Notes on Hypothesis Testing

A random sample \( X = (X_1, \ldots, X_n) \) is observed, with joint pmf/pdf \( f_{\theta}(x_1, \ldots, x_n) \). The values \( x = (x_1, \ldots, x_n) \) of \( X \) lie in some sample space \( \mathcal{X} \). The parameter (parameter tuple) \( \theta \) is unknown, varies in some parameter space \( \Theta \).

Partition \( \Theta \) into two disjoint subsets \( H_0 \) and \( H_1 \). \( H_0 \) is the null hypothesis and \( H_1 \) is the alternative hypothesis.

A hypothesis test is a partition of \( \mathcal{X} \) into two disjoint subsets, \( A \) and \( R \). \( A \) is the acceptance region and \( R \) is the rejection region.

If \( X = x \in A \) is observed, we accept \( H_0 \), and if \( X = x \in R \) is observed, we reject \( H_0 \).

If \( \theta \in H_0 \), then \( P_{\theta}(X \in R) = P_{\theta}(\text{rejection}) \) is the type I error probability.

If \( \theta \in H_1 \), then \( P_{\theta}(X \in A) = P_{\theta}(\text{acceptance}) \) is the type II error probability.

We want to choose \( A \) and \( R \) so that both error probabilities are small.

Example  A coin of unknown bias is tossed 10 times.

\[
\begin{align*}
X & \quad \text{number of heads} \\
\theta & \quad \text{probability of heads} \\
f_{\theta}(x) &= \binom{10}{x} \theta^x (1 - \theta)^{10-x} \\
H_0 & : \theta = 0.5 \\
A & = \{4, 5, 6\} \\
R & = \{0, 1, 2, 3, 7, 8, 9, 10\}
\end{align*}
\]

Type I error probability, \( \theta = 0.5 \)

\[
P_{0.5}(R) = \sum_{x=1}^{10} \binom{10}{x} \frac{1}{2^{10}}
\]

\[
= 2 \left( \binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3} \right) \frac{1}{2^{10}}
\]

\[
= \frac{11}{32} \approx 0.344
\]

Type II error probability, \( \theta \neq 0.5 \)

\[
P_{\theta}(A) = \sum_{x=4}^{6} \binom{10}{x} \theta^x (1 - \theta)^{10-x}
\]

\[
= \binom{10}{4} \theta^4 (1 - \theta)^6 + \binom{10}{5} \theta^5 (1 - \theta)^5 + \binom{10}{6} \theta^6 (1 - \theta)^4
\]

\[
= \frac{42 \theta^4 (1 - \theta)^6 (5 - 4\theta(1 - \theta))}{32} < \frac{21}{32} \approx 0.656
\]

The significance level of a hypothesis test is the highest type I error probability,

\[\alpha = \sup_{\theta \in H_0} P_{\theta}(X \in R).\]

An upper bound on \( \alpha \) is usually imposed.

To define a decision rule, we choose the form of the test. For instance, accept \( H_0 \) whenever \( c_1 < T(X) < c_2 \) or reject \( H_0 \) whenever \( T(X) > c \).

Here \( T \) is a function of the observed data, a test statistic. The constants \( c_1, c_2, c \) are critical values.
Suppose that $X_i$ are independently sampled from $\mathcal{N}(\theta, \sigma^2)$, and we want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$.

Let us choose the test statistic to be $T = \bar{X}$ and the form of the test to be accept $H_0$ whenever $c_1 < T < c_2$.

This is logical since any observation $\bar{X} \approx \theta_0$ is in favor of $H_0$.

The null hypothesis in this case is simple as it consists of only one value of $\theta$.

The alternative hypothesis is composite.

We have

$$\alpha = P_{\theta_0}(T \leq c_1 \text{ or } T \geq c_2).$$

The distribution of $T$ that is used to determine $\alpha$ is null distribution of $T$.

The tail probabilities $P_{\theta_0}(T \leq c_1)$ and $P_{\theta_0}(T \geq c_2)$ should add up to $\alpha$.

One way to achieve this is to make them of equal weight,

$$P_{\theta_0}(T \leq c_1) = P_{\theta_0}(T \geq c_2) = \alpha/2,$$

so that $c_1$ and $c_2$ are the lower and upper $\alpha/2$-quantiles of null distribution of $T$.

Since null distribution of $T$ is $\mathcal{N}(\theta_0, \sigma^2/n)$, we have

$$c_1 = \theta_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \quad \text{and} \quad c_2 = \theta_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the upper $\alpha/2$-quantile for the standard normal distribution.

For instance, if $\alpha = 0.05$, then our test is defined by the decision rule accept $H_0$ whenever $\theta_0 - 1.96 \frac{\sigma}{\sqrt{n}} < \bar{X} < \theta_0 + 1.96 \frac{\sigma}{\sqrt{n}}$.

The same test can be written in terms of the statistic

$$Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}},$$

which has null distribution $\mathcal{N}(0, 1):$

accept $H_0$ whenever $|Z| < 1.96$.

Note that the calculation of critical values $c_1, c_2$ involves only type I error probability.

When $\theta$ is a single variable, the power function of a test with rejection region $\mathcal{R}$ is

$$\text{Power}(\theta) = P_{\theta}(X \in \mathcal{R}) = P_{\theta}(\text{rejection}).$$
The plot of the power function is a power curve. In our case,

\[
\text{Power}(\theta) = \Phi \left( \frac{c_1 - \theta}{\sigma/\sqrt{n}} \right) + \Phi \left( -\frac{c_2 - \theta}{\sigma/\sqrt{n}} \right)
\]

\[
= \Phi \left( -z_\alpha + \frac{\theta_0 - \theta}{\sigma/\sqrt{n}} \right) + \Phi \left( -z_\alpha + \frac{\theta - \theta_0}{\sigma/\sqrt{n}} \right)
\]

where \( \Phi \) is the standard normal cdf. Observe that

- \( \text{Power}(\theta) \) is symmetric about \( \theta = \theta_0 \),
- \( \text{Power}(\theta_0) = \alpha \) is the minimum of \( \text{Power}(\theta) \),
- \( \text{Power}(\theta) \nearrow 1 \) as \( \theta \to \pm \infty \).

\[\text{Sampling from normal data: one-tailed } z\text{-test for mean}\]

Suppose as above that \( X_i \) are independently sampled from \( \mathcal{N}(\theta, \sigma^2) \), and we want to test

\[H_0 : \theta \leq \theta_0 \quad \text{vs} \quad H_1 : \theta > \theta_0.\]

Here both hypotheses are composite.

Choose the test statistic to be \( T = \bar{X} \) and the form of the test to be

\[\text{reject } H_0 \quad \text{whenever} \quad T > c.\]

Since the density function of \( T \) just slides to the right as \( \theta \) increases, we have

\[\alpha = \max_{\theta \leq \theta_0} P_{\theta}(T > c) = P_{\theta_0}(T > c) = P \left( Z > \frac{c - \theta_0}{\sigma/\sqrt{n}} \right) = \Phi \left( -\frac{c - \theta_0}{\sigma/\sqrt{n}} \right).\]

So

\[c = \theta_0 + z_\alpha \sigma/\sqrt{n},\]

where \( z_\alpha \) is the upper \( \alpha \)-quantile for the standard normal distribution.

The calculation of the critical value \( c \) involves only type I error probability.

The power function is

\[\text{Power}(\theta) = P_{\theta}(R) = \Phi \left( -\frac{c - \theta}{\sigma/\sqrt{n}} \right) = \Phi \left( -z_\alpha + \frac{\theta - \theta_0}{\sigma/\sqrt{n}} \right).\]
Observe that
\[
\text{Power (}\theta\text{)} \quad \text{is monotonically increasing,}
\]
\[
\text{Power (}\theta_0\text{)} = \alpha,
\]
\[
\text{Power (}\theta\text{)} \searrow 0 \quad \text{as} \quad \theta \to -\infty,
\]
\[
\text{Power (}\theta\text{)} \nearrow 1 \quad \text{as} \quad \theta \to \infty.
\]

If \( \alpha = 0.05 \), then our test is defined by the decision rule
\[
\text{reject } H_0 \quad \text{whenever} \quad \bar{X} > \theta_0 + 1.645 \frac{\sigma}{\sqrt{n}}.
\]
In terms of the statistic
\[
Z = \frac{\bar{X} - \theta_0}{\sigma/\sqrt{n}},
\]
which has null distribution \( \mathcal{N}(0,1) \), the same test can be written as
\[
\text{reject } H_0 \quad \text{whenever} \quad Z > 1.645.
\]

**Sampling from normal data: \( t \)-test for mean**

Suppose now that \( X_i \) are independently sampled from \( \mathcal{N}(\mu, \sigma^2) \), where \( \mu \) and \( \sigma^2 \) are both unknown, \( \theta = (\mu, \sigma^2) \). Consider the hypotheses
\[
H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0.
\]
Instead of using the statistic
\[
Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}},
\]
which applies when \( \sigma \) is known, we approximate \( \sigma \) by the sample standard deviation
\[
S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2}
\]
and use the Student \( t \) statistic
\[
T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}},
\]
which has null distribution \( t_{n-1} \), Student’s \( t \) distribution with \( n - 1 \) degrees of freedom. So the decision rule can be, for instance,
\[
\text{accept } H_0 \quad \text{whenever} \quad |T| < t_{n-1, \alpha/2},
\]
where $t_{n-1, \alpha/2}$ is the upper $\alpha/2$–quantile for $t_{n-1}$.

Thus a $t$-test is a refinement of the idea of testing $\mu = \mu_0$ in the case where $\sigma$ is unknown. We have $t_{n-1, \alpha} > z_{\alpha}$ (not by much for large $n$), in compensation for the increased variability of the test statistic.

If $\mu \neq \mu_0$, then

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \frac{\mathcal{N}(\delta, 1)}{\sqrt{\chi^2_{n-1}/n}}$$

where

$$\delta = \frac{\mu - \mu_0}{\sigma/\sqrt{n}}$$

is a non-centrality parameter. The distribution of $T$ when the null hypothesis is not true is called a non-central $t$ distribution. We may consider the rejection probability

$$P_{\mu, \sigma^2}^{\text{rejection}}$$

as a function of sole parameter, Power ($\delta$).

**References**

G. Lorden, *Notes on Hypothesis Testing*.

J. Rice, *Statistics and Data Analysis*.