frac{-3}{1-\hat\theta} \overset{\text{set}}{=} 0$$

$$\frac{5}{\hat\theta} = \frac{3}{1-\hat\theta}$$

$$5 - 5\hat\theta = 3\hat\theta$$

$$\Rightarrow \boxed{\hat\theta = \frac{5}{8}}$$

$\boxed{\text{in general, } \hat\theta = \frac{\sum_{i=1}^{n} k_i}{n} = \bar{X}}$

$\ln\left(\frac{AB}{C}\right) = \ln(A) + \ln(B) - \ln(C)$

**Example 4.** An experimenter has reason to believe that the pdf describing the variability in a certain type of measurement is the continuous model

$$f_Y(y;\theta) = \frac{y}{\theta^2}\, e^{-\frac{y}{\theta}}, \quad 0 < y < \infty,\ 0 < \theta < \infty.$$

If five data points have been determined $Y_1 = 9.2$, $Y_2 = 5.6$, $Y_3 = 18.4$, $Y_4 = 12.1$ and $Y_5 = 10.7$. What would be a reasonable estimate for the unknown parameter $\theta$?

Finding Estimator

Sample: $X_1, X_2 \dots X_n$

$$L(\theta) = \prod_{i=1}^{n} f_Y(y;\theta) = \prod_{i=1}^{n} \frac{y_i}{\theta^2} e^{\frac{-y_i}{\theta}}$$

$$= \frac{\prod_{i=1}^{n} y_i}{\prod_{i=1}^{n} \theta^2} \cdot \prod_{i=1}^{n} e^{\left(\frac{-y_i}{\theta}\right)} = \frac{\prod_{i=1}^{n} y_i}{\theta^{2n}} \cdot e^{\sum_{i=1}^{n} \frac{-y_i}{\theta}}$$

$$= \frac{\prod_{i=1}^{n} y_i}{\theta^{2n}} \cdot e^{\frac{-1}{\theta}\sum_{i=1}^{n} y_i}$$

$$\ell(\theta) = \ln\left(\prod_{i=1}^{n} y_i\right) - 2n\ln(\theta) - \frac{1}{\theta}\sum_{i=1}^{n} y_i$$

$$\left.\frac{d\ell}{d\theta}\right|_{\hat\theta} = 0 - \frac{2n}{\hat\theta} + \frac{1}{\hat\theta^2} \cdot \sum_{i=1}^{n} y_i \overset{\text{set}}{=} 0$$

Estimator

$$\Rightarrow 2n = \frac{\sum_{i=1}^{n} y_i}{\theta} \Rightarrow \boxed{\hat\theta = \frac{\sum_{i=1}^{n} y_i}{2n} = \frac{\overline{Y}}{2}}$$

**Example 5.** Suppose that $n$ independent (Bernoulli) trials, each of which is a success with probability $p$, are performed. What is the MLE of $p$.

$X \sim \text{Bernoulli}(p) \quad P(X=x) = p^{x}(1-p)^{1-x}$

$x = 0,1$

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

$= (p^{x_1} p^{x_2} p^{x_3} \cdots p^{x_n})((1-p)^{1-x_i} \cdots (1-p)^{1-x_n})$

$= (p^{x_1+x_2 \cdots x_n})(1-p)^{\sum_{i=1}^{n}(1-x_i)}$

$= (p^{\sum_1^1 x_i})(1-p)^{n-\sum_1^1 x_i}$

2) $\ell(p) = (\sum_1^1 x_i)\ln(p) + (n - \sum_1^1 x_i)\ln(1-p)$

$\ln(L(p))$

3) $\left.\dfrac{d\ell}{dp}\right|_{p=\hat{p}} = \dfrac{\sum_1^1 x_i}{\hat{p}} + \dfrac{-(n-\sum_1^1 x_i)}{1-\hat{p}} = \dfrac{\sum_1^1 x_i}{\hat{p}} - \dfrac{n-\sum_1^1 x_i}{1-\hat{p}}$

$\dfrac{\sum_1^1 x_i}{\hat{p}} = \dfrac{n-\sum_1^1 x_i}{1-\hat{p}} \Rightarrow (1-\hat{p})(\sum_1^1 x_i) = \hat{p}(n-\sum_1^1 x_i)$

$= \sum_1^1 x_i - \cancel{p\sum_1^1 x_i} = \hat{p}n - \cancel{\hat{p}\sum_1^1 x_i}$

$\sum_1^1 x_i = n\hat{p} \Rightarrow \boxed{\hat{p} = \dfrac{\sum_1^1 x_i}{n} = \bar{X}}$

**Example 6.** Suppose a random sample of size n is drawn from the probability model

$$p_X(k;\theta) = \frac{\theta^{2k}e^{-\theta^2}}{k!}, \quad k = 0, 1, 2, \ldots$$

Find a formula for the maximum likelihood estimator, $\hat{\theta}$.

$$L(\theta) = \prod_{i=1}^{n} \frac{\theta^{2k_i} e^{-\theta^2}}{(k_i)!}$$

$$= \frac{\prod_{i=1}^{n} \theta^{2k_i} \cdot e^{-\theta^2}}{\prod_{i=1}^{n} (k_i)!} = \frac{\left(e^{-\theta^2}\right)^n \prod_{i=1}^{n} \theta^{2k_i}}{\prod_{i=1}^{n} (k_i)!}$$

$$= \frac{\left(e^{-\theta^2}\right)^n \left(\theta^{\sum_1^n 2k_i}\right)}{\prod_{i=1}^{n} (k_i)!} = \frac{\left(e^{-n\theta^2}\right)\left(\theta^{2\sum_1^n k_i}\right)}{\prod_{i=1}^{n} (k_i)!}$$

$$\ell(\theta) = \ln\left(L(\theta)\right) = \left(2\sum_1^n k_i\right)\ln(\theta) - n\theta^2 \cancel{\ln(e)}^{1}$$
$$- \ln\left(\prod_{i=1}^{n} (k_i)!\right)$$

$$= \left(2\sum_1^n k_i\right)\ln(\theta) - n\theta^2 - \ln\left(\prod_{i=1}^{n} (k_i)!\right)$$

$\leftarrow$ constant wrt. $\theta$

$$\left.\frac{d\ell}{d\theta}\right|_{\theta=\hat{\theta}} = \left(\cancel{2}\sum_{i=1}^{n} k_i\right)\frac{1}{\hat{\theta}} - \cancel{2}n\hat{\theta} - 0 \overset{set}{=} 0$$

$$\sum_{i=1}^{n} k_i \cdot \frac{1}{\hat{\theta}} - n\hat{\theta} = 0$$

$$n\hat{\theta}^2 = \sum_{i=1}^{n} k_i \Rightarrow \hat{\theta} = \sqrt{\frac{\sum_{i=1}^{n} k_i}{n}} = \sqrt{\bar{X}}$$

## MLE when two parameters are unknown.

Steps: 1. Find the likelihood function,

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

2. To determine the values of $\theta_1$, $\theta_2$ that maximize the likelihood, Set

$$l(\theta_1, \theta_2) = \ln\ L(\theta_1, \theta_2)$$

3. Find $\dfrac{\partial\, l\,(\theta_1, \theta_2)}{\partial\, \theta_i}\Big|_{\theta_i = \widehat{\theta}_i} = 0,\ i = 1, 2$, and solve the system for $\widehat{\theta}_1$ and $\widehat{\theta}_2$.

**Example 7.** Suppose a random sample of size n is drawn from the two-parameter normal pdf

$$f_Y(y; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sqrt{\sigma^2}} e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}} \qquad -\infty < y < \infty; -\infty < \mu < \infty; \sigma^2 > 0$$

Use the method of maximum likelihood to find formulas for estimators $\mu_e$ and $\sigma_e^2$.

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2}\cdot\frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(y_i-\mu)^2}$$

$$l(\mu, \sigma^2) = \frac{-n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\cdot\frac{1}{\sigma^2}\cdot\sum_{i=1}^{n}(y_i-\mu)^2$$

linearity of derivatives

$$\frac{\partial l}{\partial \mu}\Big|_{\mu=\mu_e} = \frac{-\cancel{1}}{\cancel{2}\sigma^2}\sum_{i=1}^{n} \cancel{-2}(y_i-\mu_e) \overset{set}{=} 0$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu_e) = 0$$

$$\Rightarrow \sum_{i=1}^{n}(y_i-\mu_e) = 0$$

$$= \sum_{i=1}^{n} y_i - n\mu_e = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i = n\mu_e \Rightarrow \boxed{\mu_e = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{Y}}$$

$$\frac{\partial l}{\partial \sigma^2} = \frac{-n}{\cancel{2}}\,\frac{\cancel{2\pi}}{\cancel{2\pi}\sigma^2} - \frac{1}{\cancel{2}}\sum_{i=1}^{n}(y_i-\mu)^2 \cdot \frac{-1}{(\sigma^2)^2} \overset{set}{=} 0$$

$$\Rightarrow \frac{n}{\sigma^2} = \frac{\sum_{i=1}^{n}(y_i-\mu)^2}{(\sigma^2)^2}$$

$$\Rightarrow \boxed{\sigma_e^2 = \frac{\sum_{i=1}^{n}(y_i-\bar{Y})^2}{n}}$$

$$\sigma_e = \sqrt{\sigma_e^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{Y})^2}{n}}$$