Lecture 16: P-values

Mathematical Statistics II, MATH 60062/70062

Thursday March 17, 2022

Reference: Casella & Berger, 8.3.4

Recap: Hypothesis testing

Given $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$, consider testing

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

A **hypothesis test** finds the rejection region.

After a hypothesis test is done, conclusions may be reported in terms of

- Size α of the test
- Decision to reject H_0 or accept H_0.

Recap: Hypothesis testing

Given $\boldsymbol{X} = (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta$, consider testing

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

A **hypothesis test** finds the rejection region.

After a hypothesis test is done, conclusions may be reported in terms of

- Size α of the test
- Decision to reject H_0 or accept H_0.

Another way of reporting the results is to report the value of a certain kind of test statistic called a **p-value**.

P-values

A **p-value** $p(\boldsymbol{X})$ is a test statistic satisfying $0 \leq p(\boldsymbol{x}) \leq 1$ for every sample point $\boldsymbol{x}$. Small values of $p(\boldsymbol{X})$ give evidence against H_0. A p-value is **valid** if, for every $\theta \in \Theta_0$ and every $0 \leq \alpha \leq 1$,

$$P_\theta(p(\boldsymbol{X}) \leq \alpha) \leq \alpha.$$

- If $p(\boldsymbol{X})$ is a valid p-value, it is easy to construct a level α test based on $p(\boldsymbol{X})$.
- The test that rejects H_0 if and only if $p(\boldsymbol{X}) \leq \alpha$ is a level α test. That is, $\phi(\boldsymbol{x}) = I(p(\boldsymbol{x}) \leq \alpha)$ is a level α test,

$$\sup_{\theta \in \Theta_0} E_\theta[\phi(\boldsymbol{X})] = \sup_{\theta \in \Theta_0} P_\theta(p(\boldsymbol{X}) \leq \alpha) \leq \alpha.$$

A p-value reports the results of a test on a more continuous scale.

P-value and test statistic

Let $W = W(\boldsymbol{X})$ be a test statistic such that large values of W give evidence against H_0. For each sample point $\boldsymbol{x} \in \mathcal{X}$, define

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

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

P-value and test statistic

Let $W = W(\boldsymbol{X})$ be a test statistic such that large values of W give evidence against H_0. For each sample point $\boldsymbol{x} \in \mathcal{X}$, define

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

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

Useful result: If X have CDF $F_X(x)$, then

$$Y = F_X(X) \sim \mathrm{Unif}(0,1),$$

because

$$P(Y \leq y) = P(F_X(X) \leq y) = P(X \leq F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y.$$

Fix $\theta \in \Theta_0$. Let $F_{-W}(w \mid \theta)$ denote the CDF of $-W = -W(\boldsymbol{X})$. Consider a test rejecting H_0 for large values of W, and define

$$\begin{aligned} p_\theta(\boldsymbol{x}) &= P_\theta(W(\boldsymbol{X}) \geq W(\boldsymbol{x})) \\ &= P_\theta(-W(\boldsymbol{X}) \leq -W(\boldsymbol{x})) = F_{-W}(-W(\boldsymbol{x}) \mid \theta). \end{aligned}$$

Thus, $p_\theta(\boldsymbol{X}) = F_{-W}(-W(\boldsymbol{X}) \mid \theta)$ is a $\mathrm{Unif}(0,1)$ random variable. That is, for every $0 \leq \alpha \leq 1$,

$$P_\theta(p_\theta(\boldsymbol{x}) \leq \alpha) \leq \alpha.$$

Now, note that

$$p(\boldsymbol{x}) = \sup_{\theta' \in \Theta_0} P_{\theta'}(W(\boldsymbol{X}) \geq W(\boldsymbol{x})) \geq P_\theta(W(\boldsymbol{X}) \geq W(\boldsymbol{x})) = p_\theta(\boldsymbol{x})$$

Therefore,

$$P_\theta(p(\boldsymbol{X}) \leq \alpha) \leq P_\theta(p_\theta(\boldsymbol{X}) \leq \alpha) \leq \alpha.$$

This is true for every $\theta \in \Theta_0$ and for every $0 \leq \alpha \leq 1$; $p(\boldsymbol{X})$ is a valid p-value.

Two-sided Normal p-value

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Both parameters are unknown. Set $\boldsymbol{\theta} = (\mu, \sigma^2)$. Consider testing

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

In Lecture 12, we showed that the LRT rejects H_0 for large values of

$$W = W(\boldsymbol{X}) = \left| \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \right|.$$

This is known as **one-sample two-sided t test**.

The null parameter space is

$$\Theta_0 = \{\boldsymbol{\theta} = (\mu, \sigma^2) : \mu = \mu_0, \sigma^2 > 0\}.$$

With observed value $w = W(\boldsymbol{x})$, the p-value for the test is

$$
\begin{aligned}
p(\boldsymbol{x}) &= \sup_{\boldsymbol{\theta}\in\Theta_0} P_{\boldsymbol{\theta}}(W(\boldsymbol{X}) \geq w) \\
&= \sup_{\boldsymbol{\theta}\in\Theta_0} P_{\boldsymbol{\theta}}\left(\left|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}\right| \geq w\right) \\
&= \sup_{\boldsymbol{\theta}\in\Theta_0} P_{\boldsymbol{\theta}}\left(\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \geq w \text{ or } \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq -w\right) \\
&= \sup_{\boldsymbol{\theta}\in\Theta_0} P_{\boldsymbol{\theta}}\left(\frac{\bar{X}-\mu}{S/\sqrt{n}} \geq w + \frac{\mu_0-\mu}{S/\sqrt{n}} \text{ or } \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq -w + \frac{\mu_0-\mu}{S/\sqrt{n}}\right) \\
&= P\left(T_{n-1} \geq w \text{ or } T_{n-1} \leq -w\right) \\
&= 2P(T_{n-1} \geq |w|).
\end{aligned}
$$

W(X): the distribution of hypothesis, w: observation
see wikipedia p-value.

One-sided Normal p-value

Suppose $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Both parameters are unknown. Set $\boldsymbol{\theta} = (\mu, \sigma^2)$. Consider testing

$$H_0 : \mu \leq \mu_0 \quad \text{versus} \quad H_1 : \mu > \mu_0.$$

The LRT rejects H_0 for large values of

$$W = W(\boldsymbol{X}) = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

The null parameter space is

$$\Theta_0 = \{\boldsymbol{\theta} = (\mu, \sigma^2) : \mu \leq \mu_0, \sigma^2 > 0\}.$$

With observed value $w = W(\boldsymbol{x})$, the p-value for the test is

$$\begin{aligned}
p(\boldsymbol{x}) &= \sup_{\boldsymbol{\theta} \in \Theta_0} P_{\boldsymbol{\theta}}(W(\boldsymbol{X}) \geq w) \\
&= \sup_{\boldsymbol{\theta} \in \Theta_0} P_{\boldsymbol{\theta}}\left(\frac{\bar{X} - \mu_0}{S/\sqrt{n}} \geq w\right) \\
&= \sup_{\boldsymbol{\theta} \in \Theta_0} P_{\boldsymbol{\theta}}\left(\frac{\bar{X} - \mu}{S/\sqrt{n}} \geq w + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&= \sup_{\mu \leq \mu_0} P_{\boldsymbol{\theta}}\left(T_{n-1} \geq w + \frac{\mu_0 - \mu}{S/\sqrt{n}}\right) \\
&= P(T_{n-1} \geq w).
\end{aligned}$$