Lorenzo Beltrame 08/02/2022 MAT : 12137286

Statistic take home exam

2 test : $X_1, \dots, X_n \sim N(\mu_x, 1)$ i.i.d.

$Y_1, \dots, Y_n \sim N(\mu_y, 1)$ i.i.d.

independently from each other

def $\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, $\bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$

def test Statistic $S = \frac{(\bar{Y}_n - \bar{X}_n)}{\sqrt{\frac{2}{n}}}$

Valid p-values :

- $p_A := 1 - \Phi(S)$ for $H_0 : \mu_y = \mu_x$ against $H_1^A : \mu_y > \mu_x$

- $p_B := 1 - \Phi(-S)$ for $H_0 : \mu_y = \mu_x$ against $H_1^B : \mu_y < \mu_x$

Obj. test $H_0 : \mu_x = \mu_y$ against $H_1 : \mu_y \neq \mu_x$

$$\begin{cases} \text{if } \bar{Y}_n > \bar{X}_n & \text{report } p_A \\ \text{if } \bar{Y}_n < \bar{X}_n & \text{report } p_B \end{cases}$$ [1]

a) Is this a valid p-value?

We can write [1] as:

$$p(x) = \begin{cases} p_A = 1 - \Phi(S(x)) & \text{if } \bar{Y}_n > \bar{X}_n \\ p_B = 1 - \Phi(-S(x)) & \text{if } \bar{Y}_n < \bar{X}_n \end{cases}$$

$$= \begin{cases} 1 - \Phi\left[\dfrac{(\bar{Y}_n - \bar{X}_n)}{\sqrt{\frac{2}{n}}}\right] & \text{if } \bar{Y}_n < \bar{X}_n \\ 1 - \Phi\left[\dfrac{\bar{X}_n - \bar{Y}_n}{\sqrt{2/n}}\right] & \text{if } \bar{Y}_n < \bar{X}_n \end{cases}$$

Since we are changing the sign we can use the abs. value

$$\boxed{p(x) = 1 - \Phi\left[\,|\bar{Y}_n - \bar{X}_n| \cdot \sqrt{\frac{n}{2}}\,\right]}$$

We know that a valid p value satisfyies:

$$\alpha = P_{H_0}(p(x) < \alpha)$$

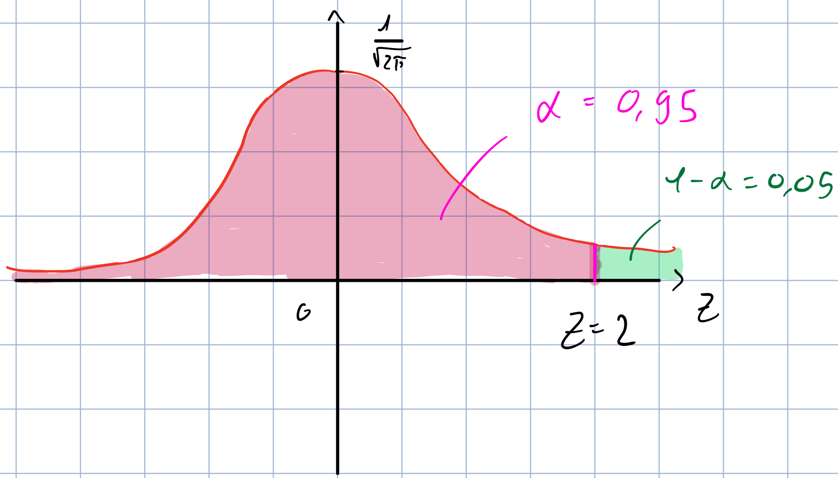
$$= P_{H_0}\left[1 - \Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right) < \alpha\right] =$$

$$= P_{H_0}\left[\Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right) > 1 - \alpha\right] =$$

$$= P_{H_0}\left[|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}} > \Phi^{-1}(1-\alpha)\right] =$$

$f(z) = N(0, 1)$

NOTE: $\alpha \in [0, 1]$



this is the valid p-value for test A

$P_{H_0}(1 - \Phi(s) > \alpha) = \alpha$

$$= P_{H_0}\left[(\bar{Y}_n - \bar{X}_n)\sqrt{\frac{n}{2}} > \Phi^{-1}(1-\alpha) \right] +$$

$$+ P_{H_0}\left[(\bar{Y}_n - \bar{X}_n)\sqrt{\frac{n}{2}} < \Phi^{-1}(1-\alpha) \right] =$$

this is the valid p-value for test B

$P_{H_0}(1 - \Phi(-s) > \alpha) = \alpha$

since p_A and p_B are valid p-values

$$= \alpha + \alpha = 2\alpha$$

We got $\alpha = 2\alpha$ which is valid only for $\alpha = 0$.

$$\Rightarrow P(x) = 1 - \Phi\left[|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}} \right]$$ is not a valid pvalue.

b) Correct the p-value:

I can correct the p value considering

$$p(x) = 2\left[1 - \Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right)\right]$$

And we can verify it:

$$\alpha = P_{H_0}(p(x) < \alpha)$$

$$= P_{H_0}\left(2\left[1 - \Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right)\right] < \alpha\right)$$

$$= P_{H_0}\left(\Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right) > 1 - \frac{\alpha}{2}\right) =$$

$$= P_{H_0}\left((\bar{Y}_n - \bar{X}_n)\sqrt{\frac{n}{2}} > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) +$$

this is the valid p-value for test A

$P_{H_0}(1 - \Phi(S) > \beta) = \beta$

$$+ P_{H_0}\left((\bar{X}_n - \bar{Y}_n)\sqrt{\frac{n}{2}} > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)$$

this is the valid p-value for test B

$P(1 - \Phi(-S) > \beta) = \beta$

I can substitute $\frac{\alpha}{2} = \beta$

$$= \beta + \beta = \alpha$$

This proves that $p(x) = 2\left[1 - \Phi\left(|\bar{Y}_n - \bar{X}_n| \sqrt{\frac{n}{2}}\right)\right]$ is a valid p-value.

c) $\Delta_\mu := (\mu_Y - \mu_X)\sqrt{\frac{2}{n}} > 0$

See Exercise_1_C.r file in the submission folder to run the code.

Exercise 2

See the file: Exercise_2.r

It prints all the required graphs

NOTE: When we order our data the median is the point in the $(n//2)+1$ position.

Since we want to compare which is the maximum local sensitivity I want to achieve it I change K entries of my data.
This can be visualize as moving the middle value to 0 or to the max value. This means that the new middle point is one of the adjacent values.
The function get_indices provides the indices of the potential new middle points

NOTE: For each potential new middle point, the local sensitivity is computed. Subsequenty we store the maximum.

Exercise 3

We have the population graph $G = (V, E)$

and we consider an estimation

$$\tau = \sum_{e \in E}' y_e$$

Using the Horowitz Thompson approach:

$$\hat{\tau} = \sum_{e \in E^*}' \frac{y_e}{\tilde{\pi}_e^{(1)}}$$

probability that the edge is sampled

$$\tilde{\pi}_e^{(1)} = P(e \in E^*) \quad [1]$$

We already computed the vertex pair inclusion probability

$$\tilde{\pi}^{(2)}_{\{i,j\}} := P(\text{"vertex pair } \{i,j\} \text{ is sampled"})$$

for all the $\{i,j\} \in V^{(2)}$.

Consider $\hat{\tau} = \sum_{e \in E^*}' \frac{y_e}{\tilde{\pi}_e^{(2)}}$ as the estimator for τ

Prove that it is in general unbiased

pf. to prove that it is not unbiased let's consider its value in expectation:

$$\mathbb{E}(\hat{\tau}) = \mathbb{E}\left(\sum_{e\in E^*}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}}\right) =$$

$$= \mathbb{E}\left(\sum_{e\in E}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}} \mathbb{I}_{E^*}\right) =$$

I can consider the sum of all the $e \in E$ and add the indicator function

$$\mathbb{I}_{E^*} = \begin{cases} 0 & e \notin E^* \\ 1 & \text{else} \end{cases}$$

linearity

$$= \sum_{e\in E}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}} \mathbb{E}(\mathbb{I}_{E^*})$$

in expectation we have that this is equal to $P(e \in E^*)$

$$= \sum_{e\in E}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}} P(e \in E^*)$$

$$\overset{[1]}{=} \sum_{e\in E}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}} \tilde{\pi}_e^{(1)}$$

To have an unbiased estimator we require:

$$\tau := \sum_{e\in E}{}' Y_e = \sum_{e\in E}{}' \frac{Y_e}{\tilde{\pi}_e^{(2)}} \tilde{\pi}_e^{(1)} := \tilde{\tau}$$

which is true only for $\boxed{\tilde{\pi}_e^{(1)} = \tilde{\pi}_e^{(2)}}$

Let's consider an example where $\tilde{\pi}_e^{(1)} \neq \tilde{\pi}_e^{(2)}$.

I choose to consider the unlabeled star sampling:

- $\tilde{\pi}_e^{(1)} = 1 - \binom{N_v - 2}{n}$

- $\tilde{\pi}_e^{(2)} = \dfrac{n(n-1)}{N_v(N_v - 1)}$

where N_v is the number of vertex.

$$1 - \binom{N_v - 2}{n} \neq \frac{n(n-1)}{N_v(N_v-1)}$$

We can see that, in a general setting:

$$\tilde{\pi}_e^{(1)} \neq \tilde{\pi}_e^{(2)} \quad \blacksquare$$