POINT ESTIMATION

Two important problems in statistical inference are (1)Estimation (2) Testing of hypothesis.

Definition:Any function of the random sample $x_1, x_2,...x_n$ that are being observed ,say $Tn(x_1, x_2,...x_n)$ is called a statistic. If it is used to estimate an unknown parameter θ of the distribution, it is called an estimator. A particular value of the estimator, say, $Tn(x_1, x_2,...x_n)$ is called an estimate of θ.

Characteristics of Estimators

(1)Unbiasedness , (2)Consistency ,(3)Efficiency ,(4) Sufficiency

(1) **Unbiasedness :** An estimator $T_n(x_1, x_2,...x_n)$ is said to be an unbiased estimator of $\upsilon(\theta)$ if

$E(T_n) = \upsilon(\theta)$ for all $\theta \in \phi$

(2) **Consistency :** An estimator $T_n = T(x_1, x_2,...x_n)$ is said to be consistent estimator of $\upsilon(\theta), \theta \in \phi$, if T_n converges to $\upsilon(\theta)$ in probability.

(3) **Efficiency** :If ,of the two consistent estimators T_1 ,T_2 of a certain parameter θ, we have $V(T_1) < V(T_2)$ for all n ,then T_1 is more efficient than T_2 for all sample sizes.

Most efficient estimator: If in a class of consistent estimators ,there exists one whose sampling variance is less than that of any such estimator, it is called the most efficient estimator.

Efficiency : If T_1 is the most efficient estimator with variance V_1 and T_2 is any other estimator with variance V_2 , then the efficiency E of T_2 is defined as $E = \frac{V_1}{V_2}$

(4) **Sufficiency:** An estimator $T = T(x_1, x_2,...x_n)$ is said to be sufficient for the parameter θ ,based on a sample $x_1, x_2,...x_n$ of size n from the population with density f(x, θ) such that the conditional distribution of $x_1, x_2,...x_n$ given T, is independent of θ, then T is the sufficient estimator for θ.

Minimum Variance Unbiased Estimator(M.V.U.E):

If a statistic $T = T(x_1,x_2,\dots x_n)$ based on sample size n is such that

(i) T is unbiased for $\nu(\theta)$,for all $\theta \in \phi$

(ii) It has the smallest variance among the class of all unbiased estimators of $\nu(\theta)$, then T is called the minimum variance unbiased estimator of $\nu(\theta)$.

Neymann Factorisation Theorem

Statement: T =T(x) is sufficient for θ if and only if the joint density function L of the sample values can be expressed in the form $L = g_{\vartheta}[t(x)].h(x)$ where $g_{\vartheta}[t(x)]$.depends on θ and x through the value of t(x) and h(x) is independent of θ

Problems

1. *The sample mean and variance are consistent and unbiased estimators of the mean and variance of the underlying distribution.*

Proof. It is easy to compute that

$$E\left[\frac{X_1+\cdots+X_n}{n}\right]=\mu$$

and

$$\begin{aligned}E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right] &= \frac{n}{n-1}E\left[(X_1-\overline{X})^2\right]\\ &= \frac{n}{n-1}\left(E[X_1^2]-2E[X_1\overline{X}]+E[\overline{X}^2]\right)\\ &= \frac{n}{n-1}\left(E[X_1^2]-\frac{2}{n}E[X_1^2]-\frac{2(n-1)}{n}E[X_1X_2]+E[\overline{X}^2]\right)\end{aligned}$$

and now expanding the $E[\overline{X}^2]$ as

$$E[\overline{X}^2]=\frac{1}{n^2}\left(nE[X_1^2]+n(n-1)E[X_1X_2]\right)$$

and also using the independence, e.g. $E[X_1X_2]=E[X_1]E[X_2]=\mu^2$ we get that the above equals to

$$E[X_1^2]-\mu^2=\sigma^2.$$

We, therefore, obtain that the sample mean and sample variance are unbiased estimators.

2.Prove that in sampling a normal population , sample mean is a consistent estimator of μ

Proof: The sample mean $\bar{x}$ is also normally distributed as N($\mu, \sigma^2/n$)

E($\bar{x}$) = μ and V($\bar{x}$) = σ^2/n

Thus as n→ ∞, E($\bar{x}$) = μ and V($\bar{x}$) = 0

Hence $\bar{x}$ is a consistent estimator for μ.

3.If $x_1,x_2,\ldots x_n$ is a random sample from a normal population N(μ ,1). Show that $t = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ is an unbiased estimator of $\mu^2 + 1$

Proof: $E(x_i) = \mu$, $V(x_i) = 1$

$E(x_i^2) = V(x_i) + [E(x_i)]^2 = 1 + \mu^2$

$E(t) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i^2) = \frac{1}{n}\sum_{i=1}^{n} 1 + \mu^2$

Hence t is an unbiased estimator of $\mu^2 + 1$

4. If T is an unbiased estimator for θ , show that T^2 is a biased estimator of θ^2.

Proof: Since T is an unbiased estimator for θ, $E(T) = \theta$

$Var(T) = E(T^2) - E(T)^2 = E(T^2) - \theta^2$

$E(T^2) = \theta^2 + Var(T)$

Since $E(T^2) \neq \theta^2$, T^2 is a biased estimator of θ^2.

5.Let X be distributed in the poisson form with parameter θ.Show that only unbiased estimator of $e^{-(k+1)\theta}$ is $T(X) = (-K)^x$

Proof: $E(T(X)) = E(-K)^x$

$$= \sum_{x=0}^{\infty} (-k)^x \left(\frac{e^{-\theta}\theta^x}{x!}\right) = e^{-\theta}\sum_{x=0}^{\infty}\frac{(-k\theta)^x}{x!} = e^{-\theta}.e^{-k\theta} = e^{-(1+k)\theta}$$

6.A random sample (x1,x2,x3,x4,x5) of size 5 is drawn from a normal population with mean μ. Consider the following estimators to estimate μ:

(1) $t_1 = (x_1 + x_2 + x_3 + x_4 + x_5)/5$

(2) $t_2 = (x_1 + x_2)/2 + x_3$

(3) $t_3 = (2x_1 + x_2 + \lambda x_3)/3$ where λ is such that t_3 is an unbiased estimator of μ. Find λ .Find the estimator which is best among t_1, t_2, t_3.

Solution:

(1) $E(t_1) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{5}\sum_{i=1}^{5} \mu = \mu$

t_1 is an unbiased estimator of μ

(2) $E(t_1 + t_2) = E(x_1 + x_2)/2 + E(x_3) = (\mu + \mu)/2 + \mu = 2\mu$

t_2 is an biased estimator of μ

(3) $E(t_3) = \mu$, since t_3 is an unbiased estimator of μ

$$\mu = E(2x_1 + x_2 + \lambda x_3)/3$$

$$2E(x_1) + E(x_2) + \lambda E(x_3) = 3\mu$$

2 μ + μ+λ μ =3 μ which gives λ =0

$$V(t_1) = \frac{1}{25}[V(x_1) + V(x_2) + V(x_3) + V(x_4) + V(x_5)] = \frac{1}{25}5\sigma^2 = \frac{1}{5}\sigma^2$$

$$V(t_2) = \frac{1}{4}\{[V(x_1) + V(x_2)] + V(x_3)\} = \frac{1}{4}(2\sigma^2) + \sigma^2 = \frac{3}{2}\sigma^2$$

$$V(t_3) = \frac{1}{9}\{4V(x_1) + V(x_2)\} = \frac{1}{9}\{(4\sigma^2) + \sigma^2\} = \frac{5}{9}\sigma^2$$

Since $V(t_1) = \frac{1}{5}\sigma^2$ is least, t_1 is the best estimator of μ.

Theorem 1. **(Cramér-Rao Inequality.)** *Assume $V(\theta)$ has continuous first derivative (except possibly at finitely many points). Then for any unbiased estimator $\hat{\theta}$,*

$$Var(\hat{\theta}) \geq \frac{1}{nI(\theta)}.$$

This is the desired theoretical bound on how efficient an estimator can be. The theorem is in fact valid under weaker assumptions (see the text), i.e., $V(\theta)$ does not need to be differentiable everywhere, but we assume this for simplicity.

To prove this result, first we need a little material from Section 11.4 of the text.

Definition 2. *If X and Y are random variables, their* **covariance** *is*

$$Cov(X, Y) = E(XY) - E(X)E(Y).$$

Note $Cov(X, X) = var(X)$. Also note if X and Y are independent, then $Cov(X, Y) = E(X)E(Y) - E(X)E(Y) = 0$ so covariance measures how dependent X and Y are.

Lemma 3. $|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}$.

Proof. Compute

$$Var(X \pm Y) = E((X \pm Y)^2) - E(X \pm Y)^2 = Var(X) \pm 2Cov(X, Y) + Var(Y).$$

Since this is always ≥ 0, we have

$$Cov(X, Y) \leq \frac{Var(X) + Var(Y)}{2}.$$

Applying this inequality to the normalized random variables $X' = \frac{X-\mu_X}{\sigma_X}$ and $Y' = \frac{Y-\mu_Y}{\sigma_Y}$ gives

$$Cov(X',Y') \leq \frac{Var(X') + Var(Y')}{2}.$$

Note $E(X') = E(Y') = 0$ and $Var(X') = Var(Y') = 1$, so we have

$$Cov(X',Y') = E(X'Y') \leq 1.$$

Since

$$\begin{aligned} E(X'Y') = \frac{E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y)}{\sigma_X \sigma_Y} &= \frac{E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y}{\sigma_X \sigma_Y} \\ &= \frac{E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}, \end{aligned}$$

Proof. (of Theorem when $n = 1$) Here $\hat{\theta} = \hat{\theta}(X_1)$ is just a function of X_1, so we may think of it as a function of X. Observe

$$Cov(V'(\theta), \hat{\theta}) = E(V'(\theta) \cdot \hat{\theta}) - E(V'(\theta))E(\hat{\theta}) = E(V'(\theta) \cdot \hat{\theta})$$

by Lemma 1. By Lemma 3, we have

$$|E(V'(\theta) \cdot \hat{\theta})| = |Cov(V'(\theta), \hat{\theta})| \leq \sqrt{Var(V'(\theta))Var(\hat{\theta})}.$$

Hence

$$Var(\hat{\theta}) \geq \frac{|E(V'(\theta) \cdot \hat{\theta})|^2}{Var(V'(\theta))} = \frac{|E(V'(\theta) \cdot \hat{\theta})|^2}{I(\theta)},$$

Rao Blackwell's Theorem:

Let $\mathbf{X} \sim f_{\mathbf{X}}(\mathrm{x}, \theta)$ and T be sufficient for θ, $\mathrm{x} \in \mathfrak{X}$ and $t \in \mathfrak{T}$. Let U be any unbiased estimator

for $g(\theta)$. Define $V_t = \mathrm{E}(U|T = t)$. Then V is an unbiased estimator for $g(\theta)$ and $\mathrm{Var}(V) \leq \mathrm{Var}(U)$ witl equality iff $V = U$ with probability one.

Proof

Since $U = U(\mathbf{X})$ is an estimator, it is also a statistic. And, since T is sufficient for θ we have

$$\begin{aligned} V &= \mathrm{E}(U|T = t) && (1) \\ &= \int_{\mathfrak{T}} u(x) f_{X|T}(x|T = t)\, dx && (2) \end{aligned}$$

By Fisher, and noting that $u(x)$ is a function of x and not θ, we see that V is θ-free. Thus, V is a statistic as well.

Further,

$$\begin{aligned}
\mathrm{E}(U) &= g(\theta) & (3)\\
&= \int_{\mathfrak{X}} u(x) f_X(x,\theta)\, dx & (4)\\
&= \int_{\mathfrak{T}} \left[\int_{X \in T=t} u(x) f_{X|T}(x|T=t)\, dx\right] f_T(t,\theta)\, dt & (5)\\
&= \int_{\mathfrak{T}} v(t) f_T(t,\theta)\, dt & (6)\\
&= \mathrm{E}(V) & (7)
\end{aligned}$$

So, V is unbiased.

$$\begin{aligned}
\mathrm{Var}(U) &= \mathrm{E}\left(U - \mathrm{E}(U)\right)^2 & (8)\\
&= \mathrm{E}\left(U - \mathrm{E}(V)\right)^2 & (9)\\
&= \mathrm{E}\left((U-V)^2\right) + \mathrm{E}\left((V-\mathrm{E}(V))^2\right) + 2\mathrm{E}\left((U-V)(V-\mathrm{E}(V))\right) & (10)
\end{aligned}$$

Since we know that $\mathrm{E}(U) = \mathrm{E}(V)$ by above,

$$\begin{aligned}
\mathrm{E}\left((U-V)(V-\mathrm{E}(V))\right) &= \int_{\mathfrak{X}} (V-\mathrm{E}(V))(U-V) f_X(x,\theta)\, dx & (11)\\
&= \int_{\mathfrak{T}} (V-\mathrm{E}(V)) \left[\int_{X \in T=t} (U-V) f_{X|T}(x|T=t)\, dx\right] f_T(t,\theta)\, dt & (12)\\
&= \int_{\mathfrak{T}} (V-\mathrm{E}(V))[0] f_T(t,\theta)\, dt & (13)\\
&= 0 & (14)
\end{aligned}$$

and thus

$$\begin{aligned}
\mathrm{Var}(U) &= \mathrm{E}\left((U-V)^2\right) + \mathrm{E}\left((V-\mathrm{E}(V))^2\right) & (15)\\
&\geq \mathrm{E}\left((V-\mathrm{E}(V))^2\right) & (16)\\
&\geq \mathrm{Var}(V) & (17)
\end{aligned}$$

with equality iff $\mathrm{E}\left((U-V)^2\right) = 0$ or $V = U$ with probability one.

METHODS OF ESTIMATION

1. Method of Maximum Likelihood Estimation
2. Method of Moments
3. Method of Minimum Chisquare
4. Method of Least Squares
5. Method of Minimum Variance
6. Method of Inverse Probability

1.**Method of Maximum Likelihood**

Definition- Likelihood function.

Let $x_1, x_2, \ldots x_n$ be a random sample of size n from a population with density function f(x, θ).Then the likelihood function is given by $L = f(x_1,\theta) f(x_2,\theta) \ldots f(x_n,\theta) = \prod_{i=1}^{n} f(x_i,\theta)$. If there exists a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots x_n)$ of the sample values which maximises L for variations in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. Thus $\hat{\theta}$ is clled maximum likelihood estimator if $\frac{\partial L}{\partial \theta} = 0, \frac{\partial^2 L}{\partial \theta^2} < 0$.

The **maximum likelihood estimator (MLE)** is the value of the parameter θ, that maximises the likelihood function, given the observed sample data, $(X_1, \ldots, X_n)$.

It is often mathematically more tractable to maximise a sum of functions, than a product of function. Therefore, instead of trying to maximise the likelihood function we prefer to maximise the **log-likelihood function**

$$\log L(\theta) = \sum_{i=1}^{n} \log f(X_i; \theta).$$

2.**Method of moments**

Let $f(x;\theta_1, \theta_2 \ldots \theta_k)$ be the density function of the parent population with k parameters $\theta_1, \theta_2 \ldots \theta_k$. The rth moment about origin is given by $\mu_r' = \int_{-\infty}^{\infty} x^r \ f(x;\theta_1, \theta_2 \ldots \theta_k)$ dx,

(r =1,2,…k)

The method of moments consists in solving k equations for θ_1, $\theta_2, \ldots \theta_k$ in terms of $\mu_1', \mu_2', \ldots \mu_k'$ and replacing these moments by the sample moments.Then by the method of moments $\hat{\theta}_1, \hat{\theta}_2 \ldots \hat{\theta}_k$ are the required estimators of $\theta_1, \theta_2 \ldots \theta_k$.

If $X_1, X_2, \ldots$ are sample data drawn from a given distribution then the k^{th} sample moment si defined as

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

and by the Law of Large Numbers (under the appropriate condition) we have that $\hat{\mu}_k$ approximates μ_k, as the sample size gets larger.

The idea behind the Method of Moments is the following: Assume that we want to estimate a parameter θ of the distribution. Then we try to express this parameter in terms of moments of the distribution.

3.**Method of minimum chi-squared**

Minimum chi-square estimation is a method of estimation of unobserved quantities based on observed data.In certain chi-square tests, one rejects a null hypothesis about a population distribution if a specified test statistic is too large, when that statistic would have approximately a chi-square distribution if the null hypothesis is true. In minimum chi-square estimation, one finds the values of parameters that make that test statistic as small as possible.

Among the consequences of its use is that the test statistic actually does have approximately a chi-square distributionwhen the sample size is large. Generally, one reduces by 1 the number of degrees of freedom for each parameter estimated by this method.

4.Method of minimum variance(Minimum variance unbiased estimates(M.V.U.E))

If $L = \prod_{i=1}^{n} f(x_i, \theta)$ is the likelihood function of a random sample of n observations from a population with probability function f(x, θ) , then the problem is to find a statistic t = in R,

$$V(T) = \int_{-\infty}^{\infty} \{t - \upsilon(\theta)\}^2 LdX \text{ is minimum}$$

Properties of the estimators determined by the method of maximum likelihood

1.The first and second order derivatives exists and are continuous functions of θ

2.The third order derivative exists such that $\left|\frac{-\partial^2 \log L}{\partial \theta^2}\right| < M(x)$

3.For every θ in R, $E\left|\frac{-\partial^2 \log L}{\partial \theta^2}\right| = \int_{-\infty}^{\infty} \frac{-\partial^2 \log L}{\partial \theta^2} Ldx = I(\theta)$ is finite and non zero.

4.The range of integration is independent of θ.

Theorem 1: With probability approaching unity as $n \to \infty$,the likelihood equation

$\frac{\partial \log L}{\partial \theta} = 0$ has a solution which converges in probability to the true value θ_0. In other words M.L.E are consistent.

Theorem 2: Any consistent solution of the likelihood equation provides a maximum of the likelihood with probability tending to unity as the sample size (n) tends to infinity.

Theorem 3: A consistent solution of the likelihood equation is asymptotically normally distributed about the true value θ_0.

Theorem 4: If M.L.E exists, it is the most ef, then ficient in the class of each estimation.

Theorem 5 : If a sufficient estimator exists, it is a function of the maximum likelihood estimator.

Theorem 6:If T is the M.L.E. of θ and $\psi(\theta)$ is 1-1 function ofθ then $\psi(T)$ is the M.L.E of$\psi(\theta)$.

Properties of the estimators determined by the method of moments

1.Sample moments are consistent estimators of the corresponding population moments.

2.Under normal conditions, the estimators obtained by method of moments ae asym $\psi(\theta)$ ptotically normal.

3.Estimators determined by method of moments are in general inefficient.

4.As the method of moments do not depend on estimation theory, it is unable to give estimatos if population moments do not exist.

5.estimators obtained by method of moments are identical with those by mehthod of maximum likelihood if the density is of the form f(x, θ) $= e^{a_0 + a_1 x + \ldots}$

PROBLEMS

7. *Suppose that the underlying distribution is a normal $N(\mu, \sigma^2)$ and we want to estimate the mean μ and variance σ^2 from sample data $(X_1, \ldots, X_n)$, using the maximum likelihood estimator.*

First, we start with the log-likelihood function, which in this case is

$$\log L(\mu, \sigma) = -n \log \sigma - \frac{n}{2} \log(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2.$$

To maximise the log-likelyhood function we differentiate with respect to μ, σ and obtain

$$\frac{\partial L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)$$

$$\frac{\partial L}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3} \sum_{i=1}^{n} (X_i - \mu)^2$$

the partials need to be equal to zero and therefore solving the first equation we get that

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i := \overline{X}.$$

Setting the second partial equal to zero and substituting $\mu = \hat{\mu}$ we obtain the maximum likelyhood estimator for the standard deviation as

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

8. *Suppose we want to estimate the parameters of a $Gamma(\alpha, \theta)$ distribution*

$$f(x;\alpha,\theta) = \frac{\theta^{-\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta}$$

The maximum likelihood equations are

$$0 = -n\log\theta + \sum_{i=1}^{n}\log X_i - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

$$0 = n\alpha\theta - \sum_{i=1}^{n} X_i$$

Solving these equations in terms of the parameters we get

$$\hat{\theta} = \frac{\overline{X}}{\hat{\alpha}}$$

$$0 = n\log\hat{\alpha} - n\log\overline{X} + \sum_{i=1}^{n}\log X_i - n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}.$$

9.Examine the parameter 'p' in sampling from binomial population with density $f(x,n,p) = nC_x p^x q^{n-x}, x = 0,1,...n$ by the method of moments.

Sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = m,$

Population mean $\mu_1^{'} = \text{E(x)} = \text{np}$

$$np = \bar{x}\ ,\ \hat{p} = \frac{\bar{x}}{n} = \frac{m}{n}$$

Estimate for parameter p is $\frac{m}{n}$.

10.By method of moments obtain the estimate for m and σ of the normal population

Sample moments $a_1 = \bar{x}$ $,a_2 = \bar{x} + \sigma^2$

Population moments $\mu_1^{'} = m, \mu_2 = \mu_2^{'} + (\mu_1^{'}) = \sigma^2 + m^2$

By the method of moments, $m = \bar{x}, a_2 = \sigma^2 + m^2$

$$\sigma^2 = a_2 - m^2 = a_2 - \bar{x}^2 = s^2$$

Hence σ = s

REFERENCES

1. Wikipedia
    a. https://en.wikipedia.org/wiki/Point_estimation
    b. https://en.wikipedia.org/wiki/Statistical_inference
2. Britannica - https://www.britannica.com/science/point-estimation