Size, P-values, and Confidence Regions

1 Size, level and UMP

Definition 1.1 *The power function with a test with rejection region R is the function of $\theta \in \Theta$, defined as*

$$\beta(\theta) = \mathbb{P}_\theta(\boldsymbol{x} \in \Theta).$$

Then, if $\theta \in \Theta_0$,

$$\text{Probability of Type I error} = \mathbb{P}_\theta(H_0 \text{ is rejected}) = \beta(\theta),$$

and if $\theta \in \Theta_1$,

$$\text{Probability of Type II error} = 1 - \mathbb{P}_\theta(H_0 \text{ is rejected}) = 1 - \beta(\theta).$$

Definition 1.2 *For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a size α test, if*

$$\sup_{\theta \in \Theta_0} \beta(\theta) = \alpha.$$

Definition 1.3 *For $0 \leq \alpha \leq 1$, a test with power function $\beta(\theta)$ is a level α test, if*

$$\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha.$$

Definition 1.4 *Let $\mathscr{C}$ be a class of tests for testing*

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_0^c.$$

A test in $\mathscr{C}$ is called uniformly most powerful (UMP), if its power function $\beta(\theta)$ satisfies

$$\beta(\theta) \geq \tilde{\beta}(\theta)$$

for all $\theta \in \Theta_0^c$ and all $\tilde{\beta}(\theta)$ is the power function of a test in $\mathscr{C}$.

We often consider $\mathscr{C} = \{\text{tests of level } \alpha\}$. In this case, we call the optimal test the UMP level α test.

Lemma 1.5 (Neyman - Pearson) *Consider a simple test*

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

Let the pdf/pmf corresponding to θ_i be $f(\boldsymbol{x}|\theta_i)$ for $i = 0, 1$. Consider a test with the rejection region R satisfying

$$\begin{cases} \text{if } f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0), \text{ then } \boldsymbol{x} \in R \\ \text{if } f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0), \text{ then } \boldsymbol{x} \in R^c \end{cases} \tag{1.1}$$

and

$$\mathbb{P}_{\theta_0}(\boldsymbol{X} \in R) = \alpha. \tag{1.2}$$

Then we have
a. (Sufficiency) Any test satisfying (1) and (2) is a most powerful level α test.
b. (Necessity) If a test exists satisfying (1) and (2) with $k > 0$, then any MP level α test satisfies (1), except perhaps on a set A satisfying

$$\mathbb{P}_{\theta_0}(\boldsymbol{X} \in A) = \mathbb{P}_{\theta_1}(\boldsymbol{X} \in A) = 0.$$

We only give the proof of sufficiency for the case of pdf.
Proof Define

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in R \\ 0 & \text{if } \boldsymbol{x} \in R^c. \end{cases}$$

Suppose the rejection region $\tilde{R}$ is a level α test, and define

$$\tilde{\phi}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} \in \tilde{R} \\ 0 & \text{if } \boldsymbol{x} \in \tilde{R}^c. \end{cases}$$

We first claim

$$\int (\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x}))(f(\boldsymbol{x}|\theta_1) - f(\boldsymbol{x}|\theta_0))d\mu(\boldsymbol{x}) \geq 0.$$

Case 1: If $f(\boldsymbol{x}|\theta_1) > kf(\boldsymbol{x}|\theta_0)$, then $\phi(\boldsymbol{x}) = 1$. By $\tilde{\phi}(\boldsymbol{x}) \leq 1$, the integrand is nonnegative.
Case 2: If $f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0)$, then $\phi(\boldsymbol{x}) = 0$. By $\tilde{\phi}(\boldsymbol{x}) \geq 0$, the integrand is nonnegative.
Case 3: If $f(\boldsymbol{x}|\theta_1) < kf(\boldsymbol{x}|\theta_0)$, then the integrand is zero.

Then we have

$$\begin{aligned} \mathbb{P}_{\theta_1}(\boldsymbol{X} \in R) - \mathbb{P}_{\theta_1}(\boldsymbol{X} \in \tilde{R}) &= \mathbb{E}_{\theta_1}(\phi(\boldsymbol{X}) - \tilde{\phi}(\boldsymbol{X})) \\ &= \int (\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x}))f(\boldsymbol{x}|\theta_1)d\mu(\boldsymbol{x}) \\ &\geq k \int (\phi(\boldsymbol{x}) - \tilde{\phi}(\boldsymbol{x}))f(\boldsymbol{x}|\theta_0)d\mu(\boldsymbol{x}) \\ &= k(\mathbb{P}_{\theta_0}(\boldsymbol{X} \in R) - \mathbb{P}_{\theta_0}(\boldsymbol{X} \in \tilde{R})) \geq 0. \end{aligned}$$

The last inequality is because R is size α while $\tilde{R}$ is level α. ■

2 P-values

Definition 2.1 *A p-value $p(\boldsymbol{x})$ is a test statistic satisfying $0 \leq p(\boldsymbol{x}) \leq 1$. Small values of $p(\boldsymbol{x})$ give evidence that H_1 is true. A p-value is valid, if, for every $\theta \in \Theta_0$ and every $0 \leq \alpha \leq 1$,*

$$\mathbb{P}_\theta(p(\boldsymbol{X}) \leq \alpha) \leq \alpha.$$

In other words, if the null hypothesis is true, $p(\boldsymbol{X})$ is stochastically larger than a uniform distribution.

Once we have a valid p-value, then a level α test based on $p(\boldsymbol{x})$ can be constructed easily. The rejection region of this test is

$$R = \{\boldsymbol{x} : p(\boldsymbol{x}) \leq \alpha\}.$$

How to construct p-values? Let's first review properties of cdf's and quantile functions.

Theorem 2.2 (Properties of cdf's) *Let $F(t) = \mathbb{P}(X \leq t)$ be the cdf of a random variable X. Then*

1. *$F(x) \leq F(y)$ for any $x \leq y$;*
2. *$\lim_{x \searrow y} F(x) = F(y)$;*
3. *$\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$.*

Definition 2.3 *For any $0 < u < 1$, define the quantile function*

$$F^{-1}(u) = \inf\{x \in \mathbb{R} : F(x) \geq u\}.$$

Lemma 2.4 (Switching lemma) *For any* $0 < u < 1$ *and* $x \in \mathbb{R}$,

$$F(x) \geq u \Longleftrightarrow x \geq F^{-1}(u).$$

In particular,

$$F(F^{-1}(u)) \geq u.$$

If, in addition, F is continuous, we have

$$F(F^{-1}(u)) = u.$$

Proof For any $u \in (0,1)$ and $x \in R$, if $F(x) \geq u$, then

$$x \geq \inf\{x : F(x) \geq u\} = F^{-1}(u).$$

On the other hand, if $x \geq F^{-1}(u)$, we have

$$F(x) \geq F(F^{-1}(u)).$$

Let $x_k \in \{x : F(x) \geq u\}$ satisfy $x_k \searrow F^{-1}(u)$. Then $F(x_k) \geq u$, and

$$F(F^{-1}(u)) = F(\lim_k x_k) = \lim_k F(x_k) \geq u,$$

which further implies $F(x) \geq u$.

If F is continuous, let $x_k \nearrow F^{-1}(u)$, by the definition of $F^{-1}(u)$, we have $F(x_k) < u$. Then

$$F(F^{-1}(u)) = F(\lim_k x_k) = \lim_k F(x_k) \leq u.$$

Together with $F(F^{-1}(u)) \geq u$, we have $F(F^{-1}(u)) = u$. ■

Theorem 2.5 (cdf representation) *Let F be a cdf of the random variable X, then*

1. *Any random variable U that is uniformly distributed on $(0,1)$ satisfies $F^{-1}(U) \stackrel{d}{=} X$.*
2. *For any $u \in (0,1)$, $\mathbb{P}(F(X) \leq u) \leq u$.*
3. *If in addition, F is continuous, then $F(X) \stackrel{d}{=} U$, which gives $\mathbb{P}(F(X) \leq u) = u$.*

Proof (1) For any x,

$$\begin{aligned} \mathbb{P}(F^{-1}(U) \leq x) &= \mathbb{P}(U \leq F(x)) \quad \text{(Switching lemma)} \\ &= F(x) = \mathbb{P}(X \leq x) \end{aligned}$$

which gives $F^{-1}(U) \stackrel{d}{=} X$.
(2) Since $F^{-1}(U) \stackrel{d}{=} X$, we have

$$\mathbb{P}(F(X) \leq u) = \mathbb{P}(F(F^{-1}(U)) \leq u).$$

By the switching lemma, $F(F^{-1}(u)) \geq u$. Then

$$\{F(F^{-1}(U)) \leq u\} \subset \{U \leq u\},$$

which implies

$$\mathbb{P}(F(F^{-1}(U)) \leq u) \leq \mathbb{P}(U \leq u) = u.$$

(3) If F is continuous, we have $F(F^{-1}(U)) = U$, so for any $u \in (0,1)$,

$$\mathbb{P}(F(X) \leq u) = \mathbb{P}(F(F^{-1}(U)) \leq u) = \mathbb{P}(U \leq u),$$

which implies that $F(X) \stackrel{d}{=} U$. ■

Theorem 2.6 (P-values construction from test statistics) *Let $W(\boldsymbol{x})$ be a test statistic such that large values of W give evidence that H_1 is true. For each sample point $\boldsymbol{x}$, define*

$$p(\boldsymbol{x}) = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(W(\boldsymbol{X}) \geq W(\boldsymbol{x})).$$

Then, $p(\boldsymbol{x})$ is a valid p-value.

Proof For any $\theta \in \Theta_0$, define

$$p_\theta(\boldsymbol{x}) := \mathbb{P}_\theta(W(\boldsymbol{X}) \geq W(\boldsymbol{x})) = \mathbb{P}_\theta(-W(\boldsymbol{X}) \leq -W(\boldsymbol{x})) = F_\theta(-W(\boldsymbol{x})),$$

where F_θ is the cdf of $-W(\boldsymbol{X})$ under θ. Therefore,

$$\mathbb{P}_\theta(p_\theta(\boldsymbol{X}) \leq \alpha) = \mathbb{P}_\theta(F_\theta(-W(\boldsymbol{X})) \leq \alpha) \leq \alpha.$$

Since

$$p(\boldsymbol{x}) = \sup_{\theta \in \Theta_0} p_\theta(\boldsymbol{x}) \geq p_\theta(\boldsymbol{x}),$$

we have

$$\{p(\boldsymbol{X}) \leq \alpha\} \subset \{p_\theta(\boldsymbol{X}) \leq \alpha\},$$

which implies that

$$\mathbb{P}_\theta(p(\boldsymbol{X}) \leq \alpha) \leq \mathbb{P}_\theta(p_\theta(\boldsymbol{X}) \leq \alpha) \leq \alpha.$$

Notice this inequality holds for all $\theta \in \Theta_0$, we have

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(p(\boldsymbol{X}) \leq \alpha) \leq \alpha,$$

which implies that $p(\boldsymbol{x})$ is a valid P-value. ■

3 Confidence regions

Definition 3.1 *For any $\boldsymbol{x}$ in the sample space, let $C(\boldsymbol{x})$ be a $1-\alpha$ confidence region for $\xi = \xi(\theta)$, if*

$$\mathbb{P}_\theta(\xi \in C(\boldsymbol{X})) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$$

Definition 3.2 *For any ξ_0, let $\mathscr{A}(\xi_0)$ be the acceptance region for a level α test of*

$$H_0 : \xi(\theta) = \xi_0 \quad vs. \quad H_1 : \xi(\theta) \neq \xi_0,$$

if it satisfies

$$\mathbb{P}_\theta\left(\boldsymbol{X} \in \mathscr{A}(\xi(\theta))\right) \geq 1 - \alpha, \quad \forall \theta \in \Theta.$$

Acceptance region to confidence region: Given the acceptance region $\mathscr{A}(\xi)$, let

$$C(\boldsymbol{x}) = \{\xi : \boldsymbol{x} \in \mathscr{A}(\xi)\}.$$

Then

$$\xi(\theta) \in C(\boldsymbol{x}) \longleftrightarrow \boldsymbol{x} \in \mathscr{A}(\xi(\theta)).$$

It follows that

$$\mathbb{P}_\theta(\xi(\theta) \in C(\boldsymbol{x})) = \mathbb{P}_\theta(\boldsymbol{X} \in \mathscr{A}(\xi(\theta))) \geq 1 - \alpha.$$

Confidence region to acceptance region: Similarly, given the confidence region $C(\boldsymbol{x})$, let

$$\xi = \{\boldsymbol{x} : \xi \in C(\boldsymbol{x})\}.$$

Again,

$$\xi(\theta) \in C(\boldsymbol{x}) \longleftrightarrow \boldsymbol{x} \in \mathscr{A}(\xi(\theta)).$$

Then

$$\mathbb{P}_\theta(\boldsymbol{X} \in \mathscr{A}(\xi(\theta))) = \mathbb{P}_\theta(\xi(\theta) \in C(\boldsymbol{x})) \geq 1 - \alpha.$$