Hypothesis Testing

A statistical hypothesis is a statement about a parameter (or a statistical functional in nonparametric models).

A hypothesis test partitions the parameter space Θ into two disjoint sets Θ_0 and Θ_1, and produces a decision rule for choosing between

$$H_0 : \theta \in \Theta_0 \quad \text{and} \quad H_1 : \theta \in \Theta_1$$

H_0 is called the null hypothesis and H_1 is called the alternative hypothesis. The possible choices are

- Retain H_0
- Reject H_0 and accept H_1

The decision of whether to reject H_0 is determined by whether the sample $X = (X_1, \ldots, X_n)$ falls into a predefined rejection region R.

Usually, the rejection region R has the form

$$R = \{x_1, \ldots, x_n : T(x_1, \ldots, x_n) > c\}$$

where T is called a test statistic and c is called a critical value.

The idea is to construct R so that the probability of the data falling into it when H_0 is true is small.

Example: Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, and let $\hat{\mu}_n$ and $\hat{\sigma}_n^2$ be the MLEs. If $H_0 : \mu = 0$, one test statistic we might consider is $T = |\hat{\mu}_n / \hat{\sigma}_n|$, reasoning that if H_0 is true, T will tend to be small.

Note: c is just a placeholder. It usually will depend on n and/or our choice of Θ_0 and Θ_1.

We evaluate a test using its power function. This is defined by

$$\beta(\theta) = P_\theta(X \in R)$$

Ideally, we would like $\beta(\theta)$ to be 0 when $\theta \in \Theta_0$ and 1 when $\theta \in \Theta_1$, but that is typically impossible to achieve.

Qualitatively, a good test has small $\beta(\theta)$ when $\theta \in \Theta_0$ and large $\beta(\theta)$ when $\theta \in \Theta_1$.

However, there are typically tradeoffs between the two, so that the researcher must choose between tests based on what kind of error probabilities he/she is willing to accept. More on this shortly.

Example 1: Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, where σ^2 is known. Consider testing $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$, using rejection region

$$R = \{x_1, \ldots, x_n : |\bar{X}_n| > c\}$$

Find and plot $\beta(\mu)$.

Example 2: Let $X \sim Bin(5, p)$. Consider testing $H_0 : p \leq 1/2$ versus $H_1 : p > 1/2$. Consider two different rejection regions:

$$\begin{aligned} R_1 &= \{x : x = 5\} \\ R_2 &= \{x : x \geq 3\} \end{aligned}$$

Plot and compare the corresponding power functions $\beta_1(p)$ and $\beta_2(p)$.

To make the problem of comparing tests better defined, we restrict ourselves to tests of a certain level, and then we try to find a test within that class that has large $\beta(\theta)$ for $\theta \in \Theta_1$.

A test is said to have level α if its size is less than or equal to α. The size of a test is defined to be

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

In words, the size of a test is the largest probability of rejecting H_0 when H_0 is true. This is called a Type I error.

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

Since H_0 usually represents a "default" hypothesis, first guaranteeing that the probability of this is small is a scientifically conservative strategy.

Continuation of Example 1: Find the size of the test as a function of c (and possibly other things). What should c be to produce a size α test?

Continuation of Example 2: Consider a rejection region of the form $R = \{x : x \geq c\}$.

- What values of c do we need to consider?

- For each of these, find the size of the corresponding test.

- What c should we choose if we want a probability of Type I error of no more than 10%?

Theorem: Correspondence between point-null tests and confidence sets

1. For each $\theta_0 \in \Theta$, let $A(\theta_0) = R^C(\theta_0)$ be the acceptance region of a level α test of $H_0 : \theta = \theta_0$. For each possible sample $x_1, \ldots, x_n$, define a set $C_n(x_1, \ldots, x_n)$ in Θ by

$$C_n(x_1, \ldots, x_n) = \{\theta_0 : x_1, \ldots, x_n \in A(\theta_0)\}$$

   Then $C_n(X_1, \ldots, X_n)$ is a $1 - \alpha$ confidence set for θ_0. That is, $P_{\theta_0}(\theta_0 \in C_n(X_1, \ldots, X_n)) \geq 1 - \alpha$.

2. Let $C_n(X_1, \ldots, X_n)$ be a $1 - \alpha$ confidence set. For any $\theta_0 \in \Theta$, define

$$A(\theta_0) = \{x_1, \ldots, x_n : \theta_0 \in C_n(x_1, \ldots, x_n)\}$$

   Then $R(\theta_0) = A^C(\theta_0)$ is the rejection region of a level α test of $H_0 : \theta = \theta_0$.

Practically speaking, this means that if we already have a $1-\alpha$ confidence interval for θ and we want to test $H_0 : \theta = \theta_0$, a level α test is just to reject H_0 if θ_0 falls outside the interval.

Often we can get *approximate* $1-\alpha$ confidence intervals using an estimator of θ that is asymptotically normal.

Wald Test:

Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$. Let $\hat{\theta}_n$ be an estimator such that $(\hat{\theta}_n - \theta_0)/\widehat{se}(\hat{\theta}_n) \xrightarrow{D} N(0,1)$. The size α Wald test rejects H_0 when $T > z_{\alpha/2}$, where

$$T = \left| \frac{\hat{\theta}_n - \theta_0}{\widehat{se}(\hat{\theta}_n)} \right|$$

We can show that asymptotically, the Wald test has size α, and that it is obtained by inverting the approximate $1-\alpha$ normal-based CI for θ.

Note that this method is quite general; we just need asymptotic normality.

For example

- Consider a multi-parameter problem in which $\hat{\theta}_n$ is the MLE and g is an invertible function. Then we can form a Wald test based on

$$\frac{g(\hat{\theta}) - g(\theta_0)}{\widehat{se}(g(\hat{\theta}_n))} \xrightarrow{D} N(0,1)$$

  where $\widehat{se}(g(\hat{\theta}_n))$ is found using the Delta method.

- Consider the case that $\theta = T(F)$ for some unknown distribution F. If T is a linear functional, the plug-in estimator is a mean of iid random variables, so we can use the CLT. In the case of a nonlinear functional, we could approximate $\widehat{se}(\hat{\theta}_n)$ using the bootstrap.

Examples

- Consider again $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, where σ^2 is known. Show that the size α Wald test for $H_0 : \mu = 0$ produces a rejection region as in Example 1 above. (Actually the size is exactly α in this case).

- Now suppose that σ^2 is unknown. Construct a size α Wald test for $H_0 : \mu = 0$.

- Suppose that $X \sim Bin(m, p_1)$ and $X \sim Bin(n, p_2)$. Construct a size α Wald test for $H_0 : p_1 = p_2$.

- Let $F(u, v)$ be the joint distribution of two r.v. U and V. Let $\theta = T(F) = \rho(U, V)$, where ρ denotes the correlation. Describe how to construct a size α Wald test for $H_0 : \rho = 0$ using the plug-in estimator and the bootstrap.