Lecture 9: Hypothesis Testing

STATS 101: Foundations of Statistics

Linh Tran
linh@thetahat.ai

January 29, 2019

Announcements

- Next '*assignment*' is posted (due 1/29 @ 9:00am)
- A *Colab* script is available for today's class.

Outline

Hypothesis testing

- Review
- Hypothesis tests
- The power function
- Two-sided test & p-values
- Common statistical tests
    - t-test
    - F-test
    - Likelihood ratio test

Recall

Given $X_n, \ldots, X_n \overset{iid}{} P_0$, we can form an estimator

$$\hat{\theta}_n = \omega(X_1, \ldots, X_n) \tag{1}$$

of some underlying parameter on P_0.

- $\hat{\theta}$ has a sampling distribution
- We can try to find estimators that reach the CRLB
- We can opt for estimators that are ranges (i.e. CI's)

Hypothesis testing

We sometimes want to make claims about our parameter space Θ.

Example: Are male and female births are equally likely?

Hypothesis testing

We sometimes want to make claims about our parameter space Θ.
Example: Are male and female births are equally likely?

The common way of making statistical claims:

1. Separate our parameter space Θ into subsets, Θ_0 and Θ_1, where $\Theta_1 = \Theta \setminus \Theta_0$.
2. Pick a rule for choosing between Θ_0 and Θ_1.
3. Apply that rule on the realized data, i.e. $X_1 = x_1, ..., X_n = x_n$.

Note: A typical formulation of a hypothesis test is

$$H_0 : \theta \in \Theta_0 \text{ vs } H_1 : \theta \in \Theta_1 \tag{2}$$

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our claim: $\mu \neq 0$.

We can separate Θ in two ways:

1. $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$
2. $H_0 : \mu > 0$ vs. $H_1 : \mu = 0$

Question: which way should we choose?

Picking a rule

Let $X_1, ..., X_n \overset{iid}{\sim} P_\theta$ where θ is unknown.

Suppose our hypothesis is set up such that

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1 \tag{3}$$

A *testing procedure* is a rule for choosing between H_0 vs H_1.

Picking a rule

Let $X_1, ..., X_n \overset{iid}{\sim} P_\theta$ where θ is unknown.

Suppose our hypothesis is set up such that

$$H_0 : \theta \in \Theta_0 \text{ vs. } H_1 : \theta \in \Theta_1 \tag{3}$$

A *testing procedure* is a rule for choosing between H_0 vs H_1.

Common approach: define a *critical region* $\mathbf{C}$, where we "reject" H_0 if $X_1, ..., X_n \in \mathbf{C}$.

Example: Normal distribution

Let $X_1, ..., X_n \stackrel{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$

An adhoc rule: reject H_0 if $\bar{X}_n$ is far away '*enough*' from 0 (e.g. at least 2).
Our critical region is therefore

$$\mathbf{C} = \left\{ (X_1, ..., X_n) : \frac{\sum_{i=1}^{n} X_i}{n} > 2 \right\} \tag{4}$$

Some notes:

- $\frac{\sum_{i=1}^{n} X_i}{n}$ is referred to as the *test-statistic*
- The value 2 is referred to as the *critical value* (or threshold)
- We typically write hypothesis tests δ as $T(X_1, ..., X_n) > c$

Types of error

In truth:

- Our data will either come from Θ_0 or Θ_1.
- Our test will either pass or fail.

This gives us four possible realizations for δ:

	Retain H_0	Reject H_0
H_0 true	Correct	*Type I error*
H_1 true	*Type II error*	Correct

Colab link

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$

- If we reject H_0 when (in reality) $\mu = 0$, then we've committed a Type I error.
- If we accept H_0 when (in reality) $\mu > 0$, then we've commited a Type II error.

Common approach: Select a critical value c such that

$$\sup_{\theta \in \Theta_0} P_\theta(\text{Type I Error}) \leq 0.05 \tag{5}$$

Colab link

The Power Function

Given a hypothesis test δ with critical region $\mathbf{C}$, the *power function* $\pi(\theta|\delta) : \Theta \to [0, 1]$ is

$$\pi(\theta|\delta) = P\left((X_1, ..., X_n) \in \mathbf{C}\right) = P_\theta(\text{reject } H_0) : \theta \in \Theta \qquad (6)$$

n.b. $\pi(\theta|\delta)$ summarizes both Type I and II errors, e.g.

- $P_\theta(\text{Type I Error}) = \pi(\theta|\delta) : \theta \in \Theta_0$
- $P_\theta(\text{Type II Error}) = 1 - \pi(\theta|\delta) : \theta \in \Theta_1$

Remark: The ideal $\pi(\theta|\delta)$ will have

- $\pi(\theta|\delta) = 0 \,\forall\, \theta \in \Theta_0$
- $\pi(\theta|\delta) = 1 \,\forall\, \theta \in \Theta_1$

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$
Our hypothesis test: reject H_0 if $\bar{X} \geq 2$

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$
Our hypothesis test: reject H_0 if $\bar{X} \geq 2$

Since $\bar{X} \sim N\left(\mu, \frac{1}{n}\right)$, we have that

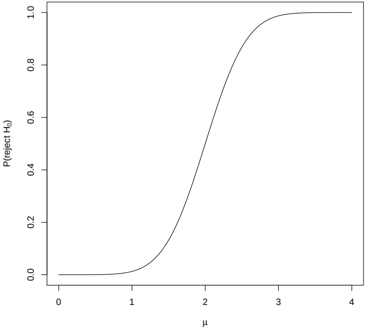
$$
\begin{aligned}
\pi(\theta|\delta) &= P_\mu(\text{reject } H_0) && (7)\\
&= P_\mu\left(\bar{X} > 2\right) && (8)\\
&= 1 - P_\mu\left(\bar{X} \leq 2\right) && (9)\\
&= 1 - F_{\bar{X},\mu}(2) && (10)\\
&= 1 - \Phi\left(\frac{2-\mu}{\sqrt{1/n}}\right) && (11)
\end{aligned}
$$

Colab link

Example: Normal distribution

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, 1) : \mu \geq 0$.

Our hypothesis: $H_0 : \mu = 0$ vs. $H_1 : \mu > 0$
Our hypothesis test: reject H_0 if $\bar{X} \geq 2$
Our power function: $1 - \Phi\left(\frac{2-\mu}{\sqrt{1/n}}\right)$



Two sided tests

Our working example is a *one-sided* hypothesis

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu > 0 \tag{12}$$

Problem: Many times, we don't have enough information to restrict our parameter space.

Two sided tests

Our working example is a *one-sided* hypothesis

$$H_0 : \mu = 0 \text{ vs. } H_1 : \mu > 0 \tag{12}$$

Problem: Many times, we don't have enough information to restrict our parameter space.

Approach: Construct hypotheses only against specific values in Θ, e.g.

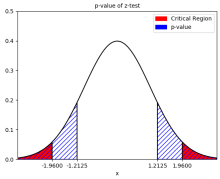
$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu \neq \mu_0 \tag{13}$$

The p-value

Recall: hypothesis tests δ can be written as $T(X_1, ..., X_n) > c$

- $T(X_1, ..., X_n)$ is random and has a sampling distribution.
- We can restrict calculation of T to $\theta \in \Theta_0$.
- The resulting statistic has pdf $f_{T,\theta}$.
- Let $t = T(X_1 = x_1, ..., X_n = x_n)$. The *p-value* is

$$P(|T(X_1, ..., X_n)| \geq |t|) \tag{14}$$



Example of a p-value for a 2-sided test.

The t-test

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

Our hypothesis: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$

The t-test

Let $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

Our hypothesis: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$
We can form a test statistic

$$U = n^{1/2} \frac{\bar{X}_n - \mu_0}{\sigma_n'} \tag{15}$$

where $\sigma_n' = \left(\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \right)^{1/2}$.

- When $\mu = \mu_0$, then $U \sim t(n-1)$.
- We therefore set up δ such that we reject H_0 if $|U| \geq T_{n-1}^{-1}(1 - \alpha_0/2)$

n.b. T is the quantile function of the $t - distribution$ with $n - 1$ degrees of freedom. *Colab link*

The t-test

Many variations exist for the t-test, e.g.

- **Two sample t-tests**: two different samples are drawn and we want to compare them to each other
- **Paired t-tests**: Matched units are compared (e.g. a before and after study on the same patients)

The t-test

Many variations exist for the t-test, e.g.

- **Two sample t-tests**: two different samples are drawn and we want to compare them to each other
- **Paired t-tests**: Matched units are compared (e.g. a before and after study on the same patients)

Many variations exist within each type of t-test as well, e.g. for the two sample t-test we have

- Equal sample sizes, equal variances
- Unequal sample sizes, equal variance
- Unequal sample sizes, unequal variances

The F-test

Commonly used for (i) testing variances of normal distributions, and (ii) testing means of more than two distributions.

The F-test

Commonly used for (i) testing variances of normal distributions, and (ii) testing means of more than two distributions.

Let $W \sim \chi^2(n)$ and $Y \sim \chi^2(m)$. Then

$$X = \frac{Y/m}{W/n} = \frac{nY}{mW} \sim F(m, n) \tag{16}$$

follows a *F-distribution* with m and n degrees of freedom.

Some notes:

- $F(m, n) \neq F(n, m)$
- If $X \sim F(m, n)$, then $1/X \sim F(n, m)$
- If $X \sim t(n)$, then $X^2 \sim F(1, n)$

The F-test

Let $X_1, ..., X_m \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, ..., Y_n \overset{iid}{\sim} N(\mu_1, \sigma_2^2)$.

Suppose we want to test: $H_0 : \sigma_1^2 = \sigma_2^2$ vs $H_1 : \sigma_1^2 \neq \sigma_2^2$.

We can form a test statistic

$$V = \frac{S_X^2/(m-1)}{S_Y^2/(n-1)} \tag{17}$$

where $S_X^2/(m-1)$ and $S_Y^2/(n-1)$ are estimators of σ_1^2 and σ_2^2, respectively.

- $V \sim F(m-1, n-1)$ when $\sigma_1^2 = \sigma_2^2$

Colab link

The Likelihood ratio test

Let $X_1, ..., X_n \overset{iid}{\sim} P_\theta$

Suppose we want to test: $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$.

Recall that the likelihood of our data is

$$L(\theta|x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i|\theta) \tag{18}$$

We can use this to form a "likelihood ratio" statistic, i.e.

$$\begin{aligned}\Gamma(x) &= -2\log\left[\frac{\sup_{\theta\in\Theta_0} L(\theta|x_1, ..., x_n)}{\sup_{\theta\in\Theta} L(\theta|x_1, ..., x_n)}\right] &(19)\\ &= 2\left[\ell(\hat{\theta}_{MLE}|x_1, ..., x_n) - \sup_{\theta\in\Theta_0} \ell(\theta|x_1, ..., x_n)\right] &(20)\end{aligned}$$

It can be shown that $\Gamma(x) \sim \chi^2(p)$. *Colab link*

References

- DeGroot & Schervish Chapters 9.1, 9.2, 9.4-9.7