The National School of Artificial Intelligence

2023-2024

Probability and Statistics
Semestre 3

Name: Corrigé type

Group:

Final Exam
(2h00)

7,75

Problem 1 *A disillusioned examiner distributes the marks randomly. He assigns each candidate a mark chosen at random from the integers $0, 1, 2, \cdots, 19$. He considers that, since perfection is not of this world, mark 20 should not be awarded. He grades the candidates independently of each other. Candidates with a mark of 10 or above, and only those with a mark of 10 or above, are admitted. Candidates who obtain a mark of 0, and only they, will lose the right to take the exam again.*

(Un examinateur désabusé distribue les notes aléatoirement. Il attribue à chaque candidat une note choisie au hasard parmi les entiers $0, 1, 2, \cdots, 19$. Il considère que, la perfection n'étant pas de ce monde, la note 20 ne doit pas être attribuée. Il note les candidats indépendamment les uns des autres. Seront déclarés admis les candidats obtenant un note supérieure ou égale à 10, et eux seulement. Les candidats obtenant la note 0, et eux seulement, perdront le droit de se présenter une nouvelle fois à cet examen.)

1. *Arslane is one of the candidates. Calculate the probability of each of the following events:*

   A : Arslane will be admitted. B : Arslane will lose the right to reapply. C : Arslane will obtain 9 or 10.

   We have $\mathbb{P}(A) = \frac{10}{20} \Rightarrow \boxed{\mathbb{P}(A) = \frac{1}{2}}$ 0,25

   $\boxed{\mathbb{P}(B) = \frac{1}{20}}$ 0,25 and $\mathbb{P}(C) = \frac{2}{20} \Rightarrow \boxed{\mathbb{P}(C) = \frac{1}{10}}$ 0,25

2. *There are exactly 100 candidates. X is the random variable equal to the number of candidates who will be admitted. Y is the random variable equal to the number of candidates who will lose the right to reapply.*

   a. *Give the probability distribution of X, its mathematical expectation and variance.*

   $\boxed{X \leadsto \mathcal{B}(100; \frac{1}{2})}$ 0,5

   $\mathbb{E}[X] = np \Rightarrow \boxed{\mathbb{E}[X] = 50}$ 0,25

   $Var(X) = npq \Rightarrow \boxed{Var(X) = 25}$ 0,25

   b. *Propose a suitable approximation for X and use this approximation to estimate the probability that at least 55 candidates will be admitted.*

   Since $n > 30$, $np > 5$ and $nq > 5$ 0,5
   then we can approximate the binomial distribution $\mathcal{B}(n, p)$ with the normal distribution $\mathcal{N}(np, \sqrt{npq}) = \mathcal{N}(50, 5)$

$\mathbb{P}(X \geq 55) = \mathbb{P}(Z \geq 55) = \cancel{\mathbb{P}(Z}$ where $Z \sim N(50; 5)$

then $\mathbb{P}(X \geq 55) = \mathbb{P}\left(\frac{Z-50}{5} \geq \frac{55-50}{5}\right) = \mathbb{P}(T \geq 1)$ 1

$= 1 - \mathbb{P}(T < 1) = 1 - 0{,}8413 \Rightarrow \boxed{\mathbb{P}(X \geq 55) \simeq 0{,}1587}$

c. *Give the probability distribution of Y, its mathematical expectation and variance.*

$\boxed{Y \sim \mathcal{B}(100; 1/20)}$ 0,5

$\mathbb{E}[Y] = n.p \Rightarrow \boxed{\mathbb{E}[Y] = 5}$ 0,25

$Var(Y) = n.p.q \Rightarrow \boxed{Var(X) = \frac{19}{4}}$ 0,25

d. *Propose a suitable approximation for Y and use this approximation to estimate the probability that there are at most 2 candidates losing the right to run again.*

Since $n = 100 > 30$, $p \leq 0{,}1$ and $np = 5 \leq 10$ then we can approximate the binomial distribution $\mathcal{B}(100; \frac{1}{20})$ with the Poisson distribution $\mathcal{P}(5)$ 0,5

let $T \sim \mathcal{P}(5)$

$\mathbb{P}(Y \leq 2) = \mathbb{P}(T \leq 2) = \mathbb{P}(T=0) + \mathbb{P}(T=1) + \mathbb{P}(T=2)$

$= e^{-5}\left(1 + 5 + \frac{5^2}{2!}\right) = 0{,}1247$ 0,75

$\Rightarrow \boxed{\mathbb{P}(Y \leq 2) = 0{,}1247}$

3. *There are exactly 100 candidates, numbered from 1 to 100. X_i is the random variable equal to the mark that will be awarded to candidate number i with $(1 \leq i \leq 100)$. $\overline{X}_{100}$ is the random variable equal to the average of the 100 marks that will be given.*

a. *Give the expectation and the variance of X_i, $(1 \leq i \leq 100)$.*

We recall that: $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$ *and* $\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$.

We have $\mathbb{E}[X_i] = \frac{\sum i}{20} = \frac{19.20}{2.20} \Rightarrow \boxed{\mathbb{E}[X_i] = 9{,}5}$ 0,5

$Var(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2$

$\mathbb{E}[X_i^2] = \frac{\sum_{i=0}^{19} i^2}{20} = \frac{19.20.39}{20.6} = \frac{19.13}{2}$ 0,5

$\Rightarrow Var(X_i) = \frac{19.13}{2} - (9{,}5)^2 \Rightarrow \boxed{Var(X_i) = 33{,}25}$ 0,25

b. *Give the expectation and the variance of $\overline{X}_{100}$.*

~~We approximate $\sum_{i=1}^{100} X_i$ by the variable $U \sim \mathcal{N}(950, 33.25)$~~

We have $\mathbb{E}[\overline{X}_{100}] = \frac{1}{100} \sum_{i=1}^{100} \mathbb{E}[X_i] \Rightarrow \boxed{\mathbb{E}[\overline{X}_{100}] = 9{,}5}$ 0,5

$Var(\overline{X}_{100}) = \frac{1}{10^4} \sum_{i=1}^{100} Var(X_i) = \frac{33{,}25}{100}$

$\Rightarrow \boxed{Var(\overline{X}_{100}) = 0{,}3325}$ 0,5

13,25 **Problem 2** *A real random variable X is said to follow an exponential distribution with parameter $\lambda\,(\lambda > 0)$, noted $\mathcal{E}(\lambda)$, if the density of the distribution of X is defined by*

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{I}_{[0,\infty[}(x).$$

A. 1. *Determine F_X the cumulative function of X. Find x_0 such that $F_X(x_0) = \frac{1}{2}$.*

* We have $F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x \lambda e^{-\lambda t}\,dt$ if $x \geq 0$

$\Rightarrow \boxed{F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-\lambda x} & \text{if } x \geq 0 \end{cases}}$ 0,5

* $F_X(x_0) = \frac{1}{2} \Rightarrow 1 - e^{-\lambda x_0} = \frac{1}{2} \Leftrightarrow e^{-\lambda x_0} = \frac{1}{2}$

$\Rightarrow -\lambda x_0 = -\ln 2 \Rightarrow \boxed{x_0 = \frac{\ln 2}{\lambda}}$ 0,25

2. *For $k \in \mathbb{N}^*$, calculate $\mathbb{E}[X^k]$. Deduce $Var(X)$.*

We have $\mathbb{E}[X^k] = \int_0^\infty x^k f_X(x)\,dx = \lambda \int_0^\infty x^k e^{-\lambda x}\,dx$

Let $u = \lambda x \Rightarrow \mathbb{E}[X^k] = \frac{1}{\lambda^k} \int_0^\infty u^k e^{-u}\,du = \frac{k!}{\lambda^k}$.

$\boxed{\mathbb{E}[X^k] = \frac{k!}{\lambda^k}}$ 0,5

$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2!}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \Rightarrow \boxed{Var(X) = \frac{1}{\lambda^2}}$ 0,25

3. *Determine φ_x the characteristic function of X.*

We have $\varphi_X(t) = \mathbb{E}[e^{itX}] = \lambda \int_0^{\infty} e^{itx} e^{-\lambda x} dx$

$= \lambda \int_0^{\infty} e^{(it-\lambda)x} dx = \lambda \left[\frac{e^{(it-\lambda)x}}{it-\lambda} \right]_0^{\infty}$

$\Rightarrow \boxed{\varphi_X(t) = \frac{\lambda}{\lambda - it}}$ 0.5

4. *For $x > 0$ and $y > 0$, calculate $\mathbb{P}(X > x + y | X > x)$.*

$\mathbb{P}(X > x+y / X > x) = \frac{\mathbb{P}(\{X > x+y\} \cap \{X > x\})}{\mathbb{P}(X > x)} = \frac{\mathbb{P}(X > x+y)}{1 - F_X(x)}$

$= \frac{1 - F_X(x+y)}{1 - F_X(x)} = \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} = e^{-\lambda x} = 1 - F_X(y)$

$\Rightarrow \boxed{\mathbb{P}(X > x+y / X > x) = \mathbb{P}(X > y)}$ 0.5

5. *Determine the distribution of $[X]$, where $[X]$ is the integer part of X.*

For $k \in \mathbb{N}$, we have $\mathbb{P}([X] = k) = \mathbb{P}(k \leq X < k+1)$

$= F_X(k+1) - F_X(k) = (1 - e^{-\lambda(k+1)}) - (1 - e^{-\lambda k}) = e^{-\lambda k} - e^{-\lambda k - \lambda}$

$\Rightarrow \boxed{\mathbb{P}([X] = k) = (1 - e^{-\lambda}) e^{-\lambda k}}$ 0.5

6. *Determine the distribution of $Z = X - [X]$.*

We remark that the values of Z are in $[0,1[$. For $z \in [0,1[$ we have:

$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}\left(\bigcup_{k=0}^{\infty} \{X \in [k, k+z[\}\right) = \sum_{k=0}^{\infty} \mathbb{P}(X \in [k, k+z[)$

$= (1 - e^{-\lambda z}) \sum_{k=0}^{\infty} e^{-\lambda k} = (1 - e^{-\lambda z}) \cdot \frac{1}{1 - e^{-\lambda}}$ 0.5

$\Rightarrow \boxed{F_Z(z) = \frac{1 - e^{-\lambda z}}{1 - e^{-\lambda}}}$

B. *Let X_1 and X_2 be two independent real random variables with respective distributions: $\mathcal{E}(\lambda)$ and $\mathcal{E}(\mu)$. We set:*

$$S = X_1 + X_2, U = \frac{X_1}{S}, V = \frac{X_2}{S}, T = \sup\{U, V\} \text{ and } W = \inf\{U, V\}.$$

1. *Determine the distribution of (S, U).*

We remark that, since X_1 and X_2 are independent then: $f_{X_1,X_2}(x_1,x_2) = \lambda\mu\, e^{-\lambda x_1 - \mu x_2}\, \mathbb{1}_{[0,\infty[}(x_1) \mathbb{1}_{[0,\infty[}(x_2)$

and U and V are in $]0,1[$. 0.25

We have $\begin{cases} S = X_1 + X_2 \\ U = \dfrac{X_1}{X_1+X_2} \\ X_1 \geq 0 ; X_2 \geq 0 \end{cases} \iff \begin{cases} X_1 = S.U \\ X_2 = S(1-U) \\ S \geq 0 \; ; \; 0 \leq U \leq 1. \end{cases}$

The Jacobian of this change of variable is:

$$J_{\varphi^{-1}}(S,U) = \begin{vmatrix} u & s \\ 1-u & -s \end{vmatrix} = -su - (1-u)s = -s.$$

then $f_{S,U}(s,u) = |-s| \, f_{X_1,X_2}(us, s(1-u))$

$= s \, \lambda\mu \, e^{-\lambda us - \mu s(1-u)}$ 1

$$\Rightarrow \boxed{f_{S,U}(s,u) = \lambda\mu s \, e^{-\lambda us - \mu s(1-u)} \mathbb{1}_{[0,\infty[}(s) \mathbb{1}_{[0,1]}(u)}$$

2. *Determine f_U the marginal distribution of U. Are S and U independent?*

$\forall u \in [0,1]$: $f_U(u) = \lambda\mu \int_0^{\infty} s e^{-\lambda us - \mu s(1-u)} ds$

$$\Rightarrow \boxed{f_U(u) = \begin{cases} 1 & \text{if } \lambda = \mu \\ \dfrac{\lambda\mu}{[\mu + (\lambda-\mu)u]^2} & \text{if } \lambda \neq \mu \end{cases}}$$ 0.5

We deduce that if $\lambda = \mu$ 0.5 S and U are independent and if $\lambda \neq \mu$ S and U are not independent.

0.5

3. *Express f_T the density of the distribution of T as a function of f_U.*

We remark that $U+V=1$. Let us determine F_T the cumulative function of T. We have

$F_T(t) = \mathbb{P}(U \leq t ; V \leq t) = \mathbb{P}(U \leq t ; 1-U \leq t)$

$= \mathbb{P}(1-t \leq U \leq t) = \int_{1-t}^{t} f_U(u)\,du$

$$= \begin{cases} 0 & \text{if } t \leq 1/2 \\ F_U(t) - F_U(1-t) & \text{if } \frac{1}{2} \leq t < 1 \\ 1 & \text{if } t \geq 1 \end{cases}$$

1

Then $\boxed{f_T(t) = [f_U(t) + f_U(1-t)]\, \mathbb{1}_{]\frac{1}{2},1[}(t)}$

4. *Express f_W the density of the distribution of W as a function of f_U.*

We have $F_W(w) = 1 - \mathbb{P}(U \geq w ; V \geq w)$

$= 1 - \mathbb{P}(U \geq w ; 1-U \geq w) = 1 - \mathbb{P}(w \leq U \leq 1-w)$

$$= \begin{cases} 0 & \text{if } w \leq 0 \\ 1 + F_U(w) - F_U(1-w) & \text{if } 0 < w < 1/2 \\ 1 & \text{if } w \geq 1/2 \end{cases}$$

1

Then: $\boxed{f_W(w) = [f_U(w) + f_U(1-w)]\, \mathbb{1}_{]0,1/2[}(w)}$

C. *Let $Y_1, \cdots, Y_n$ be independent real random variables with the same distribution $\mathcal{E}(\lambda)$.*

1. *Determine φ_{S_n} the characteristic function of $S_n = Y_1 + \cdots + Y_n$. Deduce $\mathbb{E}[S_n]$ and $Var(S_n)$.*

$\varphi_{S_n}(t) = \mathbb{E}[e^{itS_n}] = \mathbb{E}[e^{it(Y_1+\cdots+Y_n)}] = \prod_{k=1}^{n} \mathbb{E}[e^{itY_k}]$

$= \prod_{k=1}^{n} \varphi_{Y_n}(t) = \left(\frac{1}{1-i\frac{t}{\lambda}}\right)^n \Rightarrow \boxed{\varphi_{S_n}(t) = \left(\frac{1}{1-i\frac{t}{\lambda}}\right)^n}$ 1

We deduce : $\varphi'_{S_n}(t) = i\,\mathbb{E}[S_n e^{itS_n}] = \frac{in}{\lambda}\left(\frac{1}{1-i\frac{t}{\lambda}}\right)^{n+1}$

$\Rightarrow \varphi'_{S_n}(0) = i\,\mathbb{E}[S_n] = \frac{in}{\lambda}$

$\varphi''_{S_n}(t) = -\mathbb{E}[S_n^2 e^{itS_n}] = -\frac{n(n+1)}{\lambda^2}\left(\frac{1}{1-i\frac{t}{\lambda}}\right)^{n+2} \Rightarrow \varphi''_{S_n}(0) = -\mathbb{E}[S_n^2] = -\frac{n(n+1)}{\lambda^2}$.

Then $\mathbb{E}[S_n] = \frac{n}{\lambda}$ and $Var(S_n) = \frac{n}{\lambda^2}$.

05 05

2. *Determine the distibution of* $Y_{(1)} = \inf\{Y_1, Y_2, \cdots, Y_n\}$.

We have $F_{Y_{(1)}}(x) = \mathbb{P}(Y_{(1)} \le x) = 1 - \mathbb{P}(Y_{(1)} > x)$

$= 1 - (1 - F_{Y_{(1)}}(x))^n$

$= \begin{cases} 0 & \text{if } x \le 0 \\ 1 - e^{-n\lambda x} & \text{if } x > 0 \end{cases}$ 05

We deduce that $\boxed{Y_{(1)} \rightsquigarrow \mathcal{E}(n\lambda)}$

D. *Let* X *be a random variable following the* $\mathcal{E}(\lambda)$ *distribution and* Y *a random variable following, conditionally to* $\{X = x\}$, *a uniform distribution on* $[x, 1+x]$

1. *Calculate* $\mathbb{E}[Y]$ *and* $Var(Y)$.

We have $f_Y(y/x) = \mathbb{1}_{[x, x+1]}(y)$

$\mathbb{E}[Y/X=x] = \int_{-\infty}^{+\infty} y\, f_Y(y/x)\,dy = \int_x^{x+1} y\,dy = \frac{1+2x}{2}$

$\mathbb{E}[Y^2/X=x] = \int_{-\infty}^{+\infty} y^2 f_Y(y/x)\,dy = \frac{1+3x+3x^2}{3}$

We have also : $\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\,dx = \lambda\int_0^{+\infty} x e^{-\lambda x}\,dx = \frac{1}{\lambda}$

and $\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)\,dx = \frac{1}{\lambda^2}$

Using $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y/X]]$ and $\mathbb{E}[Y^2] = \mathbb{E}[\mathbb{E}[Y^2/X]]$

$\Rightarrow \mathbb{E}[Y] = \mathbb{E}\left[\frac{1+2X}{2}\right] = \frac{1+2\mathbb{E}[X]}{2} = \frac{1}{2} + \frac{1}{\lambda}$

and $\mathbb{E}[Y^2] = \mathbb{E}\left(\frac{1+3X+3X^2}{3}\right) = \frac{1}{3} + \frac{1}{\lambda} + \frac{1}{\lambda^2} = \left(\frac{1}{2} + \frac{1}{\lambda}\right)^2 + \frac{1}{12}$

1

$\Rightarrow \boxed{Var(Y) = \frac{1}{12}}$

2. *Determine the density of the distribution of Y.*

We have: $F_Y(y) = \mathbb{P}(Y \le y) = \int_{-\infty}^{+\infty} \mathbb{P}(Y \le y / x)\, f_X(x)\, dx$

$$\mathbb{P}(Y \le y/x) = \begin{cases} 0 & \text{if } y \le x \\ y & \text{if } x < y < 1+x \\ 1 & \text{if } y \ge 1+x \end{cases}$$

Thus $\mathbb{P}(Y \le y/x) = \mathbb{1}_{\{y \ge 1+x\}} + y\,\mathbb{1}_{\{x<y<1+x\}} = \mathbb{1}_{\{x \le y-1\}} + y\,\mathbb{1}_{\{y-1<x<y\}}$

$$\Rightarrow F_Y(y) = \int_{-\infty}^{+\infty} \left[\mathbb{1}_{\{x \le y-1\}} + y\,\mathbb{1}_{\{y-1<x<y\}}\right] f_X(x)\, dx$$

$$= \begin{cases} y\left(1 - e^{-\lambda y}\right) & \text{if } 0<y<1 \\ 1 - e^{-\lambda(y-1)} + y\left(e^{\lambda}-1\right)e^{-\lambda y} & \text{if } y \ge 1 \end{cases}$$ 1

$$\Rightarrow \boxed{f_Y(y) = \begin{cases} 1 - (1-\lambda y)e^{-\lambda y} & \text{if } 0<y<1 \\ \lambda e^{-\lambda(y-1)} + \left(e^{\lambda}-1\right)(1-\lambda y)e^{-\lambda y} & \text{if } y \ge 1. \end{cases}}$$

3. *Let $Y_1, \cdots, Y_n$ be independent random variables with the same distribution as Y.*

a. *Show that $T_n = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \frac{1}{2}\right)$ converges in probability to a random variable to be determined.*

For $i = 1, \dots, n$, let $U_i = Y_i - \frac{1}{2} \Rightarrow \mathbb{E}[U_i] = \frac{1}{\lambda}$ and $Var(U_i) = \frac{1}{12}$

Since $Y_1, Y_2, \dots, Y_n$ are independent, the variables $U_1, \dots, U_n$ are also independent.

Then $\mathbb{E}[T_n] = \frac{1}{\lambda}$ and $Var(T_n) = \frac{1}{12n}$.

By Tchebychev inequality: $\forall \varepsilon > 0$, $\mathbb{P}(|T_n - \mathbb{E}[T_n]| > \varepsilon) \le \frac{Var(T_n)}{\varepsilon^2}$

$\Rightarrow \forall \varepsilon > 0$: $\mathbb{P}(|T_n - \frac{1}{\lambda}| > \varepsilon) \le \frac{1}{12 n \varepsilon^2}$. Then $\boxed{T_n \xrightarrow[n\to\infty]{\mathbb{P}} \frac{1}{\lambda}}$ 0.5

b. *Show that $W_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Y_i - \frac{1}{2} - \frac{1}{\lambda}\right)$ converges in law to a random variable with normal distribution $\mathcal{N}(\mu, \sigma)$, where μ and σ are constants to be determined.*

For $i = 1, \dots, n$ let $V_i = Y_i - \frac{1}{2} - \frac{1}{\lambda}$ and $S_n = V_1 + \dots + V_n$.

We have then $\mathbb{E}[V_i] = 0$ an $Var(V_i) = \frac{1}{12}$

Since $Y_1, \dots, Y_n$ are independent, so $V_1, \dots, V_n$ are also independent then: $\mathbb{E}[S_n] = 0$ and $Var(S_n) = \frac{n}{12}$

By the central limit theorem, we have:

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{Var(S_n)}} = \frac{\sqrt{12}\, S_n}{\sqrt{n}} = \sqrt{12}\, W_n \xrightarrow[n\to\infty]{Law} Z \sim \mathcal{N}(0,1)$$

Then $\boxed{W_n \xrightarrow[n\to\infty]{law} \mathcal{N}\left(0, \frac{1}{12}\right)}$ 0.5