Project 1 - Statistical modelling and inference

November 2, 2015

Students	**Student IDs**
Max van Esso	73539
Marco Fayet	125593
Felix Gutmann	125604

1

1.1 Excercise 1

To show: For any $\alpha, c_\alpha = F_H^{-1}(1-\alpha)$

Prerequisite definitions:

- By definition: F_H is the distribution under H
- By definition: $\alpha = F_H(x) \leftrightarrow x = F_H^{-1}(\alpha)$
- By definition: α is chosen as: $\alpha = 1 - F_H(c_\alpha)$

Solution:

$$\begin{aligned}
\alpha &= 1 - F_H(c_\alpha) \\
F_H(c_\alpha) &= 1 - \alpha \\
F_H^{-1}\big(F_H(c_\alpha)\big) &= F_H^{-1}(1-\alpha) \\
c_\alpha &= F_H^{-1}(1-\alpha)
\end{aligned}$$

1.2 Excercise 2

To show: The p-value as a function of $\boldsymbol{X}$ is $p(\boldsymbol{X}) = 1 - F_H(T(\boldsymbol{X}))$

Prerequisite definitions:

- By definition: $R_\alpha = (c_\alpha, \infty)$
- By definition: $p(\boldsymbol{X}) = \inf\left\{\alpha : T(\boldsymbol{X}) \in R_\alpha\right\}$

Solution:

By definition the p-value is the infimum of alpha corresponding to a value of $T(\boldsymbol{X})$ as an element of the intervall from c_α and ∞. Hence, we a looking for the probability level for that the value of the test-statistic is larger or equal than the critical value. Combining that with the definition of alpha it follows:

$$1 - \mathbb{P}_H(c_\alpha \leq T(\boldsymbol{X})) = 1 - F_H(T(\boldsymbol{X}))$$

1.3 Excercise 3

To show: $F(y)$ and $1 - F(y)$ are **uniformly** distributed for **continous** distribution functions

Prerequisite definitions:[1]

- The probability distribution of a random variable X is: $F(x) = \mathbb{P}(X \leq x)$
- From that follows: $1 - F(x) = \mathbb{P}(X > x)$
- The quantile function of a random variable X is: $F^{-1}(x) = x$
- Let Y be eqaul to $F(y)$

Solution for $F(y)$:

$$\begin{aligned}
F(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(F(y) \leq y) \\
&= \mathbb{P}(F^{-1}(F(y)) \leq F^{-1}(y)) \\
&= \mathbb{P}(y \leq F^{-1}(y)) \\
&= F(F^{-1}(y)) \\
&= y
\end{aligned}$$

[1] c.f. Wasserman page 20

Solution for $1 - F(y)$:

$$\begin{aligned}
1 - F(y) &= \mathbb{P}(Y > y) \\
&= 1 - \mathbb{P}(Y \leq y) \\
&= 1 - \mathbb{P}(F(y) \leq y) \\
&= 1 - \mathbb{P}(F^{-1}(F(y)) \leq F^{-1}(y)) \\
&= 1 - \mathbb{P}(y \leq F^{-1}(y)) \\
&= 1 - F(F^{-1}(y)) \\
&= 1 - y
\end{aligned}$$

Final conclusion:

The previously presented solution is true for $y \in (0, 1)$. By definition 2.5 for the CDF and definition 2.16 for the quantile function in Wassermann, the CDF is $F(y) = q$, where $q \in (0, 1)$. Hence for both cases y has to be between 0 and 1 to fullfil this definition.

1.4 Excercise 4

To show: P-value under H is uniformly distributed

Prerequisite definitions:

- F_H is a and increasing continous distribution function
- The quantile function F_H^{-1} exists

Solution:

According to the last exercise any random variable with continous and increasing distribution function $1-F(y)$ has also a uniformly distributed CDF (exercise 1.3). $T(\boldsymbol{X})$ is just a transformation of our random variable and therefore it is a random variable. Applying those results to our definition of the p-value (exercise 1.2) and using the above mentioned definition for F_H (excercise sheet, page 2), we see $p(\boldsymbol{X}) = 1 - F_H(T(\boldsymbol{X}))$ is by construction uniformly distributed.

2

2.1 Excercise 1

To show: $\mathbb{P}\big[\bigcap_{i=1}^{m}\{y_i > \alpha\}\big] = (1-\alpha)^m$

Prerequisite definitions:

- $y_1, \dots, y_m$ are **independent** and **uniformly** distributed
- Hence, a test is rejected if $\mathbb{P}\big(y_i \leq \alpha\big) = \alpha$
- The complement, that is one test canot be rejected is therefore $\mathbb{P}\big(y_i > \alpha\big) = (1-\alpha)$
- For independent events the following holds: $\mathbb{P}\big(\bigcap_{i=1}^{n} A_i\big) = \prod_{i=1}^{n}\mathbb{P}(A_i)$[2]

Solution:

Applying the mentioned definitions the solution is as followed:

$$\begin{aligned}\mathbb{P}\Big[\bigcap_{i=1}^{m}\{y_i > \alpha\}\Big] &= \prod_{i=1}^{m}(1-\alpha)\\ &= (1-\alpha)^m\end{aligned}$$

2.2 Excercise 2

To show: Under the complete null rejection at least one test is $1-(1-\alpha)^m$

Prerequisite definitions:

- The complement of an Evenet A is defined as: $\mathbb{P}[A^c] = 1 - \mathbb{P}[A^c]$

Solution:

Recall from the last excersise that the propability of all y are bigger than α is $(1-\alpha)^m$ (under the complete null). Obviously the complement of that event is that at least one of them is equal or smaller than α. Hence, the desired answer is just the complement of the result of the last excersice. Therefore the solution of the problem is:

[2] Wasserman: Definition 1.9

$$
\begin{aligned}
\mathbb{P}\big[\big(\bigcap_{i=1}^{m}\{y_i > \alpha\}\big)^c\big] &= 1 - \mathbb{P}\big[\bigcap_{i=1}^{m}\{y_i > \alpha\}\big] \\
&= 1 - \prod_{i=1}^{m}(1-\alpha) \\
&= 1 - (1-\alpha)^m
\end{aligned}
$$

2.3 Excercise 3

To show: For an overall type I error α an individual test is rejected at $1-(1-\alpha)^{1/m}$

We define an overall level for a type I error of α. To show the result for an indivudal test we start from the point that all test are rejected at α. Notice that we want to define a new level for an individual test:

$$
\begin{aligned}
\mathbb{P}\big[\big(\bigcap_{i=1}^{m}\{y_i > \alpha\}\big)^c\big] &= \alpha \\
&= 1 - \mathbb{P}\big[\bigcap_{i=1}^{m}\{y_i > \alpha\}\big]
\end{aligned}
$$

Since they are by definition independent we can write this as product over all individual events.

$$
\begin{aligned}
&= 1 - \prod_{i=1}^{m}\mathbb{P}(y_i > \alpha) \\
&= 1 - \mathbb{P}(y_i > \alpha)^m
\end{aligned}
$$

In the next step we bring the probability for an individual test to the left side.

$$
\begin{aligned}
\alpha &= 1 - \mathbb{P}(y_i > \alpha)^m \\
\mathbb{P}(y_i > \alpha)^m &= (1-\alpha) \\
\mathbb{P}(y_i > \alpha) &= (1-\alpha)^{1/m}
\end{aligned}
$$

Finally we consider the complement of that event to get the result of rejecting one individual test.

$$
\mathbb{P}(y_i \le \alpha) = 1 - (1-\alpha)^{1/m}
$$

2.4 Excercise 4

The following figure 1 shows all three functions ((1),(2),(3)) with different levels of m and for $\alpha \in [0, 1]$. With m increasing f_1 and f_2 are moving more closely together. However the proposed statement ($f_2 \leq f_1 \leq f_3$) in the exercise is true in the intervall between 0 and 1.

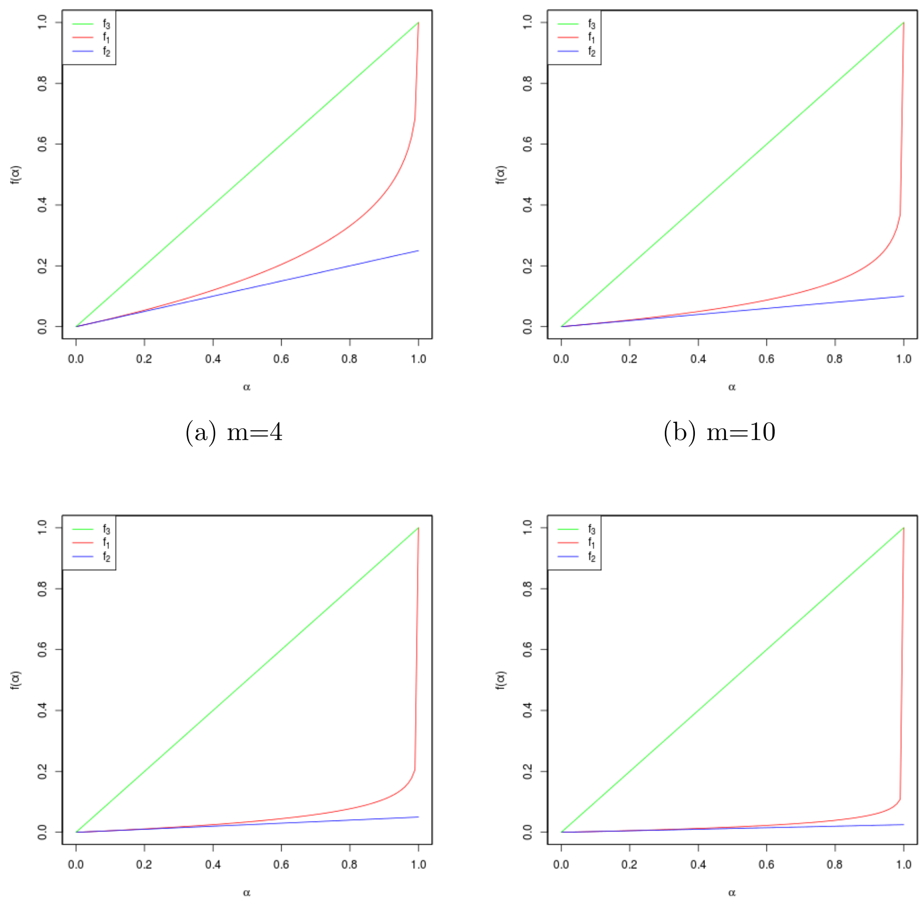
Functions :

$$f_1(\alpha) = 1 - (1 - \alpha)^{1/m} \tag{1}$$

$$f_2(\alpha) = \alpha/m \tag{2}$$

$$f_3(\alpha) = \alpha \tag{3}$$

Figure 1:



(a) m=4

(b) m=10

(c) m=20

(d) m=40

2.5 Excercise 6

To show: Conclusion

Recalling all results of the last exercises so far we can see that with increasing, that we have to correct the rejection for each test for something smaller than α. In the previous exercises we showed a possible correction. In addition figure 1 depicts the importance of such a correction, especially if ma gets larger.

3

3.1 Excercise 1

To show: $\alpha \leq \mathbb{P}_{C-H}\big[\bigcup_{i=1}^{m}\{p_i(\boldsymbol{Y}_i) < \alpha\}\big] \leq m\alpha$

To show the proposed inequality we recall the our result of the probability of at least one rejection **under the complete null**, which is $1-(1-\alpha)^m$ (exercise 2.2). Furthermore, we notice that $\alpha \in (0,1)$ and $m \leq 1$. Suppose m=1, the propability of this event becomes α. Clearly both inequalities hold. Hence, we notice that the lower bound has to be α

$$\begin{aligned}\mathbb{P}_{C-H}\big[\bigcup_{i=1}^{1}\{p_i(\boldsymbol{Y}_i) < \alpha\}\big] &= \mathbb{P}_{C-H}\big[p_1(\boldsymbol{Y}_1) < \alpha\big] \\ &= 1-(1-\alpha) = \alpha\end{aligned}$$

If we have a look on the union term, we can ask, what is the most extreme case in the other direction. Recall according to exercise 2.1, if we reject one test at α. Let's say all m events don't intersect at all, then the union is just the sum of all events m events, which is then $m * \alpha$. For disjoint events the following holds:[3]

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty}\mathbb{P}(A_i)$$

If different events intersect, we have to subtract the intersect term to obtain the union.[4] Hence the last equation is the maximum upperbound. Following that it can be shown:

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty}\mathbb{P}(A_i)$$

We are concluding that we sum over m events, the upper bound by theory has to be $m\alpha$.

[3]c.f. Definition 1.5 in Wasserman
[4]c.f. Lemma 1.6 in Wasserman

Figure 2:

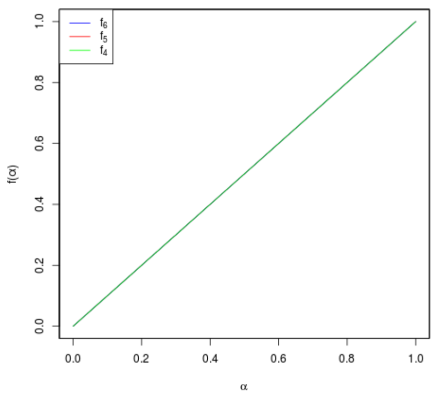
Figure 2 shows all three equations in one plot. Note that all three functions are overlapping for m=1.

Functions :

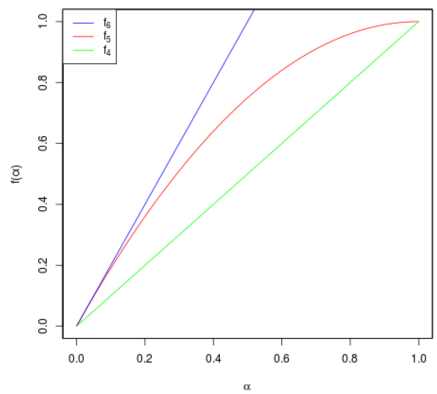
$$f_4(\alpha) = \alpha \tag{4}$$

$$f_5(\alpha) = 1 - (1 - \alpha)^m \tag{5}$$

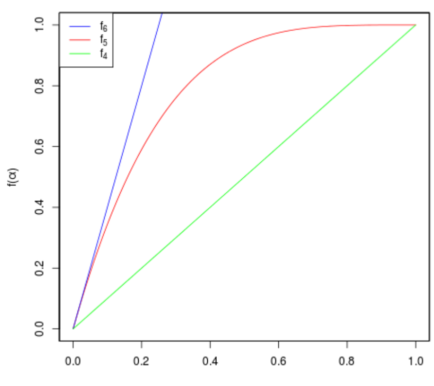
$$f_6(\alpha) = \alpha * m \tag{6}$$



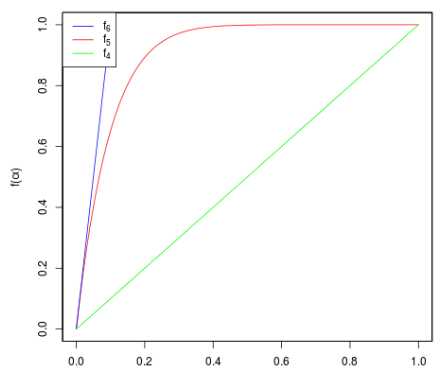
(a) m=1



(b) m=2



(c) m=4



(d) m=10

4

4.1 Excercise 1

To show: The Probability of at least one false rejection is α

Prerequisite definitions:

- $l_i = i\alpha/m$
- $\mathbb{P}\big[\bigcap_{i=1}^{m}\{y_i > l_i\}\big] = 1 - \alpha$
- $y_i \overset{iid}{\sim} Uni(0,1)$
- An individual test will be rejected if $p_{(i)} < i\alpha/m$

Solution

According to the exercise we carry out all test **independently**. Furthermore, we have showed in exercise 1.4 that the p-value is uniformly distributed under H. Connecting this result to the definitions, we notice that the probality of at least one false rejection is the **complement** event of $\mathbb{P}\big[\bigcap_{i=1}^{m}\{p_i > l_i\}\big]$, that is that all tests cannot be rejected. Due to that it follows:

$$\mathbb{P}\big[(\bigcap_{i=1}^{m}\{p_i > l_i\})^c\big] = \alpha$$