## 7.3 Concept of a P-Value

There is not always general agreement about how small $\alpha$ should be in order for rejection of $H_0$ to constitute strong evidence in support of $H_a$. Experimenter I may consider $\alpha = 0.05$ sufficiently small, while experimenter II insists on using $\alpha = 0.01$. Thus, it would be possible for experimenter I to reject when experimenter II fails to reject, based on the same data. If the experimenters agree to use the same test statistic, then this problem may be overcome by reporting the results of the experiment in terms of the observed size or P-value of the test, which is defined as the smallest size $\alpha$ at which $H_0$ can be rejected, based on the observed value of the test statistic.

Definition 7.3.1. **(P-Value)**. Smallest level of significance at which test would reject for the given observed values.

Theorem 7.3.1. **(General Method to Find P-Value)**. Suppose that the test statistic $T$ has CDF $F(t;\theta) = \{T \le t|\theta\}$ and $F$ is continuous, monotonically decreasing in $\theta$ for fixed $t$.

1. $H_0 : \theta \le \theta_0$ (or $\theta = \theta_0$) v.s. $H_a : \theta = \theta_1$, $\theta_1 > \theta_0$ (or $\theta > \theta_0$) with a critical region of form $\{t : t \ge c\}$. If $T = t$ is observed, then P-value is $1 - F(t-;\theta_0)$
2. $H_0 : \theta \ge \theta_0$ (or $\theta = \theta_0$) v.s. $H_a : \theta = \theta_1$, $\theta_1 < \theta_0$ (or $\theta < \theta_0$) with a critical region of form $\{t : t \le c\}$. If $T = t$ is observed, then P-value is $F(t;\theta_0)$.
3. $H_0 : \theta = \theta_0$ v.s. $H_a : \theta \ne \theta_0$ with a critical region of form $\{t : t \ge c_2$ or $t \le c_1, c_2 > c_1\}$. If $T = t$ is observed, then P-value is $2\min\{F(t;\theta_0), 1 - F(t-;\theta_0)\}$.

Here $F(t-;\theta_0) = \lim\limits_{\epsilon \to 0} F(t-\epsilon;\theta_0)$.

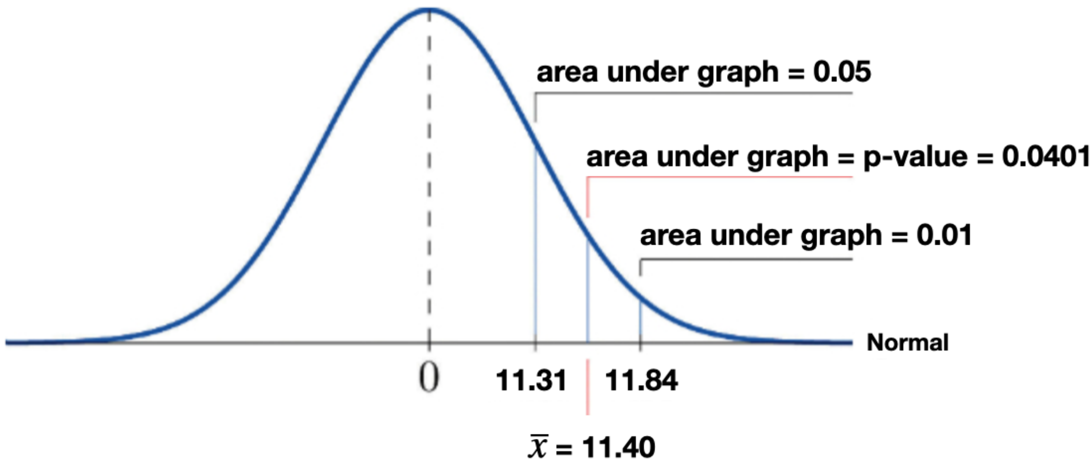
*Proof.* For 1

$$\text{P-value} = \min_{c \le t} \underbrace{\max_{\theta \le \theta_0} \mathbb{P}\{T \ge c|\theta\}}_{\text{significant level}} = \min_{c \le t} \max_{\theta \le \theta_0} 1 - F(c-;\theta)$$

$$= \min_{c \le t} 1 - F(c-;\theta_0) = 1 - F(t-;\theta_0)$$

$\square$

$\boxed{\textit{Example 7.3.1.}}$. For a random sample from $\mathcal{N}(\mu, \sigma)$, $H_0 : \mu = 10$ v.s. $H_a : \mu > 10$, $n = 25$, $\sigma^2 = 16$. Observe $\bar{x} = 11.40$. By theorem (part 1) above, the P-value is




$$\mathbb{P}\{\overline{X} \geq 11.40|\mu = 10\} = \mathbb{P}\left\{\underbrace{\sqrt{n}\frac{\overline{X} - 10}{\sigma}}_{\mathcal{N}(0,1)} \geq \sqrt{25}\frac{11.40 - 10}{4}\middle|\mu = 10\right\}$$

$$= 1 - \Phi(1.75) = 0.0401$$

P-value$= 0.0401$ is the smallest level of significance at which test would reject for the given observed values. We can verify this. If $\alpha = 0.01$, we would have rejected if $\bar{x} > \mu_0 + \frac{z_{1-0.01}}{\sqrt{n}}\sigma = 10 + \frac{2.3}{5}4 = 11.84$. And if $\alpha = 0.05$, we would have rejected if $\bar{x} > \mu_0 + \frac{z_{1-0.05}}{\sqrt{n}}\sigma = 10 + \frac{1.645}{5}4 = 11.316$. So we would reject at $\alpha = 0.05$ level, but not at $\alpha = 0.01$ level.

The theorem leads to alternative way to conduct hypothesis test: Find sufficient statistic $T$ such that $F(t;\theta) = \mathbb{P}\{T \leq t|\theta\}$ is decreasing in $\theta$ with $t$ fixed.

$\boxed{\text{Corollary 7.3.1.}}$. Under the conditions of last theorem, if $T = t$ is observed

1. $H_0 : \theta \leq \theta_0$ (or $\theta = \theta_0$) v.s. $H_a : \theta = \theta_1$, $\theta_1 > \theta_0$ (or $\theta > \theta_0$) with a critical region of form $\{t : t \geq c\}$. To reject $H_0$ if $1 - F(t-; \theta_0) \leq \alpha$ is level-$\alpha$ test.

2. $H_0 : \theta \geq \theta_0$ (or $\theta = \theta_0$) v.s. $H_a : \theta = \theta_1$, $\theta_1 < \theta_0$ (or $\theta < \theta_0$) with a critical region of form $\{t : t \leq c\}$. To reject $H_0$ if $F(t; \theta_0) \leq \alpha$ is level-$\alpha$ test.

3. $H_0 : \theta = \theta_0$ v.s. $H_a : \theta \neq \theta_0$ with a critical region of form $\{t : t \geq c_2$ or $t \leq c_1, c_2 > c_1\}$. To reject $H_0$ if $2\min\{F(t; \theta_0), 1 - F(t-; \theta_0)\} \leq \alpha$ is level-$\alpha$ test.

Here $F(t-; \theta_0) = \lim_{\epsilon \to 0} F(t - \epsilon; \theta_0)$.

Note that these are conservative if $T$ is discrete. Apply this corollary and get results of the following two theorems for testing hypotheses for means of Binomial and Poisson.

Theorem 7.3.2. **(Tests of Hypotheses for Means of Binomial).** Suppose that the test statistic $T \sim \mathcal{B}in(n, p)$, and $B(t; n, p)$ denotes a binomial CDF. Denote by $t$ an observed value of $T$.

1. A conservative size $\alpha$ test of $H_0 : p \leq p_0$ against $H_a : p > p_0$ is to reject $H_0$ if
$$1 - B\left(t - 1; n, p_0\right) \leq \alpha$$

2. A conservative size $\alpha$ test of $H_0 : p \geq p_0$ against $H_a : p < p_0$ is to reject $H_0$ if
$$B\left(t; n, p_0\right) \leq \alpha$$

3. A conservative two-sided test of $H_0 : p = p_0$ against $H_a : p \neq p_0$ is to reject $H_0$ if
$$B\left(t; n, p_0\right) \leq \alpha/2 \quad \text{or} \quad 1 - B\left(t - 1; n, p_0\right) \leq \alpha/2$$

*Proof.* Suppose $T \sim \mathcal{B}in(n, p)$, $H_0 : p = p_0$ v.s $H_a : p > p_0$. Take critical region $\{t : t \geq c\}$. Considering part 1, we reject $H_0$ if

$$1 - B(t-; n, p_0) \leq \alpha \Leftrightarrow {}^{1}1 - B(t - 1; n, p_0) \leq \alpha \Leftrightarrow B(t - 1; n, p_0) \geq 1 - \alpha$$

$$\Leftrightarrow t - 1 \geq c_{1-\alpha}$$

[1] $B(t-; n, p_0) = \lim_{\epsilon \to 0} \sum_{i=0}^{\lfloor t-\epsilon \rfloor} \binom{n}{i} p_0^i (1 - p_0)^{n-i} = \sum_{i=0}^{t-1} \binom{n}{i} p_0^i (1 - p_0)^{n-i} = B(t - 1; n, p_0)$

(where $c_{1-\alpha}$ is the smallest integer value such that $B(c_{1-\alpha}; n, p_0) \geq 1 - \alpha$)

$\Leftrightarrow t \geq c_{1-\alpha} + 1$

$$\mathbb{P}\{T \geq c_{1-\alpha} + 1 | p_0\} = 1 - \mathbb{P}\{T < c_{1-\alpha} + 1 | p_0\} = 1 - \underbrace{\mathbb{P}\{T \leq c_{1-\alpha} | p_0\}}_{\geq 1-\alpha}$$

$\leq \alpha$

$\square$

$\boxed{\text{Theorem 7.3.3.}}$ **(Tests of Hypotheses for Means of Poisson).** Let $x_1, \ldots, x_n$ be an observed random sample from $\mathcal{P}oi(\mu)$, and let $s = \sum x_i$

1. A conservative size $\alpha$ test of $H_0 : \mu \leq \mu_0$ versus $H_a : \mu > \mu_0$ is to reject $H_0$ if $1 - F(s - 1; n\mu_0) \leq \alpha$
2. A conservative size $\alpha$ test of $H_0 : \mu \geq \mu_0$ versus $H_a : \mu < \mu_0$ is to reject $H_0$ if $F(s; n\mu_0) \leq \alpha$

where $F(x; \mu)$ is the CDF of $\mathcal{P}oi(\mu)$.

## 7.4 Most Powerful Tests

The tests presented earlier were based on reasonable test statistics, but no rationale was provided to suggest that they are best in any sense. It would seem reasonable, if confronted with choosing between two or more tests of the same size, to select the one with the greatest chance of detecting when an alternative value is true. In other words, the strategy will be to select a test with maximum power for alternative values of the parameter. We will approach this problem by considering a method for deriving critical regions corresponding to tests that are most powerful tests of a given size for testing simple hypotheses.

Let $X = (X_1, \ldots, X_n)$ have joint PDF or PMF $f(x; \theta) = f(x_1, \ldots, x_n; \theta)$. Let $\mathcal{C} \subset \mathbb{R}^n$ be a critical region for testing $H_0$ v.s. $H_a$, i.e. reject $H_0$ if $X \in \mathcal{C}$. The power function corresponding to $\mathcal{C}$ is

$$\Pi_{\mathcal{C}}(\theta) = \mathbb{P}\{X \in \mathcal{C} \mid \theta\}$$

Definition 7.4.1. **(Most Powerful Test)**. Suppose $H_0 : \theta = \theta_0$ v.s. $H_a : \theta = \theta_1$. A test based on a critical region $\mathcal{C}^*$ is said to be a most powerful test of size $\alpha$ if

1. $\Pi_{\mathcal{C}^*}(\theta_0) = \alpha$
2. $\Pi_{\mathcal{C}^*}(\theta_1) \geq \Pi_{\mathcal{C}}(\theta_1)$ for any other critical region $\mathcal{C}$ for which $\Pi_{\mathcal{C}}(\theta_0) \leq \alpha$

Such a $\mathcal{C}^*$ is called a most powerful critical region of size $\alpha$.

How to find $\mathcal{C}^*$?

Theorem 7.4.1. **(Neyman-Pearson Lemma)**. Suppose that $X = (X_1, \ldots, X_n)$ have joint PDF or PMF $f(x;\theta)$ where $x = (x_1, \ldots, x_n)$. Let

$$\lambda(x;\theta_0,\theta_1) = \frac{f(x;\theta_0)}{f(x;\theta_1)}$$

and let $\mathcal{C}^*$ be the set

$$\mathcal{C}^* = \{x \mid \lambda(x;\theta_0,\theta_1) \leq k\}$$

where $k > 0$ is a constant such that

$$\mathbb{P}\{X \in \mathcal{C}^* \mid \theta_0\} = \mathbb{P}\{\lambda(X;\theta_0,\theta_1) \leq k \mid \theta_0\} = \alpha$$

Then $\mathcal{C}^*$ is a most powerful critical region of size $\alpha$ for testing $H_0 : \theta = \theta_0$ versus $H_a : \theta = \theta_1$

*Proof.* Let $\mathcal{C}$ be any other critical region such that $\Pi_{\mathcal{C}}(\theta_0) \leq \alpha$. Then for any $\theta$

$$\Pi_{\mathcal{C}^*}(\theta) - \Pi_{\mathcal{C}}(\theta) = \mathbb{P}\{X \in \mathcal{C}^* \mid \theta\} - \mathbb{P}\{X \in \mathcal{C} \mid \theta\}$$

$$\overset{1}{=} \mathbb{P}\{X \in \mathcal{C}^* \cap \mathcal{C} \mid \theta\} + \mathbb{P}\{X \in \mathcal{C}^* \cap \mathcal{C}^c \mid \theta\}$$
$$-\mathbb{P}\{X \in \mathcal{C} \cap \mathcal{C}^* \mid \theta\} - \mathbb{P}\{X \in \mathcal{C} \cap \mathcal{C}^{*c} \mid \theta\}$$
$$= \mathbb{P}\{X \in \mathcal{C}^* \cap \mathcal{C}^c \mid \theta\} - \mathbb{P}\{X \in \mathcal{C} \cap \mathcal{C}^{*c} \mid \theta\}$$

Taking $\theta = \theta_1$

$$\Pi_{\mathcal{C}^*}(\theta_1) - \Pi_{\mathcal{C}}(\theta_1) = \mathbb{P}\{X \in \mathcal{C}^* \cap \mathcal{C}^c \mid \theta_1\} - \mathbb{P}\{X \in \mathcal{C} \cap \mathcal{C}^{*c} \mid \theta_1\}$$

[1] $\mathcal{C}^* = (\mathcal{C}^* \cap \mathcal{C}) \cup (\mathcal{C}^* \cap \mathcal{C}^c)$ and $\mathcal{C} = (\mathcal{C} \cap \mathcal{C}^*) \cup (\mathcal{C} \cap \mathcal{C}^{*c})$

$$
\begin{aligned}
&= \int_{\mathcal{C}^* \cap \mathcal{C}^c} f(x;\theta_1)\,dx - \int_{\mathcal{C} \cap \mathcal{C}^{*c}} f(x;\theta_1)\,dx \\
&\geq \int_{\mathcal{C}^* \cap \mathcal{C}^c} \frac{1}{k} f(x;\theta_0)\,dx - \int_{\mathcal{C} \cap \mathcal{C}^{*c}} \frac{1}{k} f(x;\theta_0)\,dx \\
&= \frac{1}{k}\left(\mathbb{P}\left\{X \in \mathcal{C}^* \cap \mathcal{C}^c \mid \theta_0\right\} - \mathbb{P}\left\{X \in \mathcal{C} \cap \mathcal{C}^{*c} \mid \theta_0\right\}\right) \\
&= \frac{1}{k}\left(\Pi_{\mathcal{C}^*}(\theta_0) - \Pi_{\mathcal{C}}(\theta_0)\right) \geq \frac{1}{k}(\alpha - \alpha) = 0
\end{aligned}
$$

So $\Pi_{\mathcal{C}^*}(\theta_1) \geq \Pi_{\mathcal{C}}(\theta_1)$. □

*Example 7.4.1.*. We wish to determine the form of the most powerful test of $H_0 : p = p_0$ against $H_a : p = p_1 > p_0$ based on the statistic $S \sim \mathcal{B}in(n, p)$. We have

$$
\lambda(s; p_0, p_1) = \frac{\binom{n}{s} p_0^s (1-p_0)^{n-s}}{\binom{n}{s} p_1^s (1-p_1)^{n-s}}
$$

By Neyman-Pearson Lemma, optimal critical region is $\{s : \lambda(s; p_0, p_1) \leq k\} = \mathcal{C}^*$, i.e.

$$
\begin{aligned}
&\left\{s : \left(\frac{p_0}{p_1}\right)^s \left(\frac{1-p_1}{1-p_0}\right)^s \left(\frac{1-p_0}{1-p_1}\right)^n \leq k\right\} \\
&= \left\{s : s\ln\left(\frac{p_0}{p_1}\frac{1-p_1}{1-p_0}\right) + n\ln\left(\frac{1-p_0}{1-p_1}\right) \leq \ln k\right\} \\
&= \left\{s : s\ln\left(\frac{p_0}{p_1}\frac{1-p_1}{1-p_0}\right) \leq k^*\right\}
\end{aligned}
$$

Note that if $p_1 > p_0$, $0 < \frac{p_0}{p_1}\frac{1-p_1}{1-p_0} = \frac{1/p_1 - 1}{1/p_0 - 1} < 1$, $\ln\left(\frac{p_0}{p_1}\frac{1-p_1}{1-p_0}\right) < 0$, so $\mathcal{C}^* = \{s : s \geq k^{**}\}$. If $p_1 < p_0$, $\frac{p_0}{p_1}\frac{1-p_1}{1-p_0} = \frac{1/p_1 - 1}{1/p_0 - 1} > 1$, $\ln\left(\frac{p_0}{p_1}\frac{1-p_1}{1-p_0}\right) > 0$, so $\mathcal{C}^* = \{s : s \leq k^{**}\}$.

These tests are most powerful tests of their sizes provided we are willing to selection $\alpha$'s and $k^{**}$'s such that $\sum_{s=0}^{k^{**}} \binom{n}{s} p_0^s (1-p_0)^{n-s} = \alpha$ when $p_1 < p_0$ or $\sum_{s=k^{**}}^{n} \binom{n}{s} p_0^s (1-p_0)^{n-s} = \alpha$ when $p_1 > p_0$.