The Null and Alternative Hypothesis

Consider a statistical problem involving unknown param $\theta \in \Omega$.

Suppose param space Ω is partitioned into disjoint subsets Ω_0, Ω_1.

We want to decide whether $\theta \in \Omega_0$ or $\theta \in \Omega_1$.

We call $H_0: \theta \in \Omega_0$ null hypothesis, and $H_1: \theta \in \Omega_1$ alternative hypo.

Deciding between H_0, H_1 is called a problem of testing hypo.

Procedure of observing data and deciding H_0/H_1 is called a test.

If we decide H_1 is true, we are said to reject H_0.

$\quad$ . . . H_0 . . . not reject H_0.

Simple/Composite Hypo: Ω_i contains single/multiple value of θ.

Critical Region S_1: part of sample space where H_0 is rejected.

Test Statistics and Reject Region (for statistics based tests)

Let $\vec{X}$ be a random sample from a distr indexed by θ.

Test statistics: a statistics $T = t(\vec{X})$ used in test procedure

Reject region: $R \subseteq \mathbb{R}$ used in test procedure

Test procedure: Reject H_0 if $T \in R$ $\quad (R \subseteq \mathbb{R})$

Power Function

Let δ be a test procedure. It's power function $\pi(\theta|\delta)$ is defined as the probability of rejection for θ given δ.

For critical rejion based δ:

$$\pi(\theta|\delta) = \Pr(\vec{X} \in S_1 | \theta)$$

For test statistics based δ:

$$\pi(\theta|\delta) = \Pr(T \in R | \theta)$$

Type I/II Error

Type I: reject H_0 when it's true

" II: not to reject H_0 when it's false

Intuition: H_0 by convention is the base case that we trust by default, unless evidence suggest otherwise.
Thus, Type I error is more severe than Type II.
That's why we put Type I first.

Which Proposition should be chosen for H_0?

Generally, for a proposition of interest, we can make it H_0 or H_1.
To make the decision, there are two views that we may consider:

- Base case / New proposition View

  H_0: the base case that we should believe by default

  H_1: rare event that can be believed only if significant evidence is present.

- Type I/II error tradeoff

  Type I err: less palatable error that should be put tighter control on.

  Choose H_0 such that type I fits its definition.

Evaluating Quality of Tests

Strict tests will make more type II errors, loose test ... type I ...

A popular method to strike a balance:

Choose a number α_0 as threshold for type I, then among all satisfying tests, maximize π for $\theta \in \Omega_1$.

Symbolically, we require $\pi(\theta | \delta) \le \alpha_0$ for $\forall \theta \in \Omega_0$.

Interpretation: For more important errors (type I), specify a threshold for quality assurance. For the rest, do our best.

Note. there are many possible criterion, e.g. optimize a linear combination of 2 types of errors.

Level / Size of Test

A test satisfys

$$\pi(\theta | \delta) \le \alpha_0 \text{ for } \forall \theta \in \Omega_0$$

is called an level α_0 test. Or, the test has significance α_0.

(α_0: strictness, small $\alpha_0 \Rightarrow$ less type I err)

Size $\alpha(\delta)$ of a test δ is:

$$\alpha(\delta) = \sup_{\theta \in \Omega_0} \pi(\theta | \delta)$$

- Test δ is an α_0 test $\iff$ $\alpha(\delta) \le \alpha_0$

Making a Test Have Specific Significance Level

Suppose we test $H_0: \theta \in \Omega_0$, $H_1: \theta \in \Omega_1$

with test statistic T, test procedure: reject H_0 if $T \ge c$

Now, to make δ have sig α_0, or,

$$\Pr(T \ge c \mid \theta) \le \alpha_0 \quad \text{for } \forall \theta \in \Omega_0$$

we chose c s.t. $\sup_{\theta \in \Omega_0} \Pr(T \ge c \mid \theta) = \alpha_0$

p Value of a Test and Observed Data

Given a test procedure and observed data,

p-value is the smallest sig level α_0, s.t.

H_0 is rejected given observed data.

Interpretation: we are testing whether nonconventional case H_1 is true.

For any observed data, if sig level is loose (large) enough, we will be able to accept H_1.

If evidence of H_1 is significant, even strict tests will accept H_1. Thus the sig level (strictness) at phase transition can be used to characterize significance of evidence.

Calculating p-value

For tests of form "reject H_0 if $T \ge c$", and observed $T = t$,

p-value can be found by:

Let δ_t be the test "reject H_0 if $T \geq t$",

p-value is the observed size of δ_t, or,

$$p = \sup_{\theta \in \Omega_0} \pi(\theta | \delta_t) = \sup_{\theta \in \Omega_0} Pr(T \geq t | \theta)$$

Note, there are other forms of test that p-value calculation is more complex.

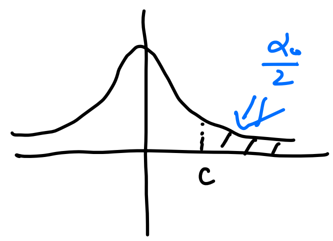
Testing Hypo about Mean of Normal w/ Known Var

$H_0: \mu = \mu_0 \qquad H_1: \mu \neq \mu_0$

Test statistics: sample, calculate sample mean $\hat{\mu}$, $T = |\hat{\mu} - \mu|$.

Test procedure: reject H_0 if $T \geq c$

Note, $\hat{\mu} \sim N(\mu, \sigma^2/n)$



To make the test have sig level α_0,

we need $\quad Pr(|\hat{\mu} - \mu_0| \geq c) \leq \alpha_0$

Then, standardized c, $c' = \Phi^{-1}(1 - \frac{\alpha_0}{2})$.

$$c = \frac{\sigma}{\sqrt{n}}(c' + \mu_0)$$

Note, when testing mean of normal, it's conventional to use statistic

$$Z = \frac{\sqrt{n}}{\sigma}(\hat{\mu} - \mu_0)$$

s.t. H_0 is rejected if $|Z| \geq \Phi^{-1}(1 - \frac{\alpha_0}{2})$.

t-test: test procedure when μ, σ^2 are unknown

$H_0: \mu \leq \mu_0 \qquad H_1: \mu > \mu_0$

Test Statistics: $\quad U = \frac{\sqrt{n}}{\sigma'}(\hat{\mu} - \mu_0)$

Test Procedure: reject H_0 if $U \geq c$, $c = \Phi^{-1}(1 - \alpha_0)$

Similarly, for $H_0: \mu \le \mu_0$, δ: reject H_0 if $U \le c$

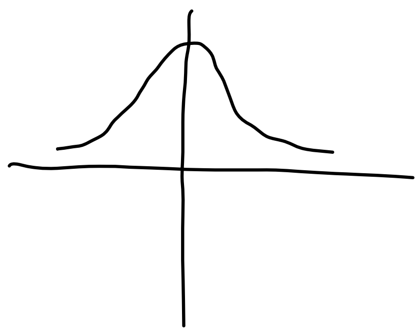
Property of t-test

Consider testing about normal distr w/ μ, σ^2

For t-test statistics U, c be $1-\alpha_0$ quantile of t distr.

δ be the t-test : reject H_0 if $U \ge c$

Power function $\pi(\mu, \sigma^2 | \delta)$ satisfies:



- $\pi(\mu, \sigma^2 | \delta) = \alpha_0$ when $\mu = \mu_0$
- $\pi(\mu, \sigma^2 | \delta) < \alpha_0$ when $\mu < \mu_0$
- $\pi(\mu, \sigma^2 | \delta) > \alpha_0$ when $\mu > \mu_0$
- $\pi \to 0$ as $\mu \to -\infty$, $\pi \to 1$ as $\mu \to \infty$

Furthermore, test δ has size α_0.

p-value for t-tests

Let u be observed value of U. T_{n-1} be cdf of T distr w/ $n-1$ dof.

p-value for $H_0: \mu \le \mu_0$ type of test is $1 - T_{n-1}^{-1}(u)$

$\ge$ $T_{n-1}^{-1}(u)$

Two-sample t-test

(Compare means of two normal)

Suppose $\{X_i\}_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\{Y_i\}_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$,

$H_0: \mu_1 \le \mu_2$, $H_1: \mu_1 > \mu_2$

Two-sample t statistic:

Define : $\bar{X}_m = \frac{1}{m}\sum X_i$, $\bar{Y}_n = \frac{1}{n}\sum Y_i$

$SXX = \sum (X_i - \bar{X}_m)^2$, $SYY = \cdots$

Test statistic

$$U = \frac{\sqrt{m+n-2}\,(\bar{X}_m + \bar{Y}_n)}{\sqrt{\frac{1}{m}+\frac{1}{n}}\,\sqrt{SXX+SYY}}$$

For $\mu_1 = \mu_2, \forall \sigma_i^2$, $U \sim$ t distr w/ $m+n-2$ dof

Properties of two-sample t-test resembles that of t-test.

e.g., level α_0 test is : reject H_0 if $U \ge T^{-1}_{m+n-2}(1-\alpha_0)$

F distr

Useful when testing hypo about var of two normal.

Definition

Y, W independent RV, $Y \sim \chi^2_m$, $W \sim \chi^2_n$,

$$X = \frac{Y/m}{W/n} \sim F \text{ distr w/ } m, n \text{ dof}$$

Properties

- If $X \sim F_{m,n}$, $1/X \sim F_{n,m}$
- If $Y \sim T_n$, $Y^2 \sim F_{1,n}$

F test (comparing var of two normal)

$\{X_i\}_m \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $\{Y_i\}_n \sim \mathcal{N}(\mu_2, \sigma_2^2)$,

$H_0 : \sigma_1^2 \leq \sigma_2^2$, $H_1 : \sigma_1^2 > \sigma_2^2$

Test statistics: $V = \dfrac{\sigma_x'^2}{\sigma_y'^2} = \dfrac{SXX/(m-1)}{SYY/(n-1)}$

Test procedure: reject H_0 if $V \geq c$.

$\dfrac{m\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{m-1}$

Consider RV $V' = \dfrac{\boxed{SXX/\sigma_1^2}/(m-1)}{\boxed{SYY/\sigma_2^2}/(n-1)}$

$\sim \chi^2_{n-1}$

Thus, if $\sigma_1^2 = \sigma_2^2$, $V = V' \sim F_{m,n}$.

Choosing c Given α_0

Denote cdf of F distr as $G(\cdot)$, $c = G^{-1}_{m-1,n-1}(1-\alpha_0)$

p-value of F test

Given observed test statistics $V = v$, $p = 1 - G^{-1}_{m-1,n-1}(v)$

(All similar to t-test)