

## Statistical Modeling and Inference – Project: Multiple Testing

NITI MISHRA · MIQUEL TORRENS · BÁLINT VÁN

November 2<sup>nd</sup>, 2015

Solution to proposed exercises.

### Exercise 1.1

Given  $\alpha = 1 - F_H(c_\alpha)$ , we have:

$$F_H(c_\alpha) = 1 - \alpha \Leftrightarrow c_\alpha = F_H^{-1}(1 - \alpha)$$

Hence proved.

### Exercise 1.2

Given that  $\mathbf{Y}$  is a dataset randomly drawn from a distribution truly consistent with  $H$ , we define the p-value by computing how likely is a test statistic drawn from  $\mathbf{Y}$ , meaning  $T(\mathbf{Y})$ , to be greater than a test statistic  $T(\mathbf{X})$ , given by a realisation of a random dataset  $\mathbf{X}$  not necessarily consistent with  $H$ . That is:

$$\text{p-value} = p(\mathbf{X}) = \mathbb{P}(T(\mathbf{Y}) > T(\mathbf{X}))$$

We use the fact that  $\alpha = 1 - F_H(c_\alpha)$  and that  $\alpha = \mathbb{P}(T(\mathbf{Y}) > c_\alpha)$  to conclude that:

$$1 - F_H(c_\alpha) = \mathbb{P}(T(\mathbf{Y}) > c_\alpha) \Leftrightarrow \mathbb{P}(T(\mathbf{Y}) > T(\mathbf{X})) = 1 - F_H(T(\mathbf{X}))$$

Therefore, we find that  $p(\mathbf{X}) = 1 - F_H(T(\mathbf{X}))$ .

### Exercise 1.3

The CDF  $F_Y(y)$  takes by definition values inside  $[0, 1]$ . Now we define  $Z \equiv F_Y(y)$  and assume continuity on the CDF of  $y$  and do:

$$F_Z(z) = \mathbb{P}(F_Y(y) \leq z) = \mathbb{P}(y \leq F_Y^{-1}(z)) = F_Y(F_Y^{-1}(z)) = z$$

Also for a uniform random variable<sup>1</sup>  $U$  between 0 and 1:

$$F_U(z) = \int f_U(u) du = \int_0^z 1 du = [u]_0^z = z - 0 = z$$

For  $z \in [0, 1]$  we have  $F_Z(z) = F_U(z)$  and as  $Z = F_Y(y)$  we know that the distributions of  $U$  and  $F_Y(y)$  are the same.

Given that the CDF gives  $F_Y(y) \in [0, 1]$ , we know that also  $1 - F_Y(y) \in [0, 1]$ . The identical proof can therefore be applied to  $1 - F_Y(y)$  to find that it is also uniformly distributed.

<sup>1</sup>Note that in this proof  $f_U = 1$  because it is a uniform distribution defined by parameters  $a = 0$  and  $b = 1$ .

### **Exercise 1.4**

We saw that  $F_Y(y)$  and  $1 - F_Y(y)$  follow a uniform distribution regardless of the distribution of  $Y$ , as long as they are continuous. The p-value is defined as:

$$p(\mathbf{X}) = 1 - F_H(T(\mathbf{X}))$$

We know that  $F_H(T(\mathbf{X}))$  and therefore  $1 - F_H(T(\mathbf{X}))$  are continuous distributions and thus it is valid to apply the proof from Exercise 1.3 to say that the p-value follows a uniform distribution.

### **Exercise 2.1**

Given that  $y_i$  is uniform between 0 and 1 for all  $i$ , and given  $\alpha \in [0, 1]$ , we know:

$$\mathbb{P}(y_i < \alpha) = \alpha$$

The complementary probability states<sup>2</sup>:

$$\mathbb{P}(y_i > \alpha) = 1 - \alpha$$

Then:

$$\mathbb{P}((y_1 > \alpha) \cap (y_2 > \alpha) \cap \dots \cap (y_m > \alpha)) = (1 - \alpha)(1 - \alpha) \dots (1 - \alpha)$$

Where  $(1 - \alpha)$  is replicated  $m$  times. Then:

$$\mathbb{P}\left(\bigcap_{i=1}^m y_i > \alpha\right) = (1 - \alpha)^m$$

Hence proved.

### **Exercise 2.2**

With a single case the null hypothesis is accepted if the p-value  $p_i(\mathbf{X}_i) > \alpha$ . Thus the complete null is satisfied with probability:

$$\mathbb{P}\left(\bigcap_{i=1}^m p_i(\mathbf{X}_i) > \alpha\right)$$

We just proved that  $\mathbb{P}(\bigcap_{i=1}^m y_i > \alpha) = (1 - \alpha)^m$ . So now we need to find the probability that at least in one of the cases the null hypothesis is rejected, and that is the complementary probability of the complete null. Thus, using  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ , we have:

$$1 - \mathbb{P}\left(\bigcap_{i=1}^m p_i(\mathbf{X}_i) > \alpha\right) = 1 - (1 - \alpha)^m$$

Hence proved.

### **Exercise 2.3**

---

<sup>2</sup>Recall that by continuity  $\mathbb{P}(y_i = \alpha) = 0$ .

We need to find an expression for the case in which the complete null is satisfied  $100 \times (1 - \alpha)\%$  of the times, meaning that the overall type I error is  $\alpha$ . At that point, each independent test will be rejected  $100 \times s\%$  of the times, so we need to find an  $s$  satisfying:

$$\mathbb{P} \left( \bigcap_{i=1}^m p_i(\mathbf{X}_i) > s \right) = 1 - \alpha$$

Then,

$$(1 - s)^m = 1 - \alpha$$

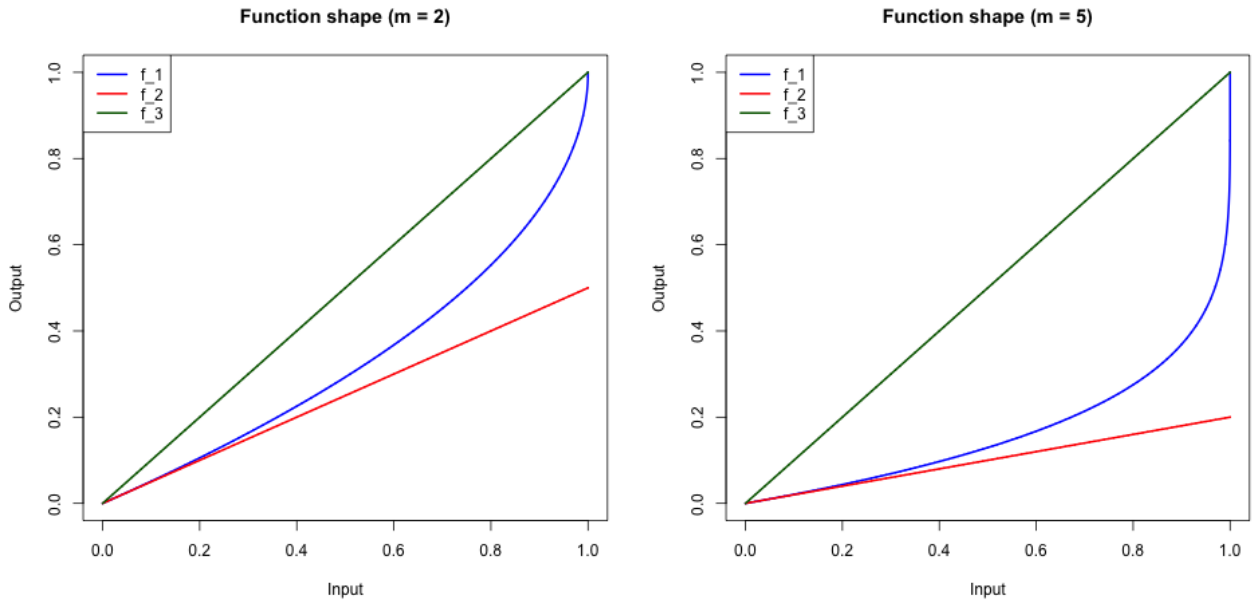
And so, the probability of rejecting each independent test is:

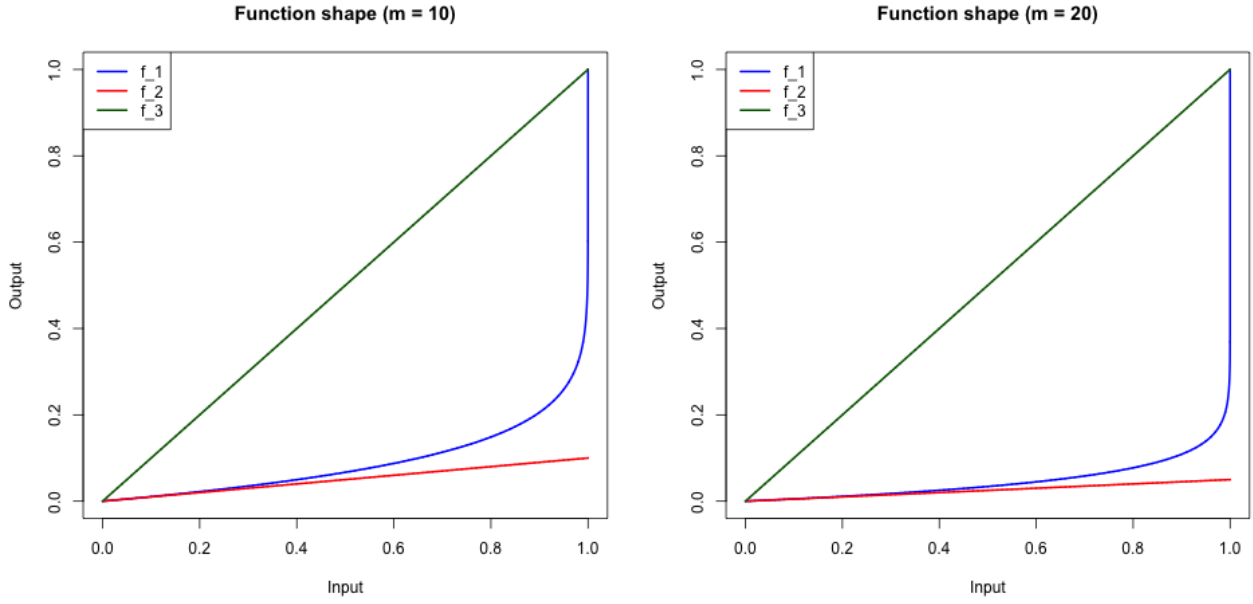
$$s = 1 - (1 - \alpha)^{1/m}$$

Thus proved.

#### **Exercise 2.4**

We plot for different values of  $m$ :





For  $m = 1$  the three curves overlap. We check graphically that  $f_2 \leq f_1 \leq f_3$  for  $\alpha \in [0, 1]$ .

### **Exercise 2.5**

We have  $\alpha \in [0, 1]$  and  $m$  as a positive integer. We first prove that  $f_2 \leq f_1$ :

$$\frac{\alpha}{m} \leq 1 - (1 - \alpha)^{1/m}$$

This would imply:

$$(1 - \alpha)^{1/m} \leq 1 - \frac{\alpha}{m}$$

It is straightforward to see that the inequality holds for  $\alpha = 0$  and  $\alpha = 1$ . To see if that is true for the all  $\alpha \in [0, 1]$ , we want to check if the slope of the function on  $\alpha$  is also steeper for  $f_2$ . We take derivatives on both sides with respect to  $\alpha$ :

$$-\frac{1}{m}(1 - \alpha)^{1/m-1} \leq -\frac{1}{m} \Leftrightarrow (1 - \alpha)^{1/m-1} \geq 1$$

These expressions are monotonic on  $\alpha$ . The exponent on LHS in the right inequality is a negative number and  $0 < (1 - \alpha) < 1$ , so the inequality holds. This concludes that  $f_2 \leq f_1$  for  $\alpha \in [0, 1]$  and  $m$  a positive integer.

For  $f_1 \leq f_3$  we need:

$$1 - (1 - \alpha)^{1/m} \leq \alpha$$

We proceed analogously. This equality holds plugging  $\alpha = 0$  and  $\alpha = 1$ . Derivatives on both sides should produce:

$$\frac{1}{m}(1 - \alpha)^{1/m-1} \leq 1$$

This holds with equality for  $\alpha = 0$  and is decreasing in  $\alpha$ , so with inequality for  $\alpha > 0$ . Thus as a whole the inequality holds. Provided monotonicity, this concludes that  $f_1 \leq f_3$ .

Hence, we conclude that  $f_2 \leq f_1 \leq f_3$  for  $\alpha \in [0, 1]$  and  $m$  a positive integer.

### **Exercise 2.6**

An approach that makes no correction for multiple testing the level is determined by the size  $\alpha \in [0, 1]$ . This level is, as proven before, necessarily higher than the one using a multiple-test sensitive approach, which leads to a significance level  $1 - (1 - \alpha)^{1/m}$ .

### **Exercise 3.1**

We shall show this tackling separately both inequalities.

As for the first inequality, the result of Exercise 2.2 concludes that:

$$\mathbb{P}_{C-H} \left[ \bigcup_{i=1}^m \{p_i(\mathbf{Y}_i) < \alpha\} \right] = 1 - (1 - \alpha)^m.$$

Given  $\alpha \in [0, 1]$  and  $m$  as a positive integer, we know:

$$(1 - \alpha)^m \leq (1 - \alpha) \Leftrightarrow \alpha \leq 1 - (1 - \alpha)^m \Leftrightarrow \alpha \leq \mathbb{P}_{C-H} \left[ \bigcup_{i=1}^m \{p_i(\mathbf{Y}_i) < \alpha\} \right]$$

The second inequality is proved using Bonferroni's inequality<sup>3</sup>. Given that the individual probability  $\mathbb{P}_{C-H}(p_i(\mathbf{Y}_i) \leq \alpha) = \alpha$ , then:

$$\begin{aligned} \mathbb{P}_{C-H} \left[ \bigcup_{i=1}^m \{p_i(\mathbf{Y}_i) < \alpha\} \right] &\leq \sum_{i=1}^m \mathbb{P}_{C-H}(p_i(\mathbf{Y}_i) < \alpha) \\ &\leq \sum_{i=1}^m \alpha \\ &\leq m\alpha \end{aligned}$$

Thus,

$$\alpha \leq \mathbb{P}_{C-H} \left[ \bigcup_{i=1}^m \{p_i(\mathbf{Y}_i) < \alpha\} \right] \leq m\alpha.$$

### **Exercise 4.1**

If we define the event that all  $y_i$  be greater than their respective  $l_i$  as  $A \equiv \bigcap_{i=1}^m \{y_i > l_i\}$ , then we have  $A^c = \bigcup_{i=1}^m \{y_i \leq l_i\}$ , because that defines the event that at least some  $y_i$  be smaller than its corresponding  $l_i$ . Since these are complementary probabilities,

$$\mathbb{P}[\bigcap_{i=1}^m \{y_i > l_i\}] + \mathbb{P}[\bigcup_{i=1}^m \{y_i \leq l_i\}] = 1$$

---

<sup>3</sup>This inequality states that  $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$ , which can be generalized to  $\mathbb{P}(\bigcup_i A_i) \leq \sum_i \mathbb{P}(A_i)$ , also known as Boole's inequality.

This implies that for  $\mathbb{P}[\cap_{i=1}^m \{y_i > l_i\}] = 1 - \alpha$ , we have:

$$\mathbb{P}[\cup_{i=1}^m \{y_i < l_i\}] = 1 - (1 - \alpha) = \alpha$$

Now given that  $p_{(i)}$  is uniformly distributed between 0 and 1, under the complete null hypothesis the probability that at least one  $p_{(i)} < l_i$  is defined by:

$$\mathbb{P}_{C-H} \left[ \bigcup_{i=1}^m \left\{ p_{(i)} < \frac{i\alpha}{m} \right\} \right] = \alpha$$

Hence proved.