Multiple Testing Project

2.2 P-value

1. Distribution of the p-value under the null

1.1 Show that for any α, $c_\alpha = F_H^{-1}(1-\alpha)$

We know that $\alpha = 1 - F_H(c_\alpha)$ from the definition of c_α.

$\alpha = 1 - F_H(c_\alpha)$

$\alpha - 1 = -F_H(c_\alpha)$

$-(\alpha - 1) = F_H(c_\alpha)$

$1 - \alpha = F_H(c_\alpha)$

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

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

Q.E.D.

1.2 Show that the p-value of the test, as a function of the data $\mathbf{X}$ used, is given by $p(\mathbf{X}) = 1 - F_H(T(\mathbf{X}))$.

The p-value is defined as $p - value = inf\{\alpha : T(X) \in R_\alpha\}$

Which is to say that the p-value is the *smallest* α for which $T(X)$ is in the region R_α of the probability distribution P_H

So the p-value is an instance of α, which is defined as $\alpha = 1 - F_H(c_\alpha)$ where c_α is chosen so that the equation is true. Therefore, if we replace c_α with our test statistic $T(\mathbf{X})$, we get a $p(\mathbf{X}) = 1 - F_H(\mathbf{X})$.

This matters because it highlights that the choice of α sets the minimum p-value for $\mathbf{X}$ as $p(\mathbf{X})$ such that one rejects the hypothesis that X is from the same distribution as Y when $T(\mathbf{X}) > c_\alpha$.

From the previous question we know, $F_H(c_\alpha) = 1 - \alpha$

$F_H(c_\alpha) = 1 - P_H[T(\mathbf{X}) > c_\alpha]$

$P_H[T(\mathbf{X}) > c_\alpha] = 1 - F_H(c_\alpha)$

$P_H[T(\mathbf{X}) > \mathbf{x}] = 1 - F_H(\mathbf{X})$

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

1.3 Show that for any univariate random variable y with continuous distribution function F, the random variables $F(y)$ and $1 - F(y)$ follow the uniform distribution.

resource resource

We know a function to be uniformly distributed if it has constant probability for any input value in the range for which it is defined. Say y_i is some value for which $F_Y(y_i)$ is defined. The set of all y_i is denoted as $\mathbf{Y}$ and the probability of $F_Y(\mathbf{Y})$ is given by $P[F_Y(\mathbf{Y})]$.

F_Y is uniformly distributed when $P[F_Y(y_1)] = P[F_Y(y_2)]$ where y_1 and y_2 are 2 distinct values in $\mathbf{Y}$.

$$P[F_Y(\mathbf{Y}) \leq y] = P[F^{-1}(F_Y(\mathbf{Y})) \leq F^{-1}(y)] = P[\mathbf{Y} \leq F^{-1}(y)]$$

We know that F_Y is strictly increasing $F_Y^{-1}(F_Y(\mathbf{Y})) = y$ by the definition of the definition of a quantile function which is the generalized inverse of a CDF. We use this property to concolude:

$$P[\mathbf{Y} \leq F_Y^{-1}(y)] = F_Y(F_Y^{-1}(y)) = y$$

This is true for all values of $\mathbf{Y}$.

By definition if a function F_Y follows the uniform distribution, so does $1 - F_Y$

1.4 Using the above results, show that the p-value follows the uniform distribution under H.

$\alpha = 1 - F_H(c_\alpha)$

c_α being a continuous random variable, we know $F_H(c_\alpha)$ to have a uniform distribution. If $F_H(c_\alpha)$ follows a uniform distribution, $1 - F_H(c_\alpha)$ does too.

MT under independence assumptions

2.1 Show that $P_H[\cap_{i=1}^m \{y_i > \alpha\}] = (1-\alpha)^m$

As $y_1, ..., y_m$ are independent unifrom random variables, the joint distribution of the probability that for every $y_i > \alpha$ is nothing but the product of the probabilities,

$$P_H[\cap_{i=1}^m \{y_i > \alpha\}] = \prod_{i=1}^m P(y_i > \alpha)$$

We know from the previous exercises that $P(y_i < \alpha) = \alpha$, or, equivalently, $P(y_i > \alpha) = 1 - \alpha$. Therefore,

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

2.2 Show that the probability of rejecting at significance level α at least one of the independent tests is $1 - (1-\alpha)^m$.

The probability of rejecting at least one of the tests could be expressed as the union,

$P_H[\cup_{i=1}^m \{y_i < \alpha\}]$

And we know that the probability of at least one rejection is the complement of the probability of having no rejections, which we have already determined as being equal to $(1-\alpha)^m$. Thus,

$P_H[\cup_{i=1}^m \{y_i < \alpha\}] = 1 - P_H[\cap_{i=1}^m \{y_i > \alpha\}]$

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

2.3 Under the above assumption, show that if we wish that the overall type I error is α, each independent test should be rejected at significance level $1 - (1-\alpha)1/m$

Type I error is rejecting the null when it is actually true. We know from 1.2 that the prbability of type one error for m hypotheses each having been tested at the confidence level α is:

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

So if we want the overall type I error to be α for m tests, each test should be rejected at a certain probability p_r that satisfies:

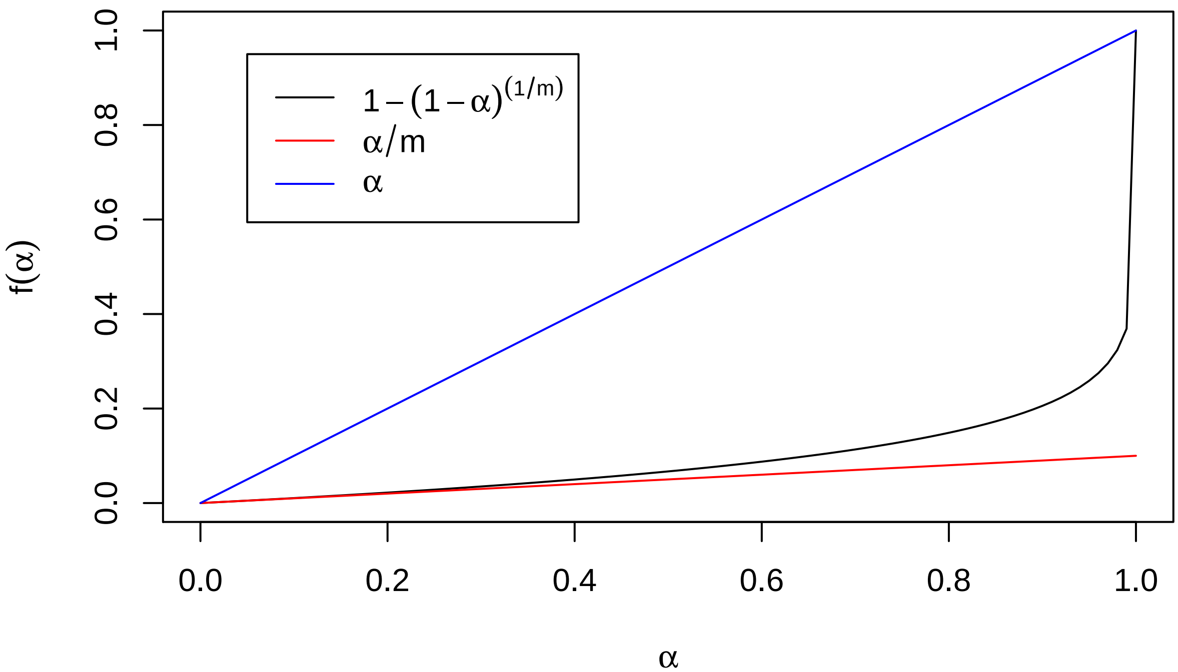
$P_H[\cup_{i=1}^m \{y_i < p_r\}] = 1 - (1-p_r)^m = \alpha$

Solving for p_r to determine the level p_r at which each test should be evaluated:

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

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

2.4 $f_1(\alpha) = 1-(1-\alpha)^{1/m},\ f_2(\alpha) = \alpha/m,\ f_3(\alpha) = \alpha.\ for\ \alpha \in [0,1]$. **Check that $f_2 \leqslant f_1 \leqslant f_3$.**



2.5 (optional)

We can say $f_1 \leq f_3$ from the following:

$1-(1-\alpha)^{frac1m} \leq \alpha$

$1-(1-\alpha)^{frac1m} - \alpha \leq 0$

We know the expression $(1-\alpha)^{frac1m}$ when $m = 1$ because $m \in [1, \inf]$ and $1-\alpha \in [0,1]$, and a fraction raised to a fraction is a greater fraction, increasing with the denominator of the exponent.

So the minimum value of $1-(1-\alpha)^{frac1m} - \alpha$ is given by $m = 1$ and

$1-1+\alpha-\alpha = 0 \leq 0$, and $f_1 \leq f_3$ Q.E.D.

We can say $f_2 \leq f_1$ from the following:

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

$0 \leq m - m(1-\alpha)^{1/m} - \alpha$

The expression $m - m(1-\alpha)^{1/m}$ will always be less than or greater than alpha because as described above, a fractaion raised a fraction is a greater fraction so the difference in these two terms m and $m - m(1-\alpha)^{1/m}$ is always greater than α and the difference increases with m. The minimum of m is 1 so for further proof, we can take the minimum to be:

$0 \leq 1-(1-\alpha)-\alpha = 0$ Q.E.D.

2.6

In 2.3 it was shown the level needed for each test is $p_r = 1-(1-\alpha)^{\frac{1}{m}}$ in order to achieve a level of α for multiple tests. The level that makes no correction for multiple testing is assuming $m = 1$:

No correction: $p_{uncorrected} = 1-(1-\alpha)$

Corrected: $p_{corrected} = 1-(1-\alpha)^{\frac{1}{m}}$

We want to show that $p_{corrected} \leq p_{uncorrected}$

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

$-(1-\alpha)^{\frac{1}{m}} \leq -(1-\alpha)$ Multiplying both sides by -1, we reverse the sign.

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

Since $m \in [1, \inf]$ and $\alpha \in [0, 1]$, this comparison is always true.

A conservative but robust test: Bonferroni **3. Upper-bounding the probability of at least one rejection.** Show that $\alpha \leqslant P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < \alpha\}] \leqslant m\alpha$

Using Boole's inequality we can say that:

$P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < \alpha\}] \leqslant \sum_{i=1}^{m} P(p_i(Y_i) < \alpha)$

From section 1 of the project we know that $P(P(p_i(Y_i) < \alpha) = \alpha$. Therefore, $\sum_{i=1}^{m} P(p_i(Y_i) < \alpha) = m\alpha$.

Hence, we have proved the upper bound $P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < \alpha\}] \leqslant m\alpha$.

The lower bound of α turns into an equality at $P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < \frac{\alpha}{m}\}]$, since, solving for p_{size},

$P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < p_{size}\}] \leqslant \sum_{i=1}^{m} P(p_i(Y_i) < p_{size}) = \alpha$

$mp_{size} = \alpha$

$p_{size} = \frac{\alpha}{m}$

This means that using a threshold probability of $\frac{\alpha}{m}$ for rejecting each individual test ensures that the overall type I error equals α.

Now, if $\alpha = P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < \frac{\alpha}{m}\}]$, for any $p_{size} \geqslant \frac{\alpha}{m}$ such as $p_{size} = \alpha$, $\alpha \leqslant P_{C-H}[\cup_{i=1}^{m}\{p_i(Y_i) < p_{size}\}]$, proving the lower bound.

Ordered p-values, family-wise error rate and a new MT correction **4.** If we reject test i when $p_{(i)} < \frac{i\alpha}{m}$, the probability of at least one false rejection is:

$P_H[\cup_{i=1}^{m}\{y_i < \frac{i\alpha}{m}\}]$

$P_H[\cup_{i=1}^{m}\{y_i < \frac{i\alpha}{m}\}] = 1 - P_H[\cap_{i=1}^{m}\{y_i > \frac{i\alpha}{m}\}]$

$1 - P_H[\cap_{i=1}^{m}\{y_i > \frac{i\alpha}{m}\}] = 1 - (1-\alpha)^m = 1 - (1-\alpha) = \alpha$