Introductory Statistics

Tatiana Tyukina
tt51@leicester.ac.uk

Topic 2 - Estimators
2.1 - *Methods of point estimation*

- **Topic 0: Introduction**
- **Topic 1: Descriptive Statistics**
- **Topic 2: Estimators**
  - 2.1 Methods of point estimation,
  - 2.2 Properties of estimators,
  - 2.3 Interval estimation,
  - 2.4 Inference based on Normal Distribution
- Topic 3: Hypothesis Testing: Test for single population mean, test for two population mean
- Topic 4: Goodness of Fit: The χ^2 test
- Topic 5: Bayesian Estimation

RANDOM VARIABLE, PARAMETER, STATISTIC, INFERENCE

$X_1, ..., X_n$ - some random variables, all have the same distribution.

A collection of probability distributions P_θ.
Each collection, called *parametric family*, is indexed by a *parameter* (or a vector of parameters), for example:
for normal distributions the parameter is $\Theta = (M, \Sigma)$,
M - the measure of central tendency, Σ - measure of variance;
for Poisson distributions the parameter is $\Theta = (\Lambda)$.
Set of all possible values of the parameter - *parametric space*, Ω_Θ.

RANDOM VARIABLE, PARAMETER, STATISTIC, INFERENCE

$X_1, ..., X_n$ - some random variables, all have the same distribution.

A collection of probability distributions P_θ.
Each collection, called *parametric family*, is indexed by a *parameter* (or a vector of parameters), for example:
for normal distributions the parameter is $\Theta = (M, \Sigma)$,
M - the measure of central tendency, Σ - measure of variance;
for Poisson distributions the parameter is $\Theta = (\Lambda)$.
Set of all possible values of the parameter - *parametric space*, Ω_Θ.

For a particular realisation, when $\Theta = \theta$, the distribution of observations is denoted by P_θ, and expectation E_θ.
If we cannot observe or obtain value of the parameter, we can propose a *model*, or strictly a *family of models*.

RANDOM VARIABLE, PARAMETER, STATISTIC, INFERENCE

Let's denote observed data obtained from a sample - X
(could be a vector of observations $X = (x_1, x_2, ...x_n)$ that are *mutually independent and identically distributed* (i.i.d.),
or some general characteristics).
The set of possible values for X is the *sample space*, $\mathcal{X} = \{X\}$.

RANDOM VARIABLE, PARAMETER, STATISTIC, INFERENCE

Let's denote observed data obtained from a sample - X
(could be a vector of observations $X = (x_1, x_2, ...x_n)$ that are *mutually independent and identically distributed* (i.i.d.),
or some general characteristics).
The set of possible values for X is the *sample space*, $\chi = \{X\}$.

The members $P_\theta(X) = f_\chi(X, \theta)$ of the parametric family are distributions over this space χ, where $\theta \in \Omega_\Theta$ is unknown.

Statistic is the estimation $\hat{\theta}$ that can be calculated from the sampling X, for example *sampling mean*, and *sampling variance* etc.

STATISTIC AS A POINT ESTIMATOR

DEFINITION

The problem of point estimation *is to determine statistics* $g_i(X_1, ..., X_n)$, $i = 1, ..., k$ *(where* k *is the dimension of* $\boldsymbol{\theta}$*), which can be used to estimate the value of each of the parameters* $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ *based on observed sample data from the population.*
These statistics are called ***estimators*** $\hat{\theta}_i$ *for the parameters, where* $\hat{\theta}_i = g_i(X_1, ..., X_n)$, $i = 1, ..., k$.
The values calculated from these statistics using particular sample data values are called ***estimates*** *of the parameters.*

STATISTIC AS A POINT ESTIMATOR

DEFINITION

The problem of point estimation *is to determine statistics* $g_i(X_1, ..., X_n)$, $i = 1, ..., k$ *(where* k *is the dimension of* $\boldsymbol{\theta}$*), which can be used to estimate the value of each of the parameters* $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ *based on observed sample data from the population.*
These statistics are called ***estimators*** $\hat{\theta}_i$ *for the parameters, where* $\hat{\theta}_i = g_i(X_1, ..., X_n)$, $i = 1, ..., k$.
The values calculated from these statistics using particular sample data values are called ***estimates*** *of the parameters.*

The estimators are random variables!

STATISTIC AS A POINT ESTIMATOR

DEFINITION

The problem of point estimation *is to determine statistics* $g_i(X_1, ..., X_n)$, $i = 1, ..., k$ *(where* k *is the dimension of* $\boldsymbol{\theta}$*), which can be used to estimate the value of each of the parameters* $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ *based on observed sample data from the population.*
These statistics are called ***estimators*** $\hat{\theta}_i$ *for the parameters, where* $\hat{\theta}_i = g_i(X_1, ..., X_n)$, $i = 1, ..., k$.
The values calculated from these statistics using particular sample data values are called ***estimates*** *of the parameters.*

The estimators are random variables!

DEFINITION

The probability distribution of a sample statistic is called its ***sampling distribution****.*

STATISTIC AS A POINT ESTIMATOR

EXAMPLE

Let the population consists of the numbers $X = \{1, 2, 3, 4, 5\}$. Consider all possible samples of three numbers randomly chosen without replacement from this population.
Obtain the distribution of the sample mean.

STATISTIC AS A POINT ESTIMATOR

EXAMPLE

Let the population consists of the numbers $X = \{1, 2, 3, 4, 5\}$. Consider all possible samples of three numbers randomly chosen without replacement from this population.

Obtain the distribution of the sample mean.

Solution.

10 samples: $(1, 2, 3)$, $(1, 2, 4)$, $(1, 2, 5)$, $(1, 3, 4)$, $(1, 3, 5)$, $(1, 4, 5)$, $(2, 3, 4)$, $(2, 3, 5)$, $(2, 4, 5)$, and $(3, 4, 5)$.

$\bar{x}$	2	$2\frac{1}{3}$	$2\frac{2}{3}$	3	$3\frac{1}{3}$	$3\frac{2}{3}$	4
$p(\bar{x})$	$1/10$	$1/10$	$2/10$	$2/10$	$2/10$	$1/10$	$1/10$

STATISTIC AS A POINT ESTIMATOR

Three of the most popular methods of estimation:

- the method of moments,
- the method of maximum likelihood,
- Bayes' method.

Criteria for choosing a desired point estimator

- unbiasedness,
- efficiency (minimal variance),
- sufficiency,
- consistency

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

DEFINITION

Let W be any random variable with pdf $f_W(w)$. For any positive integer r:

1. ***The r-th moment of W about the origin**, μ_r, is given by*

$$\mu_r = E(W^r),$$

provided $\int_{-\infty}^{\infty} |w|^r \cdot f_W(w)dw < \infty$.
(When $r = 1$, the subscript is usually omitted, i.e. $\mu_1 = \mu$)

2. ***The r-th moment of W about the mean,** μ'_r, is given by*

$$\mu'_r = E\left((W-\mu)^r\right),$$

provided fitness conditions of the part 1 hold.

3. ***The $r-$th standardized moment**, $\tilde{\mu}_r$, is a moment that is normalized, typically by the standard deviation σ^r*

$$\tilde{\mu}_r = \frac{E\left((W-\mu)^r\right)}{\sigma^r}.$$

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

Sample moments are calculated with respect to the observed sample:

DEFINITION

Let $X = X_1,\ X_2, \ldots, X_n$ be a sample of independent and identically distributed random variables. For any positive integer r the ***r-th sample moment of*** *X , m_r, is given by*

$$m_r = E(X^r) = \frac{1}{n}\sum_{i=1}^{n} X_i^r.$$

For $r = 1$, $m_1 = \bar{X}$ - the sample mean.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

The Method of Moments Procedure:
Suppose there are k parameters to be estimated: $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$.

1. Find k population moments, μ_r, $r = 1, 2, ..., k$. μ_r will contain one or more parameters $\theta_1, ..., \theta_k$.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

The Method of Moments Procedure:
Suppose there are k parameters to be estimated: $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$.

1. Find k population moments, μ_r, $r = 1, 2, ..., k$. μ_r will contain one or more parameters $\theta_1, ..., \theta_k$.
2. Find the corresponding k sample moments, $m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$, $r = 1, 2, ..., k$, n is the sample size. The number of sample moments should equal the number of parameters to be estimated.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

The Method of Moments Procedure:
Suppose there are k parameters to be estimated: $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$.

1. Find k population moments, μ_r, $r = 1, 2, ..., k$. μ_r will contain one or more parameters $\theta_1, ..., \theta_k$.
2. Find the corresponding k sample moments, $m_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$, $r = 1, 2, ..., k$, n is the sample size. The number of sample moments should equal the number of parameters to be estimated.
3. From the system of equations, $\mu_r = m_r$, $r = 1, 2, ..., k$, solved for the parameter $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$ we find a moment estimator of $\hat{\boldsymbol{\theta}}$.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let $X_1, ..., X_n$ be a random sample from a Bernoulli population with parameter p.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let $X_1, ..., X_n$ be a random sample from a Bernoulli population with parameter p.

(A) Find the moment estimator for p.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let $X_1, ..., X_n$ be a random sample from a Bernoulli population with parameter p.

(A) Find the moment estimator for p.

(B) Tossing a coin 10 times and equating heads to value 1 and tails to value 0, we obtained the following values: 0 1 1 0 1 0 1 1 1 0
Obtain a moment estimate for p, the probability of success (head).

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let $X_1, ..., X_n$ be a random sample from a Bernoulli population with parameter p.

(A) Find the moment estimator for p.

(B) Tossing a coin 10 times and equating heads to value 1 and tails to value 0, we obtained the following values: 0 1 1 0 1 0 1 1 1 0
Obtain a moment estimate for p, the probability of success (head).

$\mu = \sum_{x=0}^{1} x \cdot p^x \cdot (1-p)^{1-x} = 0 \cdot p^0 \cdot (1-p) + 1 \cdot p^1 \cdot (1-p)^0 = p$

Solution

(A) For the Bernoulli random variable, $p = E[X] = \mu$, so we can estimate p as the ratio of the total number of heads to the total number of tosses:

$m_1 = \frac{1}{n} \sum_{i=1}^{n} X_i \Rightarrow \mu = m_1 \Rightarrow m_1^* = \hat{p} = \frac{1}{n} \sum_{i=1}^{n} X_i.$, $\hat{p} = \bar{X}$

(B) The experiment results in Bernoulli random variables with $\sum_{i=1}^{n} x_i = 6$.

Thus, the moment estimate of p is $\hat{p} = 6/10 = 0.6$.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let the distribution of X be $N(\mu, \sigma^2)$.

1. For a given sample of size n, use the method of moments to estimate μ and σ^2.
2. The following data (rounded to the third decimal digit) were generated from a normal distribution with mean 2 and a standard deviation of 1.5.

$$3.163,\ 1.883,\ 3.252,\ 3.716,\ -0.049,\ -0.653,\ 0.057$$
$$4.098,\ 1.670,\ 1.396,\ 2.332,\ 1.838,\ 3.024,\ 2.706$$
$$3.830,\ 3.349,\ -0.230,\ 1.496,\ 0.231,\ 2.987$$

Obtain the method of moments estimates of the true mean and the true variance.

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE $X \sim N(\mu, \sigma^2)$

Solution

(A) 1) $\mu = E(X)$, $\sigma^2 = E(X^2) - \mu^2 \Rightarrow \mu_2 = E(X^2) = \sigma^2 - \mu^2$

2) $m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$, $m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \Rightarrow$ 3) $\begin{cases} m_1 = \mu_1 = \mu \\ m_2 = \mu_2 = \sigma^2 - \mu^2 \end{cases}$

$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$; $\hat{\sigma}^2 = m_2 - \hat{\mu}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

(B)

$X = \{$3.163, 1.883, 3.252, 3.716, −0.049, −0.653, 0.057,
4.098, 1.670, 1.396, 2.332, 1.838, 3.024, 2.706,
3.830, 3.349, −0.230, 1.496, 0.231, 2.987$\}$

$\hat{\mu} = \bar{x} = \frac{1}{20}\sum_{i=1}^{20} x_i = 2{,}005$

$\hat{\sigma}^2 = \frac{1}{20}\sum_{i=1}^{20}(x_i - \bar{x})^2 = 2{,}1025$

$\mu = 2$

$\sigma^2 = 2.25$ $\quad \hat{\sigma}^2 < \sigma^2$

METHOD OF MOMENTS FOR FINDING POINT ESTIMATORS

EXAMPLE

Let $X_1, X_2, ..., X_n$ be a sample of I.i.d. random variables with pdf,:

$$f_X(x) = \begin{cases} \theta x^{\theta-1}, & 0 < x \leq 1,\ \theta > 0, \\ 0, & otherwise. \end{cases}$$

1. Use the method of moments to estimate θ.
2. For the following observations of X calculate the method of moments estimate for θ:

$$0.3,\ 0.5,\ 0.8,\ 0.6,\ 0.4,\ 0.4,\ 0.5,\ 0.8,\ 0.6,\ 0.3$$

Solution:

$$f_X(x) = \theta x^{\theta-1}, \quad 0 < x \leq 1, \quad \theta > 0$$

1.

1) $$\mu_1 = \int_0^1 x f(x)\,dx = \int_0^1 x \cdot \theta \cdot x^{\theta-1}\,dx = \int_0^1 \theta x^{\theta}\,dx =$$

$$= \frac{\theta}{\theta+1} x^{\theta+1}\Big|_0^1 = \frac{\theta}{\theta+1}$$

2) $$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \Rightarrow$$ 3) $$\mu_1 = m_1 \Leftrightarrow \frac{\theta}{\theta+1} = \bar{X} \Rightarrow$$

$$\theta = \bar{X}(\theta+1) \Rightarrow \hat{\theta} = \frac{\bar{X}}{1-\bar{X}}$$ – method of moments estimator

2. $$X = \{0.3, 0.5, 0.8, 0.6, 0.4, 0.4, 0.5, 0.8, 0.6, 0.3\}$$

$$\bar{X} = \frac{1}{10}(0.3 + 0.5 + 0.8 + \ldots + 0.3) = \frac{5.2}{10} = 0.52$$

$$\hat{\theta} = \frac{0.52}{1-0.52} = 1.083 \text{ (3 d.p.)}$$

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

EXAMPLE

Consider a coin: $p = P(X = H)$ is unknown

<u>Question</u>: Estimate p on the basis of 3 tosses sample $X = \{HHT\}$

$$X = \begin{cases} 1, & H \quad p \\ 0, & T \quad 1-p \end{cases} \Rightarrow p(X=x) = p^x (1-p)^{1-x}$$

$x_1 = 1, \; x_2 = 1, \; x_3 = 0$

$$P(X_1 = x_1 \cap X_2 = x_2 \cap X_3 = x_3) = p \cdot p \cdot (1-p) = p^2(1-p) \xrightarrow[p]{} \max$$

$$\max_p P(X_1, X_2, X_3)$$

$$\frac{d P(X_1, X_2, X_3)}{dp} = -3p^2 + 2p = 0$$

$$\boxed{\hat{p} = \frac{2}{3}} \qquad \frac{d^2 P(\cdot)}{dp^2} < 0$$

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

EXAMPLE

$X_1 = k_1, X_2 = k_2, \dots, X_n = k_n$ – random sample of size n

$$P(X_1 = k_1 \cap X_2 = k_2 \dots \cap X_n = k_n) = \underline{p^{k_1}}(1-p)^{1-k_1} \cdot \underline{p^{k_2}}(1-p)^{1-k_2} \cdot \dots \cdot \underline{p^{k_n}}(1-p)^{1-k_n}$$

$$= p^{k_1 + k_2 + \dots + k_n}(1-p)^{1-k_1 + 1 - k_2 + \dots + 1 - k_n} =$$

$$= p^{\sum_{i=1}^n k_i}(1-p)^{n - \sum_{i=1}^n k_i}$$

$$\frac{d P(X_1, \dots, X_n)}{dp} = \sum_{i=1}^n k_i \underbrace{p^{\sum_{i=1}^n k_i - 1}} \cdot \underbrace{(1-p)^{n - \sum_{i=1}^n k_i}} +$$

$$+ \underbrace{p^{\sum_{i=1}^n k_i}} \cdot \left(n - \sum_{i=1}^n k_i\right) \cdot (-1) \cdot \underbrace{(1-p)^{n - \sum_{i=1}^n k_i - 1}} = 0$$

$p \neq 0, \; p \neq 1$ $\quad \dfrac{d P(X_1, X_2, \dots, X_n)}{dp} = \sum_{i=1}^n k_i \cdot (1-p) - \left(n - \sum_{i=1}^n k_i\right) p = 0$

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n k_i$$

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

DEFINITION

Let $f(x_1, ..., x_n,\ \theta)$, $\theta \in \Omega_\Theta \subset \mathbf{R}^k$ *, be the joint probability mass (or density) function of n random variables* $X_1, ..., X_n$ *with sample values* $x_1, ..., x_n$*.*
The ***likelihood*** *function of the sample is given by*

$$L(\theta,\ x_1, ..., x_n) = h(x_1, ..., x_n,\ \theta),\ [= L(\theta),\ \textit{in a briefer notation}].$$

We emphasize that L is a function of θ *for fixed sample values.*

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

DEFINITION

*Let $f(x_1, ..., x_n,\ \theta)$, $\theta \in \Omega_\Theta \subset \mathbf{R}^k$, be the joint probability mass (or density) function of n random variables $X_1, ..., X_n$ with sample values $x_1, ..., x_n$. The **likelihood** function of the sample is given by*

$$L(\theta,\ x_1, ..., x_n) = h(x_1, ..., x_n,\ \theta),\ [= L(\theta),\ \textit{in a briefer notation}].$$

We emphasize that L is a function of θ for fixed sample values.

If $X_1, ..., X_n$ are *discrete* i.i.d. random variables with probability mass function $p(x, \theta)$, then, the likelihood function is given by

$$L(\theta) = P(X_1 = x_1, ..., X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} p(x_i, \theta)$$

and in *continuous* case, if the probability density function is $f(x, \theta)$, then the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

DEFINITION

Let $L(\theta) = \prod_{i=1}^{n} p(x_i, \theta)$ *(or* $L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$*) be the likelihood function corresponding to random sample* $x_1, \ldots, x_n$ *drawn from the discrete pdf* $p(x, \theta)$ *(or from the continuous pdf* $p(x, \theta)$*), where* θ *is an unknown parameter.*
Let $\hat{\theta}_l$ *be a value of the parameter such that* $L(\hat{\theta}_l) \geq L(\theta)$ *for all possible values of the parameter* θ*. Then* $\hat{\theta}_l$ *is called a* ***maximum likelihood estimate*** *of* θ*.*

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

DEFINITION

Let $L(\theta) = \prod_{i=1}^{n} p(x_i, \theta)$ *(or* $L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$*) be the likelihood function corresponding to random sample* $x_1, ..., x_n$ *drawn from the discrete pdf* $p(x, \theta)$ *(or from the continuous pdf* $p(x, \theta)$*), where* θ *is an unknown parameter.*
Let $\hat{\theta}_l$ *be a value of the parameter such that* $L(\hat{\theta}_l) \geq L(\theta)$ *for all possible values of the parameter* θ*. Then* $\hat{\theta}_l$ *is called a* ***maximum likelihood estimate*** *of* θ*.*
The ***Maximum Likelihood Estimates*** *are those values of the parameters that maximize the likelihood function with respect to the parameter* θ*. That is,*

$$L(\hat{\theta}_l, x_1, ..., x_n) = max_{\theta \in \Omega_\Theta} L(\theta, x_1, ..., x_n)$$

where Ω_Θ *is the set of possible values of the parameter* θ*.*

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

Procedure to Find Maximum Likelihood Estimate(s) (MLEs):

1. Define the likelihood function, $L(\theta)$.

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

Procedure to Find Maximum Likelihood Estimate(s) (MLEs):

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm of $L(\theta)$ ($\ln(L(\theta))$).

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

Procedure to Find Maximum Likelihood Estimate(s) (MLEs):

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm of $L(\theta)$ ($\ln(L(\theta))$).
3. When applicable, differentiate $\ln L(\theta)$ with respect to θ, and then equate the derivative to zero.

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

Procedure to Find Maximum Likelihood Estimate(s) (MLEs):

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm of $L(\theta)$ ($\ln(L(\theta))$).
3. When applicable, differentiate $\ln L(\theta)$ with respect to θ, and then equate the derivative to zero.
4. Solve for the parameter θ to obtain $\hat{\theta}$.

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

Procedure to Find Maximum Likelihood Estimate(s) (MLEs):

1. Define the likelihood function, $L(\theta)$.
2. Often it is easier to take the natural logarithm of $L(\theta)$ ($\ln(L(\theta))$).
3. When applicable, differentiate $\ln L(\theta)$ with respect to θ, and then equate the derivative to zero.
4. Solve for the parameter θ to obtain $\hat{\theta}$.
5. Check whether it is a maximizer or global maximizer $\hat{\theta}_l$.

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

EXAMPLE

Suppose $X_1, ..., X_n$ are discrete i.i.d. random variables with the geometric distribution with unknown parameter p, so that $P(X = x) = (1-p)^{x-1}p$ for $x = 1, 2, ...$
Find the MLE for the parameter p for n independent observations $x_1, x_2, ..., x_n$.
Calculate the estimate $\hat{p}$ for the following set of observations:
231, 127, 60, 4, 3, 183, 2, 4, 71, 22

Solution:

1. $L(p) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} (1-p)^{x_i - 1} \cdot p = (1-p)^{\sum_{i=1}^{n}(x_i - 1)} p^n$

$$\frac{d\,L(p)}{dp} = -\sum_{i=1}^{n}(x_i - 1)(1-p)^{\sum_{i=1}^{n} x_i - n - 1} \cdot p^n + (1-p)^{\sum_{i=1}^{n} x_i - n} n p^{n-1} = 0$$

| divide by $(1-p)^{\sum_{i=1}^{n} x_i - n - 1} \cdot p^{n-1}$, $p \neq 0$, $p \neq 1$

$$\frac{dL(p)}{dp} = -\sum_{i=1}^{n}(x_i - 1) \cdot p + n \cdot (1-p) = 0 = -\sum_{i=1}^{n} x_i\, p + np + n - np = 0$$

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

METHOD OF MAXIMUM LIKELIHOOD FOR FINDING POINT ESTIMATORS

EXAMPLE

2.

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

$$\ell(p) = \ln(L(p)) = \ln (1-p)^{\sum_{i=1}^{n}(x_i-1)} + \ln p^n =$$

$$= \sum_{i=1}^{n}(x_i - 1)\ln(1-p) + n \ln p$$

$$\frac{d\,\ell(p)}{dp} = -\frac{\sum_{i=1}^{n} x_i - n}{1-p} + \frac{n}{p} = \frac{-\sum_{i=1}^{n} x_i \cdot p + np + n - np}{(1-p)\cdot p} = 0$$

$$p \neq 0, p \neq 1 \qquad -\sum_{i=1}^{n} x_i\, p + n = 0 \Rightarrow \hat{p} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

RECAP

EXAMPLE

Suppose the isolated weather-reporting station has an electronic device whose time to failure is given by the exponential model

$$f_Y(y,\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

The station also has a spare device, so the time until this instruments is not available is the sum of these two exponential pdfs, which is

$$f_Y(y,\theta) = \frac{1}{\theta^2} y e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

Five data points have been collected - $9.2, 5.6, 18.4, 12.1, 10.7$.
Find the maximum likelihood estimate for θ

RECAP

EXAMPLE

Suppose the isolated weather-reporting station has an electronic device whose time to failure is given by the exponential model

$$f_Y(y,\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

The station also has a spare device, so the time until this instruments is not available is the sum of these two exponential pdfs, which is

$$f_Y(y,\theta) = \frac{1}{\theta^2} y e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

Five data points have been collected - 9.2, 5.6, 18.4, 12.1, 10.7.
Find the maximum likelihood estimate for θ

Solution.

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

$$l(\theta) = \ln(L(\theta)) = -2n \ln(\theta) + \ln \left(\prod_{i=1}^{n} Y_i \right) - \frac{1}{\theta} \sum_{i=1}^{n} (Y_i)$$

RECAP

EXAMPLE

Suppose the isolated weather-reporting station has an electronic device whose time to failure is given by the exponential model

$$f_Y(y,\theta) = \frac{1}{\theta} e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

The station also has a spare device, so the time until this instruments is not available is the sum of these two exponential pdfs, which is

$$f_Y(y,\theta) = \frac{1}{\theta^2} y e^{-\frac{y}{\theta}},\ 0 \le y < \infty,\ 0 < \theta < \infty$$

Five data points have been collected - 9.2, 5.6, 18.4, 12.1, 10.7.
Find the maximum likelihood estimate for θ

Solution.

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

$$l(\theta) = \ln(L(\theta)) = -2n \ln(\theta) + \ln \left(\prod_{i=1}^{n} Y_i \right) - \frac{1}{\theta} \sum_{i=1}^{n} (Y_i)$$

$$\frac{dl(\theta)}{d\theta} = \frac{-2n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^{n} (Y_i) = 0 \qquad \Rightarrow \qquad \hat{\theta} = \frac{1}{2n} \sum_{i=1}^{n} Y_i = \bar{Y}/2$$

MAXIMUM LIKELIHOOD ESTIMATES FOR SEVERAL UNKNOWN PARAMETERS

If a family of probability models is indexed by two or more unknown parameters, $\theta_1,\ \theta_2,\ \ldots,\ \theta_k$, finding maximum likelihood estimates requires the solution of k simultaneous equations.
For example, for $k = 2$,

$$\left.\begin{array}{ll} \frac{\partial L(\theta_1,\theta_2)}{\partial \theta_1} & = 0 \\ \frac{\partial L(\theta_1,\theta_2)}{\partial \theta_2} & = 0 \end{array}\right\} \Longleftrightarrow \left\{\begin{array}{ll} \frac{\partial \ln L(\theta_1,\theta_2)}{\partial \theta_1} & = 0 \\ \frac{\partial \ln L(\theta_1,\theta_2)}{\partial \theta_2} & = 0 \end{array}\right.$$

Exercise: For a random sample $X_1, X_2, \ldots, X_n$ from normal distribution $N(\mu, \sigma^2)$ with the pdf:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

find the maximum likelihood estimators for μ and σ^2. Compare these estimators to the method of moments estimators discussed in the previous section.

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

THEOREM (CENTRAL LIMIT THEOREM)

Let $X_1, ..., X_n$ be a random sample of size n from a population with mean μ and variance σ^2. Then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$.

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

THEOREM (CENTRAL LIMIT THEOREM)

Let $X_1, ..., X_n$ be a random sample of size n from a population with mean μ and variance σ^2. Then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$.

PROOF.

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{\sum_{i=1}^{n} \mu}{n} = \frac{n\mu}{n} = \mu$$

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

THEOREM (CENTRAL LIMIT THEOREM)

Let $X_1, \ldots, X_n$ be a random sample of size n from a population with mean μ and variance σ^2. Then $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \frac{\sigma^2}{n}$.

PROOF.

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{\sum_{i=1}^{n} E(X_i)}{n} = \frac{\sum_{i=1}^{n} \mu}{n} = \frac{n\mu}{n} = \mu$$

$$Var(\bar{X}) = \mathrm{Var}\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

□

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

$$\mu_{\bar{X}} = E(\bar{X}) = \mu$$

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

$$\mu_{\bar{X}} = E(\bar{X}) = \mu$$

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = \frac{\sigma^2}{n}$$

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

$$\mu_{\bar{X}} = E(\bar{X}) = \mu$$

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = \frac{\sigma^2}{n}$$

DEFINITION

*The value $\sigma_{\bar{X}}$ is called the **standard error** of the mean.*

POINT ESTIMATION: STANDARD ERROR OF THE MEAN

$$\mu_{\bar{X}} = E(\bar{X}) = \mu$$

$$\sigma^2_{\bar{X}} = Var(\bar{X}) = \frac{\sigma^2}{n}$$

DEFINITION

*The value $\sigma_{\bar{X}}$ is called the **standard error** of the mean.*

From Chebyshev's inequality,

$$P\left(|\bar{X} - \mu_{\bar{X}}| < k\sigma_{\bar{X}}\right) \geq 1 - \frac{1}{k^2}$$

Let $\varepsilon = (k\sigma)/\sqrt{n} > 0$

$$P\left(|\bar{X} - \mu| < \varepsilon\right) \geq 1 - \frac{\sigma^2}{n\varepsilon^2}$$

POINT ESTIMATION: FINITE POPULATION

Let $\{c_1, c_2, ..., c_N\}$ be a finite population. Then the population mean $\mu = (1/N)\sum_{i=1}^{N} c_i$, and the population variance $\sigma^2 = (1/N)\sum_{i=1}^{n}(c_i - \mu)^2$.

THEOREM

Let $X_1, ..., X_n$ be a random sample of size n (chosen without replacement) from a finite population $\{c_1, c_2, ..., c_N\}$, then

$$E(\bar{X}) = \mu,$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right).$$

POINT ESTIMATION: FINITE POPULATION

Let $\{c_1, c_2, ..., c_N\}$ be a finite population. Then the population mean $\mu = (1/N)\sum_{i=1}^{N} c_i$, and the population variance $\sigma^2 = (1/N)\sum_{i=1}^{n}(c_i - \mu)^2$.

THEOREM

Let $X_1, ..., X_n$ be a random sample of size n (chosen without replacement) from a finite population $\{c_1, c_2, ..., c_N\}$, then

$$E(\bar{X}) = \mu,$$

$$Var(\bar{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right).$$

The factor $(N-n)/(N-1)$ is often called the **finite population correction factor**.

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;
3. The main procedure for finding estimates using the method of maximum likelihood;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;
3. The main procedure for finding estimates using the method of maximum likelihood;
4. How the moment generating function can be use for finding of moments of a population distribution;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;
3. The main procedure for finding estimates using the method of maximum likelihood;
4. How the moment generating function can be use for finding of moments of a population distribution;
5. Which characteristics of a distribution describe skewness and kurtosis;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;
3. The main procedure for finding estimates using the method of maximum likelihood;
4. How the moment generating function can be use for finding of moments of a population distribution;
5. Which characteristics of a distribution describe skewness and kurtosis;
6. Define the notion of the standard error of the mean;

QUESTIONS TO TAKE HOME

1. Difference between Estimate and Estimator;
2. The main procedure for finding estimates using the method of moments;
3. The main procedure for finding estimates using the method of maximum likelihood;
4. How the moment generating function can be use for finding of moments of a population distribution;
5. Which characteristics of a distribution describe skewness and kurtosis;
6. Define the notion of the standard error of the mean;
7. When do we need to use the finite population correction factor.