SPS 2349 Test of Hypothesis

Generalized Likelihood Ratio Test

Dr. Mutua Kilai

Kirinyaga University

Introduction

Suppose a random variable X has a pdf $f(x_1, \theta_1, \theta_2, ..., \theta_k)$ depending on the unknown parameters $\theta_1, \theta_1, ..., \theta_k$

The set of all parameters is denoted by Ω known as the parameter space.

Let Ω_0 be a subset of Ω consider the hypothesis:

$$H_0 : (\theta_1, \theta_2, ..., \theta_k) \in \Omega_0 \;\; vs \;\; H_a : (\theta_1, \theta_2, ..., \theta_k) \in \Omega - \Omega_0$$

The above case is a composite hypothesis against a composite one. In this case we do not use Neyman Pearson lemma but we use a more generalized form of it generalized likelihood ratio test(LRT)

We denote the likelihood function of a sample $x_1, x_2, ..., x_n$ by

$$L(\Omega) = \prod_{i=1}^{n} f(x_i, \theta_1, \theta_2, ..., \theta_k)$$

Let $L(\hat{\Omega}) = \max L(\Omega)$ be maximized $L(\Omega)$ and $L(\Omega_0) = \max(L(\Omega_0))$ be maximized $L(\Omega_0)$ with Ω_0 only

We take the likelihood ratio as a quotient

$$\lambda = \frac{L(\hat{\Omega_0})}{L(\hat{\Omega})}$$

Since λ is the quotient of non-negative functions it is greater than or equal to zero. Further since $\Omega_0 \in \Omega$ then

$$L(\hat{\Omega}) \leq L(\hat{\Omega})$$

hence $0 \leq \lambda \leq 1$

To test $H_0 : \theta \in \Omega_0$ against $H_a : \theta \in \Omega - \Omega_0$ the critical region for the likelihood ratio test is the set of points in the sample for which

$$\lambda = \frac{L(\hat{\Omega_0})}{L(\hat{\Omega})} \leq k$$

Where k is a constant determined so that the level of significance is α

$$P\Big[\lambda \leq k | H_0 \;\; is \;\; true\Big] = \alpha$$

Example 1

$X \sim N(\mu, \sigma^2)$ where $\mu \& \sigma^2$ are unknown. Test the hypothesis

$$H_0 : \mu = \mu_0 \;\; vs \;\; H_1 : \mu \neq \mu_0$$

Solution

Let $x_1, x_2, ..., x_n$ be a random sample of size n from X then

$$L(X, \mu, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}} \sum_{i=1}^{n} (x_i - \mu)^2$$

In

$$\Omega = \{(\mu, \sigma) : -\infty < \mu < \infty, \sigma > 0\}$$
$$\Omega_0 = \{(\mu, \sigma) : \mu = \mu_0, \sigma > 0\}$$

$L(\hat{\Omega}) = \max_{\mu,\sigma^2 \in \Omega} L(X, \mu, \sigma^2)$

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

$$\frac{\partial l}{\partial \mu} = 0 \rightarrow 2 \sum_{i=1}^{n} \frac{(x_i - \mu)}{2\sigma^2} 0 \rightarrow \hat{\mu} = \bar{x}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = 0 \rightarrow -(\frac{n}{2}) \frac{1}{\sigma^2} + \frac{\sum_{i=1}^{2} (x_i - \mu)^2}{2(\sigma^2)^2}$$

$$\begin{aligned} L(\hat{\Omega}) &= \left(\frac{1}{2\pi\hat{\sigma^2}}\right) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \bar{x})^2} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left[\frac{n}{\sum_{i=1}^{n} (x_i - \bar{x})^2}\right]^{\frac{n}{2}} e^{-\frac{-\sum_{i=1}^{n} (x_i - \bar{x})^2}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2}} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left[\frac{n}{\sum (x_i - \bar{x})^2}\right]^{\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned} \tag{1}$$

Under Ω_0 the maximum likelihood estimates $\mu \& \sigma^2$ are

$$\hat{\mu} = \mu_0, \hat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

$$\begin{aligned} L(\hat{\Omega}) &= \left(\frac{1}{2\pi\hat{\sigma^2}}\right) e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu_0)^2} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left[\frac{n}{\sum_{i=1}^{n} (x_i - \mu_0)^2}\right]^{\frac{n}{2}} e^{-\frac{-\sum_{i=1}^{n} (x_i - \mu_0)^2}{\frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2}} \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \left[\frac{n}{\sum (x_i - \mu_0)^2}\right]^{\frac{n}{2}} e^{-\frac{n}{2}} \end{aligned} \tag{2}$$

Therefore the likelihood ratio criterion is given by

$$
\begin{aligned}
\lambda &= \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})} \leq k \\
&= \frac{\left[\frac{n}{\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}}}{\left[\frac{n}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{\frac{n}{2}}} \leq k \\
&= \left[\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\mu_0)^2}\right]^{\frac{n}{2}} \leq k
\end{aligned} \tag{3}
$$

The test is to reject H_0 when $\lambda \leq c$ Where c is such that

$$P[\lambda \leq c] = \alpha$$

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu_0)^2 &= \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu_0)^2 \\
&= \sum(x_i-\bar{x})^2 + 2(\bar{x}-\mu_0)\sum(x_i-\bar{x}) + n(\bar{x}-\mu_0)^2 \\
&= \sum(X_i-\bar{X})^2 + n(\bar{x}-\mu_0)^2
\end{aligned} \tag{4}
$$

So

$$\lambda = \left[\frac{\sum(x_i-\bar{x})^2}{\sum(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2}\right]^{\frac{n}{2}}$$

$$\lambda = \left[\frac{1}{1+\frac{\sum(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}}\right]^{\frac{n}{2}}$$

$$T = \frac{\sqrt{n}(\bar{x}-\mu_0)/\sigma}{\left[\frac{\sum(x_i-\bar{x})^2}{\sigma^2(n-1)}\right]^{\frac{1}{2}}} = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}}$$

This variable has the student t distribution with $n-1$ degrees of freedom when H_0 is true.

Thus

$$T^2 = \frac{n(\bar{x}-\mu_0)^2}{\frac{\sum(x_i-\bar{x})^2}{n-1}}$$

$$\lambda = \left[\frac{1}{1+\frac{T^2}{n-1}}\right]^{\frac{n}{2}}$$

$$\lambda \leq c \to T^2 \geq k$$

We therefore look for a constant c such that

$$P[\lambda \leq c|H_0] = \alpha = P[T^2 \geq k|H_0]$$

It is enough to find a constant k such that

$$P[T^2 \geq k|H_0] = \alpha = P[|T| > \sqrt{k}|H_0] = \alpha$$

or

$$P[-k < T < k] = 1 - \alpha$$

Since t distribution is symmetric about zero we usually find k such that

$$P[T > \sqrt{k}] = \frac{\alpha}{2} = P[T < -\sqrt{k}]$$

So the test is to compute the statistic

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}}$$

And reject H_0 if $|T| \geq t_{1-\frac{\alpha}{2}(n-1)}$

Example 2

Alfafa yields of six test plots are given by

$$1.5, 1.9, 1.2, 1.4, 2.3, 1.3$$

per acre. Use a critical region of 0.05 to test

$$H_0 : \mu = 1.8 \ \ vs \ \ \mu \neq 18$$

Assume the yields have a normal distribution of mean μ

Solution

The test is to compute

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)^2}{\sqrt{\frac{(x_i - \bar{x})^2}{n-1}}}$$

and reject H_0 when $|T| > c$ Where c is such that

$$P[|T| > c|H_0] = \alpha$$

or

$$P[-c < T < c|H_0] = 1 - \alpha = 0.05$$

Here

$$n = 6$$

and

$$\bar{x} = 1.6$$

and

$$\sum(x_i - \bar{x})^2 = 0.88$$

Therefore

$$T = \frac{\sqrt{n}(\bar{x} - \mu_0)}{\sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}} = \frac{\sqrt{6}(1.6 - 1.8)}{\sqrt{\frac{0.88}{5}}} = -1.17$$

From the tables with 5 df the value of c is 2.571.

Since

$$|T| = 1.17$$

and

$$1.17 < 2.571$$

we fail to reject H_0 at 0.05 level of significance.

Example 3

$$X \sim N(\mu, \sigma^2)$$

To test

$$H_0 : \mu = \mu_0, \sigma = \sigma_0 \;\; is \;\; known$$

vs

$$H_1 : \mu = \mu_1$$

Solution

$$\Omega = \{(\mu, \sigma) : -\infty < \mu < \infty, \sigma = \sigma_0\}$$

$$\Omega_0 = \{(\mu, \sigma) : \mu = \mu_0, \sigma = \sigma_0\}$$

$$L(X, \mu, \sigma) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\mu)^2}$$

$$L(\hat{\Omega}) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2}$$

$$L(\hat{\Omega}) = \left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\mu_0)^2}$$

Hence

$$\lambda = \frac{L(\hat{\Omega_0})}{L(\hat{\Omega})} = \frac{e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\mu_0)^2}}{e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2}} = e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\mu_0)^2+\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2}$$

The critical region is

$$\lambda = e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\mu_0)^2+\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2} \leq k$$

Taking natural logarithm

$$-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2+\frac{1}{2\sigma_0^2}\sum(x_i-\mu_0)^2 \leq \ln k$$

Simplifying further we have

$$\frac{1}{2\sigma_0^2}n(\bar{x}-\mu_0)^2 \leq \ln k+\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2$$

Then

$$\frac{n(\bar{x}-\mu_0)^2}{\sigma_0^2} < k$$

Let $Z = \frac{\bar{x}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0,1)$

Hence

$$\lambda < c \Rightarrow Z^2 > k$$

or

$$P[\lambda < c|H_0] = \alpha \Rightarrow P[|Z| \geq \sqrt{k}] = \alpha$$

The test now is to compute

$$Z = \frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma_0}$$

and reject H_0 when

$$|Z| \geq \sqrt{k}$$

where k is such that

$$P[|Z| \geq \sqrt{k}] = \alpha \Rightarrow P[-\sqrt{k} \leq Z \leq \sqrt{k}] = 1-\alpha$$

Testing the Variance of Normal Population

Let $x_1, x_2, ..., x_n$ be a random sample from a normal variable X.

Test the hypothesis

$$H_0 : \sigma^2 = \sigma_0^2 \;\; vs \;\; H_a : \sigma^2 \neq \sigma_0^2$$

μ is unknown. The parameter spaces are

$$\Omega = \{(\mu,\sigma) : -\infty < \mu < \infty, \sigma > 0\}$$

$$\Omega_0 = \{(\mu,\sigma) : -\infty < \mu < \infty, \sigma = \sigma_0\}$$

$$L(X,\mu,\sigma)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2}$$

The maximum likelihood estimates under Ω are

$$\hat{\mu}=\bar{x},\hat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Hence

$$L(\hat{\Omega})=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\left[\frac{n}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]^{\frac{n}{2}}e^{-\frac{n}{2}}$$

Under Ω_0 the maximum likelihood estimates are

$$\hat{\mu}=\bar{x},\hat{\sigma^2}=\sigma_0^2$$

Hence

$$L(\hat{\Omega_0})=\left(\frac{1}{2\pi}\right)^{\frac{n}{2}}\left(\frac{1}{2\sigma_0^2}\right)e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2}$$

$$\lambda=\frac{\hat{\Omega}_0}{\hat{\Omega}}=\left[\frac{\sum(x_i-\bar{x})}{n\sigma_0^2}\right]e^{-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2+\frac{n}{2}}$$

$$\lambda=\left(\frac{U}{n}\right)^{\frac{n}{2}}e^{-\frac{1}{2}(U-n)}$$

where $U=\frac{\sum(x_i-\bar{x})^2}{\sigma_0^2}$

The test is to compute

$$U=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}$$

and reject H_0 when $0<U<b$ or $U>b$ where the constants a and b are such that

$$P[0<U<a]\ \ or\ \ U>b]=\alpha$$

When H_0 is true the statistic

$$U=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{\sigma_0^2}$$

has χ^2 distribution with $n-1$ degrees of freedom

a and b are chosen such that

$$P[0<\chi^2_{n-1}<a]=\frac{\alpha}{2}=P[\chi^2_{(n-1)}>b]$$

Consider the sample variance

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

It is possible to write the variable U in terms of S as

$$U = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2(n-1)}{\sigma_0^2(n-1)} = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{(n-1)}$$

Example

In a random sample of the weights of 19 babies of certain age. The standard deviation was found to be 2.5kgs. Test the hypothesis that

$$H_0 : \sigma = 3 \ \ vs \ \ H_1 : \sigma \neq 3$$

at $\alpha = 0.05$

Solution

The statistic is

$$U = \frac{(n-1)S^2}{\sigma_0^2} = \frac{18 \times 2.5^2}{3^2} = 12.5$$

The critical value is

$$\chi^2_{0.025,18} = 31.53 = b$$

and

$$\chi^2_{0.975,18} = 8.23 = a$$

$U = 12.5$ lies in the acceptance region therefore we do not reject H_0 at 5% level of significance.