# Chapter 6

# Estimations

## 6.1 Point Estimation

*In previous chapters, we often use sample mean* $\overline{x}$ *and sample variance* $s^2$ *from a sample to estimate the mean and the variance of a distribution. How to make these estimations? How good are these estimates? What makes an estimate good?*

*Suppose a random variable* $X$ *is known to follow certain kind of distribution. However, one parameter* $\theta$ *of the distribution is unknown, e.g. exponential distribution* $f(x;\theta) = (1/\theta)e^{-x/\theta}$, $0 < \theta < \infty$. *The space of the unknown* $\theta$ *is the* **parameter space**.

*To find* $\theta$, *we repeat the experiment* $n$ *independent times, observe the sample,* $X_1, \cdots, X_n$, *and estimate* $\theta$ *by a certain function* $u(X_1, \cdots, X_n)$ *on the observed values* $x_1, \cdots, x_n$. *The function/statistic* $u(X_1, \cdots, X_n)$ *is called an* **estimator / point estimator** *of* $\theta$.

**Ex** *Suppose* $X \sim b(1, p)$ *so that the p.m.f. of* $X$ *is*

$$f(x;p) = p^x(1-p)^{1-x}, \qquad x = 0, 1, \qquad 0 \leq p \leq 1.$$

*To estimate* $p$, *we repeat the experiment* $n$ *independent times to get a random sample* $X_1, \cdots, X_n$. *Given the observed values* $x_1, \cdots, x_n$, *the theoretical probability that* $X_1, \cdots, X_n$ *takes these particular value is*

$$\begin{aligned} P(X_1 = x_1, \cdots, X_n = x_n) &= \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum x_i}(1-p)^{n-\sum x_i} =: L(p). \end{aligned}$$

*The above joint p.m.f.* $L(p)$ *as a function of* $p$ *is called the* **likelihood function**. *We find a* $p \in [0,1]$ *that maximizes* $L(p)$. *To do so, we solve* $L'(p) = 0$ *and check if* $L''(p) < 0$. *Solving* $L'(p) = 0$ *gives* $p = (\sum_{i=1}^{n} x_i)/n = \overline{x}$, *and it satisfies that* $L''(p) < 0$. *The corresponding statistic*

$$\widehat{p} := \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

*is called the* **maximum likelihood estimator**.

*It is often easier to maximize* $\ln L(p)$ *rather than* $L(p)$. *Both produce the same maximum likelihood estimator. Note that* $[\ln L(p)]' = L'(p)/L(p)$.

*Similar method could be used to estimate several unknown parameters* $\theta_1, \cdots, \theta_n$.

**Ex** *1-3, p.276-278 (scanned file)*

**Def.** *Let* $u(X_1, \cdots, X_n)$ *be an estimator of* $\theta$. *If* $E[u(X_1, \cdots, X_n)] = \theta$, *then the statistic* $u(X_1, \cdots, X_n)$ *is called an* **unbiased estimator** *of* $\theta$. *Otherwise, it is said to be* **biased**.

**Ex** *4, p.278 (scanned file)*

*Numerically, we may use* **method of moments** *to estimate the parameters. That is, to equate the first sample moment to the first theoretical moment, then the second sample moment to the second theoretical moment if needed, and so on.*

**Ex** *5-6, p.279-281 (scanned file)*

**Homework**

§*6.1* *1, 3, 5, 9, 11, 15*

**Attachment:** *Scanned textbook pages of Section 6-1*