2nd Assignment

STATISTICAL INFERENCE FOR DATA SCIENCE

SURESH KUMAR PRAJAPATI
CUSB2302222008

Sequential Analysis :-

Introduction :-

In Neyman - Pearson theory of testing of hypothesis, n, the sample size is regarded a fixed constant and α keeping fixed we minimize β.
But in the sequential analysis theory propounded by A. Wald n, the sample size is not fixed but is regarded as a random variable whereas both α and β are fixed constant.

SEQUENTIAL PROBABILITY RATIO TEST (SPRT) :-

To test the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$ for a distribution with p.d.f. $f(x, \theta)$. For any positive integer m, the likelyhood function of a sample $x_1, x_2, \ldots x_n$ from the population with p.d.f. (p.m.f), $f(x, \theta)$ is given by

$$L_{1m} = \prod_{i=1}^{m} f(x_i, \theta_1) \quad \text{when } H_1 \text{ is true.}$$

&

$$L_{0m} = \prod_{i=1}^{m} f(x_i, \theta_0) \quad \text{when } H_0 \text{ is true.}$$

then likelyhood ratio λ_m is given by.

$$\lambda_m = \frac{L_{1m}}{L_{0m}} = \frac{\prod_{i=1}^{m} f(x_i, \theta_1)}{\prod_{i=1}^{m} f(x_i, \theta_0)} = \prod_{i=1}^{m} \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}$$

$(m = 1, 2, \ldots)$

The SPRT for testing H_0 againsts H_1 is defined as follows:

At each stage of the experiment (at the mth trial for any integral value m), the likelihood ratio λ_m, $(m = 1, 2, \ldots)$ is computed.

(i) If $\lambda_m \geq A$, we terminate the process with rejection of H_0.

(ii) If $\lambda_m \leq B$, we terminate the process with accepting H_0,

(iii) If $B < \lambda_m < A$, we continue sampling by taking an additional observation.

Here, A and B $(B < A)$ are constants which are determined by the relation

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}$$

where α, & β are the probabilities of type I error and type II error respectively.

From computational point of view, it is much convenient to deal with $\log \lambda_m$ rather than λ_m, since.

$$\log \lambda_m = \sum_{i=1}^{m} \log \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} = \sum_{i=1}^{m} z_i$$

where,

$$z_i = \log \frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}$$

In terms of z_i's SPRT is defined as follow.

(i) if $\sum z_i \geq \log A$ reject H_0.

(ii) if $\sum z_i \leq \log B$, reject H_1 (Accepting H_0)

(iii) if $\log B < \sum z_i < \log A$, continue sampling by taking an additional information.

Remark :- 1. Additional information (observations) unless the inequality

$$B < \lambda_m < A \Rightarrow \log B < \Sigma x_i < \log A$$

is violated at either end.

It has been proved that SPRT eventually terminates with probability one.

2. Saving in terms of inspection, time and money. As compared with single sampling, sequential scheme requires on the average 33% to 50% less inspection for the same degree of protection i.e. for the same value of α and β.

OPERATING CHARACTERISTIC (O.C.) FUNCTION OF SPRT :-

The O.C. function $L(\theta)$ is defined as.

$L(\theta)$ = Probability of accepting $H_0 : \theta = \theta_0$ when θ is the true value of the parameter.

Power function

$P(\theta)$ = Probability of rejecting H_0 where θ is the true value, we get

$$L(\theta) = 1 - P(\theta)$$

The O.C. function of a SPRT for testing $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1$ in sampling from a popⁿ with density function $f(x, \theta)$ is given by.

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}}$$

where for each value of θ, the value of $h(\theta) \neq 0$ is to be determined so that.

$$E\left[\frac{f(x, \theta_1)}{f(x, \theta_0)}\right]^{h(\theta)} = 1.$$

AVERAGE SAMPLE NUMBER (A.S.N.):-

"The sample size n in sequential testing is a random variable which can be determined in terms of true density function $f(x, \theta)$."

The A.S.N. function for the S.P.R.T for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is given by.

$$E(n) = \frac{L(\theta) \cdot \log B + [1 - L(\theta)] \log A}{E(Z)}$$

where $Z = \log\left(\frac{f(x, \theta_1)}{f(x, \theta_0)}\right)$, $A = \frac{1-\beta}{\alpha}$, $B = \frac{\beta}{1-\alpha}$

Neyman Pearson lemma :-

Suppose we have a problem sample $X_1, X_2, \cdots X_n$ from a probability distribution with parameter θ. Then if c is a critical region of size α and K is a constant such that

$$\frac{L(\theta_0)}{L(\theta_\alpha)} \leq K \text{ inside the critical region } c$$

and $$\frac{L(\theta_0)}{L(\theta_\alpha)} \geq K \text{ outside the critical region } c.$$

then c is the best, i.e. most powerful critical region for testing the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_A : \theta = \theta_\alpha$.

Application :-

1 ⇒ Decision Making under uncertainty.

2 ⇒ Hypothesis testing wheather a certain Hypothesis is true or not not based on observed data.

N.P lemma has its limitations and disadvantage.

1:- Binary Decision

2:- Assumption of known Parameter:-

3:- The Neyman - Pearson lemma assumes that the parameters of the probability distributions under both hypothesis known.

4:- Not suitable for all situation.

5:- Limited scope:-

The Neyman-Pearson Lemma Focuses on controlling the probability of type I error often at the expense of type II error. In some situations, minimizing type II error might be more important.

6:- Complexity:- challenging especially for individuals without a strong background in statistical theory.

Name - Suresh Kumar Prajapati ①

CUSB2302222008

Master in data Science and Applied Statistics.

2^{nd} Sem. [2023-25]

1^{st} - Assighment

To summarise the information in sample by determining a few key features of the sample values like the sample mean, the sample variance, the largest observation and the smallest observation are four computing statistics that might be used to summarize some key feature of the sample.

Here, some key point

X = Random variable = $X_1, X_2 \dots X_n$

x = Sample variable = $x_1, x_2, \dots x_n$.

$T(X)$ = data Reduction or data summary

An experimenter who use only the observed value of the statistics, $T(X)$ rather than the entire observed sample, x, will treat as equal ~~two~~ samples x and y,

that satisfy $T(x) = T(y)$ even though the actual sample value may be differents in same ways.

08/03/2024 03:58

Data Reduction in terms of a particular statistics.
As a partition of the sample space $\mathcal{X}$. Let

$$\mathcal{T} = \{t : t = T(x) \text{ for some } x \in \mathcal{X}\}$$ be the image of $\mathcal{X}$ under $T(x)$.

Then, $T(x)$ partitions the sample space into sets A_t, $t \in \mathcal{T}$, defined by,

$$A_t = \{x : T(x) = t\}$$

* **Meaning**:- if $T(x) = x_1 + x_2 + \cdots + x_n$ then $T(x)$ does not report the actual sample values but only the sum. There may be many different sample points ~~points~~ that have the same sum.

Three principle of Data Reduction:-
Methods of data reduction that do not discard important information about the unknown parameter θ and methods that successfully discard information that is irrelevant as far as gaining knowledge about θ is concerned.

(i) Sufficiency:- not discard information about θ & summarize the data.
(ii) The Likelihood Principle:- Describe the f^n of the parameter, determ by the observed sample.
(iii) Equivariance principle:- Some important featerres of the model.

The Sufficiency Principle:-

A sufficient statistic for a parameter θ is a statistic that, it contains all the information about θ contained in the sample.

Any additionally information in the sample, besides the value of sufficient statistic, does not contain any more information about θ.

These consideration lead to the data reduction technique known as sufficiency principle.

OR

if x and y are two sample points such that $T(x) = T(y)$, then the inference about θ should be the same whether $X = x$ or $X = y$ is observed.

Sufficient statistics:-

A statistic $T(X)$ is a sufficient statistic for θ if the conditional distribution of the sample X given the value of $T(X)$ does not depend on θ.

Theorem:- If $p(\mathbf{x}|\theta)$ is the joint pdf or pmf of X and $q(t|\theta)$ is the pdf or pmf of $T(X)$ then $T(X)$ is a sufficient statistic for θ if for every x in the sample space, the ration $p(\mathbf{x}|\theta) / q(T(x)|\theta)$ is a constant as a function of θ.

Theorem:- (Factorization Theorem) :- Let $f(x/\theta)$ denote the joint pdf or pmf of a sample X.
A statistic $T(X)$ is a sufficient statistic for θ if and only if there exist functions $g(t|\theta)$ and $h(x)$ such that, for all sample points X and all parameter points θ.

$$f(x/\theta) = g\left(T(x)|\theta\right) h(x)$$

Theorem:- Let $X_1, X_2 \dots X_n$ be iid observations from a pdf or pmf $f(x|\theta)$ that belongs to an exponential family given by

$$f(x|\theta) = h(x) \cdot c(\theta) \cdot \exp\left(\sum_{i=1}^{K} w_i(\theta) \cdot t_i(x)\right)$$

where

$$\theta = (\theta_1, \theta_2, \dots \theta_d), \quad d \leq K.$$

$$T(X) = \left(\sum_{j=1}^{n} t_1(x_j), \dots \sum_{j=1}^{n} t_K(x_j)\right)$$

is a sufficient statistic for θ.

Minimal sufficient statistic

A sufficient statistic $T(X)$ is called a minimal sufficient statistics if for any other sufficient statistics $T'(X)$, $T(X)$ is a function of $T'(x)$.

Theorem:- Let $f(x|\theta)$ be the pmf or pdf of a sample X. Suppose $\exists$ a function $T(x)$ such that, for every two sample points x any y, the ratio $f(x|\theta) / f(y|\theta)$ is constant as a function of θ if and only if $T(x) = T(y)$. Then $T(X)$ is a minimal sufficient statistic for θ.

$$\frac{f(x|\theta)}{f(y|\theta)} = \frac{g'(T'(x)|\theta)\,h'(x)}{g'(T'(y)|\theta)\,h'(y)} = \cancel{g(T(x)|\theta)}\frac{h'(x)}{h'(y)}$$

Ancillary Statistics:—

A statistics $S(X)$ whose distribution does not depend on the parameter θ is called as ancillary statistics.

Sufficient, Ancillary and complete statistics:-

A minimal sufficient statistics is a statistic that has achieved the maximal amount of data reduction possible while still retaining all the information about the parameter θ.
it is eliminates all the extraneous information in the sample, retaing only that piece with information about θ.
Since the distribution of an ancillary statistic does not depend on θ, it might be suspected that a minimal sufficient statistic is unrelated

An ancillary statistics

Complete statistics

Let $f(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(X)$. The family of probability distribution is called complete if

$E_\theta g(T) = 0$ for all θ implies

$P_\theta(g(T) = 0) = 1$ for all θ.

Equivalently $T(X)$ is called a complete statistic.

* Completeness is a property of a family of probability distributions, not of a particular distribution.

eg:- if X has $n(0, 1)$ distribution, then defining $g(x) = x$, we have that $Eg(X) = EX = 0$.

But the function $g(x) = x$ satisfies

$P(g(X)=0) = P(X=0) = 0$ not 1.

However, this is a particular distribution, not a family distribution. If X has a $n(\theta, 1)$ distribution $-\infty < \theta < \infty$ we shall see that no function of X except one that is 0 with probability 1 for all θ, satisfies $E_\theta g(X) = 0$ for all θ. Thus the family of $n(\theta, 1)$ distributions $-\infty < \theta < \infty$ is complete.

08/03/2024 03:59

Basu's Theorem

If $T(X)$ is a complete and minimal sufficient statistic, then $T(X)$ is independent of every ancillary statistic.

* Basu's Theorem deduce the independence of two statistics without ever finding the joint distribution of two statistics.
its proof depends on the uniqueness of a Laplace transform a property.

Complete statistics in the exponential family:-

Let X_1, X_2 ——— X_n be iid observations from an exponential with pdf or pmf of the form.

$$f(x/\theta) = h(x)\, c(\theta) \exp\left(\sum_{j=1}^{k} w(\theta_j)\, t_j(x)\right)$$

where, $\theta = (\theta_1, \theta_2, \dots \theta_k)$. Then the statistic

$$T(X) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \dots \sum_{i=1}^{n} t_k(X_i)\right)$$

is complete as long as the parameter space Θ contains an open set in R.

Theorem:- If a minimal sufficient statistic exists, then any complete statistics is also a minimal sufficient statistic.

08/03/2024 03

Basu's Theorem gives one relationship between sufficients statistics and ancillary statistics using the concept of complete statistics.

* Some relationships between sufficiency and ancillarity for these definitions are discussed by Lehmann (1981)

The Likelihood Principle :-

Used to summarize data.

Let $f(x|\theta)$ denote the joint pdf or pmf of the sample $X = (X_1, X_2 \cdots X_n)$. Then, given that $X = x$ is observed, the function of θ defined by

$$L(\theta|X) = f(X|\theta)$$

is called the likelyhood function.

If X is a discreate random vector, then $L(\theta|X) = P_\theta(X=x)$. If we compare the likelyhood function at two parameter points and find that

$$P_{\theta_1}(X=x) = L(\theta_1/x) > L(\theta_2|x) = P_{\theta_2}(X=x)$$

then the sample we actually observed is more likely to have occurred if $\theta = \theta_1$ than if $\theta = \theta_2$.

If X is a continous, real valued random variable and if the pdf of X is continous in X, then for small ϵ

08/03/2024 0

$P_\theta(x-\epsilon < X < x+\epsilon)$ is approximately

$2\epsilon f(x|\theta) = 2\epsilon L(\theta|x)$

(this follows from the definition of a derivative)

$$\frac{P_{\theta_1}(x-\epsilon < X < x+\epsilon)}{P_{\theta_2}(x-\epsilon < X < x+\epsilon)} \approx \frac{L(\theta_1|x)}{L(\theta_2|x)}$$

Comparision of the likelihood function at two parameter values again gives an approximate comparision of the probability of the observed sample value x.

The Equivariance Principle :-

A function $T(x)$ is specified but if $T(x) = T(y)$, then the equivariance Principle states that the inference made if x is observed should have a certain relationship to the inference made. if y is observed, although the two inferences may not be the same. This restriction on the inference procedure sometimes leads to a simpler analysis, just as do the data reduction principles discussed in earlier sections.

Equivariance principle :-

If $Y = g(X)$ is a change of measurement scale such that the model for Y has the same formal structure as the model for X, then an inference procedure should be both measurement equivariant and formally equivariant.

Likelihood Application :-

Statistical inference :- estimate parameter
Hypothesis testing :- Likelihood ratio test
Model selection :- Akaike Information Criterion (AIC) } complete
Bayesian Information Criterion (BIC) } model.
Machine learning :- Probabilistic models
Econometric :- estimating the parameter of economic models.
Biostatistics :- Clinical trials
Genetics :- assessing genetic linkage.
Spatial statistics :- modeling spatial patterns and dependencies.

Equivariance Application :-

Image and signal processing :- image recognition and signal processing. convolutional neural network (CNN)
Rotation - Invariant Features :- Orientation of objects may vary
Medical Image Analysis :- Patient Positioning or orientation in image.
Natural Language Processing :- (NLP) :- sentiment analysis.
Robotics :- Involve sensory data processing
Augmented Reality :- system needs to recognize and interact with objects.
Physics and Materials Science :- Involve spatial transformation and rotation.

08/03/2024 04:0