Multiple Testing Project: p-value

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 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 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.

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

If $F_H(c_\alpha)$ follows a uniform distribution, $1 - F_H(c_\alpha)$ does too, as proven in the previous exercise.