Problem Set 2

Hoang Nguyen, Huy Nguyen

January 8, 2020

Problem 1. QQ-plots

QQ-Plot 1: Laplace distribution with parameter $\sqrt{2}$ (heavier-heavier)
QQ-Plot 2: Uniform distribution on [-$\sqrt{3}$, $\sqrt{3}$] (lighter-lighter)
QQ-Plot 3: standard Gaussian distribution
QQ-Plot 4: exponential distribution with parameter 1 (lighter-heavier)
QQ-Plot 5: e Cauchy distribution (heavier-heavier)

Problem 2.

1. Mark as update

2.
Let A(t) be the cdf for $U_i's$
We have :

$$A(t) = \mathbb{P}(U_i \leq t) = \mathbb{P}\big(F(X_i) \leq t\big)$$

Because F is increasing function:

$$A(t) = \mathbb{P}\big(X_i \leq F^{-1}(t)\big) = F\big(F^{-1}(t)\big) = t$$

Notice that $0 \leq t \leq 1$
Hence, distributions of the $U_i's$ is Uniform(0,1)
Do similarly we get distributions of the $V_i's$ is also Uniform(0,1)

3.
a)

$$T_{n,m} = \sup_{t\in\mathbb{R}} \big|F_n(t) - G_m(t)\big| = \sup_{t\in\mathbb{R}} \big|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < t) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}(Y_i < t)\big|$$

b)

$$T_{n,m} = \sup_{t\in\mathbb{R}} \big|F_n(t) - G_m(t)\big| = \sup_{t\in\mathbb{R}} \big|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i < t) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}(Y_i < t)\big|$$

F, G are increasing function

$$\Rightarrow T_{n,m} = \sup_{t\in\mathbb{R}} \big|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\big(F(X_i) < F(t)\big) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\big(G(Y_i) < G(t)\big)\big|$$

$$= \sup_{t\in\mathbb{R}} \big|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\big(U_i < F(t)\big) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\big(V_i < G(t)\big)\big|$$

H_0 is true $\Rightarrow$ F(t)=G(t)=x ($0 \leq x \leq 1$):

$$T_{n,m} = \sup_{0\leq x\leq 1} \Big| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\big(U_i < x\big) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\big(V_i < x\big)\Big|$$

c)
If H0 is true, the joint distribution of the n + m random variables $U_1, ..., U_n, V_1, ..., V_m$ is the joint distribution of the n + m Uniform(0,1)
d)

$$\begin{aligned} T_{n,m} &= \sup_{0\leq x\leq 1} \Big| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\big(U_i < x\big) - \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\big(V_i < x\big)\Big| \\ &= \sup_{0\leq x\leq 1} \Big| \frac{1}{n}\sum_{i=1}^{n} Ber\big(\mathbb{P}(U_i < x)\big) - \frac{1}{m}\sum_{i=1}^{m} Ber\big(\mathbb{P}(V_i < x)\big)\Big| \\ &= \sup_{0\leq x\leq 1} \Big| \frac{1}{n}\sum_{i=1}^{n} Ber\big(x\big) - \frac{1}{m}\sum_{i=1}^{m} Ber\big(x\big)\Big| \end{aligned}$$

Hence, $T_{n,m}$ is pivotal
f) By CLT, we have

$$\begin{aligned} &\sqrt{n}\big(F_n(t) - F(t)\big) \xrightarrow[n\mapsto\infty]{(d)} \mathbb{N}\big(0, F(t)(1 - F(t))\big) \\ &\Leftrightarrow F_n(t) - F(t) \xrightarrow[n\mapsto\infty]{(d)} \mathbb{N}\big(0, \frac{F(t)(1 - F(t))}{n}\big) \end{aligned}$$

Similarly, we have:

$$Q_m(t) - Q(t) \xrightarrow[m\mapsto\infty]{(d)} \mathbb{N}\big(0, \frac{Q(t)(1 - Q(t))}{m}\big)$$

From this, if H_0 is true ($F(t) = Q(t)$), by subtracting 2 equations above, we have:

$$F_n(t) - Q_m(t) \xrightarrow[n,m\mapsto\infty]{(d)} \mathbb{N}\big(0, \frac{F(t)(1 - F(t))}{n} + \frac{Q(t)(1 - Q(t))}{m}\big)$$

H_0 is true:

$$\begin{aligned} &\Leftrightarrow F_n(t) - Q_m(t) \xrightarrow[n,m\mapsto\infty]{(d)} \mathbb{N}\big(0, \sqrt{\frac{m+n}{mn}} F(t)(1 - F(t))\big) \\ &\Leftrightarrow \sqrt{\frac{mn}{m+n}}\big(F_n(t) - Q_m(t)\big) \xrightarrow[n,m\mapsto\infty]{(d)} \mathbb{N}\big(0, F(t)(1 - F(t))\big) \\ &\Rightarrow \sqrt{\frac{mn}{m+n}} \sup_{t\in\mathbb{R}} \big|F_n(t) - Q_m(t)\big| \xrightarrow[n,m\mapsto\infty]{(d)} \sup_{t\in\mathbb{R}} \big|\mathbb{B}(F(t))\big| \end{aligned}$$

where $\mathbb{B}$ is a Brownian bridge on [0,1]

Define a test $T_n = \sqrt{\frac{mn}{m+n}} \sup_{t\in\mathbb{R}} \big|F_n(t) - Q_m(t)\big|$, by Donsker's theorem, if H_0 is true, then $T_n \xrightarrow[n,m\mapsto\infty]{(d)} Z$ where X has a known distribution

Problem 3.

1. Mark as update

2. We have: R_i, R_j take the values 1,2,..n

   With a,b=1,2,..n :

$$\mathbb{P}(R_i = a, R_j = b) = (n-2)!(\frac{n-2}{n})^{n-2}$$

$$\mathbb{P}(R_i = a) = \frac{1}{n}$$

$$\mathbb{P}(R_j = b) = \frac{1}{n}$$

$$\Rightarrow \mathbb{P}(R_i = a, R_j = b) \neq \mathbb{P}(R_i = a)\mathbb{P}(R_j = b)$$

   Hence, $R_1, ...R_n$ are not independent

3. We have:

   R_i takes the values in (1,2,..,n)

$$\Rightarrow \mathbb{P}(R_i = a) = \begin{cases} \frac{1}{n} & \text{if } a = 1,2,..,n \\ 0, & \text{Otherwise} \end{cases}$$

$$\Rightarrow R \backsim Uniform(1, n)$$

   Similarly:

$$\Rightarrow Q \backsim Uniform(1, n)$$

   Hence, R, Q do not depend on distribution of X_i's, Y_i's

4. If H_0 is true:

$$\mathbb{P}(R, Q) = \frac{1}{n^2}$$

$$\mathbb{P}(R) = \frac{1}{n}$$

$$\mathbb{P}(Q) = \frac{1}{n}$$

$$\Rightarrow \mathbb{P}(R, Q) = \mathbb{P}(R)\mathbb{P}(Q)$$

   Hence, (R_1, R_2,...,R_n) and (Q_1, Q_2,...,Q_n) are independent

5. If H_0 is true, (R_1, R_2,...,R_n, Q_1, Q_2,...,Q_n) are the iid Uniform(1,n) distributon $\Rightarrow$ joint distribution of them does not depend on the distribution of the original sample

6.
$$T_n = \frac{\sum_{i=1}^n (R_i - \bar{R_n})(Q_i - \bar{Q_n})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R_n})^2 \sum_{i=1}^n (Q_i - \bar{Q_n})^2}}$$

$$T_n = \frac{\sum_{i=1}^n R_i Q_i - \sum_{i=1}^n \bar{R_n}(R_i + Q_i) + \sum_{i=1}^n \bar{R_i}\bar{Q_i}}{\sqrt{\sum_{i=1}^n (R_i - \bar{R_n})^2 \sum_{i=1}^n (Q_i - \bar{Q_n})^2}}$$

$$T_n = \frac{12}{n(n^2-1)} \sum_{i=1}^n R_i Q_i - \frac{3(n+1)}{n-1}$$

7. From question 3 we have R, Q do not depend on distribution of X_i's, Y_i's

   $\Rightarrow$ R_i' and Q_i' are also Uniform(1, n)

   $\Rightarrow$ S_n is the same distribution as T_n