PRE1: APPLIED STATISTICS
More on hypothesis testing

Marie-Anne Poursat

AI Master
Université Paris-Saclay

October 13th 2022

The testing procedure

1. Specify the model and the hypotheses H_0 and H_1
2. Find an appropriate test statistic T :
   - T is calculable (does not depend on unknown parameters)
   - the distribution of T under H_0 is known
3. Find the form of the rejection region (*look at* H_1 *!*)
4. Calculate the p-value and make your decision
   - if you reject H_0, the risk of an incorrect decision is less than α (*type I error*)
   - if you don't reject H_0, the risk of an incorrect decision is usually unknown(*type II error*)

The **power of the test** is defined as the probability of rejecting H_0 when H_1 is true $=$ 1 - type II error.

The power depends on the actual value of the parameter under H_1 and is not calculable.

In practice, choosing the test (of given level) that maximizes the power is difficult.

Testing for a mean

one-sample t-test

1. Normal model : $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$
   $H_0 : \ \mu = \mu_0 \quad H_1 : \ \mu \neq \mu_0$
2. test statistic : $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim_{H_O} t(n-1)$
   (*exact* Student distribution under H_0)
3. Rejection region $\mathcal{R} = \{|T| > c\}$, $c = qt_{n-1}(1-\alpha)$
4. P-value $= \mathrm{P}_{H_0}(|T| \geq |T^{obs}|)$

What if the sample is not Gaussian ?

1. $X_1, \ldots, X_n$ i.i.d., $\mathrm{E}(X_i) = \mu$, $\mathrm{Var}(X_i) = \sigma^2$
   $H_0 : \ \mu = \mu_0 \quad H_1 : \ \mu \neq \mu_0$
2. test statistic : $T = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}} \sim_{H_O} \mathcal{N}(0,1)$
   (TLC : *approximate* standard normal distribution under H_0)

Testing for one parameter

$X_1, \ldots, X_n$ i.i.d. with distribution depending on θ, $\widehat{\theta}$ MLE, $\widehat{\text{s.e.}}$ estimated standard error of $\widehat{\theta}$

- $H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0$
- Test statistic $T = \dfrac{\widehat{\theta} - \theta_0}{\widehat{\text{s.e.}}} \sim \mathcal{N}(0,1)$
  (*approximate* distribution for the MLE)
- Rejection region $\mathcal{R} = \{|T| > c\}$, $c = qnorm(1-\alpha)$
- P-value $= \mathrm{P}_{H_0}(|T| \geq |T^{obs}|)$

We observe the realizations of a Gaussian sample $\mathcal{N}(\mu, \sigma^2)$:

```
> x
 [1] 7.10 7.70 8.20 7.56 7.05 7.08 7.21 7.25 7.36 6.59 6.85
     7.90 7.27 6.56 7.93 8.50
> length(x); mean(x); sd(x)
[1] 16
[1] 7.381875
[1] 0.5506023
```

1. Derive a test at level 5% of

$$H_0 : \mu = 7 \quad \text{versus} \quad H_1 : \mu > 7.$$

   What is your decision ?

2. Suppose that the value under the alternative hypothesis is known : $H_1 : \mu = 7.5$.
   In this case, what is the type II error ? Give its calculation formula.

A battery manufacturer claims that the average life of his material is 170 hours.
A consumer organization takes a random sample of 100 batteries and observes an average lifetime of 158 hours with a standard deviation of 30 hours. We suppose that the data are realizations of a sample X_i, $i = 1, \ldots, 100$, with expectation m and finite variance.

$\hookrightarrow$ Can the organization accuse the manufacturer of false advertising ?

The two-sample t-test

Comparison of two means

Is a given gene *differentially expressed* between two cell types ?

1. 2 independent normal samples $X_{11}, \ldots, X_{1n_1}$ and $X_{21}, \ldots, X_{2n_2}$ with means μ_1 and μ_2 and identical variances σ^2.

$$H_0 : \ \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \ \mu_1 \neq \mu_2$$

2. test statistic $T = \dfrac{\overline{X}_1 - \overline{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, S^2 is the *pooled estimator* of σ^2

$$S^2 = \frac{\sum_i (X_{1i} - \overline{X}_1)^2 + \sum_i (X_{2i} - \overline{X}_2)^2}{n_1 + n_2 - 2}$$

   The distribution of T under H_0 is the t-distribution with $n_1 + n_2 - 2$ df.

3. Rejection region $\mathcal{R} = \{|T| > c\}$
4. P-value $= \mathrm{P}_{H_0}(|T| \geq |T^{obs}|)$

What if the samples are not Gaussian ? or if we cannot assume that the variances are equal ?

Let s_1^2 and s_2^2 be the sample variances of the 2 samples (or other consistent estimators).

- Define $W = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
  Then the distribution of W under H_0 is approximated by the normal standard law.
- P-value $= \mathrm{P}_{H_0}(|W| \geq |W^{obs}|)$

Comparing 2 prediction algorithms

We test a prediction algorithm on a test set of size 100 and we test a second algorithm on a second test set of size 80.

X (resp. Y) $=$ number of correct predictions for algo 1 (resp algo 2)
We observe $X = 95$ and $Y = 72$: is the difference statistically significant ?

What if we used the same test set to test both algorithms ?

The Likelihood Ratio Test (LRT)

$X_1, \ldots, X_n$ i.i.d. with distribution depending on $\theta \in \mathbb{R}^p$, $\widehat{\theta}$ MLE

$$H_0 : \ \theta = \theta_0 \quad H_1 : \ \theta \neq \theta_0$$

- The Likelihood ratio test rejects H_0 if

$$\text{LRT} \ = 2 \left[\log L(\widehat{\theta}) - \log L(\theta_0) \right] \geq c$$

- In simple models, the exact distribution of a transformation of LRT can be found to calculate an exact p-value
- In more complex models, the approximate distribution of LRT under H_0 is the chi-squared distribution with p df.

The Likelihood Ratio Test (LRT)

- The LRT can be generalized to test *nested* models and is very popular.
- The degrees of freedom of the χ^2 distribution is the difference of the number of parameters under H_1 and under H_0.

Example : $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)$ and we want to test

$$H_0 : \ \theta_4 = \theta_5 = 0 \quad \text{vs} \quad H_1 : \ \theta_4 \neq 0 \text{ or } \theta_5 \neq 0$$

The approximate distribution of the LRT statistic is the χ^2-distribution with $5 - 3 = 2$ degrees of freedom