LECTURE NOTES

STATISTICAL INFERENCE II (STA 2103)

YEAR TWO SEMESTER ONE

By the end of the course unit, students should be able to;

- Estimate parameters using different methods
- Test hypotheses
- State and prove the different properties of estimators.

COURSE CONTENT

- Introduction [definition of concepts]
- Estimation methods of estimators
  - Point estimation methods [method of moments, maximum likelihood estimation method, Bayesian method, least squares estimation method ,minimum chi-square method etc]
  - Interval estimation methods
- Properties of estimators/ evaluating the goodness of estimators
  - Unbiasedness [tests for unbiasedness of estimators]
  - Sufficiency
  - Completeness
  - Best linear unbiased estimators
  - Uniform minimum variance unbiased estimators
  - Efficiency
  - Consistency
- Distributions with their derivations
  - Normal distribution [mgf,mean and variance]
  - T-distribution
  - Fisher's distribution
- Hypothesis testing
  - Definition of terms and concepts
  - Best critical region
  - Neyman Pearson Lemma for the best critical region
  - Likelihood ratio tests

REFERENCES

- Probability and mathematical statistics by Prasanna Sahoo
- Introduction to mathematical statistics by Robert V. Hogg and Allen T. Craig

CHAPTER ONE

1.0 INTRODUCTION

Statistical inference centres on the rules and processes of using sample data in order to gain more information about the underlying population from which the sample was collected. It enables us to make judgements about population parameters based on sample statistics. It requires estimation of parameters and testing hypotheses about the population parameters.

1.1 definitions

a) Parameter: numerical value describing the characteristics of a population e.g population mean and variance.
b) Statistic: numerical value describing the characteristic of a sample e,g sample mean.
c) Estimator: is a random variable whose value varies from sample to sample
d) Best unbiased estimator: an estimator that is closest to the population parameter among all the unbiased estimators.

1.2properties of estimators

The following are the desirable properties of estimators;

a) Unbiasedness : an estimator $\hat{\theta}$ is said to be unbiased if its expected value is equal to the unknown population parameter $\theta. that\ is\ E(\hat{\theta}) = \theta$.
b) Sufficiency: an estimator is said to be sufficient if it uses up all the information about a population parameter contained in the sample.
c) Efficiency : if it has minimum variance among all unbiased estimators.
d) Consistency : an estimator is consistent if it approaches the unknown population parameter as sample size increases.$\hat{\theta} \to \theta\ as\ n \to \infty$.

2.0ESTIMATION OF PARAMETERS

Parameters can be estimated using point estimation methods and interval estimation methods. The point estimation methods give a single value for the unknown population parameter while the interval estimation methods give a range of values for the unknown population parameter. The point estimation methods include;

a) Method of moments.

There are two types of moments;

i. Raw moments: these are moments about the origin or about zero. Let $x_1,x_2,\ldots,x_n$ be a random sample from a population, the r^{th} sample moment about zero is defined as the expected value of x^r denoted as μ_r *where* $\mu_r = E(x^r) = \sum_{allx} x^r f(x)$ for x discrete and $E(x^r) = \int_{all\,x} x^r f(x)dx$ for x continuous. When r = 1, $\mu_1 = E(x^1) = \bar{x}$, this is the mean of the random variable x. when r = 2, $\mu_2 = E(x^2)$, but the variance of the random variable x is $v(x) = E(x^2) - (E(x))^2 = \mu_2 - \mu_1^2$.

EXAMPLE ONE

If x has a pdf f(x)= x, 0<x<2, find the third raw moment, mean and variance of x.

SOLUTION

i. The third raw moment is $r = 3\ is\ \mu_3 = \int_{all\,x} x^3 f(x)dx = \int_0^2 x^3 * xdx = \int_0^2 x^4 dx = \frac{x^5}{5}\Big|_0^2 = \frac{2^5}{5} = \frac{32}{5} = 6.4.$

ii. Mean is when r =2, $\mu_2 = \int_0^2 x^2 * xdx = \int_0^2 x^3 dx = \frac{x^4}{4}\Big|_0^2 = \frac{2^4}{4} = \frac{16}{4} = 4.$

iii. $v(x) = E(x^2) - (E(x))^2 = \mu_2 - \mu_1^2$ where $\mu_1 = \int_0^2 x^1 * xdx = \frac{x^3}{3}\Big|_0^2 = \frac{2^3}{3} = \frac{8}{3}$, therefore the variance $v(x) = 4 - (\frac{8}{3})^2 = 4 - \frac{64}{9} = -3.11.$

EXAMPLE TWO

Find the second raw moment given the pdf f(x)= 2x, x= 0,1,2. Find mean and variance.

SOLUTION

i. Second raw moment is for r = 2, $\mu_2 = \sum_0^2 x^2 f(x) = \sum_0^2 x^2 * 2x = \sum_0^2 2x^3 = 2(0^3 + 1^3 + 2^3) = 2(1 + 16) = 2 * 17 = 34.$

ii. Mean is r =2, $\mu_2 = 34$

iii. Variance is $v(x) = E(x^2) - (E(x))^2 = \mu_2 - \mu_{1.}^2$ where $\mu_1 = \sum_0^2 x^1 * 2x = \sum_0^2 2x^2 = 2(0^2 + 1^2 + 2^2) = 2(1 + 4) = 2 * 5 = 10. therefore\ v(x) = 34 - 10^2 = -66.$

The other type of moments is;

ii. Central moment: this is the moment about the mean. The r^th moment about the mean of a random variable x is the expected value of $(x - \mu)^r\ denoted\ by\ \mu_{r.}$ that is $\mu_r = E(x - \mu)^r = \sum_{all\ x}(x - \mu)^r f(x)\ or\ \mu_r = \int_{allx}(x - \mu)^r f(x)dx$ for x is discrete or continuous respectively.
When r = 1, $\mu_1 = E(x - \mu)^1 = E(x) - \mu = \mu - \mu = 0.$ when r = 2, $\mu_2 = E(x - \mu)^2 = var(x).$

EXAMPLE THREE

Using the pdf f(x)=1, 0<x<1 . find the fourth central moment.

SOLUTION

The fourth central moment is when r = 4, $\mu_4 = \int_0^1 (x - \mu)^4 f(x)dx.$ but $\mu = \int_0^1 xf(x)dx = \frac{x^2}{2}\Big| = \frac{1}{2}$

$$\mu_4 = \frac{(x - \mu)^5}{5}\Bigg|_0^1 = \frac{(1 - 0.5)^5}{5} - \frac{(0 - 0.5)^5}{5} = 0.00625 + 0.00625 = 0.0125.$$

2.1PARAMETER ESTIMATION

a) METHOD OF MOMENTS

let $x_1,x_2,\ldots,x_n$ be a random sample from a population x with pdf $f(x;\theta_1,\theta_2,\ldots,\theta_m)\ where\ \theta_1,\theta_2,\ldots,\theta_m$ are m unknown parameters. Let $E(x^k)=\int_{-\infty}^{\infty} x^k f(x;\theta_1,\theta_2,\ldots,\theta_m)dx$ be the kth population moment about zero. Let $m_k=\frac{1}{n}\sum_{i=1}^{n} x_i^k$ be the kth sample moment about zero. The estimator for the parameters $\theta_1,\theta_2,\ldots,\theta_m$ are obtained by equating the first m population moments (if they exist) to the first m sample moments, that is;

$E(X)= M_1$, $E(x^2)=M_2$, $E(X^3)=M_3,\ldots,E(X^m)=M_m$.

EXAMPLE ONE

Let $\boldsymbol{X\sim N(\mu,\delta^2)}$ and $x_1,\ldots,x_n$ be a random sample of size n from the population x. what are the estimators of the population parameters $\mu\ and\ \delta^2$ using the method of moments.

SOLUTION

(a) For a normal population E(x)=μ and from moments $E(x)=M_1$ where $M_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$, therefore the estimator for the mean $\mu = \bar{x}$.

(b) For the variance $\delta^2 = E(x^2) - \mu^2 . making\ E(x^2) the\ subject\ and\ equating\ it\ to\ M_2.$

$$\delta^2 = M_2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$therefore, \delta^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$ as the estimator for the variance.

EXAMPLE TWO

Given the pdf $f(x;\theta) = \begin{cases} \frac{1}{\theta}\ if\ 0 < x < \theta \\ 0\ elsewhere \end{cases}$. Find an estimator for θ by the method of moments.

SOLUTION

$$\boldsymbol{E(x) = \int_{0}^{\theta} xf(x)dx = \int_0^{\theta} x * \frac{1}{\theta}dx = \frac{x^2}{2\theta}\bigg|_0^{\theta} = \frac{\theta^2}{2\theta} = \frac{\theta}{2}}$$

Equate this to the sample moment; $E(x) = M_1 = \bar{x}$

$\bar{x}=\frac{\theta}{2}, implying\ that\ \hat{\theta} = 2\bar{x}$ as the estimator for θ.

EXAMPLE THREE

Given the pdf $f(x;\theta) = \begin{cases} \frac{1}{\theta}\ if -\theta < x < \theta \\ 0\ elsewhere \end{cases}$. Find an estimator for θ by the method of moments.

SOLUTION

$$\boldsymbol{E(x) = \int_{-\theta}^{\theta} xf(x)dx = \int_{-\theta}^{\theta} x * \frac{1}{\theta}dx = \frac{x^2}{2\theta}\bigg|_{-\theta}^{\theta} = \frac{\theta^2}{2\theta} - \frac{\theta^2}{2\theta} = 0}$$

The moment does not exist at E(x), try $E(X^2)$

$E(x^2) = \int_{-\theta}^{\theta} x^2 f(x)dx = \int_{-\theta}^{\theta} x^2 * \frac{1}{\theta}dx = \frac{x^3}{3\theta}\bigg|_{-\theta}^{\theta} = \frac{\theta^2}{3} + \frac{\theta^2}{3} = \frac{2\theta^2}{3}$

Equating $E(x^2)$ to $M_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$

$\frac{1}{n}\sum_{i=1}^{n} x_i^2 = \frac{2\theta^2}{3} = \frac{3}{2n}\sum_{i=1}^{n} x_i^2 = \theta^2$, this implies that $\hat{\theta} = \sqrt{\frac{3}{2n}\sum_{i=1}^{n} x_i^2}$.

TRY: Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with pdf $\boldsymbol{f(x;\theta) = \begin{cases} \theta x^{\theta-1}, 0 < x < 1 \\ 0\ otherwise \end{cases}}$. Using the method of moments, find an estimator for θ. If $x_1= 0.2$, $x_2= 0.6$, $x_3= 0.5$, $x_4= 0.3$ is a random sample of size 4, what is the estimate of θ?

b) MAXIMUM LIKELIHOOD ESTIMATION METHOD (MLE)

Maximum likelihood estimators are the values that maximize the log likelihood function for a random sample X from a population with pdf f(x;θ) where θ is unknown. The estimator is obtained by determining the first derivative of the log likelihood function with respect to θ. The likelihood function is given as $L(x;\theta) = \prod_{i=1}^{n} f(x;\theta) = f(x_1;\theta) * f(x_2;\theta) *** f(x_n;\theta)$

The log likelihood function is; $logL(x;\theta) = \sum_{i=1}^{n} logf(x;\theta)$.

EXAMPLE ONE

If $x_1, x_2, \ldots, x_n$ is a random sample from a distribution with density function $f(x;\theta) = \begin{cases} (1-\theta)x^{-\theta}, 0 < x < 1 \\ 0\ elsewhere \end{cases}$. What is the MLE of θ?

SOLUTION

$$L(X;\theta) = \prod_{i=1}^{n} (1-\theta)x^{-\theta} = (1-\theta)^n \prod x^{-\theta}$$

Determine the log likelihood function;

$$logl = nlog(1-\theta) - \theta \sum logx$$

Maximize the log likelihood function with respect to θ.

$$\frac{dlogl(\theta)}{d\theta} = \frac{-n}{(1-\theta)} - \sum logx$$

Equate the result to zero.

$$\frac{n}{(1-\theta)} = -\sum logx, solve\ for\ \theta$$

$$\frac{n}{-\sum logx} = (1-\theta), \theta = 1 + \frac{n}{\sum \log x}$$

EXAMPLE TWO: determine the MLE for λ given a poisson distribution with pdf $f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$

SOLUTION

- Determine the likelihood function $l(x;\lambda) = \Pi f(x;\lambda) = \frac{e^{-n\lambda}\lambda^{\sum x}}{\Pi x!}$
- Determine the log likelihood function $logl(x;\lambda) = -n\lambda loge + \sum xlog\lambda - \sum logx!$
- Maximize the log likelihood function. $\frac{dlog(x;\lambda)}{d\lambda} = -n + \frac{\sum x}{\lambda} = 0$
- Solve for λ; $\lambda = \frac{\sum x}{n} = \bar{x}$, as the estimator for λ.

Try: for the random sample x from a population with pdf $\boldsymbol{f(x;\beta)} = \begin{cases} \frac{\boldsymbol{x^6 e^{-\frac{x}{\beta}}}}{\boldsymbol{\Gamma 7 \beta^7}}, \boldsymbol{if\ 0 < x < \infty}. \\ \boldsymbol{0\ elsewhere} \end{cases}$ **determine the MLE for β.**

c) LEAST SQUARES ESTIMATION METHOD (LSE)

This is suitable for estimating moments about zero of a population distribution. To derive the LSE of μ_r for a random sample of variables $x_1, x_2, \ldots, x_n$;

- Consider a random variable x and its r^{th} moment about zero; $E(x^r)=\mu_r$
- Using the sum $\sum(x^r - \mu_r)^2$, determine the value of μ_r that makes the above sum as small as possible by differentiating the sum of squares with respect to μ_r and equating the result to zero.

EXAMPLE: DETERMINE THE LSE GIVEN r =1

SOLUTION

- From $E(x^1)=\mu_1$, the sum of squares is $\sum(x-\mu_1)^2$
- Differentiate with respect to μ_1 and equate the result to zero
- $\frac{d\sum(x-\mu_1)^2}{d\mu_1} = -2\sum(x-\mu_1) = 0;\ \sum x - n\mu_1 = 0$

- $\hat{\mu} = \frac{\sum x}{n} = \bar{x}$

TRY : DETERMINE THE ESTIMATOR GIVEN r= 2.

2.2INTERVAL ESTIMATION METHODS FOR PARAMETERS

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with pdf f(x;θ) where θ is an unknown parameter. The interval estimator of θ is a pair of statistics $L=L(x_1,x_2,\ldots,x_n)$ and $U=U(x_1,x_2,\ldots,x_n)$ with L <=U such that if $x_1, \ldots, x_n$ is a set of sample data, then θ belongs to the interval $[L(x,\ldots,x_n), U(x_1,\ldots,x_n)]$. The probability of θ being on the random interval (L,U) is 1-α, that is P[L $\leq \theta \leq U$]=1-α. Where L is the lower confidence interval and U is the upper confidence interval, (1-α) is the confidence coefficient or degree of confidence.

The interval estimation methods include; pivotal quantity method, MLE, Bayesian method, Invariant method, inversion of test statistic etc.

2.2.1 PIVOTAL QUANTITY METHOD

Definition: let $x_1,x_2,\ldots,x_n$ be a random sample of size n from a population x with probability density function f(x;θ) where θ is an unknown parameter. A pivotal quantity Q is a function of $x_1,x_2,\ldots,x_n$ and θ whose probability distribution is independent of the parameter θ. It is given as $Q(x_1,x_2,\ldots,x_n, \theta)$.

If $Q= Q(x_1,x_2,\ldots,x_n,\theta)$ is a pivot, then a (1-α)100% confidence interval for θ can be constructed as follows;

- Find two values a and b such that P(a$\leq Q \leq b) = 1 - \alpha$.
- Convert the inequality $a \leq Q \leq b$ into the form $L \leq \theta \leq U$.

EXAMPLE: if x is a normal population with mean unknown μ and known variance δ^2 and pivotal quantity as $Q = \frac{\bar{x}-\mu}{\sigma}$, construct a (1-2α)100% confidence interval for μ.

Solution

Since the population x is normally distributed with mean μ and variance δ^2, the sample mean $\bar{x}$ is also normal with the same mean μ and variance $\frac{\delta}{\sqrt{n}}$. hence

$$1-2\alpha = P(-Z_\alpha \le Q \le Z_\alpha)$$

$$= P\left(-Z_\alpha \le \frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}} \le Z_\alpha\right) = P\left(-Z_\alpha \frac{\delta}{\sqrt{n}} \le \bar{x}-\mu \le Z_\alpha \frac{\delta}{\sqrt{n}}\right)$$

$=P(\bar{x} - Z_\alpha \frac{\delta}{\sqrt{n}} \le \mu \le \bar{x} + Z_\alpha \frac{\delta}{\sqrt{n}})$

Therefore, (1-2α)100% confidence interval for μ is $[\bar{x} - Z_\alpha \frac{\delta}{\sqrt{n}}, \bar{x} + Z_\alpha \frac{\delta}{\sqrt{n}}]$

a) CONFIDENCE INTERVAL FOR POPULATION MEAN

Let $x_1, x_2, \ldots, x_n$ be a random sample from a normal population with mean μ and variance δ^2 where μ is an unknown parameter and variance is a known parameter. To construct this pivotal quantity , find the likelihood estimator of the parameter μ which is $\bar{x}$. Since each of the x_i is approximately normally distributed with mean μ and variance δ^2, the distribution of the sample mean$\bar{x}$ *is normally distributed with mean μ and variance* $\frac{\delta^2}{n}$. The pivotal function is $\frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}}$ which is used to construct a confidence interval for the mean μ.

The confidence interval is ;

- $1-\alpha = P[-Z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\frac{\delta}{\sqrt{n}}} \le Z_{\frac{\alpha}{2}}]$
- $= p[-Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} \le \bar{x}-\mu \le Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}]$
- =p$[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}]$
- When x is normally distributed with known variance δ^2, confidence interval will be given by ; $[\bar{x} - Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}, \bar{x} + Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}]$

EXAMPLE: let $x_1, x_2, .., x_{40}$ be a random sample of size 40 from a distribution with known variance and unknown mean μ. If $\sum_{i=1}^{40} x_i = 286.56\ and\ \delta^2 = 10$. what is the 90% confidence interval for the population mean μ.

SOLUTION

- **Using the** interval $\bar{x} - Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}} \frac{\delta}{\sqrt{n}}]$
- **Where α=10%,** $\mathbf{z}_{0.05}$=1.64
- $\bar{x} = \frac{\sum x}{n} = \frac{286.56}{40} = 7.164$
- $7.164 - (1.64 * \frac{\sqrt{10}}{\sqrt{40}} \leq \mu \leq 7.164 + \left(1.64 * \frac{\sqrt{10}}{\sqrt{40}}\right) = [6.344\ , 7.984]$

TRY: what is the 95% confidence interval for the mean μ given a sample of size 11, $\sum_{i=1}^{11} x = 132, \delta^2 = 9.9$.

b) **CONFIDENCE INTERVAL FOR μ FOR UNKNOWN VARIANCE.**

The pivotal quantity Q($x_1, x_2, \ldots, x_n$; μ)=$\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \sim t_{n-1}$, where n-1 are the degrees of freedom. The confidence interval for the mean will be;

- $1 - \alpha = p[-t_{\frac{\alpha}{2}, n-1} \leq \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}} \leq t_{\frac{\alpha}{2}, n-1}]$
- $make\ \mu\ the\ subject$
- P[$\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}]$
- Confidence interval is [$\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}},\ \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}]$

EXAMPLE: a random sample of size 9 from a normal population yields the observed statistics $\bar{x} = 5\ and\ \frac{1}{8} \sum_{i=1}^{9} (x_i - \bar{x})^2 = 36$. what is the 95% confidence interval for the mean μ?

SOLUTION

- Using the interval [$\bar{x} - t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{\frac{\alpha}{2}, n-1} \frac{s}{\sqrt{n}}]$
- Where α=0.05 and t=2.306

- $5 - \left(2.306 * \frac{6}{\sqrt{9}}\right) \leq \mu \leq 5 + \left(2.306 * \frac{6}{\sqrt{9}}\right) = [0.388, 9.612]$

c) **CONFIDENCE INTERVAL FOR THE POPULATION VARIANCE δ².**

The pivotal quantity Q(x_1,x_2,..,x_n; δ^2)=$\sum_{i=1}^{n}(\frac{x_i-\mu}{\delta})^2$ when mean μ is known and variance is unknown. This has a chi-square distribution with n degrees of freedom. The confidence interval for the variance δ^2 is determined as;

- $1 - \alpha = P[x^2_{\frac{\alpha}{2},n} \leq \sum_{i=1}^{n}(\frac{x_i-\mu}{\delta})^2 \leq x^2_{1-\frac{\alpha}{2},n}]$
- $= p[\frac{1}{x^2_{1-\frac{\alpha}{2},n}} \leq \frac{\delta^2}{\sum_{i=1}^{n}(x_i-\mu)^2} \leq \frac{1}{x^2_{\frac{\alpha}{2}}}]$
- $= P[\frac{\sum_{i=1}^{n}(x-\mu)^2}{x^2_{1-\frac{\alpha}{2},n}} \leq \delta^2 \leq \frac{\sum_{i=1}^{n}(x-\mu)^2}{x^2_{\frac{\alpha}{2},n}}]$ as the confidence interval.

EXAMPLE : a random sample of size 9 from a normal population with mean 5 yields the observed statistics as $\frac{1}{8}\sum_{i=1}^{9} x^2 = 39.125\ and\ \sum_{i=1}^{9} x = 45$. what is the 95% confidence interval for the variance δ^2?

SOLUTION

- Using the interval $P[\frac{\sum_{i=1}^{n}(x-\mu)^2}{x^2_{1-\frac{\alpha}{2},n}} \leq \delta^2 \leq \frac{\sum_{i=1}^{n}(x-\mu)^2}{x^2_{\frac{\alpha}{2},n}}]$
- Where α=0.05, $x^2_{0.025,9} = 2.7, x^2_{0.975,9} = 19.02, \mu = \frac{45}{9} = 5$
- $\sum(x-\mu)^2 = \sum(x^2 + \mu^2 - 2x\mu) = \sum x^2 + n\mu^2 - 2\mu\sum x = 313 + (9 * 25) - (2 * 5 * 45) = 88.$
- Therefore c.i is ; $[\frac{88}{19.02}, \frac{88}{2.7}]$= [4.627, 32.593]

d) **CONFIDENCE INTERVAL FOR THE VARIANCE FOR UNKNOWN MEAN.**

The pivotal quantity is given by Q= $\frac{\sum(x-\bar{x})^2}{\delta^2} \sim x_{n-1}^2$. the (1-α)100% confidence interval for the variance δ^2 is;

- $1-\alpha = P\left[\frac{1}{x_{\frac{\alpha}{2},n-1}^2} \leq Q \leq \frac{1}{x_{1-\frac{\alpha}{2}}^2}\right] = p[\frac{1}{x_{\frac{\alpha}{2},n-1}^2} \leq \frac{\sum(x-\bar{x})^2}{\delta^2} \leq \frac{1}{x_{1-\frac{\alpha}{2}}^2}]$
- $= P[\frac{\sum(x-\bar{x})^2}{x_{1-\frac{\alpha}{2},n-1}^2} \leq \delta^2 \leq \frac{\sum(x-\bar{x})^2}{x_{\frac{\alpha}{2},n-1}^2}]$

EXAMPLE: let $x_1, x_2, .., x_n$ be a random sample of size 13 from a normal distribution with mean μ and variance δ^2. If $\sum_{i=1}^{n} x = 246.61 \; and \; \sum_{i=1}^{n} x^2 = 4806.61$. find a 90% confidence interval for δ^2?

SOLUTION

- Using the interval $[\frac{\sum(x-\bar{x})^2}{x_{1-\frac{\alpha}{2},n-1}^2}, \; \frac{\sum(x-\bar{x})^2}{x_{\frac{\alpha}{2},n-1}^2}]$
- Where $\alpha = 10\%, x_{0.95,12}^2 = 21.03 \; and \; x_{0.05,12}^2 = 5.23$
- $\sum(x-\bar{x})^2 = \sum x^2 + n\bar{x} - 2\bar{x}\sum x = 4806.61 + (13 * 18.97^2) - (2 * 18.97 * 246.61) = 128.419$
- $\left[\frac{128.419}{21.03}, \frac{128.419}{5.23}\right] = [6.107, 24.554]$

NOTE: for a random sample $x_1, x_2, \ldots, x_n$ from a distribution N(μ, δ^2) where μ and δ^2 are unknown parameters. The confidence interval for the variance is given by; $[\frac{(n-1)s^2}{x_{1-\frac{\alpha}{2},n-1}^2}, \frac{(n-1)s^2}{x_{\frac{\alpha}{2},n-1}^2}]$.

e) **CONFIDENCE INTERVAL FOR PARAMETERS OF SOME DISTRIBUTIONS**

A (1-α)100% confidence interval for the parameter θ can be constructed by taking the pivotal quantity (Q) as either; $Q = -2\sum_{i=1}^{n} lnF(x;\theta) \approx x_{,2n}^2 \; OR \; Q = -2\sum_{i=1}^{n} \ln(1-F(x;\theta)) \approx x_{,2n}^2$.

That is; $1-\alpha = P[x_{\frac{\alpha}{2},2n}^2 \leq Q \leq x_{1-\frac{\alpha}{2},2n}^2]$ and $F(x;\theta) = \int_0^x f(x)dx, is \; the \; cumulative \; distribution.$

EXAMPLE I: if $x_1, x_2, \ldots, x_n$ is a random sample from a distribution with pdf $f(x;\theta) = \begin{cases} \frac{1}{\theta}, 0 < x < \theta \\ 0\ elsewere \end{cases}$ where θ>0 is a parameter. What is the (1-α)100% confidence interval for θ?

SOLUTION

- Obtain the cumulative distribution function F(x;θ). $F(x;\theta) = \int_0^x \frac{1}{\theta} dx = \frac{x}{\theta}\Big|_0^x = \frac{x}{\theta}$.
- Taking pivotal quantity Q as $Q = -2\sum_{i=1}^{n} lnF(x;\theta) = -2\sum_{i=1}^{n} \ln(\frac{x}{\theta})$
- $\boldsymbol{= -2\sum_{i=1}^{n} lnx_i + 2nln\theta = 2nln\theta - 2\sum_{i=1}^{n} ln\, x_i}$
- $confidence\ interval\ for\ \theta\ is$;
- $1{-}\alpha = P[P\left[x^2_{\frac{\alpha}{2},2n} \leq Q \leq x^2_{1-\frac{\alpha}{2},2n}\right] = P[x^2_{\frac{\alpha}{2},2n} \leq 2nln\theta - 2\sum_{i=1}^{n} lnx_i \leq x^2_{1-\frac{\alpha}{2},2n}]$
- $= \boldsymbol{P}[x^2_{\frac{\alpha}{2},2n} + 2\sum_{i=1}^{n} lnx_i \leq 2nln\theta \leq x^2_{1-\frac{\alpha}{2},2n} + \boldsymbol{2\sum_{i=1}^{n} lnx_i}]$
- $= \boldsymbol{P}[\frac{1}{2n}(x^2_{\frac{\alpha}{2},2n} + 2\sum_{i=1}^{n} lnx_i) \leq\ ln\theta \leq \frac{1}{2n}(\, x^2_{1-\frac{\alpha}{2},2n} + \boldsymbol{2\sum_{i=1}^{n} lnx_i})]$
- **=**$\boldsymbol{P}[\boldsymbol{e}^{[\frac{1}{2n}(x^2_{\frac{\alpha}{2},2n}+2\sum_{i=1}^{n} lnx_i)} \leq \boldsymbol{\theta} \leq \boldsymbol{e}^{[\frac{1}{2n}(x^2_{1-\frac{\alpha}{2},2n}+2\sum_{i=1}^{n} lnx_i)}]$

The confidence interval is $[\boldsymbol{e}^{[\frac{1}{2n}(x^2_{\frac{\alpha}{2},2n}+2\sum_{i=1}^{n} lnx_i)}, \boldsymbol{e}^{[\frac{1}{2n}(x^2_{1-\frac{\alpha}{2},2n}+2\sum_{i=1}^{n} lnx_i)}]$

EXAMPLE II

If $x_1, x_2, \ldots, x_n$ is a random sample from a distribution with pdf $f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{\frac{-x}{\theta}}, 0 < x < \infty \\ 0\ elsewhere \end{cases}$ where θ> 0 is a parameter. What is the (1-α)100% confidence interval for θ?

SOLUTION

- $\mathrm{F(x;\theta)} = \int_0^x \frac{1}{\theta} e^{\frac{-x}{\theta}} dx = \frac{1}{\theta}\int_0^x e^{\frac{-x}{\theta}} dx = \frac{1}{\theta}\left[-\theta e^{\frac{-x}{\theta}}\right]_0^x = -e^{\frac{-x}{\theta}}\Big|_0^x = -e^{\frac{-x}{\theta}} + e^0 = 1 - e^{\frac{-x}{\theta}}$

- taking Q $= -2\sum_{i=1}^{n} \ln(1 - F(x;\theta)) = -2\sum_{i=1}^{n} \ln\left(1 - \left(1 - e^{\frac{-x}{\theta}}\right)\right) =$

  $-2\sum_{i=1}^{n} lne^{\frac{-x}{\theta}} = \frac{2}{\theta}\sum_{i=1}^{n} x_i lne = \frac{2}{\theta}\sum_{i=1}^{n} x_i$
- confidence interval for θ; $1 - \alpha = p\left[x^2_{\frac{\alpha}{2},2n} \le Q \le x^2_{1-\frac{\alpha}{2},2n}\right]$
- $= P\left[x^2_{\frac{\alpha}{2},2n} \le \frac{2}{\theta}\sum_{i=1}^{n} x_i \le x^2_{1-\frac{\alpha}{2},2n}\right] = P[\frac{2\sum_{i=1}^{n} x}{x^2_{1-\frac{\alpha}{2},2n}} \le \theta \le \frac{2\sum_{i=1}^{n} x}{x^2_{\frac{\alpha}{2},2n}}]$

Therefore; confidence interval is $[\frac{2\sum_{i=1}^{n} x}{x^2_{1-\frac{\alpha}{2},2n}}, \frac{2\sum_{i=1}^{n} x}{x^2_{\frac{\alpha}{2},2n}}]$

TRY: (i) for the pdf $f(x;\theta) = \begin{cases} \theta x^{\theta-1}, 0 < x < 1 \\ 0\ elsewhere \end{cases}$ where θ> 0 is an unknown parameter. What is the (1-α)100% confidence interval for θ?

(ii) using $\sum_{i=1}^{49} lnx_i = -0.7567$, determine 90% confidence interval for θ.

CHAPTER THREE

3.0 EVALUATING THE GOODNESS OF ESTIMATORS USING THE PROPERTIES OF ESTIMATORS

The properties of unbiasedness, consistency, efficiency, relative efficiency, uniform minimum variance unbiased estimator and sufficiency shall be considered.

3.1UNBIASEDNESS

An estimator $\hat{\theta}$is unbiased for the population parameter θ if its expected value is equal to the unknown population parameter. That is; $E\left(\hat{\theta}\right) = \theta$.

EXAMPLE I

Let $x_1, x_2, .., x_n$ be a random sample from a population with mean μ and variance $\delta^2>0$. Is the sample variance s^2 an unbiased estimator for the population variance δ^2?

SOLUTION

- Required to show that $E(s^2)=\delta^2$.
- Given $S^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$ take expectations, $E(S)^2 = E(\frac{1}{n-1}\sum(x-\bar{x})^2)$
- $=\frac{1}{n-1}E[\sum(x^2 - 2x\bar{x} + \bar{x}^2)] = \frac{1}{n-1}E[\sum x^2 - n\bar{x}^2] = \frac{1}{n-1}[\sum E(x^2) - nE(\bar{x}^2)] \ldots \ldots \ldots .(a)$
- But $E(\bar{x}) = E\left[\frac{x_1+x_2+\cdots+x_n}{n}\right] = \frac{1}{n}\sum E(x_i) = \frac{1}{n}\sum \mu = \frac{1}{n} * n\mu = \mu$............(b)
- And $V(\bar{x}) = v\left[\frac{x_1+x_2+\cdots+x_n}{n}\right] = \frac{1}{n^2}\sum v(x_i) = \frac{1}{n^2}\sum \delta^2 = \frac{n\delta^2}{n^2} = \frac{\delta^2}{n}$.............(c)
- From $V(\bar{x}) = E(\bar{x}^2) - (E(\bar{x}))^2)and\ taking\ E(\bar{x})^2 = v(\bar{x}) + \left(E(\bar{x}^2)\right), substituting\ (b) and\ (c)$
- $E(\bar{x})^2 = \frac{\delta^2}{n} + \mu^2 \ldots \ldots \ldots \ldots \ldots \ldots \ldots (d) and\ E(x^2) = v(x) + (E(x))^2 = \delta^2 + \mu^2$
- $substituting\ in\ expression\ (a)$

- $E(S^2) = \frac{1}{n-1}[\sum E(x^2) - nE(\bar{x}^2)] = \frac{1}{n-1}\left[n(\delta^2 + \mu^2) - n\left(\frac{\delta^2}{n} + \mu^2\right)\right] = \frac{1}{n-1}[(n-1)\delta^2] = \delta^2.$

Therefore, S^2 is an unbiased estimator for the population parameter δ^2.

EXAMPLE II

Let x_1,x_2,x_3 be a random sample of size 3 from a population with mean μ and variance δ^2>0. If the statistics $\bar{x}$ *and* y *where* $y = \frac{x_1+2x_2+3x_3}{6}$ are two estimators, are the estimators $\bar{x}$ *and* y *unbiased estimators for* μ?

SOLUTION

- Required to show $E(\bar{x}) = \mu$ *and* $E(y) = \mu$

  For $\bar{x}$

- $\bar{x} = \frac{x_1+x_2+x_3}{3}$, take expectations $E(\bar{x}) = E\left[\frac{x_1+x_2+x_3}{3}\right] = \frac{1}{3}E(x_1 + x_2 + x_3)$
- $= \frac{1}{3}E[\sum x_i] = \frac{1}{3}\sum E(x_i) = \frac{1}{3}\sum \mu = \frac{1}{3} * 3\mu = \mu.$
- $E(\bar{x}) = \mu$, *implying that* $\bar{x}$ *is an unbiased estimator for* μ.

For y

- $y = \frac{x_1+2x_2+3x_3}{6}$, *take expectations* $E(y) = E\left[\frac{x_1+2x_2+3x_3}{6}\right]$
- $= \frac{1}{6}E(x_1 + 2x_2 + 3x_3)$
- $= \frac{1}{6}(E(x_1) + 2E(x_2) + 3E(x_3) = \frac{1}{6}(\mu + 2\mu + 3\mu) = \mu$
- $E(y) = \mu$, *implying unbiasedness.*

TRY: let x$_1$,x_2,…,x_n be a random sample of size n from a distribution with unknown mean $-\infty < \mu < \infty$ *and* unknown variance δ^2>0. Show that the statistics $\bar{x}$ *and* $y = \frac{x_1+2x_2+\cdots+nx_n}{\frac{n(n+1)}{2}}$ *are unbiased estimators of* μ.

3.2 RELATIVELY EFFICIENT ESTIMATOR

If an estimator is unbiased and has smaller variance compared to another, then it is more efficient. Let $\hat{\theta}_1$ *and* $\hat{\theta}_2$ be two unbiased estimators of θ, the estimator $\hat{\theta}_1$ is said to be more efficient than $\hat{\theta}_2$ if variance ($\hat{\theta}_1$) is less than variance ($\hat{\theta}_2$).

The ratio η given by $\eta(\hat{\theta}_1, \hat{\theta}_2) = \frac{var(\hat{\theta}_2)}{var(\hat{\theta}_1)}$ is the relative efficiency of $\hat{\theta}_1$ *with respect to* $\hat{\theta}_2$.

EXAMPLE

Let x_1,x_2 and x_3 be a random sample of size 3 from a population with mean μ and variance δ^2>0. If the statistics $\bar{x}$ *and* $y = \frac{x_1+2x_2+3x_3}{6}$ are two unbiased estimators of the population mean μ. Which of the statistics is more efficient and what is the relative efficiency of $\bar{x}$ *with respect to* y?

SOLUTION

- Required to determine and compare the variances.

Variance for $\bar{x}$

- $\bar{x} = \frac{x_1+x_2+x_3}{3}, take\ variances\ v(\bar{x}) = v\left[\frac{x_1+x_2+x_3}{3}\right] = \frac{1}{9}v(x_1 + x_2 + x_3) = \frac{1}{9}[v(x_1) + v(x_2) + v(x_3)] = \frac{1}{9}[\delta^2 + \delta^2 + \delta^2] = \frac{1}{9} * 3\delta^2 = \frac{\delta}{3}$

Variance for y

- $y = \frac{x_1+2x_2+3x_3}{6}$, take variances $v(y) = v\left[\frac{x_1+2x_2+3x_3}{6}\right]$
- $= \frac{1}{36}[v(x_1) + 4v(x_2) + 9v(x_3)] = \frac{1}{36}[\delta^2 + 4\delta^2 + 9\delta^2] = \frac{14\delta^2}{36}$

Compare the variances; v($\bar{x}$)< v(y). Therefore $\bar{x}$ is more efficient.

Relative efficiency of $\bar{x}$ $with\ respect\ to\ y$.

- R.e = $\eta(\bar{x}, y) = \frac{v(y)}{v(\bar{x})} = \frac{\frac{14\delta^2}{36}}{\frac{\delta^2}{3}} = 7/6$.

TRY: let x_1,x_2,x_3 be a random sample of size 3 from a population with pdf $f(x;\lambda) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, x = 0,1,2,..,\infty \\ 0\ elsewhere \end{cases}$. Are the estimators $\hat{\lambda}_1 = \frac{1}{4}(x_1 + 2x_2 + x_3)$ and $\hat{\lambda}_2 = \frac{1}{9}(4x_1 + 3x_2 + 2x_3)$ unbiased? Which is more efficient and what is the relative efficiency of $\widehat{\lambda_2}$ $with\ respect\ to\ \widehat{\lambda_1}$?

3.3 **UNIFORM MINIMUM VARIANCE UNBIASED ESTIMATOR (UMVUE) , EFFICIENCY AND INFORMATION**

Definition: an unbiased estimator $\hat{\theta}$ of θ is said to be a uniform minimum variance unbiased estimator of θ if it minimizes the variance ($\hat{\theta}$) given by E$(\hat{\theta} - \theta)^2$.

The UMVUE can be obtained using Cramer Rao lower bound or fisher information inequality. For a random sample $\underline{x} = (x_1, x_2, .., x_n)$ with pdf f(x;θ) and a likelihood function L(θ) a differentiable function of θ, the Cramer Rao lower bound is $v(\hat{\theta}) \geq \frac{-1}{E[\frac{d^2 lnL(\theta)}{d\theta^2}]}$.

Definition: let x_1,x_2,..,x_n be a random sample of size n from a population x with pdf f(x;θ) where θ is a parameter. If $\hat{\theta}$ is an unbiased estimator of θ and $(\hat{\theta}) = \frac{-1}{E[\frac{d^2 lnL(\theta)}{d\theta^2}]}$, then $\hat{\theta}$ is a UMVUE of θ.

Definition: an estimator $\hat{\theta}$ is called an efficient estimator if it satisfies the cramer Rao lower bound and every efficient estimator is a uniform minimum variance unbiased estimator.

Let Y be an unbiased statistic from a random sample x for an unknown population parameter θ in a family of exponential pdfs f(x;θ). Then Y is said to be an efficient statistic for θ iff the variance obeys the Cramer Rao inequality given by $\delta_y^2 \geq \frac{1}{E[\frac{d^2 lnL(\theta)}{d\theta^2}]}$ $OR\ \ \delta_y^2 \geq \frac{1}{n*E[\frac{dlnf(x;\theta)^2}{d\theta}}$

Definition: Fisher's information in x_i ($I_{x_i}(\theta)$)is defined as the expected value of the second derivative of the log or log likelihood function, that is ($I_{x_i}(\theta) = -E[\frac{d^2 lnf(x;\theta)}{d\theta^2}]$. For a random sample $\underline{x} = x_1, x_2, \dots, x_n$, the information in the sample is sample size multiplied by the information in the first random variable, that is $I_{\underline{x}}(\theta) = n * I_{x_i}(\theta)$

EXAMPLE I

Given the pdf $f(x;\theta) = \theta x^{\theta-1}, 0 < x < 1\ and\ 0\ elsewhere$. determine (i) information in the sample (ii) the cramer rao lower bound.

SOLUTION

(i) Information in a sample $I_{\underline{x}}(\theta) = n * I_{x_i}(\theta)$ where $I_{x_i}(\theta) = -E[\frac{d^2 lnf(x;\theta)}{d\theta^2}$

- Take logs; $lnf(x;\theta) = ln\theta + \theta lnx - lnx$
- Obtain the score (differentiate with respect to θ); $\frac{dlnf(x;\theta)}{d\theta} = \frac{1}{\theta} + lnx$
- Second derivative: $\frac{d^2 lnf(x;\theta)}{d\theta^2} = \frac{-1}{\theta^2}$
- $I_{x_i}(\theta) = -E\left[\frac{d^2 lnf(x;\theta)}{d\theta^2}\right] = -E\left[\frac{-1}{\theta^2}\right] = \frac{1}{\theta^2}$
- $I_{\underline{x}}(\theta) = n * I_{x_i}(\theta) = n * \frac{1}{\theta^2} = \frac{n}{\theta^2}$

(ii) Cramer rao lower bound is $v(\hat{\theta}) \geq \frac{-1}{E\left[\frac{d^2 lnL(\theta)}{d\theta^2}\right]} or\ v(\hat{\theta}) \geq \frac{-1}{n*E[\frac{d^2 lnf(x;\theta)}{d\theta^2}]}$

- $v(\hat{\theta}) \geq \frac{-1}{n*E\left[-\frac{1}{\theta^2}\right]} = \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n}$.

EXAMPLE II

Let $x_1, x_2, .., x_n$ be a random sample of size n from a distribution with pdf

$f((x;\theta) = \begin{cases} 3\theta x^2 e^{-\theta x^3}, 0 < x < \infty \\ 0\ elsewhere \end{cases}$. What is the Cramer rao lower bound for the variance of the unbiased estimator of the parameter θ?

SOLUTION

The cramer rao lower bound is given by $v(\hat{\theta}) \geq \frac{-1}{E\left[\frac{d^2 lnL(\theta)}{d\theta^2}\right]}$

- Determine the likelihood function $L(\theta) = \Pi f(x;\theta) = \Pi 3\theta x^2 e^{-\theta x^3} = \theta^n \sum 3x^2 e^{-\theta \sum x^3}$
- Introduce logs ; $lnL(\theta) = nln\theta + \sum ln3x^2 - \theta \sum x^3$
- Differentiate with respect to θ; $\frac{dlnL(\theta)}{d\theta} = \frac{n}{\theta} - \sum x^3$
- Obtain second derivative; $\frac{d^2 \ln L(\theta)}{d\theta^2} = \frac{-n}{\theta^2}$
- $v(\hat{\theta}) \geq \frac{-1}{E\left[\frac{d^2 lnL(\theta)}{d\theta^2}\right]} = \frac{1}{E[\frac{n}{\theta^2}]} = \frac{\theta^2}{n}$

TRY: Let $x_1,x_2,..,x_n$ be a random sample from a normal population with unknown mean μ and known variance $\delta^2>0$. What is the MLE for μ? Is this MLE an efficient estimator for μ?

3.4 **SUFFICIENT STATISTICS**

Let $\boldsymbol{x \sim f(x;\theta)}$ be a population and let $x_1,x_2,\ldots,x_n$ be a random sample of size n from this population x. An estimator $\hat{\theta}$ of the parameter θ is said to be a sufficient statistic of θ if the conditional distribution of the sample given the estimator $\hat{\theta}$ does not depend on the parameter θ. That is;

$f(x_1, x_2, .., x_n | Y = y) = \frac{f(x_1, x_2, \ldots, x_n)\ and\ f(Y=y)}{f(Y=y)}$ does not depend on y.

EXAMPLE I

If $x_1,x_2,\ldots,x_n$ is a random sample from a distribution with pdf

$f(x;\theta) = \begin{cases} \theta^x (1-\theta)^{1-x}, x = 0,1. \\ 0\ elsewhere \end{cases}$ where 0 < θ <1. Show that $Y = \sum x_i$ is a sufficient statistic for θ.

SOLUTION

Using conditional distribution ; $f(x_1, x_2, .., x_n|Y = y) = \frac{f(x_1,x_2,...,x_n)\ and\ f(Y=y)}{f(Y=y)}$

- Where $f(x_1, x_2, .., x_n|Y = y) = \frac{\Pi f(x_i;\theta)}{f(Y=y)}$ and $Y = \sum x_i$ is a Bernoulli pdf given as $f(y) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$
- $\frac{\Pi f(x_i;\theta)}{f(Y=y)} = \frac{\Pi\theta^x(1-\theta)^{1-x}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}}$ but $\sum x_i = y$
- $\frac{\theta^y(1-\theta)^{n-y}}{\binom{n}{y}\theta^y(1-\theta)^{n-y}} = \frac{1}{\binom{n}{y}}$, the conditional density of the sample given the statistic Y is independent of the parameter θ. Therefore a sufficient statistic.

TRY: given the pdf $f(x;\theta) = e^{-(x-\theta)}, 0 < x < \infty\ and\ 0 < \theta < \infty\ and\ f(y) = ne^{-n(y-\theta)}, 0 < y < \infty$. show that y is a sufficient statistic for θ.

3.41 FACTORISATION THEOREM OF NEYMAN

Let $x_1, x_2, \ldots, x_n$ denote a random sample with pdf f(x;θ), which depends on the population parameter θ. The estimator $\hat{\theta}$ is sufficient for θ iff the likelihood function can be factorised into two k_1 and k_2 where k_1 has θ and k_2 does not depend on θ. that is; $k_1(x_1, x_2, \ldots, x_n; \theta)k_2(x_1, x_2, \ldots, x_n)$.

EXAMPLE

Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution with pdf

$f(x;\lambda) = \begin{cases} \lambda^x e^{-\lambda}/x! \\ 0\ elsewhere \end{cases}, x = 0,1,2, \ldots, \infty$. Where λ > 0 is a parameter. Show that the estimator is a sufficient estimator of the parameter λ.

SOLUTION

Using the factorization theorem, determine the likelihood function and factorise the function into two.

- $L(\lambda) = \frac{\Pi_{i=1}^{n}\lambda^x e^{-\lambda}}{x!} = \frac{\lambda^{\sum x_i}e^{-n\lambda}}{\Pi x!}$

- $factorise\ this;\ (\lambda^{\sum x_i} e^{-n\lambda})(\frac{1}{\Pi x!})$
- $= k_1(x;\lambda)k_2(x_i)$. Therefore it is a sufficient estimator.

TRY: let $x_1, x_2, \ldots, x_n$ be a random sample from a Bernoulli population having the density function $f(x;\theta) = \theta^x (1-\theta)^{1-x}, x = 0,1\ and\ 0\ elsewhere$. obtain the sufficient statistic for a Bernoulli.

3.41 FACTORISATION THEOREM OF NEYMAN

Let $x_1, x_2, \ldots, x_n$ denote a random sample with pdf f(x;θ), which depends on the population parameter θ. The estimator $\hat{\theta}$ is sufficient for θ iff the likelihood function can be factorised into two k_1 and k_2 where k_1 has θ and k_2 does not depend on θ. that is; $k_1(x_1, x_2, \ldots, x_n; \theta)k_2(x_1, x_2, \ldots, x_n)$.

EXAMPLE

Let $x_1, x_2, \ldots, x_n$ be a random sample from a distribution with pdf

$f(x;\lambda) = \begin{cases} \lambda^x e^{-\lambda}/x! \\ 0\ elsewhere \end{cases}, x = 0,1,2, \ldots, \infty$. Where λ > 0 is a parameter. Show that the estimator is a sufficient estimator of the parameter λ.

SOLUTION

Using the factorization theorem, determine the likelihood function and factorise the function into two.

- $L(\lambda) = \frac{\Pi_{i=1}^{n} \lambda^x e^{-\lambda}}{x!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\Pi x!}$
- $factorise\ this;\ (\lambda^{\sum x_i} e^{-n\lambda})(\frac{1}{\Pi x!})$
- $= k_1(x;\lambda)k_2(x_i)$. Therefore it is a sufficient estimator.

TRY: let $x_1,x_2,\ldots,x_n$ be a random sample from a Bernoulli population having the density function $f(x;\theta)=\theta^x(1-\theta)^{1-x}, x=0,1\ and\ 0\ elsewhere$. obtain the sufficient statistic for a Bernoulli.

3.5 **CONSISTENT ESTIMATOR**

Let $x_1,x_2,\ldots,x_n$ be a random sample from a population X with pdf f(x;θ). Let $\hat{\theta}$ be an estimator of θ based on the sample of size n denoted as $(\hat{\theta}_n)$. A sequence of estimators $(\hat{\theta}_n)$ of θ is said to be consistent for θ if and only if the sequence $(\hat{\theta}_n)$ converges in probability to θ ,that is for any $\epsilon > 0$

$$\lim_{n\to\infty} P\left[\left|\hat{\theta}_n-\theta\right| \geq \in\right]=0.$$

Consistency is a large sample property of an estimator.

To show that a sequence of estimators is consistent, we verify the limits below;

- $\lim_{n\to\infty} var(\hat{\theta}_n)=0$
- $\lim_{n\to\infty} B(\hat{\theta}_n,\theta)=0$

CHAPTER FOUR

4.0HYPOTHESIS TESTING

A hypothesis is an assertion about the underlying population. The hypothesis to be tested is the null hypothesis (H_0) and its negative is the alternative hypothesis (H_A).

A hypothesis test is an ordered sequence ($x_1,x_2,\ldots,x_n$; H_o,H_A; c) where $x_1,x_2,\ldots,x_n$ is a random sample from a population x with the pdf f(x;θ), H_0 and H_A are hypotheses concerning the parameter θ in f(x;θ) and c is the critical region. Or it is a rule that tells us for which sample values we should decide to accept H_0 as true and for which sample values we should reject H_0 and accept H_A as true.

DEFINITIONS

- Sample space: space of all possible outcomes of a statistical experiment.
- Parameter space: all possible values of the unknown parameters.
- Simple hypothesis: is one that completely specifies the density function of the population e.g H_0: μ=1.
- Composite hypothesis: doesnot completely specify the population parameter e.g μ<1.
- Critical region (c): is a subset of the sample space which is in accordance with the agreed test leading to rejection of the hypothesis under reference.
- Size of the critical region is given as $P\left[\underline{x} \ \in c/H_0\right] = \alpha$.
- Power function of a test for testing H_0 against H_A is the probability of rejecting H_0. That is; $P = prob\{x\epsilon c\}$.
- Power of a test $\Pi(\theta_A) = prob\{x\epsilon c/H_A\}$ and is yielded by particular values of θ as per hypothesis.
- Level of significance for testing H_0 versus H_A is equivalent to the size of the critical region or is the maximum value of the power function assuming H_0 is true. That is; $P\left[\underline{x} \ \in c/H_0\right] = \alpha$.
- Type I error is as a result of rejecting H_0 when it is true. It is given as α.
- Type II error is as a result of accepting H_0 when it is false. It is equal to $1 - p\{x\epsilon \, c/H_A\}$.

4.1NEYMAN-PEARSON LEMMA FOR THE BEST CRITICAL REGION

Let $x_1, x_2, \ldots, x_n$ be a random sample from population with pdf f[x;θ] and $L(\theta; x_1, .., x_n) = \prod_{i=1}^{n} f(x_i; \theta\ ) is$ the likelihood function of the sample. Then any critical region C of the form $C\{(x_1, .., x_n)/\frac{L(\theta_0; x_1, \ldots, x_n)}{L(\theta_A; x_1, \ldots, x_n)} \leq k\}$ for some constant o<k<∞ is best critical region (or uniformly most powerful) of its size for testing $H_0: \theta = \theta_0 \; versus \; H_A: \theta = \theta_A$.

Therefore, if L(θ_0) is the likelihood function for the null hypothesis and L(θ_A) is the likelihood of the alternative hypothesis; $c = \{x; \frac{L(\theta_0)}{L(\theta_A)} \leq k\}$ is the best critical region.

EXAMPLE1

Suppose x has a pdf $f(x;\theta) = \begin{cases}(1+\theta)x^{\theta}, 0 \leq x \leq 1 \\ 0 \; elsewhere\end{cases}$. Based on a single observed value of x, find the best critical region given the hypothesis H_0: θ=1 versus H_A: θ=2.

SOLUTION

- By the Neyman –Pearson lemma $c = \{x; \frac{L(\theta_0)}{L(\theta_A)} \leq k\}$
- $L(x;\theta) = \Pi f(x;\theta) = (1+\theta)x^{\theta}$
- $for \; H_0: \theta = 1; L(\theta_0) = 2x \; and \; for \; H_A: \theta = 2; L(\theta_A) = 3x^2$
- $\frac{L(\theta_0)}{L(\theta_A)} = \frac{2x}{3x^2} = \frac{2}{3x}$
- Introduce k, $\frac{2}{3x} \leq k, solve \; for \; x; x \geq \frac{2k}{3}$
- Therefore; c=$\left\{x; x \geq \frac{2k}{3}\right\}$ $is \; the \; BCR$. therefore, reject H_0 if $x \geq \frac{2k}{3}$.

EXAMPLE 2

Let x_1, x_2, x_3 denote 3 independent observations from a pdf $f(x;\theta) = \begin{cases}(1+\theta)x^{\theta}, 0 \leq x \leq 1 \\ 0 \; elsewhere\end{cases}$. determine the best critical region given the hypothesis H_o: θ=1 and H_A: θ=2.

SOLUTION

- By the Neyman –Pearson lemma $c = \{x; \frac{L(\theta_0)}{L(\theta_A)} \leq k\}$
- $L(x;\theta) = \prod_{i=1}^{3} f(x;\theta) = \prod_{i=1}^{3}(1+\theta)x^{\theta} = (1+\theta)^3 \prod_{i=1}^{3} x_i^{\theta}$
- $l(\theta_0) for \; \theta = 1 is \; (1+1)^3 \prod_{i=1}^{3} x_i^1 = 2^3 x_1 x_2 x_3 = 8x_1 x_2 x_3$
- $for \; \theta = 2, l(\theta_A) = (1+2)^3 \prod_{i=1}^{3} x_i^2 = 27 x_1^2 x_2^2 x_3^2$

- $ratio\ \frac{L(\theta_0)}{L(\theta_A)} = \frac{8x_1x_2x_3}{27x_1^2x_2^2x_3^2} = \frac{8}{27x_1x_2x_3}$
- $bcr\ c = \{x; \frac{8}{27x_1x_2x_3} \leq k\}$
- $c = \{x; \frac{1}{x_1x_2x_3} \leq \frac{27}{8}k\}$
- $c = \{x; x_1x_2x_3 \geq \frac{8}{27k}\}$ as the best critical region.

TRY: given a random variable x from a population with a normal distribution with mean θ and variance one. Determine the best critical region given the hypothesis H_o : θ=0 and H_A: θ=1for a sample of size n.

EXAMPLE

Obtain the BCR given a random sample of size n from a population with pdf f(x;θ) and the hypotheses $H_0: \frac{e^{-1}}{x!}, x = 0,1,2,3 \ldots versus\ H_A: (\frac{1}{2})^{x+1}$, x= 0,1,2….

SOLUTION

- By the Neyman –Pearson lemma $c = \{x; \frac{L(\theta_0)}{L(\theta_A)} \leq k\}$
- $L(\theta_0) = \prod_{i=1}^{n} \frac{e^{-1}}{x!} = \frac{e^{-1}}{x!} \frac{e^{-n}}{\prod_{i=1}^{n} x!}$
- $l(\theta_A) = \prod_{i=1}^{n} (\frac{1}{2})^{x+1} = \left(\frac{1}{2}\right)^n (\frac{1}{2})^{\sum x_i}$
- $ratio\ \frac{l(\theta_0)}{l(\theta_A)} = \frac{\frac{e^{-1}}{x!} \frac{e^{-n}}{\prod_{i=1}^{n} x!}}{\left(\frac{1}{2}\right)^n (\frac{1}{2})^{\sum x_i}} = \frac{(2e^{-1})^n 2^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$
- $introduce\ k > 0, \frac{(2e^{-1})^n 2^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \leq k$
- Taking logs on both sides, $log \frac{(2e^{-1})^n 2^{\sum x_i}}{\prod_{i=1}^{n} x_i!} \leq logk$
- $\sum x_i log2 + nlog2e^{-1} - log \prod_{i=1}^{n} x_i! \leq logk = c$
- Make x the subject
- $\sum x_i log2 - log \prod_{i=1}^{n} x_i! \leq \log k - nlog2e^{-1}. this\ is\ the\ BCR.$

TRY

Given $x_{1,}x_{2,\ldots}..x_{12}$ is a random sample from a normal population with mean zero and variance δ^2. What is the BCR given the hypotheses $H_0: \sigma^2 = 10 \ versus\ H_A: \sigma^2 = 5$?

4.2LIKELIHOOD RATIO TEST

Definition: the likelihood ratio test for testing a simple null hypothesis H_0 $:\theta\epsilon\theta_0$ $versus$ the composite hypothesis H_A:$\theta\epsilon\theta_A$ based on a set of random sample data $x_1,x_2,\ldots,x_n$ is defined as $w(x_1,x_2,\ldots,x_n)=\frac{L(\underset{\smile}{x};\theta_0)}{L(\underset{\smile}{x};\theta_A)}$.

A likelihood ratio test (LRT) is any test that has a critical region C (rejection region) of the form $c=\{(x_1,x_2,..,x_n)/w(x_1,x_2,\ldots,x_n)\leq k\}$ where k is a number in the unit interval (0,1) and $w(x_1,\ldots,x_n)$ is the likelihood ratio test statistic.

EXAMPLE

Let x_1, x_2, x_3 denote 3 independent observations from a pdf $f(x;\theta)=\begin{cases}(1+\theta)x^\theta, 0\leq x\leq 1\\ \quad 0\ elsewhere\end{cases}$. what is the form of the likelihood ratio test given the hypothesis H_o: θ=1 and H_A: θ=2?

SOLUTION

From definition; $c=\{(x_1,x_2,..,x_n)/w(x_1,x_2,\ldots,x_n)\leq k\}=\{(x_1,\ldots,x_n)/\frac{L(\theta_0)}{L(\theta_A)}\leq k\}$

- $L(x;\theta)=\prod_{i=1}^{3}f(x;\theta)=\prod_{i=1}^{3}(1+\theta)x^\theta=(1+\theta)^3\prod_{i=1}^{3}x_i^\theta$
- $l(\theta_0) for\ \theta=1 is\ (1+1)^3\prod_{i=1}^{3}x_i^1=2^3x_1x_2x_3=8x_1x_2x_3$
- $for\ \theta=2, l(\theta_A)=(1+2)^3\prod_{i=1}^{3}x_i^2=27x_1^2x_2^2x_3^2$
- $ratio\ \frac{L(\theta_0)}{L(\theta_A)}=\frac{8x_1x_2x_3}{27x_1^2x_2^2x_3^2}=\frac{8}{27x_1x_2x_3}$
- $c=\{x;\frac{8}{27x_1x_2x_3}\leq k\}$
- $c=\{x;\frac{1}{x_1x_2x_3}\leq\frac{27}{8}k\}$
- $c=\{x;x_1x_2x_3\geq\frac{8}{27k}\}$, reject H_0 if $x_1x_2x_3\geq\frac{8}{27k}$.

CHAPTER FIVE

5.0NON-PARAMETRIC TESTS

These are referred to as distribution free tests. They do not make an assumption about the distribution of the data, that is normality. These include; sign test, runs test, chi-square test, Kruskal wallis test, mann Whitney test etc.

5.1CHI-SQUARE TEST

This can be used as a goodness of fit test and as a test of independence. The goodness of fit test is used to determine how well an observed set of data fits an expected set of data. The null hypothesis is that, there is no difference between the observed (f_0) and the expected (f_e) frequencies OR H_0: good fit versus H_A: poor fit OR H_0: not biased versus H_A: biased

The test statistic is the chi-square given as $x^2 = \sum_{i=1}^{k} \frac{(f_0-f_e)^2}{f_e}$, where $f_e=np_i$, n is the total number of observations , p_i is the probability for each category and k are the number of categories.

The critical region is $x^2 \geq x^2_{\propto,k-1}$. (i.e reject H_0).

EXAMPLE

A gambler suspects that a normal die is biased. He throws the die 108 times and gets the following results;

Number	1	2	3	4	5	6
Frequency	25	18	15	20	22	8

What decision would he make at 0.05 level of significance?

SOLUTION

- H_0: not biased (p_1= p_2=…=p_n =1/6) versus H_A: biased
- L.o.s α = 0.05
- C.r: $x^2 \geq x^2_{\propto,k-1} = x^2 \geq x^2_{0.05,6-1} = x^2 \geq 11.07$
- $x^2 = \sum_{i=1}^{k} \frac{(f_0-f_e)^2}{f_e}, where\ f_e = np_i = 108 * \frac{1}{6} = 18$

- Compute chi-square value

F_0	25	18	15	20	22	8
f_e	18	18	18	18	18	18
$\frac{(f_0-f_e)^2}{f_e}$	2.72	0.00	0.50	0.22	0.89	5.56

$$x^2 = \sum_{i=1}^{k} \frac{(f_0-f_e)^2}{f_e} = 9.89$$

- Decision: fail to reject H_0.

EXAMPLE

It is hypothesized that an experiment results in outcomes K, L,M and N with probabilities 1/5, 3/10, 1/10,2/5 respectively. Forty independent repetitions of the experiment have results as follows.

Outcome	K	L	M	N
Frequency	11	14	5	10

If a chi-square goodness of fit test is used to test the above hypothesis at the 0.01 significance level, then what is the value of the chi-square statistic and the decision reached?

SOLUTION

- H_0: observed frequency = expected frequency VERSUS H_A: they differ
- L.o.s is 0.01
- C.r: $x^2 \geq x^2_{\propto,k-1} = x^2 \geq x^2_{0.05,6-1} = x^2 \geq 11.07$
- C.r: $x^2 \geq x^2_{\propto,k-1} = x^2 \geq x^2_{0.01,4-1} = x^2 \geq 11.35$
- $x^2 = \sum_{i=1}^{k} \frac{(f_0-f_e)^2}{f_e}, where\ f_e = np_i$, n=40

F_0	11	14	5	10
f_e	8	12	4	16
$\frac{(f_0-f_e)^2}{f_e}$	1.125	0.333	0.25	2.25

$$x^2 = \sum_{i=1}^{k} \frac{(f_0 - f_e)^2}{f_e} = 3.958$$

- Fail to reject H_0.

TRY: the distribution of the number of deaths due to a rare disease in a certain year among cities was as follows;

Deaths(x)	0	1	2	3	4	5	6
No. of cities(f)	93	70	26	8	2	0	1

Test the goodness of fit of a poisson distribution to this data at 5% level of significance. [hint: degrees of freedom are k-2 since a parameter is estimated from the sample and the pdf of a poisson distribution(p_i) is $f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$]

TRY: six coins are tossed 64 times and the following number of heads were observed.

Heads (x)	0	1	2	3	4	5	6
frequency	1	9	10	25	12	5	2

Test the goodness of fit for the data at 5% level of significance using a binomial distribution with pdf $f(x) = p_i = \binom{6}{x} p^x q^{n-x}$.[hint estimate $\bar{x} = \frac{\sum f_i x_i}{\sum f_i}, since\ \bar{x} = np\ then, p = \frac{\bar{x}}{n}, d.o.f = k - 2$.]

NOTE:

- If there are only 2 cells, the expected frequency in each cell should be greater or equal to five.
- For more than 2 cells chi-square should not be applied if more than 20% of the expected frequencies are less than 5. For this, combine the small cells into one category.

5.2CONTINGENCY TABLE

This is an array of data in rows and columns which are assumed to be independent of each other.

Any statistical experimental data which can be presented in arrays of rows and columns is a contingency table. It used to test whether the categorical data given are independent of each other. The null hypothesis is that the variables are independent (no relationship) and the alternative is that they are dependent (related).

It uses the chi-square distribution with [(r-1)(c-1)] as the degrees of freedom. The critical region is given as $x^2 \geq x^2_{\propto,(c-1)(r-1)}$ (reject H_0) and the computed chi-square value is given as $x^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right)$, where O_{ij} is the observed frequency and E_{ij} is the expected frequency given as $E_{ij} = \frac{row\ total\ X\ column\ total}{grand\ total}$.

EXAMPLE:

Given the data below for education level and social status, test whether there is a relationship between the two at 0.05 level of significance.

Education	Above average	Average	Below average	Total
College	18	12	10	40
High school	17	15	13	45
Primary	9	9	22	40
Total	44	36	45	125

SOLUTION

- H_0: independent Versus H_A: dependent
- L.o.s = 0.05
- C.r $x^2 \geq x^2_{\propto,(c-1)(r-1)} = x^2 \geq x^2_{0.05,4} = x^2 \geq 9.488$
- Compute $x^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right)$,

Observed freq. (O_{ij})	Expected freq. (E_{ij})	$\left(\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right)$
18	14.08	1.09
17	15.84	0.09
9	14.08	1.83
12	11.52	0.02
15	12.96	0.32
9	11.52	0.55
10	14.4	1.34
13	16.2	0.63
22	14.4	4.01

- $x^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{(O_{ij}-E_{ij})^2}{E_{ij}}\right) = 9.88$
- Decision: reject H_0. They are dependent.

5.3THE KRUSKAL -WALLIS TEST

This is used for simultaneous comparisons of more than two independent populations. It is an alternative to the ANOVA test. It is used to test the hypothesis that k independent samples are from the same population where the actual values are replaced by ranks.

All the samples are ranked as if they are coming from the same population.

The test statistic is H which approximates to a chi-square distribution with (k-1) degrees of freedom, that is; $H \approx X^2_{\propto,k-1}$. where H is computed as, $H = \frac{12}{N(N+1)}\left[\frac{(\sum R_1)^2}{n_1} + \cdots + \frac{(\sum R_i)^2}{n_i}\right] - 3(N+1)$.

Where ,

N= sum of all the samples (n_1+n_2+…..)

R_i is the sum of the ranks for each sample

The critical region is $H \geq X^2_{\propto,k-1}$.

EXAMPLE

Three varieties of wheat planted in 12 plots of land at random gave the following yields per acre;

Variety A	Rank A	Variety B	Rank B	Variety C	Rank C
35	1	55	12	47	6
38	2	48	7	43	4
45	5	52	9	53	10
40	3	54	4	50	8
Sum ranks	11		39		28

Apply the Kruskal wallis test to decide whether there is a significant difference among the 3 varieties of wheat at 0,05 significance level.

SOLUTION

- H_0: no difference Vs H_A: they differ
- L.o.s is 0.05
- C.r $H \geq X^2_{\propto,k-1} = H \geq X^2_{0.05\ ,3-1} = H \geq 5.99$
- Rank data
- Compute $H = \frac{12}{N(N+1)}\left[\frac{(\sum R_1)^2}{n_1} + \cdots + \frac{(\sum R_i)^2}{n_i}\right] - 3(N+1) = \frac{12}{12(12+1)}\left[\frac{11^2}{4} + \frac{39^2}{4} + \frac{28^2}{4}\right] - 3(12+1) = 7.654.$
- Reject H_0. There is a difference in the varieties.

TRY: a travel agency selected samples of hotels from three major chains and recorded the occupancy rate for each hotel on a specific date as below. The occupancy rate is the percentage of the total number of rooms that were occupied the previous night.

Best Eastern	Comfort Inn	Quality Court
58%	69%	72%
57	67	80
67	62	84
63	69	94
61	77	86
64		

Do these data suggest any difference in the occupancy rates at 0.05 level of significance assuming that the occupancy percentages are not normally distributed?