6. Point Estimation

Chapter 6: Point Estimation

- **6.1. Some General Concepts of Point Estimation**
- **6.2. Methods of Point Estimation**

6.1 Some General Concepts of Point Estimation

- In order to get some population characteristics, statistical inference needs obtain sample data from the population under study, and achieve the conclusions can then be based on the computed values of various sample quantities (statistics).

- Typically, we will use the Greek letter θ for the parameter of interest. The objective of point estimation is to select a single number, based on sample data (statistic $\hat{\theta}$), that represents a sensible value for θ.

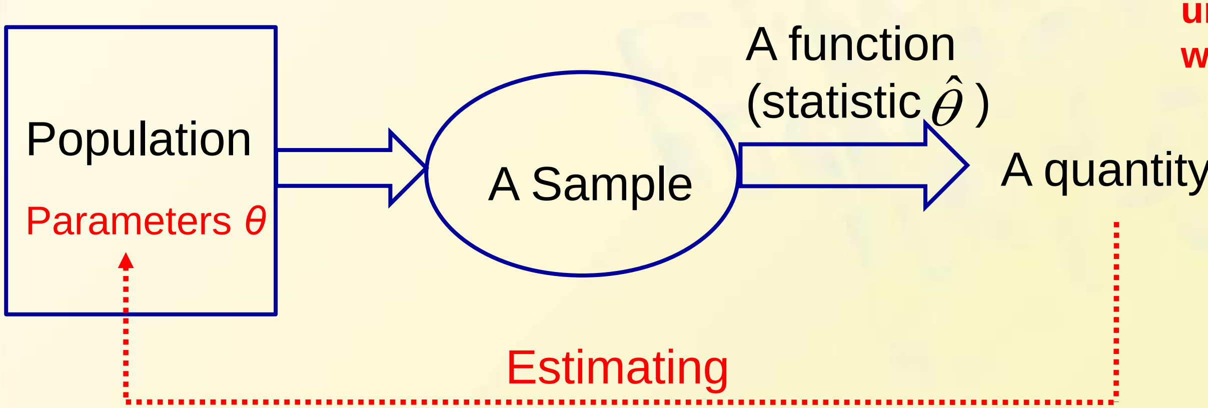
6.1 Some General Concepts of Point Estimation

- Point Estimation

A point estimate of a parameter θ is a *single* number that can be regarded as a sensible value for θ.

A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ.

Here, the type of population under study is usually known, while the parameters are unkown.



Q #1: How to get the candidate estimators based on the population?

Q #2: How to measure the candidate estimators?

6.1 Some General Concepts of Point Estimation

- Example 6.1

The manufacturer has used this bumper in a sequence of 25 controlled crashes against a wall, each at 10 mph, using one of its compact car models. Let X = the number of crashes that result in no visible damage to the automobile. What is a sensible estimate of the parameter p = the proportion of all such crashes that result in no damage

If X is observed to be $x = 15$, the most reasonable estimator and estimate are

$$\text{estimator} \quad \hat{p} = \frac{X}{n} \qquad\qquad \text{estimate} = \frac{x}{n} = \frac{15}{25} = 0.60$$

- **Example 6.2**

Reconsider the accompanying 20 observations on dielectric breakdown voltage for pieces of epoxy resin first introduced in Example 4.29 (pp. 193)

24.46	25.61	26.25	26.42	26.66	27.15	27.31	27.54	27.74	27.94
27.98	28.04	28.28	28.49	28.50	28.87	29.11	29.13	29.50	30.88

The pattern in the normal probability plot given there is quite straight, so we now assume that the distribution of breakdown voltage is **normal** with mean value μ. Because normal distribution are symmetric, μ is also the median lifetime of the distribution. The given observation are then assumed to be the result of a random sample X_1, X_2, …, X_{20} from this normal distribution.

6.1 Some General Concepts of Point Estimation

- **Example 6.2 (Cont')**

Consider the following estimators and resulting estimates for μ

a. Estimator= $\overline{X}$, estimate= $\bar{x} = \sum x_i / n = 555.86/20 = \boxed{27.793}$

b. Estimator= $\tilde{X}$, estimate= $\tilde{x} = (27.94 + 27.98)/2 = \boxed{27.960}$

c. Estimator [min(X_i) + max(X_j)]/2 = the average of the two extreme lifetimes, estimate=[min(x_i)+max(x_i)]/2 = (24.46+30.88)/2 = $\boxed{27.670}$

d. Estimator = $\overline{X}_{tr(10)}$, the 10% trimmed mean (discard the smallest and largest 10% of the sample and then average)

$$\text{estimate} = \bar{x}_{tr(10)} = \frac{555.86 - 24.46 - 25.61 - 29.50 - 30.88}{16} = \boxed{27.838}$$

6.1 Some General Concepts of Point Estimation

- **Example 6.3**

In the near future there will be increasing interest in developing low-cost Mg-based alloys for various casting processes. It is therefore important to have practical ways of determining various mechanical properties of such alloys. Assume that the observations of a random sample X_1, X_2, …, X_8 from the population distribution of elastic modulus under such circumstances. We want to estimate the population variance σ^2

Method #1: sample variance

$$\hat{\sigma}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1} = \frac{\sum X_i^2 - \left(\sum X_i\right)^2 / n}{n-1} \qquad \hat{\sigma}^2 = S^2 = \frac{\sum X_i^2 - \left(\sum X_i\right)^2 / 8}{7} \boxed{\approx 0.251}$$

Method #2: Divided by n rather than n-1

$$\hat{\sigma}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n} = \frac{\sum X_i^2 - \left(\sum X_i\right)^2 / n}{n} \qquad \hat{\sigma}^2 = S^2 = \frac{\sum X_i^2 - \left(\sum X_i\right)^2 / 8}{8} \boxed{\approx 0.220}$$

Estimation Error Analysis

Note that $\hat{\theta}$ is a function of the sample X_i's, so it is a random variable.

$$\hat{\theta} = \theta + \text{error of estimation}$$

Therefore, an accurate estimator would be one resulting in small estimation errors, so that estimated values will be near the true value θ (unkown).

A good estimator should have the two properties:

1. unbiasedness (*i.e.* the average error should be zero)
2. minimum variance (*i.e.* the variance of error should be small)

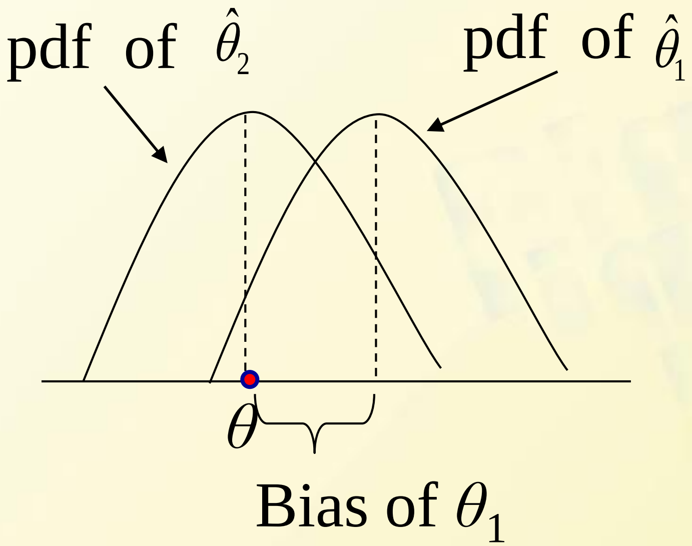
6.1 Some General Concepts of Point Estimation

- Unbiased Estimator (无偏估计)

A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if

$E(\hat{\theta}) = \theta$ for every possible value of θ.

If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the bias of $\hat{\theta}$



Note: "centered" here means the expected vaule, not the median, of the distribution of $\hat{\theta}$ is equal to θ

6.1 Some General Concepts of Point Estimation

- Proposition

When X is a binomial rv with parameters n and p, the sample proportion $\hat{p}$ $=X/n$ is an unbiased estimator of p.

Refer to Example 6.1, the sample proportion X/n was used as an estimator of p, where X, the number of sample successes, had a binomial distribution with parameters n and p, thus

$$E(\hat{p}) = E(\frac{X}{n}) = \frac{1}{n}E(X) = \frac{1}{n}(np) = p$$

6.1 Some General Concepts of Point Estimation

- **Example 6.4**

Suppose that X, the reaction time to a certain stimulus, has a uniform distribution on the interval from 0 to an unknown upper limit θ. It is desired to estimate θ on the basis of a random sample $X_1, X_2, \ldots, X_n$ of reaction times. Since θ is the largest possible time in the entire population of reaction times, consider as a first estimator the largest sample reaction time:

$$\hat{\theta}_1 = \max(X_1, X_2, \cdots, X_n)$$ biased estimator, why?

Since $E(\hat{\theta}_1) = \dfrac{n}{n+1} \cdot \theta < \theta$ (refer to Ex. 32 in pp. 279)

Another estimator $\hat{\theta}_2 = \dfrac{n+1}{n} \cdot \max(X_1, X_2, \cdots, X_n)$ unbiased estimator

$$E(\hat{\theta}_2) = \frac{n+1}{n} \cdot \frac{n}{n+1} \cdot \theta = \theta$$

6.1 Some General Concepts of Point Estimation

- **Proposition**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the estimator

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

is an unbiased estimator of σ^2, namely $E(S^2) = \sigma^2$

Refer to pp. 245 for the proof.

However,

$$E\left(\frac{\sum (X_i - \bar{X})^2}{n}\right) = \frac{n-1}{n} E(S^2) = \frac{n-1}{n}\sigma^2 < \sigma^2$$

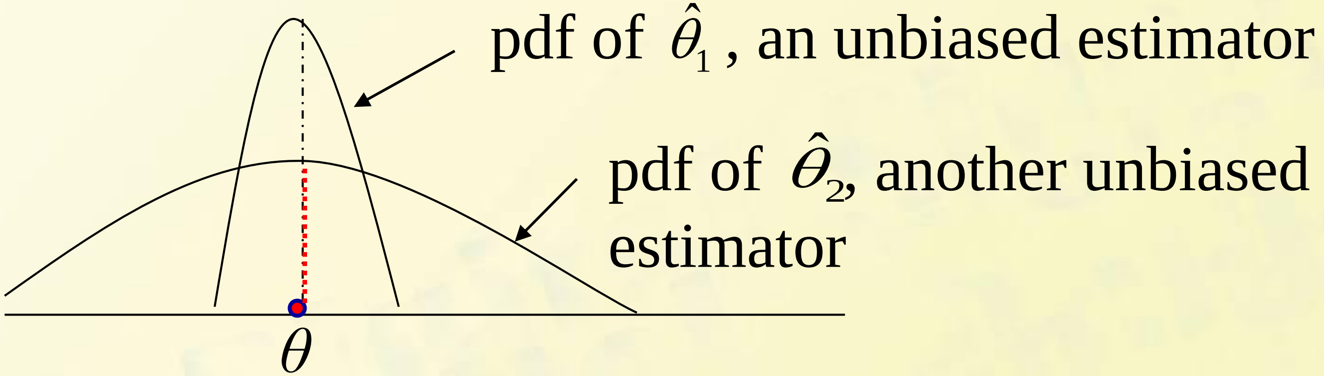
6.1 Some General Concepts of Point Estimation

- **Proposition**

If X_1, X_2,…X_n is a random sample from a distribution with mean μ, then $\overline{X}$ is an unbiased estimator of μ. If in addition the distribution is **continuous and symmetric**, then $\tilde{X}$ and any trimmed mean are also **unbiased** estimator of μ

Refer to the estimators in Example 6.2

- **Estimators with Minimum Variance**

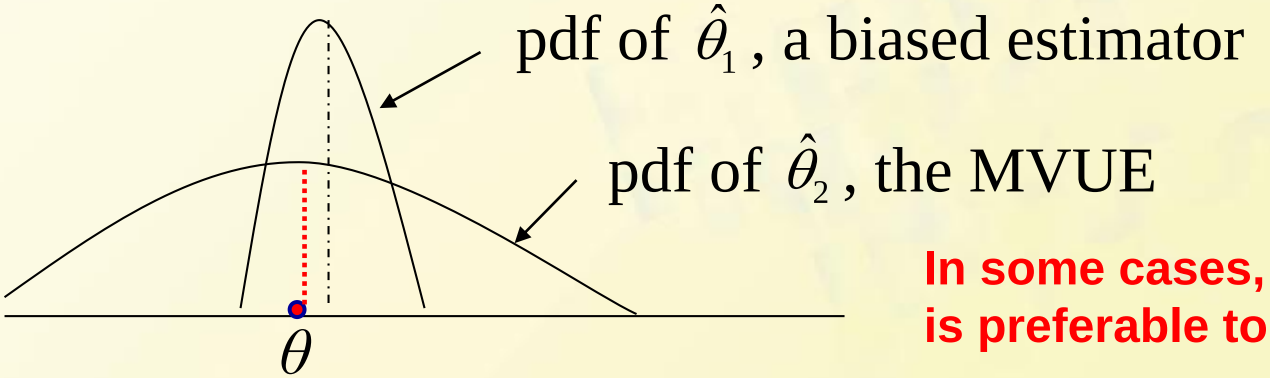


Obviously, the estimator $\hat{\theta}_1$ is better than the $\hat{\theta}_2$ in this example

6.1 Some General Concepts of Point Estimation

- **Estimator Selection**

  - When choosing among several different estimators of θ, select one that is unbiased.
  - Among all estimators of θ that are unbiased, choose the one that has **minimum variance**. The resulting $\hat{\theta}_1$ is called the minimum variance unbiased estimator (MVUE) of θ.



In some cases, a biased estimator is preferable to the MVUE

6.1 Some General Concepts of Point Estimation

- **Example 6.6 (Ex. 6.4 Cont')**

When $X_1, X_2, \ldots X_n$ is a random sample from a uniform distribution on $[0, \theta]$, the estimator

$$\hat{\theta} = \frac{n+1}{n} \cdot \max(X_1, X_2, \ldots, X_n)$$

is unbiased for θ

It is also shown that $\hat{\theta} = \frac{n+1}{n} \cdot \max(X_1, X_2, \ldots, X_n)$ is the MVUE of θ.

6.1 Some General Concepts of Point Estimation

- **Theorem**

  Let $X_1, X_2, \ldots, X_n$ be a random sample from a **normal distribution** with parameters μ and σ. Then the estimator

$$\hat{\mu} = \overline{X}$$

is the MVUE for μ.

How about those un-normal distributions?

Example 6.7

Suppose we wish to estimate the thermal conductivity μ of a certain material. We will obtain a random sample X_1, X_2, …, X_n of n thermal conductivity measurements. Let's assume that the population distribution is a member of one of the following three families:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

Gaussian Distribution

$$f(x) = \frac{1}{\pi[1+(x-\mu)^2]} \qquad -\infty < x < \infty$$

Cauchy Distribution

$$f(x) = \begin{cases} \dfrac{1}{2c} & -c \le x - \mu \le c \\ 0 & \text{otherwise} \end{cases}$$

Uniform Distribution

6.1 Some General Concepts of Point Estimation

1. If the random sample comes from a normal distribution, then $\overline{X}$ is the best of the four estimators, since it is the MVUE.

2. If the random sample comes from a Cauchy distribution, then $\overline{X}$ and $\overline{X}_e$ (the average of the two extreme observations) are terrible estimators for μ, whereas $\tilde{X}$ is quite good; $\overline{X}$ is bad because it is very sensitive to outlying observations, and the heavy tails of the Cauchy distribution make a few such observation likely to appear in any sample.

3. If the underlying distribution is uniform, the best estimator is $\overline{X}_e$; this estimator is greatly influenced by outlying observations, but the lack of tails makes such observations impossible.

4. The trimmed mean is best in none of these three situations, but works reasonably well in all three. That is, $\overline{X}_{tr(10)}$ does not suffer too much in any of the three situations.

A Robust estimator

6.1 Some General Concepts of Point Estimation

- The Standard Error

  The standard error of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$.

6.2 Methods of Point Estimation

- Two “constructive” methods for obtaining point estimators

➢ **Method of Moments**

➢ **Maximum Likehood Estimation**

6.2 Methods of Point Estimation

- **The Method of Moment**

- **The basic idea of this method is equate certain sample characteristics, such as the mean, to the corresponding population expected values. Then solving theses equations for unknown parameter values yields the estimators.**

6.2 Methods of Point Estimation

- **Moments**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a pmf or pdf $f(x)$. For $k = 1, 2, 3, \ldots$, the kth population moment, or kth moment of the distribution $f(x)$, is $E(X^k)$. The kth sample moment is .

$$(1/n)\sum_{i=1}^{n} X_i^k$$

6.2 Methods of Point Estimation

- **Moment Estimator**

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f(x;\theta_1,\ldots,\theta_m)$, where $\theta_1,\ldots,\theta_m$ are parameters whose values are unknown. Then the moment estimators $\hat{\theta}_1,\ldots,\hat{\theta}_m$ are obtained by equating the first m sample moments to the corresponding first m population moments and solving for $\theta_1,\ldots,\theta_m$.

With the given sample x_i

$$(1/n)\sum_{i=1}^{n} X_i^k \xrightarrow{\text{n is large}} E(X^k)$$

With unkonwn θ_i

K-th sample moment

K-th population moment

6.2 Methods of Point Estimation

General Algorithm :

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \ldots, \theta_m) \\ \mu_2 = \mu_2(\theta_1, \theta_2, \ldots, \theta_m) \\ \cdots \\ \mu_m = \mu_m(\theta_1, \theta_2, \ldots, \theta_m) \end{cases} \Rightarrow \begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \ldots, \mu_m) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \ldots, \mu_m) \\ \cdots \\ \theta_m = \theta_m(\mu_1, \mu_2, \ldots, \mu_m) \end{cases}$$

The first m population moments

The solution of equations

Use the first m sample moment $A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l, l = 1, 2 \ldots m$

to represent the population moments μ_l

$$\hat{\theta}_i = \hat{\theta}_i(A_1, A_2, \ldots, A_m), i = 1, \ldots, m$$

- **Example 6.12**

Let $X_1, X_2, \ldots, X_n$ represent a random sample of service times of n customers at a certain facility, where the underlying distribution is assumed exponential with parameter λ. How to estimate λ by using the method of moments?

Step #1: The 1st population moment $E(X) = 1/\lambda$

then we have $\lambda = 1/E(X)$

Step #2: Use the 1st sample moment $\overline{X}$ to represent 1st population moment $E(X)$, and get the estimator

$$\hat{\lambda} = 1/\overline{X}$$

6.2 Methods of Point Estimation

- **Example 6.13**

  Let $X_1, \ldots, X_n$ be a random sample from a gamma distribution with parameters α and β. Its pdf is

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  There are two parameter need to be estimated, thus, consider the first two moments

- **Example 6.13 (Cont')**

Step #1: $E(X) = \mu = \alpha\beta$

$$E(X^2) = V(X) + [E(X)]^2 = \alpha\beta^2 + \alpha^2\beta^2 = \alpha\beta^2(1+\alpha)$$

$$\Rightarrow \quad \alpha = \frac{E(X)^2}{E(X^2) - E(X)^2}, \beta = \frac{E(x^2) - E(X)^2}{E(X)}$$

Step #2: $\bar{X} \to E(X), \frac{1}{n}\sum X_i^2 \to E(X^2)$

$$\hat{\alpha} = \frac{\bar{X}^2}{\frac{1}{n}\sum X_i^2 - \bar{X}^2} \qquad \hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - \bar{X}^2}{\bar{X}}$$

6.2 Methods of Point Estimation

- **Example 6.14**

  Let $X_1, \ldots, X_n$ be a random sample from a generalized negative binomial distribution with parameters r and p. Its pmf is

$$nb(x; r, p) = \binom{x+r-1}{r-1} p^r (1-p)^x, \quad x = 0,1,2,\ldots.$$

Determine the moment estimators of parameters r and p.

Note: There are two parameters needs to estimate, thus the first two moments are considered.

- **Example 6.14 (Cont')**

Step #1: $E(X) = r(1-p)/p$

$$E(X^2) = r(1-p)(r - rp + 1)/p^2$$

$$\Rightarrow \quad p = \frac{E(X)}{\mathrm{E(X^2)\text{-}E(X)^2}}, r = \frac{E(X)^2}{E(X^2) - E(X)^2 - E(X)}$$

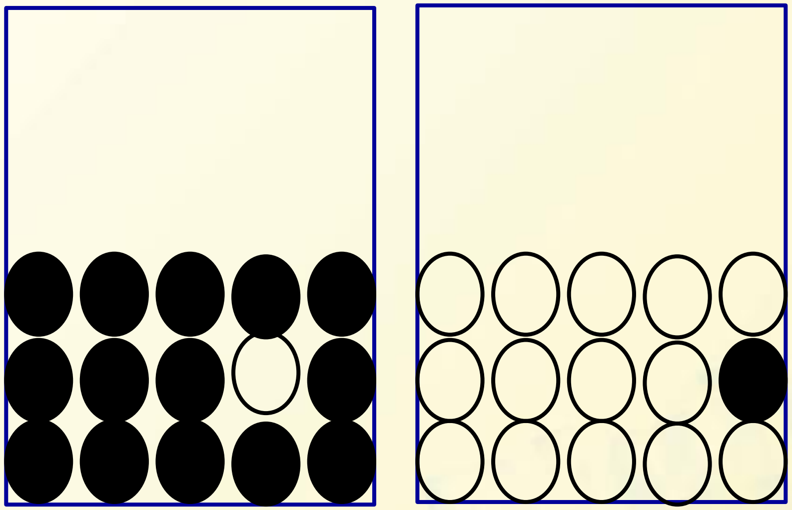
Step #2:

$$\bar{X} \to E(X), \frac{1}{n}\sum X_i^2 \to E(X^2)$$

$$\hat{p} = \frac{\bar{X}}{\frac{1}{n}\sum X_i^2 - \bar{X}^2} \qquad \hat{r} = \frac{\bar{X}^2}{\frac{1}{n}\sum X_i^2 - \bar{X}^2 - \bar{X}}$$

6.2 Methods of Point Estimation

- Maximum Likelihood Estimation (Basic Idea)



Box 1 Box 2

Experiment:
We firstly randomly choose a box,
And then randomly choose a ball.

Q: If we get a white ball, which box has the Maximum Likelihood being chosen?

$$P(W \mid Box1) = 1/15$$
$$P(W \mid Box2) = 14/15$$

Maximum Likelihood Estimation (Basic Idea)

The basis idea of the method of maximum likelihood is that we look at the sample values and then choose as our estimates of the unknown parameters the values for which the probability or probability density of getting the sample values is a maximum.

6.2 Methods of Point Estimation

- Maximum Likelihood Estimation (Basic Idea)

Q: What is the probability p of hitting the target?

$$f(\mathrm{p}) = p^3(1-p)^{5-3} = p^3(1-p)^2$$

$$f(0.2) \approx 0.0051 \quad f(0.4) \approx 0.0230 \quad f(0.6) \approx 0.0346 \quad f(0.8) \approx 0.0205 \quad \cdots$$

The best one among the four options

6.2 Methods of Point Estimation

- **Example 6.15**

A sample of ten new bike helmets manufactured by a certain company is obtained. Upon testing, it is found that the first, third, and tenth helmets are flawed, whereas the others are not. Let $p = P(\text{flawed helmet})$ and define $X_1, \ldots, X_{10}$ by $X_i = 1$ if the ith helmet is flawed and zero otherwise. Then the observed x_i's are 1,0,1,0,0,0,0,0,0,1.

The Joint pmf of the sample is

$$f(x_1, x_2, \ldots, x_{10}) = p(1-p)p\cdots p = p^3(1-p)^7$$

For what value of *p* is the observed sample most likely to have occurred? Or, equivalently, what value of the parameter *p* should be taken so that the joint pmf of the sample is maximized?

6.2 Methods of Point Estimation

- **Example 6.15 (Cont')**

$$f(x_1, x_2, ..., x_{10}) = p(1-p)p \cdots p = p^3(1-p)^7$$

$$\ln[f(x_1, x_2, ...x_{10}; p)] = 3\ln(p) + 7\ln(1-p)$$

Equating the derivative of the logarithm of the pmf to zero gives the maximizing value **(why?)**

$$\frac{d}{dp}\ln[f(x_1, x_2, ...x_{10}; p)] = \frac{3}{p} - \frac{7}{1-p} = 0 \Rightarrow p = \frac{3}{10} = \frac{x}{n}$$

where x is the observed number of successes (flawed helmets). The estimate of p is now $\hat{p} = 3/10$. It is called the **maximum likelihood estimate** because for fixed $x_1, \ldots, x_{10}$, it is the parameter value that maximizes the likelihood of the observed sample.

6.2 Methods of Point Estimation

Maximum Likelihood Estimation

Let $X_1, X_2, \ldots, X_n$ have joint pmf or pdf

$$f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

where the parameters $\theta_1, \ldots, \theta_m$ have unknown values. When $x_1, \ldots, x_n$ are the observed sample values and f is regarded as a function of $\theta_1, \ldots, \theta_m$, it is called the likelihood function.

The maximum likelihood estimates(mle's) $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are those values of the θ_i's that maximize the likelihood function, so that

$$f(x_1, \ldots, x_n; \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m) \quad \text{for all } \theta_1, \ldots, \theta_m$$

When the X_i's are substituted in place of the x_i's, the maximum likelihood estimators result.

6.2 Methods of Point Estimation

- **Three steps**

1. Write the joint pmf/pdf (i.e. Likelihood function)

$$f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m) = \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_m)$$

2. Get the ln(likelihood) **(if necessary)**

$$\ln[f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)] = \sum_{i=1}^{n} \ln(f(x_i; \theta_1, \ldots, \theta_m))$$

3. Take the partial derivative of ln(f) with respect to θ_i, equal them to 0, and solve the resulting m equations.

$$\frac{d}{d\theta_i} \ln[f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)] = 0$$

6.2 Methods of Point Estimation

- **Example 6.16**

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from an exponential distribution with the unknown parameter λ. Determine the maximum likelihood estimator of λ.

The joint pdf is (independence)

$$f(x_1,\ldots,x_n;\lambda) = (\lambda e^{-\lambda x_1})\cdots(\lambda e^{-\lambda x_n}) = \lambda^n e^{-\lambda \sum x_i}$$

The ln(likelihood) is $\ln[f(x_1,\ldots,x_n;\lambda)] = n\ln(\lambda) - \lambda \sum x_i$

Equating to zero the derivative w.r.t. λ:

$$\frac{d\ln[f(x_1,\ldots,x_n;\lambda)]}{d\lambda} = \frac{n}{\lambda} - \sum x_i = 0 \Rightarrow \lambda = \frac{n}{\sum x_i} = \frac{1}{\bar{x}} \Rightarrow \hat{\lambda} = 1/\bar{X}$$

The estimator

6.2 Methods of Point Estimation

- **Example 6.17**

  Let X_1, X_2, …, X_n is a random sample from a normal distribution $N(\mu, \sigma^2)$. Determine the maximum likelihood estimator of μ and σ^2.

  The joint pdf is

$$f(x_1,\ldots,x_n;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\cdot\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}\cdots\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}=\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}e^{-\frac{\sum(x_i-\mu)^2}{2\sigma^2}}$$

$$\ln[f(x_1,\ldots,x_n;\mu,\sigma^2)]=-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2$$

Equating to 0 the partial derivatives w.r.t. μ and σ^2, finally we have

$$\hat{\mu}=\bar{X},\ \ \sigma^2=\frac{\sum(X_i-\bar{X})^2}{n}$$

Here the mle of σ^2 is not the unbiased estimator.