# Section 4.1 (con't.)

01/10/2019

## 1 Point Estimators

Note: The MLE works for more than one parameter: $\theta = [\theta_1, \theta_2]$

Example 4.1.3

Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$. Let $\theta = [\theta_1, \theta_2] = \left[\mu, \sigma^2\right]$. Then

$$f(x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L(\theta) = \prod f(x_i; \theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}$$

$$l(\theta) = \ln L(\theta) = \prod f(x_i; \theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum(x_i - \mu)^2$$

Taking partial derivatives with respect to $\mu$ and $\sigma^2$, respectively

1. $\frac{\partial}{\partial\mu} l(\theta) = -\frac{1}{2\sigma^2}\sum 2(x_i - \mu)(-1)$
2. $\frac{\partial}{\partial\sigma^2} l(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum(x_i - \mu)^2$

Setting both equations to 0 and solving for $\hat{\mu}$ and $\hat{\sigma}^2$,

$$\sum(x_i - \hat{\mu}) = 0 \quad \Rightarrow \quad \sum x_i = n\hat{\mu} \quad \Rightarrow \quad \hat{\mu} = \overline{x}$$

$$\frac{n}{\hat{\sigma}^2} = \frac{1}{(\hat{\sigma}^2)^2}\sum(x_i - \overline{x})^2 \quad \Rightarrow \quad \hat{\sigma}^2 = \frac{\sum(x_i - \overline{x})^2}{n}$$

so the MLE is

$$\hat{\theta} = \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \overline{x} \\ \frac{\sum(x_i - \overline{x})^2}{n} \end{bmatrix}$$

Comments:

- MLE of $\sigma^2$ is not unbiased.
- If we let $\theta_2 = \sigma$ instead of $\theta_2 = \sigma^2$, we can show that ...

$$\text{MLE of } \hat{\sigma} = \sqrt{\frac{\sum(x_i - \overline{x})^2}{n}}$$

so that the MLE of $\sigma^2 = [\text{MLE of } \sigma]^2$.

This is a general property of the MLE, i.e.

$$\text{MLE of } g(\theta) = g(\text{MLE of } \theta)$$

This is useful because the quantities we want to estimate often turn out to be functions of simple parameters.

Example:

Given a random sample $x_1, \ldots, x_n$, estimate the percentage of population less than a certain value, say, 10.0

1. Exponential

   $F(10) = 1 - e^{-10/\theta}$ which is a function $g(\theta)$.

   MLE of $F(10)$ is then $g(\hat{\theta}) = 1 - e^{-10/\overline{x}}$

2. $N(\mu, \sigma^2)$

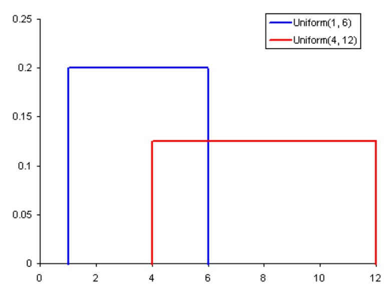
   $F(10) = \Phi\left(\frac{10-\mu}{\sigma}\right)$ which is a function $g(\theta_1, \theta_2)$.

   MLE of $F(10)$ is $g(\hat{\theta}_1, \hat{\theta}_2) = \Phi\left(\frac{10-\overline{x}}{\sqrt{\sum(x_i-\overline{x})^2/n}}\right)$

Comment: Be careful about taking derivatives

Example 4.1.4

Let $(x_1, x_2, x_3, x_4) = (7.9, 10.5, 4.2, 7.1)$ be a random sample from Unif(0, $\theta$). Find MLE of $\theta$.




$$f(x_i; \theta) = \begin{cases} \frac{1}{\theta}, & \text{if } 0 \leq x_i \leq \theta \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned}
L(\theta) &= \begin{cases} \left(\frac{1}{\theta}\right)^4, & \text{if } 0 \leq x_1 \leq \theta,\ 0 \leq x_2 \leq \theta,\ 0 \leq x_3 \leq \theta,\ 0 \leq x_4 \leq \theta, \\ 0, & \text{otherwise} \end{cases} \\
&= \begin{cases} 0, & \text{if } \theta = 9.0 \\ 0, & \text{if } \theta = 10.0 \\ (1/11)^4, & \text{if } \theta = 11.0 \\ (1/12)^4, & \text{if } \theta = 12.0 \end{cases} \\
&= \begin{cases} 0, & \text{if } \theta < 10.5 \\ (1/\theta)^4, & \text{if } \theta =\geq 10.5 \end{cases}
\end{aligned}$$

which is maximum at $\hat{\theta} = 10.5$. In general, for $x_1, \ldots, x_n$ from Unif(0, $\theta$), the MLE is

$$\hat{\theta}_{\text{MLE}} = \max(x_1, \ldots, x_n)$$

## 2 Estimating a density (nonparametric estimates vs MLE)

Case 1: Discrete $p(x)$

Example 4.1.6

For $j = 1, 2, 3, 4, 5, 6$

$$\begin{aligned} \hat{p}(j) &= \text{proportion of sample equal to } j \\ &= \frac{\#\{x_i = j\}}{n} \\ &= \frac{\sum \mathbf{I}(x_i = j)}{n} \end{aligned}$$

where $\mathbf{I}(E) = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{otherwise} \end{cases}$

For example, $p(3) = 5/30 = .167$. Note that this is not the same as the MLE estimate. (HW 2: Prob. 4.1.8.)

Case 2: Continuous $f(x)$.

$$\begin{aligned} P[x - h < X < x + h] &= \int_{x-h}^{x+h} f(t)dt \\ &= 2hf(\epsilon) \text{ for some } \epsilon \text{ in the interval } [x - h, x + h] \\ &\doteq 2hf(x) \end{aligned}$$

Therefore

$$\begin{aligned} \hat{f}(x) &= \frac{\hat{P}(x - h < X < x + h)}{2h} \\ &= \frac{\#\{x - h < x_i < x + h\}}{n2h} \\ &= \frac{1}{n2h} \sum_{i=1}^{n} \mathbf{I}(x - h \le x_i \le x + h) \end{aligned}$$

Comments:

1. This is a kernel density estimator (KDE) using a rectangular kernel
2. May be generalized to nonrectangular 'smoother' kernels
3. $2h$ is called the *bandwidth*. There is a lot of research on optimal choice of bandwidth.

(Example 4.1.7 of 7th Ed)

```
[Workspace loaded from ~/.RData]

> ex417<-c(63,58,60,60,39,41,57,49,44,36,52,48,44,19,42,67,44,64,34,46)
> hist(ex417)
> density(ex417)

Call:
density.default(x = ex417)

Data: ex417 (20 obs.); Bandwidth 'bw' = 5.92

       x                  y
 Min.   : 1.241   Min.   :3.788e-05
 1st Qu.:22.120   1st Qu.:2.084e-03
 Median :43.000   Median :7.240e-03
 Mean   :43.000   Mean   :1.196e-02
 3rd Qu.:63.880   3rd Qu.:2.351e-02
 Max.   :84.759   Max.   :3.109e-02
> plot(density(ex417))
```