LIKELIHOOD RATIO TEST AS AN ALTERNATIVE TO CHI-SQUARE TEST (LARGE SAMPLE APPROXIMATION)

[*1]Amalare A.A, [2] Nurudeen T.S, [3]Adeniyi M.O

[*1,2,3] Department of Mathematics, Lagos State Polytechnic, Ikorodu, Nigeria
+2348023998403 (e-mail: amalareasm@yahoo.com)

ABSTRACT

In this paper , different methods of arriving at a reliable and authentic decision in hypothesis testing are discussed. The choice of the critical region, level of significance, testing equality of variances, maximum likelihood estimate and multinomial distribution are highlighted. The test Statistic $T(\underline{X})$ of likelihood ratio test and its asymptotic approximation, $-\ 2\log T(\underline{X})$, are also carefully studied. Our study revealed that the proposed chi-square method $2\sum 0_i \log\left(\frac{0_i}{E_I}\right)$ as against the conventional chi-square $\sum \frac{(0i-E_i)^2}{E_i}$ gives a better approximation of Likelihood ratio test ($-2\log T(\underline{X})$) .

Keywords: Hypothesis testing, maximum likelihood estimate, likelihood ratio test, equality of variances, multinomial distribution, restricted MLE

1.0 INTRODUCTION

The method of maximum Likelihood was first introduced by Fisher [1]. Giving a collection of r observations each of which is normal with population mean μ and population variance σ^2. We want to estimate μ and σ^2, where both are unknown. The problem has been addressed in literature by several researchers [2-5]. In this paper a very simple example of inconsistency of maximum likelihood method is presented that shows clearly one danger to be wary of in an otherwise regular looking situation. The discussion of this paper is centred on a sequence of independent, identically distributed and for the sake of convenience, real random variables, $X_1, X_2, ..., X_n$, distributed according to a distribution $F(X|\theta)$ for some θ in a fixed parameter space Θ.

Various problems that will lead to a reliable and authentic decision in hypothesis testing shall be discussed and the choice of critical region testing, equality of variances, maximum likelihood estimate and multinomial distribution shall also be considered.

2.0 DEFINITION OF TERMS

Maximum Likelihood Estimate (MLE): A maximum likelihood Estimate $\hat{\theta}$ of parameter θ in the frequency function $f(X_i;\ \theta)$ is an estimate that maximizes the likelihood function $L_n(\theta) = \prod_{i=1}^{n} f(X_i;\ \theta)$ that is, the parameter values that agree most closely with the observed data.

Restricted / Unrestricted M.L.E**:**
Let $\underline{x} = (x_1,\ldots,x_n)$ be a sample drawn from a particular distribution whose probability density function is a function of parameter θ. We say $\hat{\theta}$ is restricted MLE of θ if and only if $\hat{\theta}$ subjects to some particular conditions (restriction), otherwise it is an unrestricted MLE of θ.

PROPERTIES OF MLE (ASYMPTOTIC)

- Under regularity conditions, maximum likelihood estimates are consistent.
- Under regularity condition, MLE is asymptotically normal with $\hat{\theta}$ (asymptotic unbiasedness) and variance equal to the crammer – Rao Lower bound [6].
- θ_n^x, the MLE of θ based on sample of size n is weakly consistent that is, converges in probability to θ as n tends to infinity provided that $f(x;\theta)$ is such that

$$E\left[\frac{1}{n}\sum_{i=1}^{n} Log\ f(X_i,\ \theta)\right] \text{exists}$$

2.1 LIKELIHOOD RATIO TESTS

The method of maximum likelihood estimate discussed earlier in 1.0. is a constructive method of obtaining estimators which have desirable properties under certain conditions earlier stated. Likelihood ratio test is a testing procedure closely allied to MLE. This method was proposed by Neyman and Scott [2] . In the case of simple null hypothesis H_0 and alternative hypothesis H_1, a likelihood ratio test is often most powerful.

The procedure of the likelihood ratio test (LRT) applies to the testing of hypothesis.

Suppose that the family $[f(x;\theta),\ \theta \in W]$ of probability density function of a random variable X is defined. If we intend to test the hypothesis H_0: $\theta \in$ W against the alternative H_I: $\theta \in \int - W$.

Where W and $\int - W$ are the distribution under H_0 and H_1 respectively.

The procedure is formulated as follows:

$$T(\underline{X}) = \frac{\underset{\theta \in W}{Sup}\ f(x;\theta)}{\underset{\theta \in \int - W}{SUP} f(x;\theta)} \qquad (1)$$

The distribution of $T(\underline{X})$ under the null hypothesis must be determined using the above test procedure.

Example 1

Let $\underline{X} = [x_1,\ \ldots,\ x_n]$ be a random sample from normal distribution $N(\mu,\ \sigma^2)$, μ *and* σ^2 are both unknown. To Test the hypothesis; H_0: $\mu = \mu_0$ against

H_1: $\mu \neq \mu_0$

Considering the Normal distribution, Let

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

$$= \frac{1}{\sigma^n}(2\pi)^{\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum(x_i - x)^2 + n(\bar{x} - \mu)^2\right] \qquad (2)$$

Under H_0;

$$f\left(x, \mu_0, \sigma^2\right) = \frac{1}{\sigma^n}\left(2\pi\right)^{n/2} \exp\left[-\frac{1}{2\sigma^2}\sum\left(x_i - x\right)^2 + n\left(\bar{x} - \mu_0\right)^2\right] \qquad (3)$$

Now taking the logarithm of the likelihood function in (3) and maximizing with respect to σ, the maximum likelihood estimate $\hat{\sigma}^2$ of σ^2 will be

$$\hat{\sigma}^2 = 1/n\left[\sum\left(xi - \bar{x}\right)^2 + n\left(\bar{x} - \mu_0\right)^2\right] \qquad (4)$$

Similarly,
Under H_1

$$\hat{\sigma}.^2 = 1/n\sum\left(x_i - \bar{x}\right)^2 \qquad (5)$$

By substituting $\hat{\mu} = \bar{x}$ where $\hat{\mu}$ represent the maximum likelihood estimate of μ. and putting (4) and (5) into (1), we get

$$T\left(\underline{X}\right) = \left[\frac{\hat{\sigma}.^2}{\hat{\sigma}^2}\right]^{n/2} \qquad (6)$$

Putting (4) and (5) into (6), yields

$$T\left(\underline{X}\right) = \left(\frac{\sum\left(x_i - \bar{x}\right)^2}{\sum\left(x_i - \overline{x^2}\right) + n\left(\bar{x} - \mu_0\right)^2}\right)^{n/2} \qquad (7)$$

Which is equivalent to

$$T\left(\underline{X}\right) = \frac{1}{\left[1 + t^2 / n - 1\right]^{n/2}}$$

where

$$t = \sqrt{n}\left(\bar{x} - \mu_0\right) / \left(\left(\sqrt{\sum\left(x_i - \bar{x}\right)^2 / n - 1}\right)\right)$$

Thus we accept H_0 if $T\left(\underline{X}\right) \leq k$

Since $T\left(\underline{X}\right)$ is a monotonic decreasing function of t^2, H_0 is rejected if $t^2 \geq k^1$ or equivalently if $|t| \geq k$

Example 2

Let x_1, ……………, x_n be a random sample from normal distribution N(θ, 1). We wish to test the hypothesis

H_0: $\theta = \theta_0$ against
H_1: $\theta \neq \theta_0$

Considering the normal distribution with mean θ and variance of 1, the density function is

$$f(x\,;(\theta)) = \exp\left(-\tfrac{1}{2}(x-\theta)^2 / \sqrt{2\pi}\right)$$

$$L_0 f(x\,;(\theta)) = \exp\left(-\tfrac{1}{2}\sum(x_i-\theta)^2 \;/(2\pi)^{n/2}\right)$$

$$d\log f(x,\theta)/d\theta = \sum_{i=1}^{n}(x_i-\theta) = 0$$

$$\Rightarrow \qquad \theta = \sum_{i=1}^{n}\frac{X_i}{n} = \overline{X} \tag{8}$$

Under H_0

$$\underset{\theta \varepsilon W}{\text{Sup}}\; f(x;\theta) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2}\sum_{i=1}^{n}(x_i-\theta_0)^2 = 0\right]$$

Under H_1,

$$\underset{\theta \in \int - W}{\text{Sup}}\; f(x;\theta) = (2\pi)^{-n/2} \exp\left[\left(-\frac{1}{2}\sum_{i=1}^{n}x_i - \overline{x}\right)^2\right]$$

Upon simplifying this it becomes

$$T(\underline{X}) \quad = \quad \exp\left[-\tfrac{n}{2}\left(\overline{x}-\theta_0\right)^2\right] \tag{9}$$

Thus 5% critical region for the likelihood ratio test is equivalent to the two equal tails of the $\overline{X}$ distribution given by the familiar inequality $\left|\overline{X}-\theta_0\right|\sqrt{n} > 1.96$

Example 3 Testing Equality of Variances

Let x_1, x_2, .. x_k be k independently normally distributed variables with μ_1, $\mu_2 \ldots \mu_k$ means and variances σ_1^2, σ_2^2, …, σ_k^k. Let random sample of sizes n_1, …, n_k be drawn from those populations and let the hypothesis to be tested be

H_0: $\sigma_1^2 = \sigma_2^2 = .. \sigma_k^2$

H_1: $\sigma_1^2 \neq \sigma_2^2 \neq \ldots\ldots \neq \sigma_k^k$

The random variable corresponding to the j^{th} observation for the variable x_i is represented by x_{ij}. Thus, there are $\sum_i^k n_i = n$ random variables.

Considering the Normal distribution function with parameters μ and σ^2 as follows

$$f(x_{ij},\, \mu_i,\, \sigma_i) = \frac{\exp\left[-\tfrac{1}{2}\left[(x_{ij}-\mu_i)/\sigma_i\right]^2\right]}{\sigma_i\sqrt{2\Pi}}$$

The likelihood function is

$$L\;(x,\mu,\sigma) = \frac{\exp\left[-\frac{1}{2}\left[\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij}-\mu_i)/\sigma_i\right]^2\right]}{(2\pi)^{n/2}\sigma_1^{\,n1}\ldots\sigma_k^{\,nk}}$$

$$L_0\, f\left(x,\mu,\sigma^2\right) = \frac{\exp\left[-\frac{1}{2}\left[\sum^{k}\sum^{ni}\left(x_{ij}-\overline{x_i}\right)/\sigma\right]^2\right]}{(2\pi)^{n/2}\ \sigma^n} \tag{10}$$

Where σ^n denotes the common value of all σ_i^2 when the null hypothesis is considered.

Similarly,
Under H_1

$$L_n f\left(x,\mu,\sigma_2\right) = \frac{\exp\left[-\frac{1}{2}\left(\sum^{k}\sum^{ni}\left(x_{ij}-\mu_i/\sigma_i\right)\right)^2\right]}{(2\pi)^{n/2}\ \pi_{i=1}^{k}\sigma_i^{i}} \tag{11}$$

Maximizing (11) with respect to the parameters μ_i and σ_i, the resulting expression will give

$$\hat{\mu}_i = \sum_r^{ni}\frac{x_{ij}}{ni} \tag{12}$$

and

$$\hat{\sigma}^2 = \sum_{i=1}^{k}\sum_{j=1}^{ni}\left(x_{ij}-\overline{x_i}\right)^2/n, \tag{13}$$

respectively. Where $\sum^{k} n_i = n$.

However, $\hat{\sigma}_i^2 = s_i^2$

therefore,

$$\hat{\sigma}^2 = \sum_{j=1}^{k}\sum_{j=1}^{ni}\left(x_{ij}-\overline{x_i}\right)^2/n$$

$$\hat{\sigma}^2 = \sum_{j=1}^{k} n_i\, s_i^2/n$$

$$\frac{Sup\ L_0}{Sup\ L_i} = \frac{\exp\left(-n/2\right)\sum\left(x_{ij}-\overline{x_i}\right)^2/\ (2\pi)^{n/2}\ \left[\left(n_1 s_1^2+..+n_k s_k^2\right)/n\right]^{n/2}}{\exp\left(-n/2\right)\sum\sum\left(x_{ij}-\overline{x_i}\right)^2/(2\pi)^{n/2}\ \left[s_1^{n1}\ldots\ldots\ldots s_k^{nk}\right]^{n/2}}$$

Where L_0 and L_1 represent the likelihood functions for null and alternative hypotheses respectively . Which reduces to

$$T\left(\underline{X}\right) = \frac{\pi_{i=1}^{k}\left(s_i^2\right)^{\frac{ni}{2}}}{\left[\sum n_i s_i^{\,2}/\sum^{k} n_i\right]^{\sum\frac{ni}{2}}} \tag{14}$$

3.0 Large – Sample Approximation

3.1 Application of Large-sample approximations

We shall state the Cramer theorem without proof.

Theorem: Cramer [7]

Subject to regularity conditions necessary for asymptotic results for maximum likehood estimates, the likelihood ratio statistic $T(\underline{x})$ for testing

H_0: W: $\theta = \theta_{r0}$ against the alternative

H_1: $\int - w: \theta_r \neq \theta_{r0}$ (where r stands for restriction imposed on

Parameter θ is such that as the sample size n becomes very large (i.e. $n \to \infty$),
$-2\log T(\underline{X})$ is asymptotically Chi – square with r degrees of freedom for all $\theta \varepsilon W$. We now apply the theorem as follows:

By taking the logarithm of both sides of (7), the expression becomes

$$-2\log T \; T(\underline{X}) = n log[1 + n(\overline{x} - \mu_0)^2 / \sum(x_i - \overline{x})^2)] \quad (15)$$

The test statistics approximate to χ^2 table of critical values. The critical value say C_0 depends on the degree of freedom r and level of significance α chosen (r stands for the number of restrictions imposed on the parameter.

If the calculated value is greater than the value C_0; we reject the hypothesis H_0.
Recall (9) from example 2 and taking logarithm of both sides, we obtained

$$-2\log T(\underline{X}) = n(\overline{x} - \theta_0 \;)^2 \quad (16)$$

For the sample size n, $\overline{x}$ being the mean of the observations and given the value of θ_0; we reject H_0 if table value say C_0 degree of freedom r and level of significance α is smaller than the calculated value of $-2\log T(\underline{X})$

Recall from example 3.That

$T(X)$ in (14) can be reduced to $\pi_{i=1}^{k}\left(s^2\sqrt{s^2}\right)^{n\frac{1}{2}}$ and taking the logarithm of both sides, we have,

$$-2\log T(\underline{x}) = \sum n_i \log s_i^2 - \sum n_i \log s^2 \quad (17)$$

The value on the R.H.S. is equally computable since we know that $\sum^{k} n_i = n$

$s_i^2 = \sum (x_{ij} - x_i)^2 / n_i$

And $s^2 = \sum n_i s_i^2 / n$

Then, the value thus calculated is compared with the critical value from the χ^2 table with known degrees of freedom r and level of significance $\propto$. We reject H_0 if the calculated value is greater than the table critical value.

3.2 **Approximation of – 2Log T($\underline{X}$) to Chi – Square Statistics:**

3.2.1 Multinomial Case:

Considering the multinomial distribution

$$f(x,\theta) = \frac{n!}{n_1!\ldots\ldots n_s!} \quad \theta_1^{ni}\ldots\ldots\theta_s^{ns} \qquad \textbf{(18)}$$

where $\sum^{s} n_i = n$

$$\sum_{i=1}^{s}\theta_i = 1$$

The likelihood function of (18) gives

$$\text{L } (\theta) = f(x,\theta) = \left(\frac{n!}{\prod_{i=1}^{5} n_i!}\right)\prod_{i=1}^{s}\left(\theta_i^{ni}\right)$$

By taking the logarithm of both of sides, the expression becomes

$$\text{Log } (\theta) = \log n! - \sum_{i=1}^{s}\log n_i! + \sum_{i=1}^{s-1} n_i \log\theta_i + n_s \log\left(1-\sum\theta_i\right) \qquad (19)$$

Then generalizing and maximizing (19) with respect to θ_i we have

$\hat{\theta}_i = \frac{n_i}{n}$ for any i such that 1≤ i ≤ s. Thus $\hat{\theta}_i$ is MLE of θ for a multinomial distribution

Again,

$$\text{T}(\underline{X}) = \frac{\underset{\theta\,\varepsilon W}{Sup}\left(f(x,\theta)\right)}{\underset{\theta\,\varepsilon \int -W}{Sup}\, f(x,\theta)} = \frac{\frac{n!}{\Pi n!}\Pi\left(P_i^{ni}\right)}{\frac{n!}{\Pi n!}\Pi\left(\frac{n_i}{n}\right)^{ni}}$$

$$\Rightarrow T\ (\underline{X}) = \frac{\left[{p_i}/{n_1}\right]^{n1}\ldots\ldots\ldots\left[{P_s}/{n_s}\right]^{ns}}{\left[{1}/{n}\right]^{n1}\ldots\ldots\left[{1}/{n}\right]^{ns}} \qquad (20)$$

By taking the logarithm of (20) and using the theorem by Cramer (1946), we have

$$-2\log T(\underline{X}) = 2\sum_{i=1}^{s} 0_i \log {0_i}/{E_i} \qquad (21)$$

But $\log\left({0_i}/{E_i}\right) = \log\left(1+\frac{0_i - E_i}{E}\right)$ (22)

Thus, expanding (22) , using maclaurin series gives

$$\log\left(1+\frac{0_i - E_i}{E_i}\right) = \frac{0_i - E_i}{E_i} - \frac{1}{2}\frac{(0_i - E_i)^2}{E_i^2} + \frac{1}{3}\frac{(0_i - E_i)^3}{E_i^{\ 3}} - \frac{1}{4}\frac{(0_i - E_i)^4}{E_i^4} + \ldots \qquad (23)$$

By substituting (23) into (21), the resulting expression yields

$$-2\log\ T(X) = \sum\frac{(0_i - E_i)^2}{E_i} \approx \chi^2 \qquad (24)$$

We therefore concluded that

$-2\log\ T(\underline{x}) \approx \sum\frac{(0_i - E_i)^2}{E_i}$ Subject to some conditions as stated in (25) , where O_i is the observed frequency and E_i is the expected frequency

3.3 **Conditions for Approximation**

According to Perlman [8] and Rice [9], the series in (23) converges if and only if

$$\left|\frac{0_i - E_i}{E_i}\right| < 1 \qquad (25)$$

And $O_i > 0$ or $O_i < 2E_i$.

It follows that

$- 2 \log T(X) = 2\sum 0_i \log\ {}^{0_i}/_{E_i}$, should be used instead of

$$\chi^2 = \sum\frac{(0_i - E_i)^2}{E_i}$$

Since

$-2\sum 0_i \log\ {}^{0_i}/_{E_i}$ is computationally convenient and valid for all values of 0_i and E_i.

This statistic should replace the conventional Neyman – Pearson's Statistic.

$$\sum\frac{(0_i - E_i)^2}{E_i} \text{ which is valid only if } 0_i < 2\ E_i$$

If this condition does not hold for even one cell, it should not be used.

3.4 Conclusion

Whenever the conditions for regularity are satisfied and the sample size n is a large one, the approximation $- 2\log\ T(\underline{x})$. to χ^2 can always be used.

This "large – sample" approximation proves very helpful for easier determination of critical region (critical value) especially when we do not know the distribution of the test statistic and this critical value is needed for the decision making.

Moreso, since $-2\log\ T(\underline{x})$ approximates to $\chi^2_{(r)}$ (r stands for restrictions imposed on parameter) the critical value with a fixed level of significance α , r degrees of freedom is read from the χ^2 table, thus we reject H_0 if the value computed from the test statistic $-2\log\ T(\underline{x})$ is greater than the table value, otherwise we do not have any basis to reject H_0.

For multinomial distribution

$$-2\log\ T(\underline{x}) = 2\sum O_i \log\ (O_i / E_i)$$

should be used instead of

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Since 2 $\sum O_i \log (O_i / E_i)$ is computationally convenient and valid for all values of 0_i and E_i. The statistic should replace the conventional Neyman – Pearson's statistic $\sum (O_i - E_i)^2 / E_i$ which is valid only if $0_i < 2E_i$.

References

[1] FISHER R.A (1956)" Statistical Method and Scientific Inference (3rd edition) New York, Hafner press

[2] NEYMAN, J and SCOTT, E (1948) " Consistent Estimator Based on partially consistent observations" Econometrica , Vol. 16 , 1-32 .

[3] BASU ,D (1955) " An inconsistency of the method of maximum Likelihood" Annals of Mathematical Statistics , Vol.26, 144-145

[4] KRAFT ,C.H and LECAM L.M (1956) " A Remark on the roots of the maximum Likelihood Equation. Annals of Mathematical Statistics Vol. 27, ,1174-1177 .

[5] BAHADUR R.R (1958) "Examples of inconsistency of maximum Likelihood Estimates" sankhya vol. 20, 207-210.

[6] CRAMER H and RAO C.R (1965) linear statistical inference and its Applications, New York: John Wiley.

[7] CRAMER H (1946)"Mathematical Methods of Statistics, Princeton: Princeton University Press

[8] PERLMAN , M.D (1972) " On the strong consistency of Approximate maximum Likelihood Estimates Proceedings of the sixth Berkeley symposium on Mathematical statistics and Probability, Vol.1,263-281 .

[9] RICE John A (1967) "Rates of Convergence of Estimates and Test Statistics" Annals of mathematical statistics, vol. 38 , 303-324.