More on Hypothesis Testing

Introduction

Consider the test of the hypotheses

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1$$

A 'best test' at significance level α would be the test with the greatest power. Our quest is to find such a test.

① Most powerful Critical Region [MPCR]:

Suppose, we want to test ~~Simple~~ the simple null hypothesis $H_0: \theta = \theta_0$ against the simple alternative hypothesis $H_1: \theta = \theta_1$ $(\theta_1 \neq \theta_0)$ in our testing problem. Then, the most appropriate CR ~~corre~~ corresponding to this test will be called the MPCR.

Definition: A critical region ω_0 is said to be an MPCR of size 'α' for testing $H_0: \theta = \theta_0$ vs. $H_1: \theta = \theta_1 (\neq \theta_0)$ if

i) $P(\omega_0 | H_0) = \alpha \cdots ①$, $\forall \theta = \theta_1 \neq \theta_0$ [size condition]

& ii) $P(\omega_0 | H_1) \geq P(\omega | H_1) \cdots ②$, $\forall \theta = \theta_1 \neq \theta_0$ [power condition],

holds, what ever be the other CR 'ω' satisfying ① may be.

(2) Uniformly most powerful critical region [UMPCR]:

Suppose, we want to test the null hypothesis $H_0: \theta = \theta_0$ against the composite alternative hypothesis $H_1: \theta \neq \theta_0$ in our testing problem. Here, the most appropriate CR will be called the UMPCR.

Definition: A CR ω_0 is said to be UMPCR of size α for testing

$$\left.\begin{array}{l} H_0: \theta = \theta_0 \\ \text{vs. } H_1: \theta \neq \theta_0 \end{array}\right\}$$ ~~of size α~~, if, i) $P(\omega_0 \mid H_0) = \alpha$ [Size condition]

ii) $P(\omega_0 \mid H_1) \geq P(\omega \mid H_1)$, $\forall \theta \neq \theta_0$, [Power condition]

what ever the other CR 'ω' ~~may be~~ satisfying (i) may be.

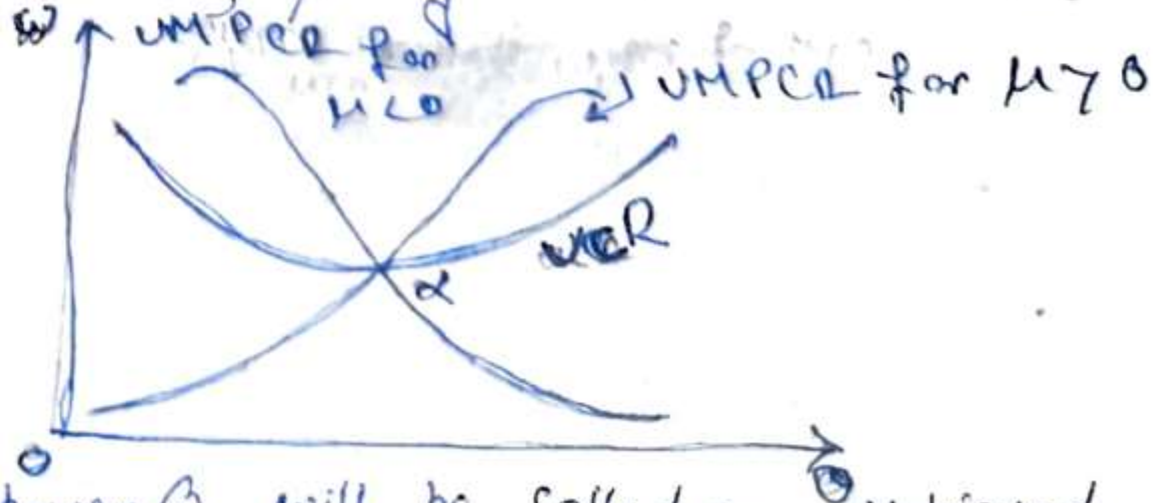
Similarly we can define UMPCR for one sided alternative hypothesis like,

1) $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$ & 2) $H_0: \theta = \theta_0$ vs. $H_1: \theta < \theta_0$.

③ Unbiased critical region :

In most of the both sided testing problems ∄ any UMPCR. In those situations, we have to introduce some additional conditions or criterion over the size & the power conditions.

This criteria is called the unbiasedness criteria.



Definition: A 'CR' ω_0 of level α & power β will be called an unbiased 'CR' if $\beta \geq \alpha$ i.e. $P[W_0 | H_1] \geq P[W_0 | H_0]$ & it is quite obvious that for a testing problem $\beta \geq \alpha$ will be a desirable situation.

(4) Uniformly ~~p~~ Most powerful Unbiased Critical Region [UMPUCR].

Definition: A CR ω_0 of size 'α' for testing $H_0: \theta = \theta_0$ against $H_1: \theta \neq \theta_0$

if i) $P(\omega_0 | H_0) = \alpha$ ... [Size condition]

ii) $P(\omega_0 | H_1) \geq P(\omega_0 | H_0)$ [unbiasedness criteria]

& iii) $P(\omega_0 | H_1) \geq P(\omega | H_1)$ [Power condition], whatever the other CR 'ω' satisfying (i) & (ii) may be.

Neyman – Pearson Lemma

Suppose that $X_1, \ldots, X_n$ have joint pdf $f(x_1, \ldots, x_n \,|\, \theta_0)$ under H_0 and $f(x_1, \ldots, x_n \,|\, \theta_1)$ under H_1. Define

$$\lambda(x_1, \ldots, x_n; \theta_0, \theta_1) \quad = \quad \frac{f(x_1, \ldots, x_n \,|\, \theta_0)}{f(x_1, \ldots, x_n \,|\, \theta_1)}. \tag{4.1}$$

Then $\lambda(x_1, \ldots, x_n; \theta_0, \theta_1)$ is the ratio of the likelihoods under H_0 and H_1. Let the critical region $C^* \subseteq \Omega$ be

$$C^* \quad = \quad \{(x_1, \ldots, x_n) : \lambda(x_1, \ldots, x_n; \theta_0, \theta_1) \leq k\} \tag{4.2}$$

where k is a constant chosen to make the test have significance level α, that is

$$P\{(X_1, \ldots, X_n) \in C^* \,|\, H_0 \text{ true}\} \quad = \quad \alpha.$$

The test based on the critical region $C^* = \{(x_1, \ldots, x_n) : \lambda(x_1, \ldots, x_n; \theta_0, \theta_1) \leq k\}$ has the largest power (smallest type II error) of all tests with significance level α.

Thus, among all tests with a given probability of a type I error, the likelihood ratio test minimises the probability of a type II error.

Test of mean, when Variance is known

Suppose that $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$ random quantities with σ^2 known. We shall apply the Neyman-Pearson lemma to construct the best test of the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu = \mu_1$$

where $\mu_1 > \mu_0$. From equation (4.1) we have that

$$\begin{aligned}
\lambda(x_1, \ldots, x_n; \mu_0, \mu_1) &= \frac{f(x_1, \ldots, x_n \,|\, \theta_0)}{f(x_1, \ldots, x_n \,|\, \theta_1)} \\
&= \frac{L(\mu_0)}{L(\mu_1)} \\
&= \frac{\exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2\right\}}{\exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2\right\}} \\
&= \exp\left\{\frac{1}{2\sigma^2}\left(\sum_{i=1}^n (x_i - \mu_1)^2 - \sum_{i=1}^n (x_i - \mu_0)^2\right)\right\}. \qquad (4.3)
\end{aligned}$$

Now,

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\mu_1)^2-\sum_{i=1}^{n}(x_i-\mu_0)^2 &= \sum_{i=1}^{n}(x_i^2-2\mu_1 x_i+\mu_1^2)-\sum_{i=1}^{n}(x_i^2-2\mu_0 x_i+\mu_0^2)\\
&= -2\mu_1 n\overline{x}+n\mu_1^2+2\mu_0 n\overline{x}-n\mu_0^2\\
&= n(\mu_1^2-\mu_0^2)-2n\overline{x}(\mu_1-\mu_0). \qquad (4.4)
\end{aligned}$$

Substituting equation (4.4) into (4.3) gives

$$\lambda(x_1,\ldots,x_n;\mu_0,\mu_1) \quad = \quad \exp\left\{\frac{1}{2\sigma^2}\left(n(\mu_1^2-\mu_0^2)-2n\overline{x}(\mu_1-\mu_0)\right)\right\}.$$

Using the Neyman-Pearson Lemma, see Lemma 1, the critical region of the most powerful test of significance level α for the test $H_0 : \mu = \mu_0 \text{versus} H_1 : \mu = \mu_1 \ (\mu_1 > \mu_0)$ is

$$
\begin{aligned}
C^* &= \left\{(x_1, \ldots, x_n) : \exp\left\{\frac{1}{2\sigma^2}\left(n(\mu_1^2 - \mu_0^2) - 2n\overline{x}(\mu_1 - \mu_0)\right)\right\} \leq k\right\} \\
&= \{(x_1, \ldots, x_n) : n(\mu_1^2 - \mu_0^2) - 2n\overline{x}(\mu_1 - \mu_0) \leq 2\sigma^2 \log k\} \\
&= \{(x_1, \ldots, x_n) : -2n\overline{x}(\mu_1 - \mu_0) \leq 2\sigma^2 \log k + n(\mu_0^2 - \mu_1^2)\} \\
&= \left\{(x_1, \ldots, x_n) : \overline{x} \geq \frac{-\sigma^2}{n(\mu_1 - \mu_0)} \log k + \frac{(\mu_0 + \mu_1)}{2}\right\} \qquad (4.5) \\
&= \{(x_1, \ldots, x_n) : \overline{x} \geq k^*\}. \qquad (4.6)
\end{aligned}
$$

$$
k^* = \mu_0 + z_{(1-\alpha)} \frac{\sigma}{\sqrt{n}},
$$

where $P(Z < z_{(1-\alpha)}) = 1-\alpha$. This is also written as $\Phi^{-1}(1-\alpha)$. $z_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of Z, the standard normal distribution.

A Practical Example of the Neyman-Pearson lemma

Suppose that the distribution of lifetimes of TV tubes can be adequately modelled by an exponential distribution with mean θ so

$$f(x \,|\, \theta) \quad = \quad \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right)$$

for $x \geq 0$ and 0 otherwise. Under usual production conditions, the mean lifetime is 2000 hours but if a fault occurs in the process, the mean lifetime drops to 1000 hours. A random sample of 20 tube lifetimes is to taken in order to test the hypotheses

$$H_0 : \theta = 2000 \quad \text{versus} \quad H_1 : \theta = 1000.$$

Use the Neyman-Pearson lemma to find the most powerful test with significance level α.

Note that

$$
\begin{aligned}
L(\theta) = \prod_{i=1}^{20} f(x_i \,|\, \theta) &= \frac{1}{\theta^{20}} \exp\left(-\frac{1}{\theta}\sum_{i=1}^{20} x_i\right) \\
&= \frac{1}{\theta^{20}} \exp\left(-\frac{20\overline{x}}{\theta}\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\lambda(x_1, \ldots, x_{20}; \theta_0, \theta_1) &= \frac{L(2000)}{L(1000)} \\
&= \frac{\frac{1}{2000^{20}} \exp\left(-\frac{20\overline{x}}{2000}\right)}{\frac{1}{1000^{20}} \exp\left(-\frac{20\overline{x}}{1000}\right)} \\
&= \left(\frac{1000}{2000}\right)^{20} \exp\left(-\frac{20\overline{x}}{2000} + \frac{20\overline{x}}{1000}\right) \\
&= \frac{1}{2^{20}} \exp\left(\frac{\overline{x}}{100}\right).
\end{aligned}
$$

Using the Neyman-Pearson lemma, the most powerful test of significance α has critical region

$$\begin{aligned} C^* &= \left\{(x_1, \dots, x_2) : \frac{1}{2^{20}} \exp\left(\frac{\overline{x}}{100}\right) \leq k\right\} \\ &= \left\{(x_1, \dots, x_2) : \frac{\overline{x}}{100} \leq \log 2^{20} k\right\} \\ &= \{(x_1, \dots, x_2) : \overline{x} \leq k^*\}. \end{aligned}$$

That is, a test of the form reject H_0 if $\overline{x} \leq k_1$. To find k^*, we need to know the sampling distribution of $\overline{X}$ when $X_1, \dots, X_{20}$ are iid exponentials with mean $\theta = 2000$ as

$$P(\overline{X} \leq k^* \mid \theta = 2000) = \alpha.$$

Example ③: Let X be a discrete r.v. Suppose,

$H_0: X \sim f(x) = \frac{e^{-1}}{x!}, \quad x = 0, 1, 2, \ldots$

Ag. $H_1: X \sim f(x) = \frac{1}{2^{x+1}}, \quad x = 0, 1, 2, \ldots$

Let, $k = 1$, for MPCR by NP lemma. Find the MPCR & show that it is unbiased.

Solution: Using NP lemma, here, the MPCR is, $W = \left\{ \underset{\sim}{x} : \frac{f(x, H_1)}{f(x, H_0)} \geq 1 \right\}$.

Now, $\frac{f(x, H_1)}{f(x, H_0)} \geq 1 \Rightarrow f(x, H_1) \geq f(x, H_0) \Rightarrow \frac{1}{2^{x+1}} \geq \frac{e^{-1}}{x!} \Rightarrow \frac{x!}{2^{x+1}} \geq \frac{1}{e}$.

$\therefore W = \left\{ \underset{\sim}{x} : \frac{x!}{2^{x+1}} \geq \frac{1}{e} \right\} = \left\{ x : \frac{x!}{2^x} \geq \frac{2}{e} \right\} = \left\{ \underset{\sim}{x} : \frac{e}{2} \geq \frac{2^x}{x!} \right\}$.

$\therefore W = \{0\}$. is the MPCR.

Now, $P_{H_1}(W) = P_{H_1}(X = 0) = \frac{1}{2^{0+1}} = \frac{1}{2}$; $P_{H_0}(W) = P_{H_0}(X = 0) = \frac{1}{e}$

Since, $\frac{1}{2} > \frac{1}{e}$, show, $P_{H_1}(W) > P_{H_0}(W) = \alpha$.

So, Power > Size. $\therefore$ The test is unbiased. [Proved]

Power function :

Definition: Let, γ be a test of the null hypothesis H_0. Then, the power function of the test γ, denoted by $\beta(\theta)$ is defined to be the probability that H_0 is rejected when the distribution from which the sample was obtained was parameterized by θ. As such, $\beta(\theta)$ is given by

$$\beta(\theta) = \begin{cases} \text{Probability of type-I error associated with } \theta, \text{ if } \theta \in \Theta \\ 1 - \text{Probability of type-II error associated with } \theta, \text{ if } \theta \notin \Theta. \end{cases}$$

In other words, $\beta(\theta) = P_{H_1}(W_0)$ is the probability of rejecting a false null hypothesis, i.e. the probability of accepting a true null hypothesis & thus, taking the right decision.

Use of power function:

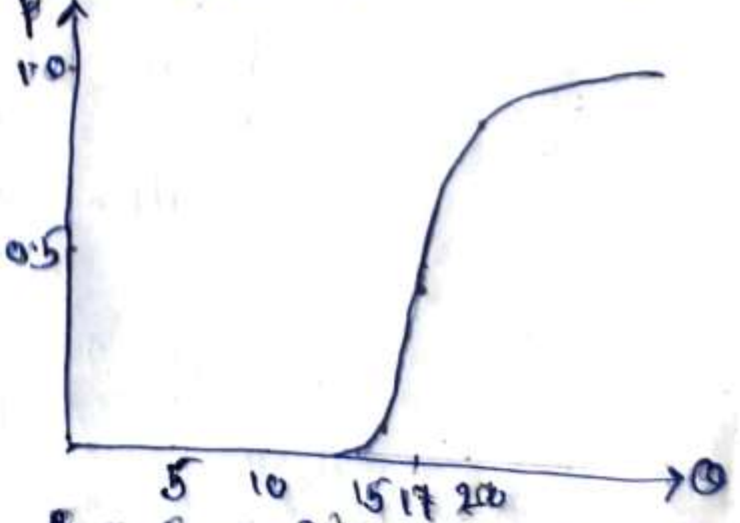
The whole nature of the test can be judged by looking at the power function. The power function plays the same role in hypothesis testing that meansquare error plays in estimation. It is usually the standard in as assessing the goodness of a test or in comparing two competing tests.

An ideal power function is a function that assumes the value '0' for those θ corresponding to ~~the~~ H_0 & is unity for those θ corresponding to ~~H~~ the alternative hypothesis H_1. As such, power function is useful in telling how good a particular test is.

Example: Let, $X_i \overset{iid}{\sim} N(\theta, 25)$, $\forall\, i = 1(1)n$. To test $H_0 : \theta \le 17$. Let, it is decided to reject the test iff $\bar{x} > 17 + \frac{5}{\sqrt{n}}$.

$$\therefore \beta(\theta) = P_\theta\left[\bar{x} > 17 + \frac{5}{\sqrt{n}}\right] = 1 - \Phi\left(\frac{17 + 5/\sqrt{n} - \theta}{5/\sqrt{n}}\right)$$

Here, if $\theta > 20$, the test is almost certain ~~not~~ to reject H_0, as it should. If $\theta < 16$, the test is almost certain not to reject H_0 & if $17 < \theta < 18$, (so, H_0 is false) the test has less than $1/2$ a chance of rejecting H_0.



Power Curve ~~with~~ for $n = 25$