

## 1 Distribution of p-value under the null

### 1.1

## 2 Multiple testing under independence assumptions

### 2.1

Using the formula for distribution complements  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ , along with the definition for a given random variable  $y$  drawn from a uniform distribution spanning 0 to 1, and  $\alpha \in [0, 1]$ :

$$\mathbb{P}(y \leq \alpha) = \alpha$$

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

The joint probability of a series of independent events is given by the product of their individual probabilities. If  $\mathbb{P}(A_m)$  for all values of  $m$  are equal, this is equivalent to:

$$\mathbb{P}(\cap_{i=1}^m A_i) = \mathbb{P}(A)^m$$

From which follows that for any series of independent draws from a uniform distribution,  $y_1, \dots, y_m$ :

$$\mathbb{P}(\cap_{i=1}^m \{y_i > \alpha\}) = \mathbb{P}(y_1 > \alpha)^m$$

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

### 2.2

We reject a single null hypothesis at significance level  $\alpha$  with probability  $\alpha$ . The complement of rejecting a single null hypothesis is to accept a single null hypothesis, which we therefore do with probability  $1 - \alpha$ . To accept ALL null hypotheses, from a set of  $m$  of hypothesis tests with I.I.D. uniformly distributed p-values, is therefore given by:

$$\mathbb{P}(\text{accepting all null hypotheses}) = (1 - \alpha)^m$$

The complement of accepting every null hypothesis, would be to reject one or more null hypothesis:

$$\mathbb{P}(\text{reject one or more null hypotheses}) = 1 - \mathbb{P}(\text{accepting all null hypotheses})$$

$$\mathbb{P}(\text{reject one or more null hypotheses}) = 1 - (1 - \alpha)^m$$

## 2.3

Let  $\alpha_0$  represent the significance level with which we reject each individual null hypothesis and  $\alpha$  represent the probability that we reject one or more null hypotheses from a set of hypotheses with i.i.d. uniform p-values. Starting from our equation above, we can find  $\alpha_0$  in terms of  $\alpha$ :

$$\alpha = 1 - (1 - \alpha_0)^m$$

$$1 - \alpha = (1 - \alpha_0)^m$$

$$(1 - \alpha)^{\frac{1}{m}} = 1 - \alpha_0$$

$$\alpha_0 = 1 - (1 - \alpha)^{\frac{1}{m}}$$

## 2.4

plot!!!!

## 2.5

We consider all situations where  $\alpha \in [0, 1]$  and all  $m \geq 1$  to prove that  $f_2(\alpha) \leq f_1(\alpha)$ :

$$\begin{aligned}
f_2(\alpha) &\leq f_1(\alpha) \\
\frac{\alpha}{m} &\leq 1 - (1 - \alpha)^{\frac{1}{m}} \\
1 - \alpha &\leq (1 - \frac{\alpha}{m})^m \\
1 - \alpha &\leq 1 - m\left(\frac{\alpha}{m}\right) + \left(\frac{\alpha}{m}\right)^m \\
1 - \alpha &\leq 1 - \alpha + \left(\frac{\alpha}{m}\right)^m
\end{aligned}$$

For all  $m \geq 1$  and all positive real values for  $x$ , we can say that  $x \geq \frac{x}{m}$ . From this it follows that  $f_3(\alpha) \geq f_2(\alpha)$ . We can also say that for all positive real values of  $x$ ,  $x^{\frac{1}{m}} \geq x$  for all  $m \geq 1$ . From here we can easily prove our remaining inequality:

$$\begin{aligned}
f_3(\alpha) &\geq f_1(\alpha) \\
\alpha &\geq 1 - (1 - \alpha)^{\frac{1}{m}} \\
1 - \alpha &\geq (1 - \alpha)^{\frac{1}{m}}
\end{aligned}$$

## 2.6

Under independence assumptions, as shown in (2.3), the significance level needed for each test,  $\alpha_0$ , is shown to be  $1 - (1 - \alpha)^{\frac{1}{m}}$ . Making now correction for multiple testing, we follow the generic per-comparisson error rate (PCER), and accept or reject each individual test at  $\alpha_0 = \alpha$ . As shown in (2.5):

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

And from that we can see that the significance level needed for each individual test if we consider each test p-value as being I.I.D. from a uniform distribution, is much lower if we consider the joint probability of all our tests and aim to keep the probability of rejecting one or more null hypothesis at our significance level, then if we make no correction for the joint probability of our multiple tests.

### 3 Proof of Bonferroni Correction

Using Boole's Inequality, we can rewrite the probability of rejecting one or more null hypotheses to be:

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

If we consider  $\mathbf{Y}_1, \dots, \mathbf{Y}_m$  to be p-values drawn from a uniform distribution from 0 to 1, we can state that the probability of any given  $\mathbf{Y}_i \leq \alpha$ , where  $\alpha \in [0, 1]$ , is equal to  $\alpha$ . This leads to the following:

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

### 4 Ordered p-values