Section 4.1

01/08/2019

1 Sampling and Statistics

Two sets of assumptions

Case 1: $f(x)$ or $p(x)$ is completely unknown

Case 2: $f(x)$ or $p(x)$ is known up to a set of unknown parameters

Example: Maternal Flu Vaccine (handout)

1. $p(x) \sim \text{Binomial}(n, p) \quad \leftarrow$ "Birthweight < 2.5 kg"
2. $f(x) \sim N(\mu, \sigma^2) \quad \leftarrow$ "Birthweight"
3. $f(x) \sim \text{Exp}(\lambda)$

Definition 4.1.1
Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables. Then $X_1, X_2, \ldots, X_n$ is called a *random sample* of size n from the common distribution $f(\cdot)$.

Definition 4.1.2
Any function $T(X_1, X_2, \ldots, X_n)$ is called a *statistic*. Note that $T(X_1, X_2, \ldots, X_n)$ cannot be a function of unknown parameters.

Definition 4.1.3
If $E[T(X_1, X_2, \ldots, X_n)] = \theta$ then T is an *unbiased estimator of* θ.

Example: Sample mean is unbiased for population mean

$$\begin{aligned}
E(T(X_1, X_2, \ldots, X_n) &= E\left(\frac{X_1 + \cdots + X_n}{n}\right) \\
&= E\left(\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n\right) \\
&= \frac{1}{n}E(X_1) + \cdots + \frac{1}{n}E(X_n) \\
&= \frac{1}{n}\mu_1 + \cdots + \frac{1}{n}\mu_n \\
&= \mu
\end{aligned}$$

2 Maximum likelihood estimation

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with distribution $f(x;\theta)$. The joint distribution of the sample is

$$\begin{aligned} f(x_1,\ldots,x_n) &= f(x_1;\theta)\cdots f(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta) \end{aligned}$$

The *likelihood function* is

$$\begin{aligned} L(\theta) &= f(x_1;\theta)\cdots f(x_n;\theta) \\ &= \prod_{i=1}^{n} f(x_i;\theta) \end{aligned}$$

which looks the same, but is seen as a function of the parameter θ.
The *maximum likelihood estimate* is the specific value of θ, denoted $\hat{\theta}$, that maximizes $L(\theta)$.

Example 4.1.1 (p.205/227)
Suppose $X \sim \text{Exp}(\theta)$, i.e.

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

and is zero otherwise. Consider the air conditioner data from textbook

$$359, 413, 25, 130, 90, 50, 50, 487, 102, 194, 55, 74, 97$$

where the values represent days of operation until service is needed.

```
ophrs<-c(359,413,25,130,90,50,50,487,102,194,55,74,97)
> mean(ophrs)
[1] 163.5385
> dexp(ophrs,rate=1/150)
 [1] 0.0006088317 0.0004247675 0.0056432115 0.0028023359 0.0036587442
 [6] 0.0047768754 0.0047768754 0.0002593578 0.0033774466 0.0018290316
[11] 0.0046202708 0.0040705847 0.0034919256
> prod(dexp(ophrs,rate=1/150))
[1] 3.592636e-35
> prod(dexp(ophrs,rate=1/160))
[1] 3.76489e-35
> prod(dexp(ophrs,rate=1/163))
[1] 3.776437e-35
> prod(dexp(ophrs,rate=1/163.5385))
[1] 3.776704e-35
> prod(dexp(ophrs,rate=1/164))
[1] 3.776509e-35
> prod(dexp(ophrs,rate=1/165))
```

```
[1] 3.774767e-35
>
> # Using FOR loop
>
> store<-rep(0,170)
> for (i in 150:170){store[i]<-prod(dexp(ophrs,rate=1/i))}
> cbind(seq(150,170),store[150:170])
      [,1]         [,2]
 [1,]  150 3.592636e-35
 [2,]  151 3.619624e-35
 [3,]  152 3.644391e-35
 [4,]  153 3.666951e-35
 [5,]  154 3.687319e-35
 [6,]  155 3.705517e-35
 [7,]  156 3.721568e-35
 [8,]  157 3.735499e-35
 [9,]  158 3.747341e-35
[10,]  159 3.757127e-35
[11,]  160 3.764890e-35
[12,]  161 3.770669e-35
[13,]  162 3.774504e-35
[14,]  163 3.776437e-35
[15,]  164 3.776509e-35
[16,]  165 3.774767e-35
[17,]  166 3.771256e-35
[18,]  167 3.766024e-35
[19,]  168 3.759119e-35
[20,]  169 3.750591e-35
[21,]  170 3.740489e-35
```

We can analytically, rather than numerically, show that the likelihood function is maximized at the sample mean. The MLE maximizes

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} (\frac{1}{\theta}) e^{-x_i//theta}$$
$$= (1/\theta)^n e^{-\sum_{i=1}^{n} x_i//theta}$$

or maximizes $l(\theta) = \ln[L(\theta)]$

$$l(\theta) = -n\ln(\theta) - \frac{\sum_{i=1}^{n} X_i}{\theta}$$
$$\frac{\partial}{\partial\theta} l(\theta) = \frac{-n}{\theta} + \frac{\sum_{i=1}^{n} X_i}{\theta^2}$$

Equating the RHS to 0, and denoting the solution by $\hat{\theta}$,

$$\frac{-n}{\hat{\theta}} = -\frac{\sum_{i=1}^{n} X_i}{\hat{\theta}^2}$$

Hence

$$\hat{\theta} = \frac{\sum_{i=1}^{n} X_i}{n} = \overline{X}$$