Statistical Modeling and Inference – Project: Multiple Testing

Niti Mishra · Miquel Torrens · Bálint Ván

November 2nd, 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[1] U between 0 and 1:

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

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.

[1]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 distributuon between 0 and 1.

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[2]:

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

Then by independence:

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

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

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 α. 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,

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

[2]Recall that by continuity $\mathbb{P}(y_i = \alpha) = 0$.

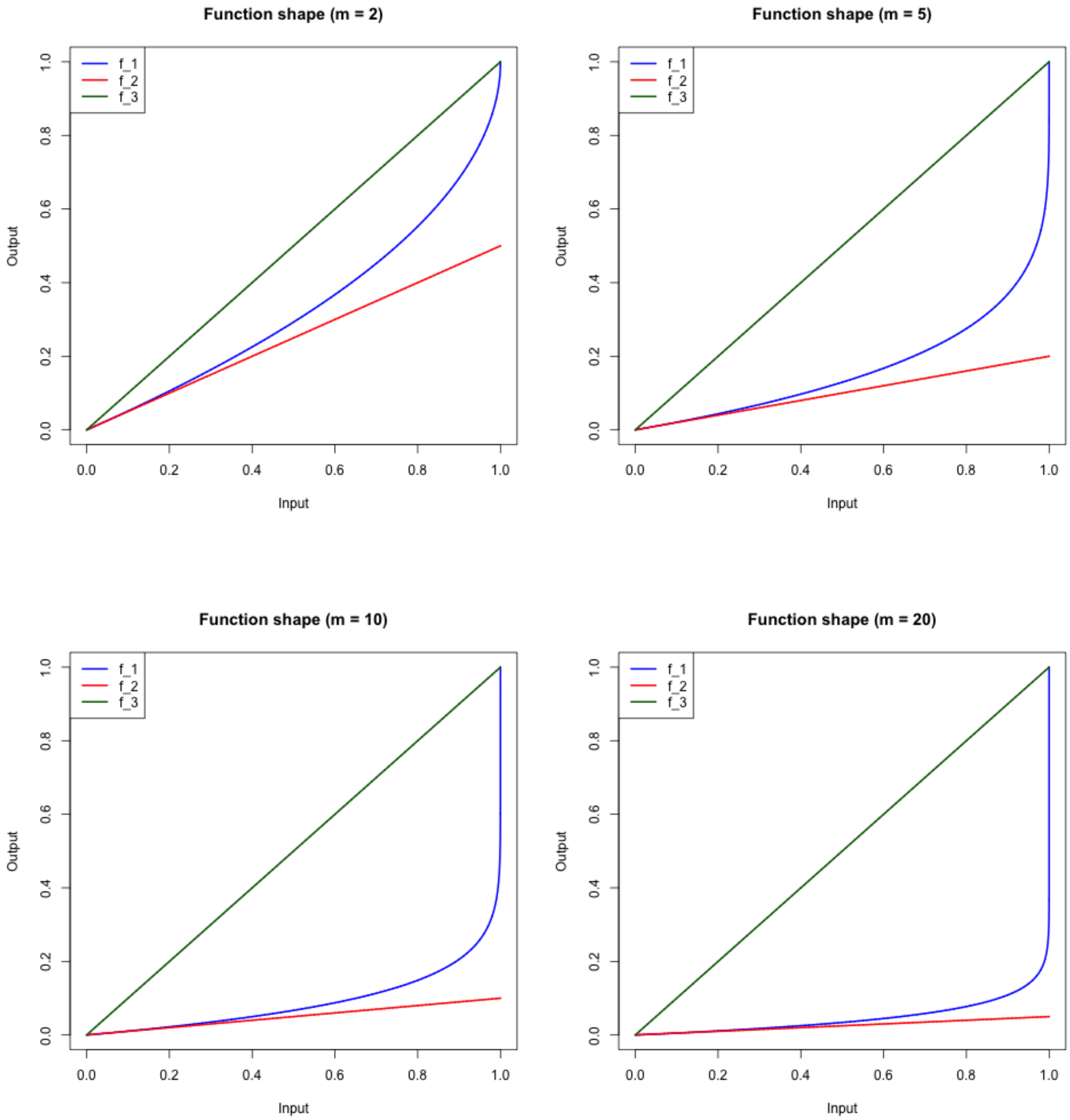
And so, the probability s of rejecting each independent test is:

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

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 **(Optional)**

We first prove that $f_2 \le f_1$. We need to show that:

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

Observe that this would imply:

$$\frac{\alpha}{m} \le 1 - (1-\alpha)^{1/m} \Leftrightarrow (1-\alpha)^{1/m} \le 1 - \frac{\alpha}{m} \Leftrightarrow 1 - \alpha \le \left(1 - \frac{\alpha}{m}\right)^m$$

Now let's define $x \equiv -\alpha/m$. Given $\alpha < m$, we know that $x \ge -1$. Now we can use Bernouilli's inequality, that says that $(1+x)^r > (1+rx)$ for $r > 0$ and $x \ge -1$. Applying this inequality, we get:

$$\left(1 - \frac{\alpha}{m}\right)^m \ge 1 - m\frac{\alpha}{m} = 1 - \alpha$$

This proves the expression above and shows that $f_2 \le f_1$. For the second inequality we need to show $f_1 \le f_3$, that is that $1 - (1-\alpha)^{1/m} \le \alpha$. We proceed analogously:

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

When $\alpha \in [0,1]$ and m as a positive integer, it is straightforward to see that this second inequality holds. This proves that $f_1 \le f_3$.

Therefore, we conclude $f_2 \le f_1 \le f_3$ when $\alpha \in [0,1]$ and m is a positive integer.

Exercise 2.6

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

Exercise 3.1

We shall show this tackling separately both inequalities.

As for the first inequality, we use the basic property $\mathbb{P}[A_1] \leq \mathbb{P}[A_1 \cup A_2]$ to more generally see that:

$$\mathbb{P}[A_i] \leq \mathbb{P}\left[\bigcup_{i=1}^{m} A_i\right]$$

Given the individual probability $\mathbb{P}_{C-H}[p_i(\mathbf{Y}_i) \leq \alpha] = \alpha$, in our setting this implies that:

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

This proves the lower bound inequality. The second inequality is proved using the following rule:

$$\mathbb{P}[A_1 \cup A_2] = \mathbb{P}[A_1] + \mathbb{P}[A_2] - \mathbb{P}[A_1 \cap A_2] \leq \mathbb{P}[A_1] + \mathbb{P}[A_2]$$

This can be generalized[3] to:

$$\mathbb{P}[\cup_i A_i] \leq \sum_i \mathbb{P}[A_i]$$

Given that the individual probability $\mathbb{P}_{C-H}[p_i(\mathbf{Y}_i) \leq \alpha] = \alpha$. Then:

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

Thus, we conclude:

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

[3] We find this to be referred in some contexts as Bonferroni's inequality or Boole's inequality.

Exercise 4.1

If we define the event that all i.i.d. $y_i \sim \text{Unif}(0,1)$ be greater than their respective l_i as $A \equiv \cap_{i=1}^{m}\{y_i > l_i\}$, then we have $A^c = \cup_{i=1}^{m}\{y_i < l_i\}$, because that defines the event that at least some y_i be smaller than its corresponding l_i. Since these are complementary probabilities,

$$\mathbb{P}[\cap_{i=1}^{m}\{y_i > l_i\}] + \mathbb{P}[\cup_{i=1}^{m}\{y_i < l_i\}] = 1$$

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 the set of ordered $p_i(\mathbf{X}_i)$ is uniformly distributed[4] between 0 and 1, under the complete null hypothesis the probability that at least one $p_{(i)} < l_i$ is defined by:

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

Thus proved.

[4] See exercises 1.3 and 1.4