Appendix G — Previous exams

G.1 Partial 2022-2023

Exercise 1

Let A and B be events such that $P(A) = 0.2$ and $P(B) = 0.3$. Find the probability that at least one of the events A and B occurs if

1. A and B are mutually exclusive.
2. A and B are independent.

Exercise 2

An urn contains one ball of number -1, two of number 0 and two of number 4. We pick randomly two balls from this urn.

1. Determine the probability distribution of X which is the sum of the obtained numbers.
2. Calculate $E(X)$.

Exercise 3

Let (X, Y) a pair of random variables with the following joint distribution:

$X\backslash Y$	-1	0	1
0	1/9	2/9	1/9
1	1/9	2/9	1/9
2	0	1/9	0

1. Calculate the marginal distributions of X and Y.
2. Calculate the probability that Y is greater of equal to zero given that X is 0.
3. Calculate the covariance of X and Y.
4. Are X and Y independent ?

Exercise 4

A box contains three transistors A, B and C. Transistor A was made by a machine that produces 3% defected transistors, transistor B by a machine that produces 5% defected transistors, and transistor C by a machine that produces 7% defected transistors.

1. Let p the probability that a randomly chosen transistor is defected. Show that $p = 0.05$.

2. Ten transistors are chosen randomly (with replacement) from the box. Let N be the number of defected transistors. What is the distribution of N? (The answer must be justified and detailed).

3. Calculate $P(N = 0)$.

4. What is the expected value of N.

Exercise 5

Let X a random variable following a Normal distribution $\mathcal{N}(4, 2^2)$.

1. Calculate $P(X < 3)$.

2. Find x_0 such that $P(X < x_0) = 0.9$.

Now let X following a Uniform distribution on $[0; 8]$.

1. Calculate the density function of $Y = e^X$.

2. Calculate the expected value of Y^2.

Exercise 6

A Poisson random variable X of parameter $\lambda > 0$ has

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad \text{if } k \in \mathbb{N}$$

Let $(X_n)_{n\in\mathbb{N}^*}$, independent random variables, each follows Poisson distribution of parameter 1. For every n of $\mathbb{N}^*$, let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S_n = X_1 + \ldots + X_n$.

1. Apply the Central Limit Theorem (CTL) in this case to deduce the probability distribtion of $\overline{X}_n$ when n is big.

2. What is the probability distribution of S_n if n is big.

3. Deduce that $\lim_{n\to+\infty} P(S_n \leq n) = \frac{1}{2}$.

Exercise 7

Let X a random variable of density function $f(x) = (\theta + 1)x^\theta \times \mathbb{1}_{]0,1[}(x)$. What is the Maximum Likelihood Estimator of θ ?

Exercise 8

Let X_1 and X_2 two independent random variables of expected value μ and variance σ^2. Let the following estimators of μ:

$$\theta_1 = \frac{X_1 + X_2}{2} \quad \text{and} \quad \theta_2 = \frac{X_1 + 3X_2}{4}$$

1. Check if these estimators are biased.

2. Which one between θ_1 and θ_2 is better to estimate μ?

Solution (Click to expand)

Maths for DS
CC 23-09-2022

(2)

Ex1 $P(A) = 0,2 \quad P(B) = 0,3$

1) $P(A\cup B) = P(A) + P(B) - \underbrace{P(A\cap B)}_{=\phi} = 0,5.$ 1

2) $P(A\cup B) = P(A) + P(B) - P(A\cap B) = P(A) + P(B) - P(A)P(B) =$ 1
$0,2 + 0,3 - 0,06 = 0,44.$

(3)

Ex2 $|\Omega| = \binom{5}{2} = 10$

$1\times$ (-1)
$2\times$ (0)
$2\times$ (4)

1) $X(\Omega) = \{-1, 0, 3, 4, 8\}$ 1

x_i	-1	0	3	4	8	
$P(X=x_i)$	$2/10$	$1/10$	$2/10$	$4/10$	$1/10$	1

1

* $P(X=-1) = \dfrac{\binom{1}{1}\binom{2}{1}}{\binom{5}{2}} = 2/10$ * $P(X=0) = \binom{2}{2}/\binom{5}{2} = 1/10 \ldots$

2) $E(X) = \sum x_i P(X=x_i) = \dfrac{-2+6+16+8}{10} = 28/10 = 2,8.$ 1

(4)

Ex3

$X\backslash Y$	-1	0	1	Marg. de X
0	1/9	2/9	1/9	4/9
1	1/9	2/9	1/9	4/9
2	0	1/9	0	1/9
Marg. de Y	2/9	5/9	2/9	1

1) Voir en tableau 1

2) $P(Y\geq 0 / X=0) = \dfrac{P(X=0 ; Y\geq 0)}{P(X=0)} = \dfrac{P(X=0, Y=0) + P(X=0, Y=1)}{P(X=0)}$

$= \dfrac{2/9 + 1/9}{4/9} = 3/4.$ 1

3) $Cov(X,Y) = E(XY) - E(X)E(Y)$

* $E(XY) = \sum_i\sum_j x_i y_j P(X=x_i, Y=y_j) = (0)(-1)(1/9) + \ldots = 0$
* $E(X) = \sum x_i P(X=x_i) = 6/9 = 2/3.$
* $E(Y) = \sum y_j P(Y=y_j) = 0$ donc $Cov(X,Y) = 0$ 1

(un exemple suffit pour prouver)

4) Non , $P(X=0, Y=0) = 2/9 \neq P(X=0)P(Y=0) = \frac{4}{9} \times \frac{5}{9}$ 1

4

Ex 4

1) $p = P(\text{"transistor tiré est défectueux"}) = P(\text{"déf et A" ou "déf et B" ou "déf et C"})$

$= \frac{1}{3} \times 0{,}03 + \frac{1}{3} \times 0{,}05 + \frac{1}{3} \times 0{,}07 = 0{,}05.$ 1

2) * transistor tiré est défectueux (succès) ou Non ⇒ épreuve de Bernoulli
* répétiti 10 fois (n) de l'épreuve.
* les épreuves sont indépendantes (car remise). 1,5
* On s'intéresse au nb de succès parmi n.

Donc $N \sim B(10; 0{,}05)$

$\forall\, 0 \le k \le 10,\ P(N=k) = C_{10}^{k} (0{,}05)^{k} (0{,}95)^{10-k}.$

3) $P(N=0) = C_{10}^{0} (0{,}05)^{0} (0{,}95)^{10-0} = (0{,}95)^{10} \simeq 0{,}5987.$ 0,5

4) $E(N) = n \times p = 10 \times 0{,}05 = 0{,}5.$ 1

4

Ex 5

$X \sim \mathcal{N}(4, 2^2)$

1) $P(X<3) = P(\underbrace{\frac{X-4}{2}}_{\to \mathcal{N}(0,1)} < \frac{3-4}{2}) = P(Z < -0{,}5) = \Phi(-0{,}5)$

$= 1 - \Phi(0{,}5) = 1 - 0{,}6915 = 0{,}3085$ 1

2) $P(X < x_0) = 0{,}9$ So $P(Z < \frac{x_0 - 4}{2}) = 0{,}9$

$\Phi(\frac{x_0-4}{2}) = 0{,}9$ So $\frac{x_0 - 4}{2} = 1{,}285 \Rightarrow x_0 = 6{,}57$ 1

$X \sim U([0, 8])$

1) $Y = e^X$ So $Y(\Omega) = [1, e^8]$

* $F_Y(y) = \begin{cases} 0 \text{ si } y \le 1 \\ 1 \text{ si } y \ge e^8 \\ \text{et si } y \in [1, e^8]: \\ P(Y \le y) = P(e^X \le y) = P(X \le \ln y) = \int_1^{\ln y} 1/8\, dx = \frac{1}{8} \ln y - \frac{1}{8} \end{cases}$ 1

* $f_Y(y) = 1/8y \times \mathbb{1}_{[1, e^8]}(y)$

2) $E(Y^2) = \int_{\mathbb{R}} y^2 f_Y(y)\,dy = \int_1^{e^8} y^2 \cdot \frac{1}{8y}\,dy = \frac{1}{8}\left[\frac{y^2}{2}\right]_1^{e^8}$

$= \frac{1}{8}\left(\frac{e^{16}}{2} - \frac{1}{2}\right) = \frac{e^{16}}{16} - \frac{1}{16} = \frac{1}{16}(e^{16}-1).$ 1

3

Ex6 $X_i \to \mathcal{P}(1)$ $\overline{X_n} = \frac{1}{n}\sum X_i$ $S_n = \sum X_i$

1) TCL: $\overline{X_n} \to \mathcal{N}(E(\overline{X_n}), V(\overline{X_n}))$

On $E(\overline{X_n}) = 1$ et $V(\overline{X_n}) = 1/n$ 1

So $\overline{X_n} \sim \mathcal{N}(1, 1/n)$

2) $S_n = n\overline{X_n}$ So $S_n \to \mathcal{N}(E(n\overline{X_n}), V(n\overline{X_n}))$

* $E(S_n) = nE(\overline{X_n}) = n$ * $V(S_n) = V(n\overline{X_n}) = n^2 V(\overline{X_n})$
$= n^2 \times 1/n = n.$

So $S \sim \mathcal{N}(n, n)$ when $n >>>$ 1

3) $\lim_n P(S_n \le n) = \lim P\Big(\underbrace{\frac{S_n - n}{\sqrt{n}}}_{\mathcal{N}(0,1)} \le \frac{n-n}{\sqrt{n}}\Big) = \Phi(0) = 1/2.$ 1

3

Ex7 $f(x) = (\theta+1)x^\theta \times \mathbb{1}_{]0,1[}(x)$

* $L(\theta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n (\theta+1)x_i^\theta = (\theta+1)^n \prod_{i=1}^n x_i^\theta$ 1

* $\ln L(\theta) = n\ln(\theta+1) + \sum_{i=1}^n \theta \ln x_i = n\ln(\theta+1) + \theta\sum_{i=1}^n \ln x_i$ 1

* $\frac{\partial \ln L(\theta)}{\partial\theta} = \frac{n}{\theta+1} + \sum_{i=1}^n \ln x_i = 0$ So

$\frac{n}{\hat\theta+1} = -\sum \ln x_i \Rightarrow \hat\theta + 1 = -\frac{n}{\sum \ln x_i}$

$\hat\theta_{MLE} = -1 - \frac{n}{\sum \ln x_i}$ 1

3

Ex 8 $\Theta_1 = \frac{X_1+X_2}{2} = \overline{X_2}$ $\Theta_2 = \frac{X_1+3X_2}{4}$

1 1) $\star E(\Theta_1) = E\left(\frac{X_1+X_2}{2}\right) = \frac{1}{2}(E(X_1)+E(X_2)) = \frac{1}{2}(2\mu) = \mu$ non biasé

$\circ E(\Theta_2) = E\left(\frac{X_1+3X_2}{4}\right) = \frac{1}{4}E(X_1) + \frac{3}{4}E(X_2) = \mu$ non biasé.

2) $\circ MSE(\Theta_1) = V(\Theta_1) + \underbrace{bias^2(\Theta_1)}_{=0} = V(\Theta_1) = \frac{1}{4}\overbrace{(V(X_1)+V(X_2))}^{X_1 \perp X_2 \text{ ind.}}$

1 $= \frac{1}{4}(2\sigma^2) = \sigma^2/2$

$\circ MSE(\Theta_2) = V(\Theta_2) + \underbrace{bias^2(\Theta_2)}_{=0} = V(\Theta_2) = \frac{1}{16}V(X_1) + \frac{9}{16}V(X_2)$

$= \frac{10}{16}\sigma^2 > \sigma^2/2 = V(\Theta_1)$

1 So Θ_1 is better.

G.2 Final Exam 2022-2023

Exercise 1

Let X be the duration of use (in hours) of a computer terminal during a working day.

1. *Suppose that X follows a Normal distribution $\mathcal{N}(4, 2)$.*
    1. *Calculate $P(X < 3)$.*
    2. *Calculate the number x_0 such that, for 90% of the days, X is less than x_0.*
2. *Suppose that X follows a Uniform distribution on $[0; 8]$.*
    1. *Determine the density function of $Y = e^X$.*
    2. *Calculate the expected value of Y^2.*

Exercise 2

Some birds fly after making a few jumps on the ground. It is assumed that the number X of jumps can be modeled by a Pascal (Geometric) distribution on $\mathbb{N}^*$:

$$P(X = x) = p(1-p)^{x-1} \quad x \geq 1$$

For $n = 130$ birds of this type, we collected the following data:

Number of jumps x	1	2	3	4	5	6	7	8	9	10	11	12
Occurence	48	31	20	9	6	5	4	2	1	1	2	1

1. *What is the Maximum Likelihood Estimator (MLE) of p?*
2. *Calculate a value of this estimator using the collected data.*

Exercise 3

Nine pallets of bricks of the same manufacture were weighed. We obtained an average of 500 kg and a variance of 16.

We admit that the random variable "Weight of a pallet of bricks" follows a Normal distribution of expectation μ and variance σ^2.

1. *Give an unbiased estimators of μ and σ^2.*
2. *Construct a confidence interval of μ of 90% confidence.*
3. *Now suppose that we know the variance, which is given by the constructor σ^2 = 25. Recalculate the 90% confidence interval of μ.*
4. *In this case, what confidence level to choose if we want to have a confidence interval of length 10.*

Exercise 4

A car manufacturer called A decides to market 100% electric cars. He claims that the average range (autonomy) of its new model is 200 km. The European standard considers a vehicle to be "compliant" if its autonomy is greater than 175 km. A competitor B would like to verify, on average, the autonomy of manufacturer A's cars. He rents 16 cars. Let take X the random variable corresponding to the range of each tested car. We will suppose that X follows *a Normal distribution of mean μ and variance σ^2. With the sample data, the manufacturer obtains a mean range of 190 km and an unbiased estimate of the variance of 9.*

1. *Describe the random experiment, the population, and the studied random variable.*
2. *What does the parameter μ represent? Give an estimator and then an estimation of μ.*
3. *Test the hypothesis that the average range is 200 km with confidence 95%.*
4. *Explain how to answer the previous question with a two-sided confidence interval for μ.*
5. *Can we say, with a 2.5% risk of error, that the average range of this model is strictly over 175 km..*
6. *Give a confidence interval at 10% for the variance σ^2.*

   The proportion of cars tested with a range of more than 175 km in this sample is of 0.88.

7. *Describe the new random variable studied Y. What is its distribution?*
8. *Give a point estimate of the proportion p of cars that are compliant to the European standard.*
9. *Assuming that the sample is large enough to apply the central limit theorem, can we say, with a 5% risk of error, that the proportion p of compliant cars in the entire production is strictly greater than 0.90?*

Solution (Click to expand)

MDS DE 2022-2023

4

Ex1 $X \rightarrow \mathcal{N}(4,2)$

1 1) $P(X<3) = P\left(\frac{X-4}{\sqrt{2}} < \frac{3-4}{\sqrt{2}}\right) = P(Z<-0{,}707) = \Phi(-0{,}707)$ with $\frac{X-4}{\sqrt{2}} \rightarrow \mathcal{N}(0,1)$

$= 1-\Phi(0.707) = 1-\frac{\Phi(0.70)+\Phi(0.71)}{2} = 1-\frac{0.758+0.76}{2}$
$\simeq 0.2404$

1 2) $P(X<x_0)=0{,}9$ So $P\left(Z<\frac{x_0-4}{\sqrt{2}}\right)=0{,}9$

$\Phi\left(\frac{x_0-4}{\sqrt{2}}\right)=0{,}9$ So $\frac{x_0-4}{\sqrt{2}} \simeq 1{,}285 \Rightarrow x_0 \simeq 5{,}817$

$X \rightarrow U([0,8])$

1,5 1) $Y=e^X$ so $Y(\Omega)=[1,e^8]$

* $F_Y(y) = \begin{cases} 0 \text{ si } y\leq 1 \\ 1 \text{ si } y\geq e^8 \\ \text{et si } y\in[1,e^8]: \\ P(Y\leq y) = P(e^X\leq y) = P(X\leq \ln y) = \int_0^{\ln y} 1/8\, dx = \frac{1}{8}\ln y \end{cases}$

* $f_Y(y) = 1/8y \times \mathbb{1}_{[1,e^8]}(y)$

0,5 2) $E(Y^2) = \int_{\mathbb{R}} y^2 f_Y(y)dy = \int_1^{e^8} y^2 \cdot \frac{1}{8y}dy = \frac{1}{8}\left[\frac{y^2}{2}\right]_1^{e^8}$

$= \frac{1}{8}\left(\frac{e^{16}}{2}-\frac{1}{2}\right) = \frac{e^{16}}{16} - 1/16 = \frac{1}{16}(e^{16}-1).$

3

Ex2

2 1) $X \rightarrow G(p)$ on $\mathbb{N}^*$
$P(X=x)=p(1-p)^{x-1}$ $x\in\mathbb{N}^*$

* Likelihood
$\mathcal{L}(p;x_1,\dots,x_n) = \prod_{i=1}^{n} p(1-p)^{x_i-1}$
$= p^n(1-p)^{\sum_{i=1}^{n}(x_i-1)}$

* log Likelihood:
$\ln\mathcal{L}(p) = \ln\left[p^n(1-p)^{\sum(x_i-1)}\right]$
$= n\ln p + \ln(1-p)\sum_{i=1}^{n}(x_i-1)$

* derivative:
$\frac{\partial \ln\mathcal{L}(p)}{\partial p} = \frac{n}{p} - \frac{\sum(x_i-1)}{1-p} = \frac{n(1-p)-p\sum(x_i-1)}{p(1-p)}$

$= \frac{n-np-p\sum x_i+np}{p(1-p)} = \frac{n-p\sum x_i}{p(1-p)}$

* $\frac{\partial\ln\mathcal{L}(p)}{\partial p}=0$ So $n-\hat{p}\sum x_i=0$
$\hat{p}=\frac{n}{\sum x_i} = 1/\overline{x}_n.$

1 2/ $\hat{p} = \frac{1}{\overline{x}_n}$

where $\overline{x}_n = \frac{(1\times 48)+(2\times 31)+\cdots}{136}$

$\overline{x}_n = 2{,}7923$

So $\hat{p} = \frac{1}{2{,}7923} = 0{,}3581$

5

Ex3 Let X = weight of a pallet of bricks in kg.

$X \rightarrow \mathcal{N}(\mu, \sigma^2)$, $n=9$, $\bar{x}_n = 500$ kg , $\hat{\sigma}^2 = 16$

1 1) * $\hat{\mu} = \bar{X}_n = \frac{1}{n}\sum X_i$ * $\hat{\sigma}^2 = S^{*2} = \frac{1}{n-1}\sum(X_i - \bar{X}_n)^2$

2) * estimator of μ is $\bar{X}_n \rightarrow \mathcal{N}(\mu, \sigma^2/n)$

σ is unknown so $\frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}} \rightarrow St(n-1)$

2 * IC s.t $P(t_{n-1,\frac{\alpha}{2}} < \frac{\bar{X}_n - \mu}{\hat{\sigma}/\sqrt{n}} < t_{n-1,1-\frac{\alpha}{2}}) = 1-\alpha$

So $\mu \in [\bar{X}_n \pm t_{n-1,1-\frac{\alpha}{2}} \cdot \hat{\sigma}/\sqrt{n}]$

$n=9$
$\bar{x}_n = 500$ So $\mu \in [497{,}52 ; 502{,}47]$
$\hat{\sigma} = \sqrt{16}$
$t_{n-1,1-\frac{\alpha}{2}} = 1{,}86$

1 3) Now $\sigma^2 = 25$ So $\sigma = 5$. In this case $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow \mathcal{N}(0,1)$

So $\mu \in [\bar{X}_n \pm z_{1-\frac{\alpha}{2}} \sigma/\sqrt{n}]$

$z_{1-\frac{\alpha}{2}} = z_{0.95} = 1{,}64$ So $\mu \in [497{,}25 ; 502.74]$.

1 4) $L = 2 z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 10 \Rightarrow z_{1-\frac{\alpha}{2}} = \frac{10\sqrt{n}}{2\sigma} = \frac{10 \times 3}{2 \times 5} = 3$

$P(Z < z_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2} = 0.9987 \Rightarrow \alpha = 0.0026$

Confidence level $= 99{,}74\ \%$.

8 +2

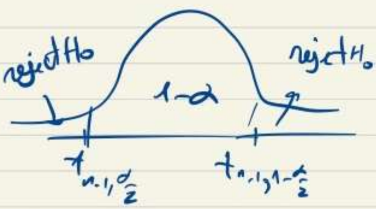
Ex4 $n=16$, $X \sim \mathcal{N}(\mu,\sigma^2)$, $\bar{x}_n = 190$ km, $\hat{\sigma}^2 = 9$

1 1) $\mathcal{E}$: "Randomly choose a car".
$\Omega = \{$cars manufactured by A$\}$
$X: \Omega \to \mathbb{R}$
$x \to$ Autonomy of car x.

1 2) $\mu =$ average autonomy of cars manufactured by A.
* estimator: $\bar{X}_n = \frac{1}{n}\sum X_i$
* estimation: $\bar{x}_n = 190$ km.

3) $H_0: \mu = 200$ vs $H_1: \mu \neq 200$

- Under H_0: $T_0 = \dfrac{\bar{X}_n - \mu_0}{S^*/\sqrt{n}} \to St(n-1)$



1,5 - we reject H_0 if $|T_0| > t_{n-1,1-\frac{\alpha}{2}}$

- $T_0 = \dfrac{190-200}{\sqrt{9/16}} = -13{,}33$; $t_{n-1,1-\frac{\alpha}{2}} = 2.13$

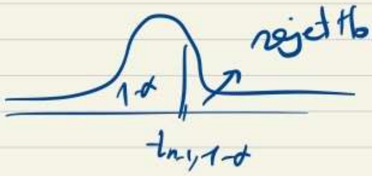
$T_0 > t_{n-1,1-\frac{\alpha}{2}}$ so we reject H_0.

1 4) $IC_\alpha(\mu) = \bar{X}_n \pm t_{n-1,1-\frac{\alpha}{2}} \dfrac{S^*}{\sqrt{n}} = [188.40 ; 191{,}59]$

$200 \notin IC_\alpha(\mu) \Rightarrow$ we reject H_0

5) One sided test: $H_0: \mu \leq 175$ vs $H_1: \mu > 175$

$T_0 = \dfrac{\bar{X}_n - \mu_0}{S^*/\sqrt{n}} \to St(n-1)$



1,5 we reject H_0 if $T_0 > t_{n-1,1-\alpha}$

$t_{n-1,1-\alpha} = 2.13$ and $T_0 = \dfrac{190-175}{3/4} = 20$

we reject $H_0 \Rightarrow$ average is > 175

6) $IC_\alpha(\sigma^2) = \left[\frac{(n-1)S^{*2}}{\chi^2_{n-1,\,1-\alpha/2}} \,;\, \frac{(n-1)S^{*2}}{\chi^2_{n-1,\,\alpha/2}} \right]$

1 where $S^{*2} = 16$, $\chi^2_{1-\alpha/2} = 25$; $\chi^2_{\alpha/2} = 7.26$

So $IC_\alpha(\sigma^2) = [5.4 \,;\, 18.59]$

7) $Y : \Omega \to \mathbb{R}$

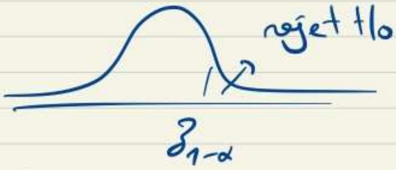
1 $x \to Y(x) = \begin{cases} 1 & \text{if autonomy of car } x > 175 \\ 0 & \text{if not} \end{cases}$

$Y \sim \mathcal{B}(p)$ where $p = P(Y=1)$

0.5 8) $\hat{p} = 0.88$ $\hat{p} = \frac{1}{n}\sum Y_i = \overline{Y_n}$

1,5 9) $H_0 : p \leq 0.9$ vs $H_1 : p > 0.9$

$Z_0 = \frac{\overline{Y_n} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \to \mathcal{N}(0,1)$



we reject Ho if $Z_0 > z_{1-\alpha}$

$Z_0 = \frac{0.88 - 0.9}{\sqrt{\frac{0.9(1-0.9)}{16}}} = -0.26$ and $z_{1-\alpha} = 1,64$

$\Rightarrow$ we can't reject Ho.

G.3 Final Exam 2023-2024

Exercise 1

It has been established that the duration X (in days number units) of a device follows a Poisson distribution with parameter $\lambda > 0$:

$$P(X = k) = \frac{\lambda^k \exp(-\lambda)}{k!}, \ k \in \mathbb{N}$$

We consider a random iid sample $(X_1, \ldots, X_{100})$ following the same variable X.

1. Establish the expression for the maximum likelihood estimator of λ.

2. We suppose suppose now that the λ is fixed and is equal to 1. So $E(X) = V(X) = 1$:

   a. What is the approximate probability distribution of $\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i$.

   b. Calculate $P(\overline{X} > 1.2)$.

Exercise 2

Let $X_1, X_2, ., ., . X_n$ a random sample (iid), of size $n > 10$. The population mean is μ and variance is noted σ^2 . We define the following estimators of mu:

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad T_2 = \frac{3X_1 - 2X_{10} + 2X_4}{3}$$

1. Using the CLT, give the distribution of T_1 for large values of n.

2. Show that T_1 et T_2 are unbiased estimators of μ.

3. Between T_1 and T_2 which one is better for estimating μ.

Exercise 3

A student has been asked to randomly select 30 books from different shelves and heights in the library, open them at random, and count the words on the right-hand page. The student must then calculate a 95% confidence interval based on his 30 chosen books.

1. Define the random experiment ε, the population Ω and the random variable X under study.
2. What are we trying to estimate with this experiment?
3. Propose an estimator of it and give its probability distribution according to the central limit theorem.
4. The student has observed a sample with mean 302.4 and sample standard deviation 60.5. Construct a 95% confidence interval for the mean based on this sample.

We now ask 120 students to do the same experiment and calculate a 95% confidence interval based on the 30 books chosen.

5. Will all 120 intervals be centered on the same value? Justify your answer.
6. Will all 120 intervals have the same length? Justify your answer.
7. Approximately how many students should have the true value of the parameter in their interval?
8. Approximately how many students will significantly overestimate the true value of the parameter, i.e. both bounds of their interval will be greater than the true value?

Exercise 4

Two types of exercise machines for disabled people, A and B, are used to measure the effect of a particular exercise on heart rate (in beats per minute). Normal distribution is considered a good model for both types.

Seven subjects took part in a study to determine whether the two types of device have the same effect on heart rate. The results were as follows.

Sujet	1	2	3	4	5	6	7	$\overline{x}$	$\sum_{i=1}^n (x_i-\overline{x})^2$
Type A	162	163	140	191	160	158	155	**161.29**	**1391.43**
Type B	161	187	199	206	161	160	162	**176.57**	**2449.71**

1. To determine whether the heart rate differs significantly from one device to another, we first check whether the variances are equal. What is the R function used to perform this test?

2. The variance comparison test on R gives the following result. What are the test hypotheses and what decision at threshold $\alpha = 5\%$ can be made on the basis of this result?

```
	F test to compare two variances

data:  x1 and x2
F = 0.568, num df = 6, denom df = 6, p-value = 0.509
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.09759794 3.30560002
sample estimates:
ratio of variances
         0.5679963
```

3. Perform a hypothesis test to determine whether the heart rate differs significantly from one device to another at the 5% threshold, specifying: the hypotheses of the test, the estimator chosen and its probability distribution, the statistic used for this test and its distribution, the H_0 rejection rule and the final decision.

Exercise 5

A company operates on four machines with three shifts of employees every day. Production reports show the following data on the number of breakdowns

	Machines			
Equipe	**A**	**B**	**C**	**D**
1	41	20	12	16
2	31	11	9	14
3	15	17	16	10

We want to check whether the breakdowns are independent of the team in place, using a χ^2 test.

1. What are the hypotheses H_0 and H_1?
2. What is the test statistic U_0? What is its distribution?
3. What is the rule for rejecting the null hypothesis, using a threshold of 1%?
4. Given that the observed value of the test statistic is 11.649, what can you conclude?

Solution (Click to expand)

Barème Examen 23-24 - MDS

5 **Ex 1** 1) Let $n=100$; Sample $(X_1, \dots, X_n)$ iid $X_i \to P(\lambda)$

* Likelihood: $L(\lambda) = \prod_{i=1}^{n} P(X = x_i) \quad x_i \in \mathbb{N}$

1 $= \prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$

* Log-likelihood: $\ln L(\lambda) = \ln\left(\prod_{i=1}^{n} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}\right) = \sum_{i=1}^{n} \ln\left(e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}\right)$

$= \sum_{i=1}^{n} \left[\ln e^{-\lambda} + x_i \ln \lambda - \ln(x_i!)\right]$

1 $= -n\lambda + \ln\lambda \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \ln(x_i!)$

0,5

* $\frac{\partial \ln L(\lambda)}{\partial \lambda} = -n + \frac{\sum x_i}{\lambda} - 0 = 0$ So $\frac{\sum x_i}{\hat{\lambda}} = n \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$\boxed{\hat{\lambda} = \overline{X_n}}$

2) $\lambda = 1$, $E(X) = V(X) = \lambda = 1$, $n = 100$

1 a) $\overline{X} = \frac{1}{100}\sum_{i=1}^{100} X_i \longrightarrow \mathcal{N}(\lambda, \lambda/n) = \mathcal{N}(1, 1/100)$

1,5 b) $P(\overline{X} > 1{,}2) = 1 - P(\overline{X} \leq 1{,}2) = 1 - P\Big(\underbrace{\frac{\overline{X}-1}{\sqrt{1/100}}}_{\to \mathcal{N}(0,1)} \leq \frac{1{,}2-1}{\sqrt{1/100}}\Big)$

$= 1 - \Phi\left(\frac{0{,}2}{\sqrt{1/100}}\right) = 1 - \Phi(2) = 1 - 0{,}9772 = 0{,}0228.$

5 **Ex 2** $T_1 = \frac{1}{n}\sum X_i = \overline{X_n} \qquad T_2 = \frac{3X_1 - 2X_{10} + 2X_4}{3}$

0,5 1) $T_1 \rightsquigarrow \mathcal{N}(\mu, \sigma^2/n)$

1 2) * $E(T_1) = E\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum E(X_i) = \frac{1}{n} \cdot n \cdot \mu = \mu$

1 * $E(T_2) = E\left(\frac{3X_1 - 2X_{10} + 2X_4}{3}\right) = \frac{3}{3}E(X_1) - \frac{2}{3}E(X_{10}) + \frac{2}{3}E(X_4)$

$= \mu - \frac{2}{3}\mu + \frac{2}{3}\mu = \mu$

1 3) * $MSE(T_1) = V(T_1) + bias^2(T_1) = V(T_1) = V(\overline{X_n}) = \sigma^2/n.$

1 * $MSE(T_2) = V(T_2) + bias^2(T_2) = V(T_2) = V\left(\frac{3X_1 - 2X_{10} + 2X_4}{3}\right)$

X_i are iid $= V(X_1) + \frac{4}{9}V(X_{10}) + \frac{4}{9}V(X_4) = \sigma^2 + \frac{8}{9}\sigma^2 = \frac{17}{9}\sigma^2$

0,5 $MSE(T_1) < MSE(T_2)$ so T_1 is better.

6 Ex3 1) $\mathcal{E}$: "Choose a book randomly and open a random page"

1 $\Omega = \{$ all the books of the library $\}$

X = nb of words of the randomly opened right page of a book.

2) * the average number of words of the right pages of the books of this library

0,5 or * we are estimating $\mu = E(X)$.

0,5 3) $\hat{\mu} = \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $n = 30$; $\overline{X}_n \rightarrow \mathcal{N}(\mu, \sigma^2/n)$
where $\mu = E(X)$ and $\sigma^2 = V(X)$.

4) σ is unknown so we construct $CI(\mu)$ s.t

1 $P(t_{\alpha/2} < T < t_{1-\frac{\alpha}{2}}) = 1 - \alpha$ s.t $T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \rightarrow St(n-1)$

0,5 $\mu \in [\overline{X} \pm t_{n-1, 1-\frac{\alpha}{2}} \, S/\sqrt{n}]$

$\overline{x} = 302.4$ $s = 60,5$ $n = 30$ $t_{29, 1-\frac{\alpha}{2}} = 2,04$

0,5 so $\mu \in [279,87 ; 324,93]$

0,5 5) No, each time we have a new $\overline{x}_n$.

0,5 6) No, " " " " " " S.

0,5 7) Approx. $0,95 \times 120 = 114$ students.

0,5 8) Approx $\underbrace{0,025}_{\alpha/2} \times 120 = 3$ students

4 Ex4 1) 0,5 H_0 : $\sigma_A^2 = \sigma_B^2$ or $\sigma_A^2/\sigma_B^2 = 1$

0,5 H_1 : $\sigma_A^2 \neq \sigma_B^2$ or $\sigma_A^2/\sigma_B^2 \neq 1$

0,5 P-value $= 0.509 > \alpha = 0,05$ so we don't reject H_0 so we consider the variances equal.

2) * hyp : $H_0 : \mu_A = \mu_B \iff \mu_A - \mu_B = 0$

0,5 $H_1 : \mu_A \neq \mu_B \iff \mu_A - \mu_B \neq 0$

0,5 * Est : $\hat{\mu}_A - \hat{\mu}_B = \bar{X}_A - \bar{X}_B \rightarrow \mathcal{N}\left(\mu_A - \mu_B ; \frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}\right)$

* Statistic : Since σ_A and σ_B are unknown but suppositdly equal we use

0,5 $T_0 = \frac{\bar{X}_A - \bar{X}_B - \overbrace{(\mu_A - \mu_B)}^{=0 \text{ under } H_0}}{S_p\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}}$ where $S_p^2 = \frac{(n_A - 1)S_A^2 + (n_B - 1)S_B^2}{n_A + n_B - 2}$

$T_0 \rightarrow St(n_A + n_B - 2)$

reject $\frac{\alpha}{2}$ reject $\frac{\alpha}{2}$ $t_{1-\frac{\alpha}{2}}$

0,5 * Decision rules : we reject H_0 if $|T_0| > t_{n_A + n_B - 2 ; 1 - \frac{\alpha}{2}}$

Numeric
* Application : * $\bar{X}_A = 161,29$ * $\bar{X}_B = 176,57$

* $S_A^2 = 1391,43/(7-1) = 231,905$ * $S_B^2 = \frac{2449,71}{7-1} = 408,285$

* $S_p^2 = \frac{6 \times S_A^2 + 6 \times S_B^2}{(7+7-2)} = 320,095$ * $S_p = 17,8912$

So $t_0 = \frac{161,286 - 176,571}{17,8912\sqrt{2/7}} = -1,598$

and $t_{12; 1-\frac{\alpha}{2}} = 2,18$ the $|t_0| < t_{1-\frac{\alpha}{2}}$ So we don't reject H_0

0,5 we can't say the is a difference.

2 Ex5 1) H_0 : nb of breakdowns are type of machines are indep.
0,5 H_1 : " " " " " " are not indep.

0,5 2) $U_0 = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^{3}\sum_{j=1}^{4}\frac{\left(n_{ij} - \frac{n_{i.} n_{.j}}{n}\right)^2}{\frac{n_{i.} n_{.j}}{n}} \rightsquigarrow \chi^2((3-1)(4-1) = 2\times3 = 6)$

0,5 3) We reject H_0 if $u_0 > \chi^2_{6; 1-\alpha} = \chi^2_{6, 0.99} = 16,81$

0,5 4) $U_0 = 11,649 < 16,81$ So we can't reject H_0

they are independant.