BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Statistical Inference-II

Course Code: 23ST11CC

Unit-V

Sequential Procedures

Dr. T. Jai Sankar

Associate Professor and Head

Department of Statistics

Ms. I. Angel Agnes Mary

Guest Faculty

Department of Statistics

UNIT - V

Sequential Procedures

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

Stopping Times

An integer valued random variable N is said to be a stopping time for the sequence of $\{X_i\}$ if the event $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, X_{n+3}, \ldots\ldots (n = 1, 2, 3, \ldots.)$

Stopping time is a specific type of "random time": a random variable whose value is interpreted as the time at which a given stochastic process exhibits a certain behavior of interest. A stopping time is often defined by a stopping rule, a mechanism for deciding whether to continue or stop a process on the basis of the present position and past events, and which will almost always lead to a decision to stop at some finite time.

Example

Consider a coin toss experiment. Let the outcome of the ith toss s denoted by $X_i = 1 \text{ or } 0$.

$$P(X_i = 1) = p = 1 - P(X_i = 0)$$

Then $E(X_i) = p$. The sum of $S_n = x_1, x_2, \ldots, x_n$ is the cumulative number of heads n the first tosses.

Suppose m is positive integer the $N = min\{n; S_n = m)$ s a stopping time. Consider $N(t)$ is the number of renewals by time t with respect to sequence of inter arrival times $\{X_i\}$.

Now $N(t) = n$ whenever $S_n \le t$ and $S_{n+1} > t$ that is the event $N(t) = n$ depends not only on $x_1, x_2, \ldots, x_n$.

Wald's Equation

Statement

If $X_1, X_2, X_3, \ldots$ be independent identically distributed with finite mean $E(X)$, and N is a stopping time with $E[N] < \infty$, then $E[X_1 + \cdots + X_n] = E[X_1]\, E[N]$.

Proof

$$\text{Let } I_n = \begin{cases} 1, & if,\ N \ge n \\ 0, & Otherwise \end{cases}$$

$$E\left(\sum_{n=1}^{N} X_n\right) = E\left(\sum_{i=1}^{\infty} X_n \,.\, I_n\right)$$

$$= \sum_{n=1}^{\infty} E(X_n . I_n)$$

X_n and I_n are independent.

$$E\left(\sum_{n=1}^{N} X_n\right) = \sum_{n=1}^{\infty} E(X_n) \ E(I_n)$$

Where,

$$\sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} \sum_{k=n}^{\infty} P(N = n)$$

$$= \sum_{n=1}^{k} \sum_{k=1}^{\infty} P(N = n)$$

$$= \sum_{k=1}^{\infty} k . P(N = n)$$

$$= E(N)$$

Therefore,

$$E\left(\sum_{n=1}^{N} X_n\right) = E(X_1) E(N)$$

Hence prove.

Wald's Fundamental Identity

Statement

If $X_1, X_2, X_3, \ldots$ be independent identically distributed with common finite mean with $E[X_n] = E[X_1] < \infty$, and T to be a stopping rule which is independent of $X_{T+1}, X_{T+2}, \ldots.$ for which $E(T) < \infty$ then

$$E[X_1 + \cdots + X_t] = E[X_1]\, E[T].$$

Proof

We begin by inductively defining a sequence of stopping times $T_1, T_2, \ldots$ from our initial sequence $X_1, X_2, \ldots$ of random variables.

Let $T_1 = T$

$P(T_1 = n | X_{T+1}, X_{T+2}, \ldots) = P(T_1 = n)$

$P(T_2 = n | X_{T1+T2+1}, X_{T1+T2+2}, \dots) = P(T_2 = n)$

We obtain the desired sequence $T_1, T_2, \dots$ We can then easily verify that

$$P(T_{n+1} = \alpha_{n+1} \mid T_1 = \alpha_1, \dots, T_n = \alpha_n) = \frac{P(T_{n+1} = \alpha_{n+1}, T_1 = \alpha_1, \dots, T_n = \alpha_n)}{P(T_1 = \alpha_1, \dots, T_n = \alpha_n)}$$

$$= \frac{P(T_{n+1} = \alpha_{n+1})\, P(T_1 = \alpha_1, \dots, T_n = \alpha_n)}{P(T_1 = \alpha_1, \dots, T_n = \alpha_n)}$$

$$= P(T_{n+1} = \alpha_{n+1})$$

Therefore $T_1, T_2, \dots$ is an i.i.d. sequence of random variables. Having established independence of the sequence of stopping times, we now define another sequence of random variables. Put

$$S_1 = X_1 + \dots\dots + X_{T_1}$$

$$S_2 = X_{T_1+1} + \dots\dots + X_{T_1+T_2}$$

$$S_3 = X_{T_1+T_2+1} + \dots\dots + X_{T_1+T_2+T_3}$$

.
.
.

By reasoning similar to the sequence of T's, we see that $S_1, S_2, \dots$ is an i.i.d. sequence of random variables. We would like to be able to apply the Strong Law of Large Numbers to this sequence, but we do not yet know that $E|S_n| = E|S_1| < \infty$, we merely know that they have common mean. If we were to show that we could apply the Converse to $S_1, S_2, \dots$, we would then be able to apply the Strong Law. To do this, we consider the following:

$$\lim_{k\to\infty} \frac{S_1 + \cdots + S_k}{k} = \lim_{k\to\infty} \frac{S_1 + \cdots + S_k}{N_1 + \cdots + N_k} \cdot \lim_{k\to\infty} \frac{N_1 + \cdots + N_k}{k} \qquad (2)$$

Where we have multiplied by 1 within the limit and separated the product. By definition of $S_1, \dots, S_k$, we observe that the leftmost term in (2) is equal to

$$\lim_{k\to\infty} \frac{X_1 + \cdots + X_{\tau_1 + \cdots + \tau_k}}{N_1 + \cdots + N_k},$$

which is merely a subsequence of the sequence of averages of the X's. By the Strong Law, we know that

$$\mathbb{P}\left(\lim_{k\to\infty} \frac{X_1 + \cdots + X_k}{k} = \mathbb{E}(X_1)\right) = 1,$$

and so the subsequence above must also converge to a common limit with probability 1. Likewise, applying the Strong Law to the rightmost term in (2), we find that

$$\mathbb{P}\left(\lim_{k\to\infty}\frac{N_1+\cdots+N_k}{k}=\mathbb{E}(\tau)\right)=1.$$

Together, these facts imply that

$$\mathbb{P}\left(\lim_{k\to\infty}\frac{S_1+\cdots+S_k}{k}=\mathbb{E}(\tau)\mathbb{E}(X_1)\right)=1, \tag{3}$$

Hence the limit exists with probability 1. By the Converse to SLLN, we find that $E|S_n| = E|S_1| < \infty$ and so we can apply the Strong Law to our sequence of sums $S_1, S_2, \ldots,$ obtaining

$$\mathbb{P}\left(\lim_{k\to\infty}\frac{S_1+\cdots+S_k}{k}=\mathbb{E}(S_1)\right)=1. \tag{4}$$

Putting (3) and (4) together, we are left to conclude that

$$\mathbb{E}(X_1+\cdots+X_\tau)=\mathbb{E}(\tau)\mathbb{E}(X_1)$$

Hence proved.

Sequential Probability Ratio Test (SPRT)

According to the Neyman-Pearson fundamental lemma, the best procedure for testing the simple hypothesis H that the probability density of X is p_o against the simple alternative that it is p, accepts or rejects H as

$$\frac{p_{1n}}{p_{0n}}=\frac{p_1(x_1)\ldots\ldots\ldots p_1(x_n)}{p_0(x_1)\ldots\ldots\ldots p_0(x_n)}$$

is less or greater than a suitable constant C. However, further improvement is possible if the sample size is not fixed in advance but is permitted to depend on the observations. The best procedure, in a certain sense, is then the following sequential probability ratio test.

Termination Property

Suppose that to test the hypothesis, H_0: $\theta = \theta_0$ against the alternative hypothesis, H_1: $\theta = \theta_0$, for a distribution with probability density function $f(x, \theta)$. For any positive integer m, the likelihood function of a sample $x_1, x_2, \ldots, x_m$ from the population with probability density function $f(x, \theta)$ is given by

$L_{1m}=\prod_{i=1}^{m} f(x_i,\theta_1)$ When H_1 is true and

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

and the likelihood ratio λ_m is given by

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

Where, m = 1, 2, ...

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

At each stage of the experiment (at the m^{th} trial for any integral value m), the likelihood ratio λ_m (m = 1,2, . . .) is computed.

- If $\lambda_m \geq A$, we terminate the process with the rejection of H_0.
- If $\lambda_m \leq B$, we terminate the process with the acceptance of H_0.
- If $B < \lambda_m < A$, we continue sampling by taking an additional observation.

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

$$A = \frac{1-\beta}{\alpha}, B = \frac{\beta}{1-\alpha} \quad ------\rightarrow (2)$$

Where α and 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 λ_m rather than λ_m, since

$$log\lambda_m = \sum_{i=1}^{m} log\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} = \sum_{i=1}^{m} z_i \quad ------\rightarrow (3)$$

Where,

$$z_i = log\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} \quad ------\rightarrow (4)$$

In terms of z_i's, SPRT is termed as follows:

- If $\sum z_i \geq log\ A$ reject H_0
- If $\sum z_i \leq log\ B$ reject H_1
- If $\log\ B < \sum z_i < log\ A$ continue sampling by taking an additional observation.

Note:

- In SPRT, we continue taking additional observations unless the inequality $B < \lambda_m < A \Rightarrow \log B < \sum z_i < log\ A$ is violated at either end. It has been proved that SPRT eventually terminates with probability one.
- Sequential schemes provide for a minimum amount of sampling and thus result is considerable 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 values of α and β.

Approximation to Stopping Bounds

As in classical hypothesis testing, SPRT starts with a pair of hypotheses, say H_0 and H_1 for the null hypothesis and alternative hypothesis respectively. They must be specified as follows:

$H_0: p = p_0$

$H_1: p = p_1$

The next step is to calculate the cumulative sum of the log-likelihood ratio, $\log \lambda_i$, as new data arrive: with $S_0 = 0$, then, for $i = 1, 2, \ldots$

The stopping rule is a simple thresholding scheme:

- $a < S_i < b$: continue monitoring (critical inequality)
- $S_i \geq b$: Accept H_1
- $S_i \leq a$: Accept H_0

Where, a and b $(a < 0 < b < \infty)$ depend on the desired type I and type II errors, α *and* β. They may be chosen as follows:

$$a = \log \frac{1-\beta}{\alpha} \text{ and } B = \log \frac{\beta}{1-\alpha}$$

In other words, α *and* β must be decided beforehand in order to set the thresholds appropriately. The numerical value will depend on the application. The reason for being only an approximation is that, in the discrete case, the signal may cross the threshold between samples. Thus, depending on the penalty of making an error and the sampling frequency, one might set the thresholds more aggressively. The exact bounds are correct in the continuous case.

Applications to Standards Distributions

The application of Parameter estimation of a probability distribution function. Consider the exponential distribution:

$$f_\theta(x) = \theta^{-1} e^{-\frac{x}{\theta}}$$

The hypotheses are

$H_0: \theta = \theta_0$

$H_1: \theta = \theta_1 \qquad (\theta_1 > \theta_0)$

Then the log-likelihood function (LLF) for one sample is

$$\log\lambda(x) = log\left(\frac{\theta_1{}^{-1}e^{-\frac{x}{\theta_1}}}{\theta_0{}^{-1}e^{-\frac{x}{\theta_0}}}\right)$$

$$= log\left(\frac{\theta_1}{\theta_0}e^{\frac{x}{\theta_0}-\frac{x}{\theta_1}}\right)$$

$$= log\left(\frac{\theta_1}{\theta_0}\right) + \log\left(e^{\frac{x}{\theta_0}-\frac{x}{\theta_1}}\right)$$

$$= -log\left(\frac{\theta_1}{\theta_0}\right) + \left(\frac{x}{\theta_0} - \frac{x}{\theta_1}\right)$$

$$= -log\left(\frac{\theta_1}{\theta_0}\right) + \left(\frac{\theta_1 - \theta_0}{\theta_0\theta_1}\right)$$

The cumulative sum of the LLFs for all *x* is

$$S_n = \sum_{i=1}^{n} log\lambda(x_i) = -nlog\left(\frac{\theta_1}{\theta_0}\right) + \left(\frac{\theta_1 - \theta_0}{\theta_0\theta_1}\right)\sum_{i=1}^{n} x_i$$

Accordingly, the stopping rule is:

$$a < -nlog\left(\frac{\theta_1}{\theta_0}\right) + \left(\frac{\theta_1 - \theta_0}{\theta_0\theta_1}\right)\sum_{i=1}^{n} x_i < b$$

After re-arranging we finally find

$$a + nlog\left(\frac{\theta_1}{\theta_0}\right) < \left(\frac{\theta_1 - \theta_0}{\theta_0\theta_1}\right)\sum_{i=1}^{n} x_i < b + nlog\left(\frac{\theta_1}{\theta_0}\right)$$

The thresholds are simply two parallel lines with slope $log\left(\frac{\theta_1}{\theta_0}\right)log\left(\frac{\theta_1}{\theta_0}\right)$. Sampling should stop when the sum of the samples makes an excursion outside the *continue-sampling region*.

Operating Characteristic (OC)

The O.C. function L(θ) is defined as

L(θ) = Probability of accepting H_0: $\theta = \theta_0$ when θ is the true value of the parameter, and since the power function

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

L(θ) = 1- P(θ) ---------> (5)

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 population with density function *f(x, θ)* is given by

$$L(\theta) = \frac{A^{h(\theta)} - 1}{A^{h(\theta)} - B^{h(\theta)}} \qquad ------\rightarrow (6)$$

Where, for each value of θ, the value of h(θ) ≠ 0, is to be determined so that

$$E\left[\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}\right]^{h(\theta)} = 1 \qquad ------\rightarrow (7)$$

Where, the constants A and B have already been defined in equation (2). It has been proved that under very simple conditions on the nature of the function *f(x, θ)*, there exists a unique value of h(θ) ≠ 0, such that equation (7) is satisfied.

Average Sample Number (ASN) Functions

The sample size n in sequential testing is a random variable which can be determined in terms of the true density function *f(x, θ)*. The Average Sample Number function for the Sequential Probability Ratio Test for testing H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ is given by

$$E(n) = \frac{L(\theta)\log B + [1 - L(\theta)]\log A}{E(z)} \qquad ------\rightarrow (8)$$

Where,

$$z = \log\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)}, A = \frac{1-\beta}{\alpha} \text{ and } B = \frac{\beta}{1-\alpha} \qquad ------\rightarrow (9)$$

Example:

Give the Sequential Probability Ratio Test for testing H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ ($> \theta_0$), in sampling from a normal density.

$$f(x, \theta) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right], -\infty < x < \infty$$

Where, σ is known. Also obtain its Operating Characteristic function and Average Sample Number function.

Solution

The likelihood ratio *f(x, θ)* is

$$\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} = exp\left[-\frac{1}{2\sigma^2}((x_i - \theta_1)^2 - (x_i - \theta_0)^2)\right]$$

$$= exp\left[-\frac{1}{2\sigma^2}\{(\theta_0 - \theta_1)(2x_i - \theta_0 - \theta_1)\}\right] \quad ----\rightarrow (*)$$

$$\therefore \quad z_i = \log\frac{f(x_i, \theta_1)}{f(x_i, \theta_0)} = \frac{\theta_1 - \theta_0}{\sigma^2}\left(x_i - \frac{\theta_0 - \theta_1}{2}\right) \quad ---------\rightarrow (**)$$

$$\Rightarrow \qquad \log\lambda_m = \sum_{i=1}^{m} z_i = \frac{\theta_1 - \theta_0}{\sigma^2}\left(\sum_{i=1}^{m} x_i - \frac{m(\theta_0 - \theta_1)}{2}\right)$$

Hence the S.P.R.T. for H_0: $\theta = \theta_0$ against H_1: $\theta = \theta_1$ is given by

(i) Reject H_0 if

$$\frac{\theta_1 - \theta_0}{\sigma^2}\left(\sum_{i=1}^{m} x_i - \frac{m(\theta_0 - \theta_1)}{2}\right) \geq \log\left(\frac{1-\beta}{\alpha}\right)$$

$$\sum_{i=1}^{m} x_i \geq \frac{\sigma^2}{\theta_1 - \theta_0}\log\left(\frac{1-\beta}{\alpha}\right) + \frac{m(\theta_0 - \theta_1)}{2}\ ;(\theta_1 > \theta_0)$$

(ii) Accept H_0 if

$$\frac{\theta_1 - \theta_0}{\sigma^2}\left(\sum_{i=1}^{m} x_i - \frac{m(\theta_0 - \theta_1)}{2}\right) \leq \log\left(\frac{\beta}{1-\alpha}\right)$$

$$\sum_{i=1}^{m} x_i \leq \frac{\sigma^2}{\theta_1 - \theta_0}\log\left(\frac{\beta}{1-\alpha}\right) + \frac{m(\theta_0 - \theta_1)}{2}\ ;(\theta_1 > \theta_0)$$

(iii) Continue taking additional observations as long as

$$\log\left(\frac{\beta}{1-\alpha}\right) < \frac{\theta_1 - \theta_0}{\sigma^2}\left(\sum_{i=1}^{m} x_i - \frac{m(\theta_0 - \theta_1)}{2}\right) < \log\left(\frac{1-\beta}{\alpha}\right)$$

$$\frac{\sigma^2}{\theta_1 - \theta_0}\log\left(\frac{\beta}{1-\alpha}\right) + \frac{m(\theta_0 - \theta_1)}{2} < \sum_{i=1}^{m} x_i < \frac{\sigma^2}{\theta_1 - \theta_0}\log\left(\frac{1-\beta}{\alpha}\right) + \frac{m(\theta_0 - \theta_1)}{2} ;(\theta_1 > \theta_0)$$

Operating Characteristic function

To determine $h = h(\theta) \neq 0$ from equation (7) i.e., from

$$\int_{-\infty}^{\infty}\left[\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)}\right]^h f(x,\theta)\,dx = 1$$

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2\right].\left[exp\left[-\frac{1}{2\sigma^2}\{(\theta_0 - \theta_1)(2x_i - \theta_0 - \theta_1)\}\right]\right]^h dx = 1$$

On using (*)

$$\Rightarrow \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2\sigma^2}\left\{x^2 - 2x((\theta_1 - \theta_0)h + \theta) + \theta^2 + \left(\theta_1{}^2 - \theta_0{}^2\right)h\right\}\right] dx = 1$$

If we take

$$\lambda = (\theta_1 - \theta_0)h + \theta$$

$$\lambda^2 = \left(\theta_1{}^2 - \theta_0{}^2\right)h + \theta^2$$

Then L.H.S becomes

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} exp\left[-\frac{1}{2}\left(\frac{x-\lambda}{\sigma}\right)^2\right]dx$$

Which being the total area under normal probability curve with mean μ and variance σ^2 is always unity, as desired. Thus h = h(θ) is given by

$$({\theta_1}^2 - {\theta_0}^2)h + \theta^2 = [(\theta_1 - \theta_0)h + \theta]^2$$

$$\Rightarrow \left({\theta_1}^2 - {\theta_0}^2\right)h = (\theta_1 - \theta_0)^2 h^2 + 2\theta(\theta_1 - \theta_0)h$$

Since, h = h(θ) ≠ 0 and $\theta_1 \neq \theta_0$ on dividing throughout by $(\theta_1 - \theta_0)$ h.

We get

$$(\theta_1 + \theta_0) = (\theta_1 - \theta_0)h + 2\theta$$

$$\Rightarrow h(\theta) = \frac{\theta_1 - \theta_0 - 2\theta}{\theta_1 - \theta_0}$$

Substituting for h(θ) in equation (6) we get the required O.C. function,

$$\Rightarrow L(\theta) = \frac{A^{\frac{\theta_1-\theta_0-2\theta}{\theta_1-\theta_0}} - 1}{A^{\frac{\theta_1-\theta_0-2\theta}{\theta_1-\theta_0}} - B^{\frac{\theta_1-\theta_0-2\theta}{\theta_1-\theta_0}}}$$

$$\Rightarrow L(\theta) = \frac{A^{\frac{\theta_1-\theta_0-2\theta}{\theta_1-\theta_0}} - 1}{(A-B)^{\frac{\theta_1-\theta_0-2\theta}{\theta_1-\theta_0}}}$$

Average Sample Number (ASN) Functions

$$z = log\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)} = \frac{\theta_1 - \theta_0}{\sigma^2}\left(x - \frac{\theta_0 - \theta_1}{2}\right)$$

$$E(z) = \frac{\theta_1 - \theta_0}{2\sigma^2}[2E(z) - \theta_0 - \theta_1]$$

$$= \frac{\theta_1 - \theta_0}{2\sigma^2}[2\theta - \theta_0 - \theta_1]$$

Substituting for *E(z)* in equation (8), we get the required A.S.N. function,

$$E(n) = \frac{L(\theta)\log B + [1 - L(\theta)]\log A}{\frac{\theta_1 - \theta_0}{2\sigma^2}[2\theta - \theta_0 - \theta_1]}$$