Report 4

Klaudia Balcer

12/17/2021

Contents

Introduction

In this report, we will consider multiple hypotheses testing problem and the properties of the Kolmogorov-Smirnov test. First, we will take a look at some multiple comparison procedures. Then, we discuss the relation between the KS statistics and the Brownian Bridge. The last section is devoted to the closure procedure, another way to handle multiple hypothesis testing.

Each of the topics is discussed theoretically and illustrated with simulations.

Multiple Testing

Problem Definition

Let's consider n testing problems:

$$H_{0,i} : \mu_i = 0 \quad vs \quad H_{1,i} : \mu_i \neq 0$$

Multiple testing means testing many individual hypotheses ($H_{0,i}$ vs. $H_{1,i}$) at the same time. To simply statistically test each hypothesis separately doesn't lead to satisfying results. We use **multiple comparison procedures** (MCPs) instead. MCPs are used to improve the quality of the tests.

The outcome of an MCP can be presented in the following form:

	accepted	rejected	total
true	TN	FP	n_0
false	FN	TP	n - n_0
total	n - R	R	n

The symbols used:

- TN - True Negatives - the null hypothesis is true, and it was accepted (U in *Candes*),
- FP - False Positives - the null hypothesis is true, but it was rejected (V in *Candes*),
- FN - False Negatives - the null hypothesis is false, and it was accepted (T in *Candes*),
- TP - True Positives - the null hypothesis is false, and it was rejected (S in *Candes*),
- n_0 - the number of true null hypotheses,
- R - the number of rejected hypotheses,
- n - the number of hypotheses.

Note: n and n_0 are the numbers. n is a known number; n_0 is an unknown number. TN, FP, FN, TP, R are random variables. R is an observed random variable; TN, FP, FN, TP are unobserved random variables.

Test Quality Measures

The definition of the Type I Error does not simply propagate to multiple testing problem. On the one hand, allowing a single false discovery with the probability α is a very strict condition. On the other hand, allowing the probability of false discovery in each test to be α, leads to many false discoveries. Hence we need to introduce new quality measures: FWER and FDR.

FWER

FWER stands for **F**amili**w**ise **E**rror **R**ate.

In **strong** sense: it is the probability of making any false dicoveries:

$$FWER = \mathbb{P}(FP \geq 1)$$

In **weak** sense: it is the probability of making any false discoveries if all the global null hypothesis is true:

$$FWER = \mathbb{P}(FP \geq 1 | \forall_i H_{0,i})$$

FDR

FDR stands for **F**alse **D**iscovery **R**ate id the expected value of FDP (**F**alse **D**iscovery **P**roportion) - the ratio between the numbers of false discoveries and all rejections:

$$FDR = \mathbb{E}\left[\frac{FP}{max(R,1)}\right]$$

Under the global null hypothesis, FDR and FWEAR are equivalent.

Power

Power is the probability of rejecting the null hypothesis when it is false. In the above terms, we can express it as the expected value of the ratio of TP and the number of false hypotheses:

$$power = \mathbb{E}\left[\frac{TP}{n - n_0}\right]$$

Tests

For all below procedures, first we calculate the p-values of single tests:

- p-values: $p_1, p_2, \ldots, p_n$,
- ordered p-values: $p_{(1)}, p_{(2)}, \ldots, p_{(n)}$.

Bonferroni's procedure

Reject $H_{0,i}$ if:

$$p_i < \frac{\alpha}{n}$$

This method is very conservative. We know from the lecture that Bonferroni's method controls FWER in a strong sense. In fact,

$$FWER \leq \frac{n_0}{n}\alpha$$

Sidak's procedure

Reject $H_{0,i}$ if:

$$p_i < \frac{\alpha_n}{n}$$

where $(1 - \frac{\alpha_n}{n})^n = 1 - \alpha$.

Sidak's procedure is slightly less conservative than Bonferroni's. There is a small difference between those two tests. For large n, both procedures give the same results. Sidak's procedure also controls FWER in a strong sense. In fact,

$$FWER = 1 - (1 - \alpha_n)^{n_0}$$

Holm's procedure

Reject $H_{0,(i)}$ if:

$$\forall_{(j \leq i)} \quad p_{(j)} < \frac{\alpha}{n - j + 1}$$

Holm's method is a step-up procedure. Step-up procedures (requiring all rejections) are more conservative and less powerful than step-down procedures (requiring any rejection). It is less conservative than Bonferroni's procedure. Holm's procedure also controls FWER strongly. Thus, it can be used instead of Bonferroni's method.

Hochberg's procedure

Reject $H_{0,(i)}$ if:

$$\exists_{(j \geq i)} \quad p_{(j)} < \frac{\alpha}{n - j + 1}$$

Hochberg's method is a step-down procedure. It is more powerful than Holm's method, but still controls FWER (in a strong sense under independence).

Benjamini-Hochberg's procedure

Reject $H_{0,(i)}$ if:

$$\exists_{(j \geq i)} \quad p_{(j)} < \frac{j}{n}\alpha$$

Benjamini-Hochberg's method is a step-down procedure. This procedure controls FDR under independence. Thus, it controls FWER weakly. It does not control FWER in a strong sense. It is much more liberal than Hochberg's procedure (more powerful and leads to more false discoveries).

Simulations

Simulation in both tasks consist of 1000 repetitions.

Task 1

In the first task, we will consider a low-dimensional case. Let n = 20.

Table 2: $\mu_1 = 1.2\sqrt{2logn}, \quad \mu_2 = \ldots = \mu_20 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.049	0.050	0.051	0.051	0.072
FDR	0.037	0.037	0.038	0.038	0.048
power	0.477	0.481	0.477	0.477	0.485

Table 3: $\mu_1 = \cdots = \mu_5 = 1.02\sqrt{2log(\frac{n}{10})}, \quad \mu_6 = \ldots = \mu_20 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.035	0.036	0.035	0.035	0.046
FDR	0.033	0.034	0.033	0.033	0.036
power	0.033	0.033	0.033	0.033	0.038

Table 4: $\mu_i = \sqrt{2log(\frac{20}{i})}, i = 1, \dots 10, \quad \mu_1 1 = \dots = \mu_2 0 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.020	0.020	0.020	0.020	0.081
FDR	0.010	0.010	0.010	0.010	0.026
power	0.112	0.113	0.115	0.115	0.174

Task 2

In the second task, we will repeat the above comparison in the high-dimensional case. let n = 5000.

Table 5: $\mu_1 = 1.2\sqrt{2logn}, \quad \mu_2 = \dots = \mu_5 000 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.048	0.050	0.048	0.048	0.084
FDR	0.036	0.037	0.036	0.036	0.054
power	0.698	0.699	0.698	0.698	0.701

Table 6: $\mu_1 = \dots = \mu_1 00 = 1.02\sqrt{2log(\frac{n}{200})}, \quad \mu_1 01 = \dots = \mu_5 000 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.047	0.050	0.047	0.047	0.415
FDR	0.012	0.012	0.012	0.012	0.047
power	0.034	0.035	0.034	0.034	0.105

Table 7: $\mu_1 = \dots = \mu_1 00 = \sqrt{2log(\frac{n}{200})}, \quad \mu_1 01 = \dots = \mu_5 000 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.047	0.049	0.047	0.047	0.403
FDR	0.015	0.016	0.015	0.015	0.054
power	0.030	0.030	0.030	0.030	0.089

Table 8: $\mu_1 = \dots = \mu_1 000 = 1.002\sqrt{2log(\frac{n}{2000})}, \quad \mu_2 001 = \dots = \mu_5 000 = 0$

	Bonferroni's	Sidak's	Holm's	Hochberg's	Benjamini-Hochberg's
FWER	0.044	0.045	0.044	0.044	0.174
FDR	0.027	0.027	0.027	0.027	0.040
power	0.001	0.001	0.001	0.001	0.003

All procedures controlling FWER (Bonferroni's, Sidak's, Holm's, Hochberg's) give similar results. Sidak's and Holm's methods are at least that good as Bonferroni's. Hochberg's procedure is at least as good as Holm's. Benjamini-Hochberg's method does control FDR but it doesn't control FWER. In all cases, it has greater power (at the expense of more false discoveries - greater FDR and FWER).

Empirical CDF Properties

In this task, we will compare the Brownian Bridge and an empirical process.

Theoretical Definitions

Empirical Process

The considered empirical process is a function of the ECDF. ECDF stands for the **E**mpirical **C**umulative **D**istribution **F**unction. The ECDF of the p-values is defined as:

$$F_n(t) = \frac{1}{n}\#\{i : p_i \leq t\}$$

We will consider the process:

$$U_n(t) = \sqrt{n}(F_n(t) - t), \quad t \in [0, 1]$$

which describes the difference between the empirical and theoretical CDFs of p-values under the null hypothesis. We will compare it with the Brownian Bridge.

Brownian Bridge

The Brownian bridge is a stochastic process. We will show its derivative from normally distributed random variables.

Let's start with $Z_1, Z_2, \ldots, Z_n \sim \mathcal{N}(0, 1)$.

The Wiener process is a stochastic probess with below properties:

- w(0) = 0,
- w(t) - w(s), w(t') - w(s') - are independent random variables for all $s' < t'$, $s < t$,
- w(t) - w(s) $\sim \mathcal{N}(0, t - s)$.

So using $Z_1, Z_2, \ldots, Z_n$ we can construct a sample from Wiener process like:

$$w_n(\frac{i}{n}) = \sum_{1 \leq k \leq n} Z_k$$

The Brownian bridge can be constructed from Wiener process:

$$B(t) = w(t) - tw(1), \quad t \in [0, 1]$$

So in the discrete case, we can construct the sample like:

$$B_n\left(\frac{i}{n}\right) = w_n\left(\frac{i}{n}\right) - \frac{i}{n}w_n(1)$$

T Statistics

We will compare quantiles of two statistics. The first of them is:

$$T = sup_{t \in [0,1]}|B_n(t)|$$

K-S Statistics

The second statistics for quantile comparison:

$$KS = sup_{t\in[0,1]}|U_n(t)|$$

Under H_0, it is when p-values are uniformly distributed, the KS statistic converges to the T statistic in probability.

Simulation

Task 3

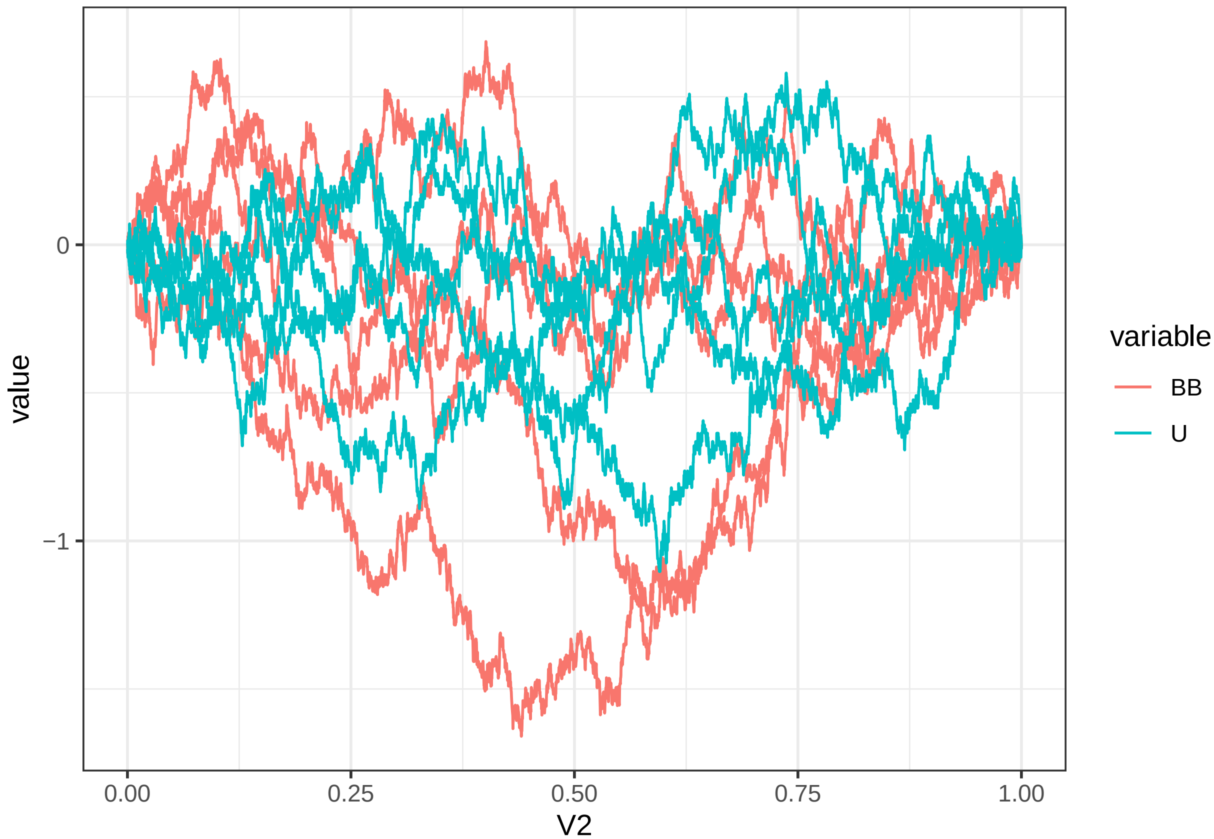

Table 9: T quantiles

80%	90%	95%
1.074	1.242	1.393

Table 10: KS quantiles

80%	90%	95%
1.061	1.23	1.344

As expected, the quantiles are similar.

Closure Procedure

In the last task, we will compare FWER, FDR, and power of closure tests (Bonferroni and χ^2). Closure procedure is a way to apply global tests for multiple testing problem.

Closure Tests

In the last task, we will use the closures of Bonferroni and χ^2 tests for multiple hypotheses testing.

The closure of a test means that we run the global tests on all subsets of hypotheses. To determine, which single hypothesis to reject, we look at all subsets containing those single hypotheses. When the global hypotheses for all subsets containing the hypothesis were rejected, we reject the single hypothesis. Otherwise, we have no evidence to reject the single hypothesis.

The closure procedure is computationally expensive. In n-dimensional case, we need to consider $2^n - 1$ global hypotheses.

The closure procedure of Bonferroni's method is Holm's procedure. The closure of the χ^2 procedure needs calculating from the definition.

Simulation

Task 4

Table 11: Needle in the haystack: $\mu_1 = 1.02\sqrt{2logn}, \quad \mu_2 = \ldots = \mu_{10} = 0$

	bonferroni_closure	chisq_closure
FWER	0.048	0.006
FDR	0.040	0.006
power	0.272	0.100

Table 12: Many small effects: $\mu_1 = \ldots = \mu_5 = 1, \quad \mu_6 = \ldots = \mu_{10} = 0$

	bonferroni_closure	chisq_closure
FWER	0.026	0.009
FDR	0.024	0.008
power	0.039	0.009

It is to see, that the global null tests hold their properties when used in the closure procedure. The closure of Bonferroni's test deals better with the needle in the haystack problem. Fisher gives better results when detecting distributed effects.