Problem Set 2

Hoang Nguyen, Huy Nguyen

December 30, 2019

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.

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

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

Because F is increasing function:

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

Notice that $0 \le t \le 1$
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)

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

Hence, $T_{n,m}$ is pivotal

4. $(\mathbb{R}, (N(\mu, \sigma^2))_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+})$. These parater are identified.

5.
$$\mathbb{P}(N(\mu, \sigma^2) > 0) = \mathbb{P}\Big(N(0, 1) > \frac{-\mu}{\sigma^2}\Big) = \phi(\frac{\mu}{\sigma^2})$$

Hence, the statistical model is: $(\{0, 1\}, (Ber(\phi(\frac{\mu}{\sigma^2}))_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+})$. This model depends on $\frac{\mu}{\sigma^2}$ $\Rightarrow$ these parameters are not identified.

6. Same for 3.

7. Let X $\sim Exp(\lambda) \Rightarrow \mathbb{P}(X > 20) = e^{-20\lambda}$. Hence, the statistical model is:

$$(\{0, 1\}, (Ber(e^{-20\lambda}))_{\lambda>0})$$

This parameter is identified.

8. Let X $\sim Ber(p)$ such that:

$$\begin{cases} X_i = 1 \text{ if machine i has timelife less than 500 days} \\ X_i = 0 \text{ otherwise} \end{cases} \tag{1}$$

Hence:
$$p = \mathbb{P}(X_i = 1) = 1 - e^{-500\lambda}$$

The number of machines that have stopped working before 500 days is a binominal random variable with parameter (67, $1 - e^{-500\lambda}$)
The statistical model is $(\{1, 2, 3, .., 67\}, (Binominal(67, 1 - e^{-500\lambda}))_{\lambda>0})$. This parameter is identified.

Problem 3.

1. By central limit theorem (CLT), we have:

$$\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \sim (N(0, 1))$$

Hence, $(a_n)_{n\in\mathbb{N}}$ can be $\frac{\sqrt{n}}{\sigma}$ and $(b_n)_{n\in\mathbb{N}}$ can be μ.

2. We have: $Z \backsim N(0, 1)$
Hence, $\mathbb{P}[|Z| \leqslant t] = \mathbb{P}[-t \leqslant Z \leqslant t] = \phi(t) - \phi(-t) = \phi(t) - (1 - \phi(t)) = 2\phi(t) - 1 = 2\mathbb{P}[Z \leqslant t] - 1$.

3. From part 1 we get:

$$\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \sim (N(0,1))$$

from part 2 we get:

$$\mathbb{P}[|Z| \leqslant t] = 2\mathbb{P}[Z \leqslant t] - 1$$

Substitution:

$$\mathbb{P}\Big[|\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}| \leqslant t\Big] = 2\mathbb{P}[Z \leqslant t] - 1$$

We have $2\mathbb{P}[Z \leqslant t] - 1 = 0.95 \Rightarrow t = \phi^{-1}(\frac{0.95+1}{2}) = 1.96$.
Hence,

$$\mathbb{P}\Big[|\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}| \leqslant 1.96\Big] = 0.95$$

$$\mathbb{P}\Big[-\frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma} \leqslant 1.96 \leqslant \frac{\sqrt{n}(\bar{X}_i - \mu)}{\sigma}\Big] = 0.95$$

Because X_i is Poisson random variable with parameter λ, so $\mu = \lambda$ and $\sigma = \sqrt{\lambda}$
We get:

$$\mathbb{P}\Big[-\frac{\sqrt{n}(\bar{X}_i - \lambda)}{\sqrt{\lambda}} \leqslant 1.96 \leqslant \frac{\sqrt{n}(\bar{X}_i - \lambda)}{\sqrt{\lambda}}\Big] = 0.95$$

$$\mathbb{P}\Big[\bar{X}_i - \frac{1.96\sqrt{\lambda}}{\sqrt{n}} \leqslant \lambda \leqslant \bar{X}_i + \frac{1.96\sqrt{\lambda}}{\sqrt{n}}\Big] = 0.95$$

We know: $\bar{X}_i \xrightarrow{P} \mathbb{E}[\bar{X}_i] = \lambda$
Hence,

$$\mathbb{P}\Big[\bar{X}_i - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}} \leqslant \lambda \leqslant \bar{X}_i + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}\Big] \geqslant 0.95$$

$\Rightarrow$ L=$[\bar{X}_i - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}, \bar{X}_i + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}]$
4. We can easily see $\min(X_i) \leqslant \bar{X}_i \leqslant \max(X_i)$. Hence, a new interval can be:

$$L = [min(X_i) - \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}, max(X_i) + \frac{1.96\sqrt{\bar{X}_i}}{\sqrt{n}}]$$

Problem 4. We have X_i is IID. Hence, $\mathbb{P}(M_n \leqslant t) = \prod_{n=1}^{n} \mathbb{P}(X_i \leqslant t)$.
By uniform distribution, the CDF of M_n:

$$\mathbb{P}(M_n \leqslant t) = F(t) = \Big(\frac{t}{\theta}\Big)^n$$

Hence, the PDF of M_n is:

$$f(t) = \frac{dF}{dt} = n\theta^{-n}t^{n-1}$$

We can easily get:

$$\mathbb{E}[M_n] = \int_0^\theta tn\theta^{-n}t^{n-1}dt = \frac{n}{n+1}\theta \to \theta \textbf{ as n} \to \infty$$

By Markov's Inequality:

$$\mathbb{P}\Big[|M_n - \theta| > \epsilon\Big] \leqslant \mathbb{P}[M_n - \theta > \epsilon] \leqslant \frac{\mathbb{E}[M_n - \theta]}{\epsilon} = \frac{\mathbb{E}[M_n] - \theta}{\epsilon} \to 0$$

Hence, M_n converages in probility to θ.

2. From part 1 we get: M_n: $\mathbb{P}[M_n \leqslant t] = \left(\frac{t}{\theta}\right)^n$. Hence, CDF of $n(1 - \frac{M_n}{\theta})$ is:

$$P\Big[n(1 - \frac{M_n}{\theta}) \leqslant t\Big] = \mathbb{P}\Big[M_n \geqslant \frac{(n-t)\theta}{n}\Big] = 1 - \Big(\frac{n-t}{n}\Big)^n \to 1 - e^{-t} \textbf{ as n} \to \infty$$

Hence, $n(1 - \frac{M_n}{\theta})$ converages in distribution to an exponential random variable with parameter 1.

3. Let A is an exponential random variable with parameter 1. Because $n(1 - \frac{M_n}{\theta})$ converages in distribution to X, we have:

$$\mathbb{P}\Big[n(1 - \frac{M_n}{\theta}) \leqslant t\Big] \to \mathbb{P}[X \leqslant t] = 1 - e^{-t}$$

$1 - e^{-t} = 0.95 \Rightarrow t = 3$. We have:

$$\mathbb{P}\Big[n(1 - \frac{M_n}{\theta}) \leqslant 3\Big] \to 0.95$$

which is:

$$\mathbb{P}\Big[\theta \leqslant \frac{nM_n}{n-3}\Big] \to 0.95$$

On the other hand, we always have $\theta \geqslant M_n$ (uniform distribution). Hence, we get:

$$\mathbb{P}\Big[M_n \leqslant \theta \leqslant \frac{nM_n}{n-3}\Big] \to 0.95 \textbf{ as n} \to \infty$$

We conclude $L = \Big[M_n, \frac{nM_n}{n-3}\Big] = \Big[M_n, M_n + \frac{3M_n}{n-3}\Big]$.

4. $bias(M_n) = \mathbb{M}_{\ltimes} - \theta = \frac{n}{n+1}\theta - \theta \neq 0$. Hence, M_n is biased.