Math 4501 - Probability and Statistics II

8.7 - Hypothesis testing: best critical regions

Best critical region of size α

Definition

Consider the test of the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$.

Let C be a critical region of size α; that is, $\alpha = P(C \,|\, \theta = \theta_0)$.

We say that C is a *best critical region of size* α if, for every other critical region D of size $\alpha = P(D \,|\, \theta = \theta_0)$, we have

$$P(C \,|\, \theta = \theta_1) \geq P(D \,|\, \theta = \theta_1) \ .$$

Notes:

- the condition in the definition says that when $H_1 : \theta = \theta_1$ is true, the probability of rejecting $H_0 : \theta = \theta_0$ with the use of the critical region C is at least as great as the corresponding probability with the use of any other critical region D of the same size α.
- a best critical region of size α is the critical region that has the greatest power among all critical regions of size α.

Neyman-Pearson Lemma

Theorem (Neyman-Pearson Lemma)

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with pdf or pmf $f(x;\theta)$, where θ_0 and θ_1 are two possible values of θ.

Denote the joint pdf or pmf of $X_1, X_2, \ldots, X_n$ by the likelihood function

$$L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = f(x_1;\theta)\, f(x_2;\theta) \cdots f(x_n;\theta) \ .$$

If there exist a positive constant k and a subset C of the sample space such that

(a) $P\left[(X_1, X_2, \ldots, X_n) \in C \,|\, \theta = \theta_0\right] = \alpha$,

(b) $\dfrac{L(\theta_0)}{L(\theta_1)} \leq k$ *for* $(x_1, x_2, \ldots, x_n) \in C$, *and*

(c) $\dfrac{L(\theta_0)}{L(\theta_1)} \geq k$ *for* $(x_1, x_2, \ldots, x_n) \in C'$,

then C is a best critical region of size α for testing the simple null hypothesis $H_0 : \theta = \theta_0$ against the simple alternative hypothesis $H_1 : \theta = \theta_1$.

Proof :

We will discuss here the case of random - variables of continuous - type.
The case of discrete - type random variables is analogous with some natural adjustments (such as replacing integrals by summations).

For the sake of simplicity, we will use the following notation:

(1) $L(\theta)$ will stand for $L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$

(2) $\int_B L(\theta)\, d^n x$ will stand for $\int \cdots \int_B L(\theta; x_1, \dots, x_n)\, dx_1 \dots dx_n$

Let C be as given in the statement (ie, satisfying (a), (b), (c))

Assume there exists another critical region of size α, denoted D here. Then:

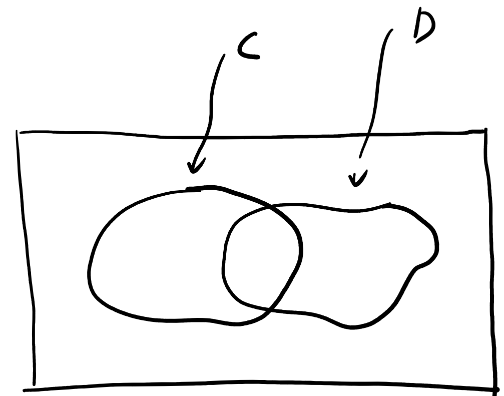
$$\alpha = \int_C L(\theta_0)\, d^n x = \int_D L(\theta_0)\, d^n x$$

and so

$$0 = \int_C L(\theta_0)\, d^n x - \int_D L(\theta_0)\, d^n x$$

$$= \int_{C\cap D} L(\theta_0)\, d^n x + \int_{C\cap D'} L(\theta_0)\, d^n x - \int_{C\cap D} L(\theta_0)\, d^n x - \int_{C'\cap D} L(\theta_0)\, d^n x$$

cancel



$$C = (C\cap D) \cup (C\cap D')$$

$$D = (C\cap D) \cup (C'\cap D)$$

Thus, we find that

$$0 = \int_{C\cap D'} L(\theta_0)\, d^n x - \int_{C'\cap D} L(\theta_0)\, d^n x$$

Hypothesis (b) $\Rightarrow$ $L(\theta_0) \leq k\, L(\theta_1)$ on C (and, thus, on $C \cap D' \subseteq C$)

Hypothesis (c) $\Rightarrow$ $L(\theta_0) \geq k\, L(\theta_1)$ on C' (and, thus, on $C' \cap D \subseteq C'$)

As a consequence, we get that

$$\int_{C \cap D'} L(\theta_0)\, d^n x \leq k \int_{C \cap D'} L(\theta_1)\, d^n x \qquad \text{(from hypothesis (b))}$$

and

$$\int_{C' \cap D} L(\theta_0)\, d^n x \geq k \int_{C' \cap D} L(\theta_1)\, d^n x \qquad \text{(from hypothesis (c))}$$

We conclude that

$$0 = \int_{C \cap D'} L(\theta_0)\, d^n x - \int_{C' \cap D} L(\theta_0)\, d^n x \leq k \left(\int_{C \cap D'} L(\theta_1)\, d^n x - \int_{C' \cap D} L(\theta_1)\, d^n x \right)$$

But this means that

$$0 \leq K \left(\int_{C\cap D} L(\theta_1)\, d^n x + \int_{C\cap D'} L(\theta_1)\, d^n x - \int_{C\cap D} L(\theta_1)\, d^n x - \int_{C'\cap D} L(\theta_1)\, d^n x \right)$$

or

$$0 \leq K \left(\int_{C} L(\theta_1)\, d^n x - \int_{D} L(\theta_1)\, d^n x \right)$$

Hence, we obtain that

$$\int_{C} L(\theta_1)\, d^n x \geq \int_{D} L(\theta_1)\, d^n x$$

that is $\quad P(C \mid \theta = \theta_1) \geq P(D \mid \theta = \theta_1)$

Since this holds for any critical region of size α, then C is the best critical region of size α.

Note: The inequality

$$\frac{L(\theta_0)}{L(\theta_1)} \leq k$$

can often be expressed in terms of a function $u(x_1, x_2, \ldots, x_n)$, say, as

$$u(x_1, \ldots, x_n) \leq c_1$$

or as

$$u(x_1, \ldots, x_n) \geq c_2 \ ,$$

where c_1 or c_2 is selected so that the size of the critical region is α.

Thus, the test can be based on the statistic $u(x_1, x_2, \ldots, x_n)$.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\mu, 36)$. Determine a statistic on which to base a best critical region for testing

$$H_0 : \mu = 50 \qquad \text{against} \qquad H_1 : \mu = 55$$

In the case where $n = 16$, find a best critical region with size $\alpha = 0.05$.

We will employ the Neyman-Pearson lemma, that is, we will look for all the points in the sample space such that the ratio of likelihood functions $\frac{L(50)}{L(55)}$ is less than or equal to some constant k.

Observe that

$$\frac{L(50)}{L(55)} = \frac{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(36)}} \exp\left(-\frac{(x_i-50)^2}{2(36)}\right)}{\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi(36)}} \exp\left(-\frac{(x_i-55)^2}{2(36)}\right)} = \frac{\exp\left(-\frac{1}{72}\sum_{i=1}^{n}(x_i-50)^2\right)}{\exp\left(-\frac{1}{72}\sum_{i=1}^{n}(x_i-55)^2\right)}$$

$$= \exp\left\{ -\frac{1}{72}\left(\underbrace{\sum_{i=1}^{n}(x_i-50)^2}_{\sum_{i=1}^{n} x_i^2 - 100x_i + 50^2} - \underbrace{\sum_{i=1}^{n}(x_i-55)^2}_{\sum_{i=1}^{n} x_i^2 - 110x_i + 55^2} \right) \right\}$$

$110-100$

x_i^2 terms cancel

$$= \exp\left\{ -\frac{1}{72}\left(10\sum_{i=1}^{n} x_i + n\left(50^2 - 55^2\right) \right) \right\}$$

want to make this $\leq k$

The inequality $\dfrac{L(50)}{L(55)} \leq k$ is then equivalent to (taking ln on both sides)

$$-10\sum_{i=1}^{n} x_i - n\left(50^2 - 55^2\right) \leq 72\ln(k)$$

or, equivalently:

$$\frac{1}{m}\sum_{i=1}^{m} x_i \geq -\frac{1}{10m}\left(m(56^2 - 55^2) + 72 \ln k \right)$$

which is of the form

$$\bar{x} \geq c$$

where $c \in \mathbb{R}$. Hence, by the Neyman-Pearson Lemma, a best critical region is

$$C = \left\{ (x_1, \ldots, x_m) : \bar{x} \geq c \right\}$$

where c is to be selected so that the size of C is α. In particular, the critical region is based on the sample mean a

If $n = 16$ and we are looking for a critical region of size $\alpha = 0.05$:

$$0.05 = \alpha = P(\text{Rej } H_0 \mid H_0 \text{ true}) = P(\bar{X} \geq c \mid \mu = 50)$$

Under H_0, $\mu = 50$ and so $X_1, \dots, X_{16} \sim N(50, 36) \Rightarrow Z = \dfrac{\bar{X} - 50}{6/4} \sim N(0,1)$

Hence, we find that

$$0.05 = P\left(\frac{\bar{X} - 50}{3/2} \geq \frac{c - 50}{3/2}\right) \iff 0.05 = 1 - \Phi\left(\frac{c-50}{3/2}\right)$$

$$\iff \Phi\left(\frac{c - 50}{3/2}\right) = 0.95 \iff \frac{c - 50}{3/2} = 1.645 \iff c = 50 + \frac{3}{2}(1.645) \simeq 52.47$$

We conclude that the best critical region of size $\alpha = 0.05$ is

$$C = \{(x_1, \dots, x_n) : \bar{x} \geq 52.47\}.$$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean λ. Show that a best critical region for testing

$$H_0 : \lambda = 2 \qquad \text{against} \qquad H_1 : \lambda = 5$$

is given by

$$C = \left\{ (x_1, x_2, \ldots, x_n) : \sum_{i=1}^{n} x_i \geq c \right\}$$

and determine its size α if $n = 4$ and $c = 13$.

A best critical region for testing $H_0 : \lambda = 2$ against $H_1 : \lambda = 5$ is given by

$$\frac{L(2)}{L(5)} \leq k$$

Observe that

$$\frac{L(2)}{L(5)} = \frac{\prod_{i=1}^{n} \frac{2^{x_i} e^{-2}}{x_i!}}{\prod_{i=1}^{n} \frac{5^{x_i} e^{-5}}{x_i!}} = \left(\frac{2}{5}\right)^{\sum_{i=1}^{n} x_i} \cdot e^{3n}$$

Hence, the inequality $\dfrac{L(2)}{L(5)} \leq k$ may be rewriten as

$$\left(\frac{2}{5}\right)^{\sum_{i=1}^{n} x_i} \cdot e^{3n} \leq k \,,$$

which is equivalent to

$$\underbrace{\ln\left(\frac{2}{5}\right)}_{\substack{\text{negative} \\ \text{because } \frac{2}{5} < 1}} \sum_{i=1}^{n} x_i \leq \ln k - 3n$$

or $$\sum_{i=1}^{n} x_i \geq \frac{\ln k - 3n}{\ln 2/5} \,,$$ which is of the form $\sum_{i=1}^{n} x_i \geq c$

We conclude that the desired best critical region is of the form:

$$C = \left\{ (x_1, \dots, x_n) : \sum_{i=1}^{n} x_i \geq c \right\}$$

If $n = 4$ and $c = 13$, the size of the critical region C is

$$\alpha = P(\text{Rej } H_0 \mid H_0 \text{ true}) = P\left(\sum_{i=1}^{4} X_i \geq 13 \mid \lambda = 2 \right)$$

Under $H_0 : \lambda = 2$, we have that $X_1, \ldots, X_4 \sim \text{Poisson}(4)$ and so $\sum_{i=1}^{4} X_i \sim \text{Poisson}(8)$.

Thus, we conclude that

$$\alpha = P\left(\sum_{i=1}^{4} X_i \geq 13 \mid \lambda = 2 \right) = 1 - P\left(\sum_{i=1}^{4} X_i < 13 \mid \lambda = 2 \right)$$

$$= 1 - \sum_{x=0}^{12} \frac{8^x e^{-8}}{x!} \approx 0.064$$

Most powerful test

- When $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ are both simple hypotheses, a critical region of size α is a best critical region if the probability of rejecting H_0 when H_1 is true is a maximum compared with all other critical regions of size α.
- The test using the best critical region is called a *most powerful test* because it has the greatest value of the power function at $\theta = \theta_1$ compared with that of other tests with significance level α.
  - If H_1 is a composite hypothesis, the power of a test depends on each simple alternative in H_1.

Definition

A test defined by a critical region C of size α is a *uniformly most powerful test* if it is a most powerful test against each simple alternative in H_1.

The critical region C is called a *uniformly most powerful critical region of size* α.

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\mu, 36)$.

We have already seen that when testing

$$H_0 : \mu = 50 \qquad \text{against} \qquad H_1 : \mu = 55$$

a best critical region is defined by

$$C = \{(x_1, x_2, \ldots, x_n) : \bar{x} \geq c\} \ ,$$

where c is selected so that the significance level is α.

Show that such critical region is a uniformly most powerful test, and C is a uniformly most powerful critical region of size α, when testing

$$H_0 : \mu = 50 \qquad \text{against} \qquad H_1 : \mu > 50$$

For each simple hypothesis in H_1, $\mu = \mu_1$, the ratio of the likelihood functions is

$$\frac{L(50)}{L(\mu_1)} = \frac{\prod_{i=1}^{n} (2\pi(36))^{-1/2} \cdot \exp\left(-(x_i - 50)^2/2(36)\right)}{\prod_{i=1}^{n} (2\pi(36))^{-1/2} \cdot \exp\left(-(x_i - \mu_1)^2/2(36)\right)}$$

$$= \exp\left\{ -\frac{1}{72} \left(\sum_{i=1}^{n} (x_i - 50)^2 - \sum_{i=1}^{n} (x_i - \mu_1)^2 \right) \right\}$$

$$= \exp\left\{ -\frac{1}{72} \left(2(\mu_1 - 50) \sum_{i=1}^{n} x_i + n(50^2 - \mu_1^2) \right) \right\}$$

We then observe that $\dfrac{L(50)}{L(\mu_1)} \leq k$ if and only if

$$-\frac{1}{72} \left(2(\mu_1 - 50) \sum_{i=1}^{n} x_i + n(50^2 - \mu_1^2) \right) \leq \ln k$$

taking ln on both sides

which reduces to

$$\bar{x} \geq \frac{-72 \ln k - n(50^2 - \mu_1^2)}{2(\mu_1 - 50)\, n}$$

that is,

$$\bar{x} \geq \underbrace{-\frac{72 \ln k}{2n(\mu_1-50)} + \frac{(50+\mu_1)}{2}}_{c}$$

Thus, the best critical region of size α for testing $H_0: \mu = 50$ vs $H_1: \mu = \mu_1$, where $\underbrace{\mu_1 > 50}$, is given by

Note this is true for any such μ_1!!!

$$C = \{ (x_1, \ldots, x_n) : \bar{x} \geq c \},$$

where c is selected such that $P(\bar{X} \geq c \mid \underbrace{\mu = 50}_{H_0 \text{ true}}) = \alpha$

The same value of c can be used for each $\mu_1 > 50$ (even though k changes)

Since C defines a test that is most powerful against each simple alternative $\mu_1 > 50$, this is a uniformly most powerful test, and C is a uniformly most powerful critical region of size α.

Example

Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter $p \in (0,1)$.

Find a uniformly most powerful test of

$$H_0 : p = p_0 \qquad \text{against} \qquad H_1 : p > p_0 \ .$$

For each simple hypothesis in H_1, $p = p_1$ with $p_1 > p_0$, the ratio of the likelihood functions is

$$\frac{L(p_0)}{L(p_1)} = \frac{\prod_{i=1}^{n} p_0^{x_i}(1-p_0)^{1-x_i}}{\prod_{i=1}^{n} p_1^{x_i}(1-p_1)^{1-x_i}} = \left[\frac{p_0(1-p_1)}{p_1(1-p_0)}\right]^{\sum_{i=1}^{n} x_i} \cdot \left[\frac{1-p_0}{1-p_1}\right]^{n}$$

We then observe that $\dfrac{L(p_0)}{L(p_1)} \leq k$ reduces to

$$\left[\frac{p_0(1-p_1)}{p_1(1-p_0)}\right]^{\sum_{i=1}^{n} x_i} \leq k \left[\frac{1-p_0}{1-p_1}\right]^{-n}$$

solving for $\sum_{i=1}^{n} x_i$ yields

$$\underbrace{\ln\left[\frac{p_0(1-p_1)}{p_1(1-p_0)}\right]}_{<0 \text{ because } p_1 > p_0 \text{ and so } (1-p_1) < 1-p_0 \text{ so that } \frac{p_0}{p_1}\cdot\frac{1-p_1}{1-p_0} < 1} \sum_{i=1}^{n} x_i \leq \ln k - n \ln\left(\frac{1-p_0}{1-p_1}\right)$$

and so

$$\sum_{i=1}^{n} x_i \geq \frac{\ln k - n\ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left[\frac{p_0(1-p_1)}{p_1(1-p_0)}\right]}$$

or even

$$\bar{x} \geq \frac{\ln k - n\ln\left(\frac{1-p_0}{1-p_1}\right)}{\ln\left(\frac{p_0(1-p_1)}{p_1(1-p_0)}\right)} = c \qquad \text{for each } p_1 > p_0$$

Thus, the best critical region to test $H_0 : p = p_0$ vs $H_1 : p = p_1$, with $p_1 > p_0$, is of the form

$$C = \left\{ (x_1, \ldots, x_n) : \bar{x} \geq c \right\},$$

where c is to be selected so that $P(\bar{X} \geq c \mid H_0) = \alpha$

In particular, the same C may be used for each choice of $p_1 > p_0$.

Since C defines a test that is most powerful against each simple alternative $p_1 > p_0$, the test associated with C is a uniformly most powerful test

Relation with sufficient statistics

If a sufficient statistic $Y = u(X_1, X_2, \ldots, X_n)$ exists for an unknown parameter θ of a distribution with pdf/pmf $f(x;\theta)$ then, by the factorization theorem,

$$\frac{L(\theta_0)}{L(\theta_1)} = \frac{\phi[u(x_1, x_2, \ldots, x_n); \theta_0]\, h(x_1, x_2, \ldots, x_n)}{\phi[u(x_1, x_2, \ldots, x_n); \theta_1]\, h(x_1, x_2, \ldots, x_n)}$$
$$= \frac{\phi[u(x_1, x_2, \ldots, x_n); \theta_0)}{\phi[u(x_1, x_2, \ldots, x_n); \theta_1]}.$$

Thus, the condition

$$\frac{L(\theta_0)}{L(\theta_1)} \leq k$$

provides a critical region that is a function of the observations $x_1, x_2, \ldots, x_n$ only through the observed value of the sufficient statistic

$$Y = u(X_1, X_2, \ldots, X_n)\ .$$

Hence, best critical and uniformly most powerful critical regions are based upon sufficient statistics when they exist.