

PC8 – Hypothesis testing

Octobre 16 2018

In PC7, we have given six steps to build hypothesis tests and study properties of these tests (as the significance level, and the power function). The steps are

- (1) write the statistical model,
- (2) choose and write the hypotheses,
- (3) choose an estimator/a test statistic,
- (4) choose the shape of the rejection region,
- (5) compute the boundaries of the rejection region using the chosen significance level,
- (6) summarize the previous steps by precisising the rule of rejection for H_0 and then conclude considering the observed values.

In the following, we are going to define a p-value associated to a test (it could be a 7-th step) and how to conclude from the p-value. Then we will consider a particular test, namely the likelihood ratio test which enables to automatically do steps (3) and (4). We will discuss uniformly most powerful test at the end.

P-values

Let $(Z, \mathcal{Z}, \{\mathbb{P}_\theta; \theta \in \Theta\})$ be a statistical model and z_{obs} the observed value, a realization of Z . We want to test $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$. Consider that the estimator $S(Z)$ is the test statistic we have chosen and \mathcal{S}_R the rejection region for this statistic. The p-value is the probability to observe a $S(Z)$ as extreme or more extreme than the observed one $S(z_{obs})$ under H_0 . For instance if the rejection region is one sided and has this shape: $\mathcal{S}_R = [a, +\infty)$ then the p-value is

$$\mathbb{P}_{H_0} (S(Z) \geq S(z_{obs})).$$

Equivalently, the p-value is the smallest significance level at which the null hypothesis would be rejected. Then the chosen test is equivalent to reject H_0 iff the p-value is smaller than the chosen significance level.

Exercise 1: P-values and doping controls

We consider the same doping control as in Exercise 1 of PC6. Namely, during a sport meeting, J.C. and S.R. are subject to an unannounced doping control. Doctors measure their hematocrit levels in their blood. Normally this rate is equal to $\tau_0 = 45\%$ but it can be increased by taking some drug. The measure of this rate is assumed to be Gaussian with a standard deviation of 2. The observed value of J.C. is 48 and the one of S.R. is 50. We want to know if these values are abnormal, i.e. that they have taken a drug, or if they just are the result of the imprecision of the measurements. In PC6, we chose to reject the H_0 hypothesis if the measurement x of the hematocrit level is larger than $\tau_c = 45 + 2q_{1-\alpha} \simeq 48.29$ and we don't reject H_0 otherwise. So we didn't reject H_0^{JC} for J.C. and we have rejected H_0^{SR} for S.R..

1. How do you compute the p-value in the doping experiment?
2. What are the p-values for J.C. and S.R.?
3. What are your conclusions using the p-values?

Exercise 2: p-values and Student test

Let

$$X_1, \dots, X_n \stackrel{i.i.d.}{\sim} \mathcal{N}(\mu, \sigma^2).$$

1. We assume that σ is not known and we want to test $H_0: \mu = \mu^*$ versus $H_1: \mu < \mu^*$ for $\mu^* \in \mathbb{R}$.

- Propose a test at level 0.05 (steps 1 to 6 and as step 7 the computation of the p-value) and create an R function `rej_reg` that takes as input a vector of observations $x = (x_1, \dots, x_n)$, the parameter μ^* and a level of significance $\alpha \in (0, 1)$, and outputs a vector of size four containing the upper bound of the acceptance region, the statistics of the test, the result of the test: 0 if H_0 is not rejected and 1 if H_0 is rejected and the value of the p-value.
- Check that it is indeed equivalent to the rejection of H_0 when the test statistic is in the rejection region and when the p-value is smaller than the significance level through simulations when $\sigma = 2$, $\alpha = 0.05$, $\mu^* = 0$, $\mu \in \{0, -0.5\}$ and $n = 100$.
- Check that under H_0 , the p-values are distributed uniformly on $(0, 1)$ through simulations when $\sigma = 2$, $\alpha = 0.05$, $\mu^* = 2$ and $n = 20$.

Likelihood ratio test and Neyman-Pearson

In PC7, we saw that the power function was a good way to compare two hypothesis tests with the same significance level. Here we present a test, namely the likelihood ratio test, which is automatically the best, i.e. with the smallest type II error for a fixed significance level when the hypotheses to test are simple, i.e. $\text{Card}(\Theta_0) = \text{Card}(\Theta_1) = 1$ (Neyman-Pearson Lemma).

Exercise 3: Binomial example

Imagine that there are two coins. The probability of heads of coin 1 is 0.4 and the probability of heads of coin 2 is 0.5. A friend is tossing one of them 5 times but you don't know which coin it is and you would like to find it out.

- Write the statistical model.
- In a dataframe, write the likelihood to obtain x heads for $x \in \{0, 1, \dots, 5\}$ for the both dice.
- Compute the likelihood ratio $\mathcal{L}(p = 0.5, x) / \mathcal{L}(p = 0.4, x)$ for $x \in \{0, 1, \dots, 5\}$. What does it mean if $\mathcal{L}(p = 0.5, x) / \mathcal{L}(p = 0.4, x) = 2$? What do you observe for the likelihood ratio?
- Show that $x \in \{0, 1, \dots, 20\} \mapsto \mathcal{L}(p = 0.5, x) / \mathcal{L}(p = 0.4, x)$ is increasing.
- Enumerate all the possible rejection regions to test $H_0 : p = 0.4$ against $H_1 : p = 0.5$, where p is the probability to obtain a head.
- Compute the significance level and the power of all possible tests. Among the tests with a level smaller than 0.0103, 0.0871 and 0.3175 which one has the largest power?
- Among the tests with a level smaller than 0.25 which one has the largest power? Compute the level and the power of the randomized test

$$\phi(X) = \begin{cases} 1 & \text{if } X > 3 \\ \frac{0.25 - \mathbb{P}_{Z \sim \text{Bin}(5, 0.4)}(Z > 3)}{\mathbb{P}_{Z \sim \text{Bin}(5, 0.4)}(Z = 3)} & \text{if } X = 3 \\ 0 & \text{otherwise} \end{cases}.$$

- Give the most powerful test at level 0.05 to test $H_0 : p = 0.4$ against $H_1 : p = 0.5$.

Exercise 4: Swimming pool

A swimming pool seller wants to compare two products which kills bacteria. Both products guarantee that 95% of bacteria are eliminated. Yet the pH may become more basic so more harmful for users. Thus he does an experiment with two swimming pools in his shop. He puts the product *A* in pool 1 and product *B* in pool 2. He has a pH meter but he doubts its reliability. Therefore he does 10 measurements in each swimming pool.

In pool 1, he observes x_1, \dots, x_{10} with respective values

7.33 ; 6.17 ; 7.46 ; 8.13 ; 6.68 ; 6.76 ; 7.97 ; 6.76 6.81 ; 8.40 .

The empirical mean is $\bar{x}_{10} = 7.247$ and the empirical variance (the one divided by $n - 1$) is $\sqrt{\sigma_x^2} = 0.73$.

In pool 2, he observes y_1, \dots, y_{10} with respective values

10.40 ; 7.27 ; 8.99 ; 7.28 ; 9.18 ; 9.10 ; 7.96 ; 7.71 ; 9.59 ; 9.61

The empirical mean is $\bar{y}_{10} = 8.709$ and the empirical variance (the one divided by $n - 1$) is $\sqrt{\sigma_y^2} = 1.08$.

The seller does not have any preference beforehand and wants to give the best advice to his customers. He wants to clearly say “prefer A”, “prefer B” or “do as you wish”. We know that $pH = 7$ is neutral for the skin while $pH = 9$ is basic and harmful.

1. Assume that the observations are realizations of independent Gaussian random variables with variance σ^2 and mean m_1 for pool 1, and m_2 for pool 2. Give the statistical model considering a Gaussian vector. Compute the mean and the covariance matrix.
2. Assume that $\sigma^2 = 1$ is known.
 - a. Give a likelihood ratio test $H_0 : m_1 = m_2 = 7$ against $H_1 : m_1 = 7, m_2 = 9$ at level α . Simplify the result as much as possible. Compute the p-value.
 - b. Explain why this test is uniformly most powerful among the hypothesis tests at level α which test the same hypotheses.
3. We still assume that $\sigma^2 = 1$ is known, test $H_0 : m_1 = m_2$ against $H_1 : m_1 < m_2$ at level α .
4. Now, we assume that σ^2 is unknown (more realistic for this problem statement).
 - a. Give the distribution of $Z = \bar{Y} - \bar{X}$ and of $W = 9\sigma_x^2 + 9\sigma_y^2$. We admit that W and Z are independent.
 - b. Give a hypothesis test $H_0 : m_1 = m_2$ against $H_1 : m_1 < m_2$ at level α . Compute the p-value.
 - c. What should be the conclusion of the seller?

UMP hypotheses tests for composite hypotheses

Exercise 5: Test for the variance in Gaussian models

Let (X_1, \dots, X_n) be a n -sample of a normal distribution $\mathcal{N}(\mu_0, \sigma^2)$, where μ_0 is known and σ^2 is unknown.

1. Give the statistical model.
2. Compute explicitly the likelihood ratio test of size at most α for testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$, where $\sigma_0^2 < \sigma_1^2$.
 - a. Is this test $\text{UMP}(\alpha)$ for testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 = \sigma_1^2$ when $\sigma_0^2 < \sigma_1^2$?
 - b. Is this test $\text{UMP}(\alpha)$ for testing $H'_0 : \sigma^2 \leq \sigma_0^2$ against $H'_1 : \sigma^2 > \sigma_0^2$?

Exercise 6: Likelihood ratio test in exponential models

Let (X_1, \dots, X_n) be n i.i.d. random variables distributed from an exponential distribution with parameter $\theta \in \Theta = (0, +\infty) =: \mathbb{R}_+^*$,

$$p(\theta; x) = \theta^{-1} \exp(-x/\theta) \mathbb{1}_{\mathbb{R}_+}(x).$$

Let $\theta_0 > 0$. We want to test

$$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0.$$

We first consider

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta = \theta_1,$$

where $\theta_1 > \theta_0$.

1. Show that the likelihood ratio is an increasing function of some statistic, determine what the statistic is.

2. Give a UMP(α) test. The constants, in the definition of the test, have to be explicitly given. You can use that if $X_i \sim \text{Gamma}(a_i, b)$ independently and c is a real constant then $\sum_i X_i \sim \text{Gamma}(\sum_i a_i, b)$ and $cX_i \sim \text{Gamma}(a_i, b/c)$, that a $\exp(\lambda)$ distribution is a $\text{Gamma}(1, 1/\lambda)$ distribution and that a $\text{Gamma}(n, 1/2)$ distribution is a $\chi^2(2n)$ distribution.
3. Compute the power function associated to this test.
4. Is the test built in Question 2 also a UMP(α) test for testing

$$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta = \theta_1 .$$

5. Is the test built in Question 2 also a UMP(α) test for

$$H_0 : \theta = \theta_0 \text{ against } H_1 : \theta > \theta_0 .$$

6. Is the test built in Question 2 also a UMP(α) test for

$$H_0 : \theta \leq \theta_0 \text{ against } H_1 : \theta > \theta_0 .$$