

Rappels de statistiques ; nuage de points et corrélation

EXERCICE 1: [Indications] [Correction] - Cours - Calculer -

L'espérance de vie à la naissance des français (hors Mayotte) est résumée approximativement ici de 1950 à 2015 (données suivant INSEE). (Les années, en ordre croissant, sont résumées par les deux derniers chiffres.)

Année	50	55	60	65	70	75	80	85	90	95	00	05	10	15
Hommes	63	65.5	67	67.5	68.5	69	70	71.5	73	74	75.5	77.5	78	80
Femmes	69	71.5	74	75	76	77	78.5	79.5	81	82	82.5	84	84.5	85.5

On note A , H et F les différentes valeurs respectives correspondant aux années, à l'espérance de vie des hommes et à celle des femmes.

1. a) Déterminer les moyennes et les écarts-type de F et H .
b) Déterminer $\sigma_{A,H}$ la covariance de A, H ainsi que $\sigma_{A,F}$ et expliquer son signe.
c) En déduire les coefficients de corrélation $r_{A,H}$ et $r_{A,F}$.
2. a) Déterminer les droites d'ajustement linéaire $y = ax + b$ et $y = a'x + b'$ exprimant respectivement H en fonction de A et F en fonction de A .
b) Ces droites sont-elles pertinentes ?
c) Peut-on en déduire une espérance de vie éventuelle pour 2030 ?
3. a) Y-a-t-il un rapport entre l'évolution des hommes et des femmes ?
b) Déterminer le coefficient directeur d'une des deux droites d'ajustement. Qu'en déduire ?

EXERCICE 2: [Indications] [Correction] - Calculer - Chercher -

D'après le rapport de concours de 2021, (période où l'oral de maths au concours Agro était réparti en 2 matières maths/info distinctes) la répartition des notes (en pourcentage de candidats admissibles) était la suivante : (les notes sont sur 10 pour chaque partie Mathématiques et Informatique, et sur 20 au total.)

classe	[0; 2[	[2; 4[	[4; 6[	[6; 8[	[8; 10]
Maths	10	26	30	23	11
Info	3	21	37	31	9

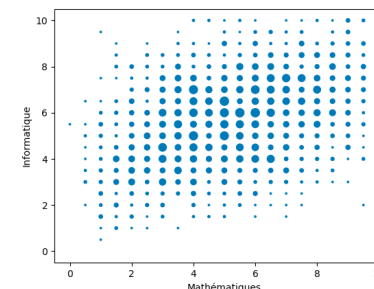
et

classe	[0; 4[	[4; 8[	[8; 12[	[12; 16[	[16; 20]
Note Finale	2	20	42	29	7

1. a) Calculer la moyenne, l'écart-type, la classe médiane et une médiane (on supposera que les effectifs sont répartis uniformément à l'intérieur de chaque classe).
Les résultats annoncés par le concours sont les suivants :

	Moyenne	Écart-type	médiane
Maths	5,28	2,24	5
Info	5,72	1,85	6
Final	11	3,44	11

- b) Expliquez la différence avec les moyennes et écart-types obtenus dans le calcul précédent.
- c) Commentez les résultats sur la médiane par-rapport à ceux obtenus.
2. Le nuage suivant donne la corrélation entre les notes de mathématiques et d'informatique. La taille du cercle de centre (x, y) est proportionnelle au nombre de candidats ayant eu la note x en mathématiques et y en Informatique.



- a) Le coefficient de corrélation est-il, à votre avis, très proche de ± 1 ?
- b) Le coefficient de corrélation est-il positif ou négatif ?
- c) Les candidats ayant obtenu une bonne note en Maths ont-ils généralement obtenus une bonne note en Informatique ?
- d) Les candidats ayant obtenu une bonne note en Informatique ont-ils généralement obtenus une bonne note en maths ?

EXERCICE 3: [Indications] [Correction] – Mobiliser – Modéliser – ☆
 Montrer en utilisant l'inégalité de Bienaymé-Tchebychev que, pour tout $x > 0$,

$$\frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du \geq \frac{1}{2} \left(1 - \frac{1}{x^2}\right)$$

Pour quels x cette inégalité a-t-elle un sens ?

EXERCICE 4: [Indications] [Correction] – Cours – Mobiliser – Calculer –
 Une urne contient une proportion inconnue de boules blanches. On y effectue une série de n tirages d'une boule avec remise et on approche p par la proportion Y_n de boules blanches obtenue dans les n tirages.

- Montrer que, pour tout $\epsilon > 0$, on a $P(|Y_n - p| \geq \epsilon) \leq \frac{1}{4n\epsilon^2}$.
- En déduire une condition suffisante sur n pour que l'approximation de p par Y_n donne une valeur approchée de p à 0,01 près avec une probabilité supérieure ou égale à 95%.
- a) Montrer qu'en utilisant le théorème central limite, on trouve comme condition suffisante

$$n \geq 196^2 p(1-p)$$

- En prenant n_0 minimal vérifiant la condition, montrer que $n_0 \leq 9604$. Commentez.

Approximations

EXERCICE 5: [Indications] [Correction] – Cours – Modéliser – Calculer – Dans les ruches de Robert, la longévité d'une abeille ouvrière née au printemps suit une variable aléatoire T définie par la densité de probabilités suivante :

$$\forall t \geq 0, \quad f(t) = \lambda t^2 e^{-\alpha t}$$

- Sachant que la longévité moyenne est de 50 jours, calculer λ et α .
- Soit Z la longévité moyenne des 2000 ouvrières nées le premier mai. Déterminer la valeur de t pour laquelle :

$$P(|Z - 50| \geq t) \leq 0,01$$

- en utilisant l'inégalité de Bienaymé-Tchebychev.
- en effectuant une approximation grâce au TCL.

EXERCICE 6: [Indications] [Correction] – Modéliser – Cours –

Les fourmis rouges du haut plateau du Katanga ont les yeux bleus, excepté les albinos qui ont les yeux jaunes. Chaque fourmi a une chance sur dix d'être albinos. Un éléphant vient s'abreuver au bord du lac et écrase malencontreusement 453 fourmis rouges.

Soit X le nombre de fourmis aux yeux jaunes victimes du pachyderme.

- Quelle est la loi de probabilité de X ? Déterminer $\mathbb{E}[X]$ et $\sigma(X)$.
- Par quelle loi de probabilité peut-on approcher la loi de X ?
- a) Calculer la probabilité $P(20 < X < 50)$
 b) Déterminer un intervalle I fermé et centré au tour de $E[X]$ tel que $P(X \in I) \simeq 0,95$.

EXERCICE 7: [Indications] [Correction] – Modéliser – Cours –

Dans l'eau rejetée par une centrale nucléaire, un pisciculteur daltonien croit élever des poissons rouges. En réalité, 2% de cette espèce de poissons sont verts. Un incident de fonctionnement du réacteur provoque une brusque élévation de la température qui est fatale à certains poissons.

Le pisciculteur ramasse 150 poissons morts au hasard pour en mesurer la radioactivité. Soit X la variable aléatoire égale au nombre de poissons verts de l'échantillon.

- Quelle est la loi de probabilité de X ? Déterminer son espérance.
- Par quelle(s) probabilité(s) peut-on approcher la loi de X ?
- Calculer $P(X > 3)$ en utilisant l'approximation précédente. (Vérifiez à la calculatrice ou avec Python si ce résultat est proche du résultat sans approximation.)

EXERCICE 8: [Indications] [Correction] – Modéliser – Calculer –

Lors d'une soirée d'Halloween, un couple a décidé d'offrir au choix des bonbons au chocolat et au citron. On admet qu'un enfant a une probabilité de 0,6 d'avoir envie d'un bonbon au chocolat (et qu'il ne peut choisir qu'un seul bonbon). Sachant qu'il y a 200 enfants dans le quartier, en utilisant une approximation de loi bien choisie, combien faut-il acheter au minimum de bonbons au chocolat pour avoir

- 90% de chances de contenter tous les enfants qui aiment le chocolat ?
- 99% de chances de contenter tous les enfants qui aiment le chocolat ?
- Et en ce qui concerne les bonbons au citrons ?
- Commenter.

EXERCICE 9: [Indications] [Correction] – Modéliser – Calculer –

Les "fortune Cookies" contiennent des messages qui annoncent la richesse une fois sur 10. Combien doit-on acheter de ces petits gateaux pour avoir 90% de chance de tomber au moins 5 fois sur l'annonce de la fortune ?

EXERCICE 10: [Indications] [Correction] – Chercher – Modéliser – ☆☆

Soit $0 < p < 1$, $p + q = 1$.

- Montrer que si $p > q$, alors, pour n assez grand,

$$\sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} p^k (1-p)^{n-k} \simeq 1$$

- Supposons qu'une élection oppose deux candidats A et B . Chaque électeur n'ayant pas voté blanc a une probabilité p de voter pour A et une probabilité q de voter pour B . Interpréter le résultat de la question précédente par-rapport à cette situation.

EXERCICE 11: [Indications] [Correction] – Chercher – Mobiliser –
Soit n un entier naturel non nul et λ un nombre réel fixé. On pose

$$P_n(\lambda) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}$$

Soient $X_1, \dots, X_n$ n variables aléatoires définies sur le même espace probabilisé, indépendantes et suivant toutes la loi de Poisson de paramètre 1.

1. Exprimer $P_n(n)$ en fonction de la fonction de répartition de $S_n = X_1 + \dots + X_n$.
2. A l'aide du théorème de la limite centrée, démontrer que $\lim_{n \rightarrow \infty} P_n(n) = \frac{1}{2}$.

Complément : convergence en loi

EXERCICE 12: [Indications] [Correction] – Cours – Raisonner – Calculer –

Soit $(X_i)_{i \in \mathbb{N}^*}$ une suite de variables aléatoires indépendantes de lois uniformes sur $[0; 1]$. Soit N_n une variable aléatoire de loi binomiale $\mathcal{B}(n, p)$ et indépendante des X_i . Soit $T_n = n \min_{1 \leq i \leq N_n} X_i$ et $X \hookrightarrow \mathcal{E}(p)$. On note F_n la fonction de répartition de T_n et F la fonction de répartition de X .

1. Calculer $F_n(x)$ pour tout $x \in \mathbb{R}$.
2. Montrer que pour tout $x \in \mathbb{R}$, $F_n(x) \xrightarrow{n \rightarrow +\infty} F(x)$. (Ceci s'appelle la convergence en loi.)
3. En déduire que pour tout $a, b \in \mathbb{R}$, $P(a < T_n \leq b) \xrightarrow{n \rightarrow +\infty} P(a < X \leq b)$

EXERCICE 13: [Indications] [Correction] – Cours – Raisonner – Calculer –

Chaque côté de la galère royale est occupé par n rameurs. Ces galériens rament sans arrêt, jusqu'à épuisement de toute l'équipe, pendant X_n jours.

La variable aléatoire X_n suit une loi exponentielle de paramètre $\frac{1}{n}$. (Attention, X_n n'est pas forcément un nombre entier !)

La galère quitte Toulon pour un long voyage le 31 du mois d'août à minuit. On note Y_n l'heure de la journée à laquelle les galériens cessent de ramer, traduite en jour décimal. (par exemple, midi = 0.5 jour ...)

1. Justifier que

$$Y_n \leq t \Leftrightarrow (\text{il existe } k \in \mathbb{N} \text{ tel que } k \leq X_n \leq k + t)$$

2. Montrer que la fonction de répartition F_n de Y_n est définie par :

$$F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{1 - e^{-\frac{t}{n}}}{1 - e^{-\frac{1}{n}}} & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases}$$

3. En déduire que Y_n converge en loi vers une variable aléatoire de loi uniforme sur $[0; 1]$.

test de conformité

EXERCICE 14: [Indications] [Correction] – Cours – Calculer – Raisonner –

Le propriétaire d'un bar soupçonne que sa machine à café est dérégulée. Sa machine ne délivre jamais exactement la même quantité de café, mais ses expressos devraient faire en moyenne 40ml. Au service du midi, il mesure chacun des $n = 50$ cafés apportés au client. On note v le volume moyen en ml que délivre en réalité la machine. On pose les hypothèses suivantes :

$$\mathcal{H}_0 : "v = 40" \quad ; \quad \mathcal{H}_1 : "v \neq 40"$$

On note V_n le volume empirique moyen observé sur les n cafés et S_n la variance observée.

1. Sous l'hypothèse $\mathcal{H}_0$, déterminer x tel que

$$P\left(\left|\frac{V_n - 40}{\sqrt{S_n}/\sqrt{n}}\right| \leq x\right) \simeq 0.95$$

2. Les mesures prises expérimentalement donnent $S_n = 2$, En déduire un intervalle I sous la forme $I = [40 - a, 40 + a]$ tel que

$$P(V_n \in I) \simeq 0.95$$

3. Les données expérimentales donnent également $V_n = 39$. Peut-on conclure sur le fait que la machine est dérégulée ?

EXERCICE 15: [Indications] [Correction] – Cours – Calculer – Raisonner –

On souhaite tester si une pièce est truquée. On note p_0 la probabilité pour cette pièce de tomber sur "Pile". On pose les hypothèses suivantes :

$$\mathcal{H}_0 : "p_0 = \frac{1}{2}" \quad ; \quad \mathcal{H}_1 : "p_0 \neq \frac{1}{2}"$$

On note F_n la proportion empirique de "Pile" observée dans une série de n lancers.

1. Sous l'hypothèse $\mathcal{H}_0$, déterminer x tel que

$$P\left(\left|\frac{F_n - \frac{1}{2}}{\sqrt{\frac{1}{4}/\sqrt{n}}}\right| \leq x\right) \simeq 0.95$$

2. En déduire a tel que

$$P\left(F_n \in \left[\frac{1}{2} - a, \frac{1}{2} + a\right]\right) \simeq 0.95$$

3. a) Les données expérimentales sont 29 piles. Peut-on conclure sur le fait que la pièce est truquée ?
b) Un autre dé donne 32 Piles. Peut-on conclure ?

EXERCICE 16: [Indications] [Correction] – Cours – Modéliser –

Dans un bar, le manager aimerait connaître la probabilité p pour un client de laisser un pourboire. Il fait une étude sur une journée de 100 consommateurs. On note X_i la variable de Bernoulli dont le succès est "le $i^{\text{ème}}$ client laisse un pourboire." et X le nombre total de client qui laissent un pourboire.

Le serveur constate que sur le 100-échantillon de clients, 38 ont laissé un pourboire.

1. Donner une variable aléatoire qui approche p (justifier vos dires) ?
2. Déterminer la moyenne et l'écart-type empiriques des X_i .
3. a) Déterminer un nombre u tel que $P\left(\overline{X_n} - u \frac{S_n}{\sqrt{n}} \leq p \leq \overline{X_n} + u \frac{S_n}{\sqrt{n}}\right) = 0,99$
 b) En déduire un intervalle I tel que $P(p \in I) = 0,99$.
 c) Chercher de même un intervalle I tel que $P(p \in I) = 0,95$.
4. Peut-on conclure que la proportion de clients laissant un pourboire est de 50% ?
5. On raisonne maintenant avec une confiance de 95% et On se place le lendemain de la journée test. On note X le nombre de clients laissant un pourboire sur les $n = 120$ clients. ☆
 a) Encadrez la probabilité pour que moins de 100 clients laissent un pourboire (et en déduire une valeur approchée de cette probabilité)
 b) Encadrez de même s'il s'agit de moins de la moitié des clients.

EXERCICE 17: [Indications] [Correction] – Cours – Calculer –

On souhaite comparer les tailles d'un même type de plants de tomates effectués sur des sols de différentes natures.

On désigne par A, B, C, D les populations constituées par les plants sur le sol de type respectifs 1,2,3,4. Constitués chacun de plus de 10 000 plants. On note X_i la variable statistique : *taille de l'individu (en mm) i à 5 mois après plantation.*

L'ensemble des questions suivantes sera résolu en utilisant une approximation bien choisie et en justifiant les théorèmes utilisés.

1. Calculs sur la population A :
 Les données sur la population A sont déjà connues. On sait que la moyenne et l'écart type (théoriques) de la taille d'un plant de A sont : $\mu_A = 570$ et $\sigma_A = 152$.
 On mesure au hasard dans A un échantillon de $n = 250$ individus. Soit M_n la moyenne aléatoire associée à X .
 a) Calculer la probabilité d'avoir une moyenne M_n supérieure ou égale à 550.
 b) Déterminer le nombre h tel que $P[M_n > h] = 0,95$.
 c) Déterminer le nombre a tel que $P(|M_n - \mu_A| < a) = 0,95$ et en déduire l'intervalle I_A centré autour de μ_A pour lequel $P(M_n \in I_A) = 0,95$.
 (Intervalle de confiance de M_n de niveau de confiance 0,95.)

- d) Déterminer de même l'intervalle de confiance I'_A pour la probabilité $p = 0,90$.
- e) Quelle devrait être la taille minimale de l'échantillon à considérer pour que M_n soit une estimation de μ_A à 10mm près avec une confiance $\alpha = 0,99$?

2. La population B :

Dans la population B , la taille moyenne empirique d'un échantillon $\mathcal{E}_B$ de 250 individus est $m_B = 555$.

Peut-on dire a priori qu'il y a une différence significative entre les tailles moyennes des deux catégories A et B ? (Utiliser c)

3. La population C :

La taille moyenne μ_C dans la population C n'est pas connue. Pour l'estimer, on a choisi au hasard un échantillon $\mathcal{E}_C$. Le tableau suivant représente la distribution statistique sur $\mathcal{E}_C$ de la taille en mm.

x	100 – 300	300 – 400	400 – 500	500 – 600	600 – 700	700 – 900
n_i	10	19	30	38	31	22

- a) Quel est le nombre total d'individus étudiés ?
- b) Calculer m_C et s_C les moyennes et écarts-type de C .
- c) Donner une estimation de la taille moyenne μ_C de la population C avec la confiance $\alpha = 0,95$.
- d) Peut-on dire, avec une confiance de 95% qu'il y a-t-il une différence significative de taille entre les populations A et C ?
4. la population D :
 Concernant la population D , quelqu'un de bien intentionné avait déjà fait les calculs avant nous. La personne en question a obtenu une moyenne $m_D = 507$, un écart-type $s_D = 170$ et un intervalle de confiance à 95% de la moyenne théorique $\mu_D : I_D = [480, 534]$.
 Peut-on affirmer qu'il y a une différence de taille moyenne entre les populations C et D ?

Indications

Exercice 3 [Correction]

Le membre de gauche est positif...

Exercice 4 [Correction]

1. On montrera que $p(1-p) \leq \frac{1}{4}$ pour tout $p \in [0, 1]$.

Exercice 5 [Correction]

1. On peut reconnaître le moment d'ordre 2 d'une variable suivant une loi exponentielle.

Exercice 11 [Correction]

2. Interpréter à l'aide d'une moyenne empirique.

Exercice 1

1. a) Les calculs donnent

$$m_F \simeq 78,6 \quad m_H \simeq 71,3 \quad \sigma_F \simeq 4,86 \quad \sigma_H \simeq 4,9$$

- b) On peut utiliser $V(A + H) = V(A) + V(H) + 2Cov(A, H)$ ou alors $Cov(A, H) = \mathbb{E}[AH] - \mathbb{E}[A]\mathbb{E}[H]$.

La première est rapide avec Python et `np.var`. Le deuxième est plus rapide pour nous. On obtient

$$\sigma_{A,H} \simeq 98,21 \quad \sigma_{A,F} \simeq 96,97$$

Les deux sont positifs, ce qui signifie que les espérances de vie augmentent avec les années.

- c) On a

$$r_{A,H} = \frac{\sigma_{A,H}}{\sigma_A \times \sigma_H}$$

et de même pour A et F . On obtient :

$$r_{A,H} \simeq 0,993 \quad r_{A,F} \simeq 0,989$$

2. a) Pour la droite de H par-rapport à A :

$$a = \frac{\sigma_{A,H}}{\sigma_A^2}, \text{ d'où } a_H \simeq 0,24176$$

de plus, la droite passe par le point (m_A, m_H) , d'où

$$b_H = m_H - am_A \simeq -407,85$$

d'où la droite de régression de H par-rapport à A :

$$y = 0,24176x - 407,85$$

Pour la droite de H par-rapport à A :

$$a_F = \frac{\sigma_{A,F}}{\sigma_A^2}, \text{ d'où }$$

$$a_F \simeq 0,23868$$

de plus, la droite passe par le point (m_A, m_F) , d'où

$$b_F = m_F - am_A \simeq -394,6$$

d'où la droite de régression de H par-rapport à A :

$$y = 0,23868x - 394,6$$

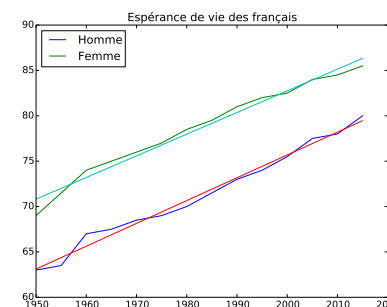
- b) Le coefficient de corrélation est très proche de 1, on peut donc penser que les droites correspondent assez bien aux variables.
c) Les droites étant pertinentes, on peut espérer une espérance de vie de

$$a_H * 2030 + b_H \simeq 82,9$$

pour les hommes, et

$$a_F * 2030 + b_F \simeq 89,9$$

pour les femmes. Ce qui peut éventuellement donner lieu au graphique suivant :



3. a) On a

$$r_{F,H} \simeq 0,983.$$

On peut donc estimer que l'évolution des hommes et des femmes entre eux se fait de manière linéaire.

- b) Pour la droite d'ajustement de H par-rapport à F , on trouve

$$a \simeq 0,992.$$

Ceci nous confirme que l'évolution des hommes et des femmes se fait à la même vitesse. (Information déjà eu avec les coefficients identiques pour les droites d'ajustement déjà obtenues.)

Exercice 2

1. a) En prenant comme repère le centre des classes, on trouve

	Moyenne	écart-type	Classe Médiane	Médiane
Maths	4,98	2,31	[4, 6[	4,98
Info	5,44	1,94	[4, 6[	4,6
Final	10,76	3,61	[8, 12[	9,3

- b) Les résultats du concours sont calculés avec précision sur l'ensemble des notes précises, alors que nous ne disposons ici que des classes pour lesquelles nous avons résumé les notes en centrant la classe. Or, les notes ne sont pas réellement réparties de manière uniforme au sein d'une même classe.
- c) Le problème de calcul est le même pour la médiane. En réalité, on ne devrait parler que de "classe médiane". Dans ce cas, on retrouve bien que la médiane annoncée par le concours est bien dans la même classe médiane que nous avons trouvée.
2. a) Le coefficient de corrélation n'est pas proche de 1. En effet, le nuage est très loin d'être concentré autour d'une droite.
- b) Si on devait tracer une allure générale, le nuage semblerait toutefois se diriger vers le haut quand x augmente. Le coefficient de corrélation est donc positif.

Pour information, les vrais résultats indiquent d'ailleurs un coefficient de

$$r = 0,38$$

- c) En prenant une note constante en mathématiques, on constate que les notes d'informatique sont plutôt correctes, voir bonnes. De plus, plus les notes de mathématiques augmentent, plus les notes d'informatique semblent globalement augmenter aussi. On peut donc dire que globalement, avoir une bonne note en mathématiques entraîne une note plutôt bonne en Informatique.
- d) On constate une très légère tendance, mais rien qui ne permette d'affirmer ceci. On observe un certain nombre d'étudiants ayant eu une excellente note en informatique alors que la note de maths n'est pas franchement exceptionnelle, cette tendance s'observant beaucoup moins dans l'autre sens.

Exercice 3

$$\begin{aligned} \frac{1}{\sqrt{2\pi}} \int_0^x e^{-u^2/2} du &= P(0 < Z \leq x) \quad \text{où } Z \sim \mathcal{N}(0; 1) \\ &= 1 - P(Z \leq 0) - P(Z > x) \\ &= \frac{1}{2} - \frac{1}{2}P(|Z| > x) \end{aligned}$$

Or, d'après le théorème de B-T, on a $P(|Z| > x) \leq \frac{V(Z)}{x^2} = \frac{1}{x^2}$. C'est fini !

Exercice 4

1. On pose S_n le nombre de boules blanches obtenues dans l'échantillon. C'est une v.a. suivant une loi $\mathcal{B}(n, p)$ et on a $Y_n = S_n/n$. D'après l'inégalité de Bienaymé-Thébychev, on obtient

$$\forall \epsilon > 0, \quad P(|Y_n - E(Y_n)| \geq \epsilon) \leq \frac{V(Y_n)}{\epsilon^2}$$

Or, $E(Y_n) = E(S_n)/n = np/n = p$ et

$$V(Y_n) = V(S_n/n) = \frac{V(S_n)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Mais par l'étude de la fonction $x \in [0, 1] \mapsto x(1-x)$ (tableau de variation entre 0 et 1), on peut voir que

$$x(1-x) \leq \frac{1}{4} \quad \forall x \in [0, 1]$$

donc

$$V(Y_n) \leq \frac{1}{4n}$$

2. D'après la première question, en posant $\epsilon = 0,01$, on a

$$P(|Y_n - p| \leq \epsilon) = 1 - P(|Y_n - p| \geq \epsilon) \geq 1 - \frac{1}{4n\epsilon^2}$$

On répond donc à la question en cherchant n tel que $1 - \frac{1}{4n\epsilon^2} \geq 0,95$, i.e

$$\boxed{n \geq 50000.}$$

Ce n'est pas très pratique !

3. a) S_n suit une loi binomiale $\mathcal{B}(n, p)$. Ainsi, par le TCL, on sait que, pour $n \geq 30$, $S_n^* = \frac{S_n - np}{\sqrt{np(1-p)}}$ suit une loi normale $\mathcal{N}(0; 1)$. Ainsi,

$$\begin{aligned} 0,95 \leq P(|Y_n - p| \leq \epsilon) &= P\left(\left|\frac{S_n}{n} - p\right| \leq \epsilon\right) \\ &= P\left(\left|\frac{S_n - np}{n}\right| \leq \epsilon\right) \\ &= P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| \leq \frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \\ &= 2\phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) - 1 \end{aligned}$$

D'où

$$\phi\left(\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}}\right) \geq \frac{1,95}{2} = 0,975$$

qui entraîne

$$\frac{\sqrt{n}\epsilon}{\sqrt{p(1-p)}} \geq 1,96$$

puis

$$\sqrt{n} \geq \frac{1,96}{\epsilon} \sqrt{p(1-p)} = 196 \sqrt{p(1-p)}$$

- b) Tout d'abord, ne sachant pas si $196^2 p(1-p)$ sera entier, il faut prendre ses précautions. Ainsi,

$$n_0 < \lfloor 196^2 p(1-p) \rfloor + 1$$

De plus, par une simple étude de fonction, on constate que

$$p(1-p) \leq \frac{1}{4}$$

pour tout $p \in [0, 1]$, ce qui donne le résultat

$$n_0 < \lfloor 196^2 p(1-p) \rfloor + 1 \leq 9604 + 1$$

et donc

$$\boxed{n_0 \leq 9604}$$

Pour $n = 9604$, l'approximation par la loi normale est très précise. On peut donc penser que c'est exactement une borne supérieure du nombre de cas à étudier. Il est nettement plus bas que celui obtenu avec la méthode B-T, ce qui paraît censé.

Exercice 5

1. • Utilisons que f est une densité de probabilité pour obtenir une première équation :

On peut reconnaître en f le moment d'ordre 2 d'une variable Z suivant une loi exponentielle de paramètre α :

$$\int_0^{+\infty} \lambda t^2 e^{-\alpha t} dt = \frac{\lambda}{\alpha} \mathbb{E}[Z^2]$$

Or,

$$\mathbb{E}[Z^2] = V(Z) + \mathbb{E}[Z]^2 = \frac{2}{\alpha^2} \quad \text{ou la formule de cours} \quad \mathbb{E}[Z^n] = \frac{n!}{\alpha^n}$$

D'où

$$\int_0^{+\infty} \lambda t^2 e^{-\alpha t} dt = \frac{2\lambda}{\alpha^3}$$

On obtient donc comme première équation et on pose

$$\lambda = \alpha^3/2$$

- Utilisons l'espérance pour obtenir α :

On a

$$50 = E(T) = \int_0^{+\infty} t f(t) dt = \int_0^{+\infty} \lambda t^3 e^{-\alpha t} dt = \frac{\lambda}{\alpha} \mathbb{E}[Z^3] = \frac{6\lambda}{\alpha^4} = \frac{3}{\alpha}$$

Ainsi,

$$\boxed{\alpha = \frac{3}{50}, \quad \lambda = \frac{27}{250\,000}}$$

2. a) L'inégalité de B-T appliqué à T donne

$$\forall t > 0, \quad P(|Z - E(Z)| > t) < \frac{V(Z)}{t^2}$$

L'application numérique au cas présent donne

$$\forall t > 0, \quad P(|Z - 50| > t) < \frac{V(Z)}{t^2}$$

Z est la moyenne de 2000 variables aléatoires indépendantes de même loi T . C'est-à-dire que l'on peut écrire

$$Z = \frac{T_1 + \dots + T_{2000}}{2000}$$

On a donc

$$E(Z) = E(T) = 50, \quad V(Z) = \frac{V(T)}{2000}$$

Le calcul de la variance donne

$$V(T) = \frac{2500}{3} \quad \text{d'où} \quad V(Z) = \frac{25}{60} = \frac{5}{12}$$

On a alors

$$\forall t > 0, \quad P(|Z - 50| > t) < \frac{25}{60t^2}$$

Pour être certain d'avoir $P(|Z - 50| > t) < 0.01$, il faut

$$\frac{25}{60t^2} < 0.01$$

Autrement dit,

$$t \geq \sqrt{\frac{25}{60 \times 0.01}} \simeq 6.45$$

- b) D'après le TCL, on sait que la variable $Z^* = \frac{Z - E(Z)}{\sigma(Z)}$ suit approximativement une loi normale centrée réduite. D'où

$$\begin{aligned} P(|Z - 50| > t) &= P(|Z - E(Z)| > t) \\ &= P\left(\frac{|Z - E(Z)|}{\sigma(Z)} > \frac{t}{\sigma(Z)}\right) \\ &= P\left(Z^* > \frac{t}{\sigma(Z)}\right) \\ &= 1 - P\left(-\frac{t}{\sigma(Z)} \leq Z^* \leq \frac{t}{\sigma(Z)}\right) \\ &= 2 - 2\Phi\left(\frac{t}{\sigma(Z)}\right) \end{aligned}$$

On en déduit que

$$\begin{aligned} P(|Z - 50| > t) < 0.01 &\Leftrightarrow 2 - 2\Phi\left(\frac{t}{\sigma(Z)}\right) < 0.01 \\ &\Leftrightarrow \Phi\left(\frac{t}{\sigma(Z)}\right) > 1 - \frac{0.01}{2} = 0.995 \end{aligned}$$

D'après les tables de la loi normale inversée, on a

$$\Phi(2, 575) \simeq 0.995$$

D'où

$$\begin{aligned} P(|Z - 50| > t) < 0.01 &\Leftrightarrow \Phi\left(\frac{t}{\sigma(Z)}\right) > 0.995 \\ &\Leftrightarrow \frac{t}{\sigma(Z)} > 2.575 \\ &\Leftrightarrow t > 2.575\sigma(Z) = 2.575 \times \sqrt{\frac{25}{60}} \simeq 1.662 \end{aligned}$$

Sous l'hypothèse que Z suit une loi normale,

$$P(|Z - 50| > t) < 0.01 \quad \Leftrightarrow \quad t > 1.662$$

Exercice 6

1. On a $X \sim \mathcal{B}(n, p)$, avec $p = 0.1$, $n = 453$
on sait alors que

$$E(X) = np = 45.3 \quad \sigma(X) = \sqrt{np(1-p)} = \sqrt{40.77} \simeq 6.4$$

2. Étant donné que $n \geq 30$, $np = 45.3 \geq 5$, $n(1-p) \simeq 407 \geq 5$, par le théorème de Moivre Laplace, on peut approcher la loi de X par

$$\mathcal{N}(\mu, \sigma^2), \quad \text{avec } \mu = E(X) = 45.3, \quad \sigma^2 = V(X) = np(1-p) \simeq 40.77$$

3. a) • Méthode 1 :

Directement avec la calculatrice avec la fonction de répartition d'une variable de loi $\mathcal{N}(\mu, \sigma^2)$, on obtient

$$P(20 < X < 50) \simeq 0.7691$$

ou encore, avec correction de continuité :

$$P(20 < X < 50) = P(20,5 < X < 49,5) \simeq 0.7446$$

• Méthode 2 : on se ramène à la loi $\mathcal{N}(0, 1)$:

X pouvant être approchée par une loi normale, on aimerait se ramener à la loi $\mathcal{N}(0, 1)$ dont on connaît les tables :

$$\frac{X - \mu}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$$

Avec correction de continuité :

$$\begin{aligned} P(20 < X < 50) &= P(20.5 < X \leq 49.5) = P\left(\frac{20.5 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{49.5 - \mu}{\sigma}\right) \\ &= P\left(\underbrace{\frac{20.5 - 45.3}{\sqrt{40.77}}}_{\alpha \simeq -3.884} < \frac{X - \mu}{\sigma} \leq \underbrace{\frac{49.5 - 45.3}{\sqrt{40.77}}}_{\beta \simeq 0.6578}\right) \\ &= \Phi(\beta) - \Phi(\alpha) = \Phi(\beta) - 1 + \Phi(-\alpha) \simeq 0.7446 \end{aligned}$$

Conclusion,

$$P(20 < X < 50) = P(20,5 < X < 49,5) \simeq 0.7446$$

b) On cherche u tel que

$$P(|X - \mathbb{E}[X]| < u) \simeq 0,95.$$

On peut encore centrer et réduire pour obtenir :

$$P\left(\left|\frac{X - np}{\sqrt{np(1-p)}}\right| \leq u\right) \simeq 0,95.$$

Or, X pouvant être approché par une loi normale, on en déduit

$$2\Phi(u) - 1 \simeq 0,95.$$

i.e

$$\Phi(u) = 0,975$$

et donc,

$$u \simeq 1,96$$

D'où

$$P([E[X] - u \cdot \sigma(X); E[X] + u \cdot \sigma(X)]) \simeq 0,95$$

Un intervalle possible est donc

$$[32, 78; 57, 8]$$

Ou encore, comme X est entier :

$$[33; 57]$$

Exercice 7

1. On peut donc commencer par dire que X suit une loi binomiale

$$\mathcal{B}(n, p), \quad \text{avec } n = 150, p = 0,02$$

$$E(X) = np = 3$$

2.

$n \geq 30$ et $p \leq 0.1$. on peut donc continuer avec une loi de Poisson de paramètre λ

$$\mathcal{P}(\lambda)$$

Avec $\lambda = E(X) = np = 3$. On peut donc estimer que

$$X \rightsquigarrow \mathcal{P}(3)$$

3. Grâce à l'approximation, on peut estimer que

$$\begin{aligned} P(X > 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - e^{-3} \left(\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!} + \frac{3^3}{3!} \right) \\ &= 1 - e^{-3} \left(1 + 3 + \frac{9}{2} + \frac{9}{2} \right) = 1 - 13e^{-3} \simeq 0.35 \end{aligned}$$

D'où

$$\boxed{P(X > 3) \simeq 0.353}$$

Remarque : par calcul direct avec ordinateur, on obtient comme résultat environ 0,353. L'approximation est donc très précise.

Exercice 8

On note S le nombre d'enfant préférant le chocolat. On sait que

$$S \rightsquigarrow \mathcal{B}(200; 0,6).$$

On note N le nombre de bonbons à acheter. On souhaite que

$$P(S \leq N) \geq 0,90 \quad \text{ou} \quad P(S \leq N) \geq 0,99$$

Il est impossible ici d'utiliser la loi binomiale pour faire le calcul, étant donné qu'on ne connaît pas N . On va donc faire une approximation. La probabilité d'être contenté proche de 0,5, ça sera certainement une loi normale. Vérifions.

$$200 \geq 30, \quad 200 \times 0,6 = 120 \geq 5, \quad 200 \times 0,4 = 80 \geq 5.$$

On sait donc, d'après le TCL, que

$$S^* = \frac{S - E(S)}{\sigma(S)} \rightsquigarrow \mathcal{N}(0; 1)$$

ou encore que, de manière équivalente :

$$S \rightsquigarrow \mathcal{N}(E(S); V(S)) = \mathcal{N}(120; 48)$$

i.e. ici, avec $E(S) = 120$ et $\sigma(S) = \sqrt{48}$.

$$S^* = \frac{S - 120}{\sqrt{48}} \rightsquigarrow \mathcal{N}(0; 1)$$

D'où, par **correction de continuité**, on cherche N tel que

$$P(S \leq N) = P(S \leq N + 0,5) = 0,9 \text{ ou } 0.99$$

ou du moins, par logique de l'exercice, sinon ne tombe pas exactement sur un entier :

$$P(S \leq N) = P(S \leq N + 0,5) \geq 0,9 \text{ ou } 0.99$$

- Méthode 1 : avec $\mathcal{N}(120; 48)$:

$$P(S \leq N) = P(S \leq N + 0,5) = 0.9$$

à la calculatrice, on trouve

$$N + 0,5 \simeq 128.88$$

d'où

$$N \simeq 128.38$$

On prend donc

$$N = 129 \text{ bonbons}$$

- Méthode 2 : avec $\mathcal{N}(0, 1)$:

$$P(S \leq N) = P(S \leq N + 0,5) = P\left(S^* \leq \frac{N + 0,5 - 120}{\sqrt{48}}\right) = \Phi\left(\frac{N - 119,5}{\sqrt{48}}\right)$$

1. D'après les tables de la loi normale inversée, on a

$$\phi(1,282) = 0,90$$

Il faut donc

$$\frac{N - 119,5}{\sqrt{48}} \geq 1,282$$

i.e.

$$N \geq \underbrace{1,282 \sqrt{48} + 119,5}_{\simeq 128,38}$$

Ainsi, il faut acheter au moins

129 bonbons au chocolat pour avoir 90% de chances de contenter les enfants.

2. De même, avec correction de continuité, par la calculatrice directement mais avec une probabilité de 0.99 :

$$N + 0.5 \simeq 136.11$$

et donc

$$N = 135,61$$

Ainsi, il faut acheter au moins

136 bonbons au chocolat pour avoir 99% de chances de contenter les enfants.

3. Attention, la réponse n'est pas $200 - 129$ (ou $200 - 136$). Ce n'est pas parce que vous aurez contenté les enfants qui souhaitent du chocolat que vous avez aussi contenté ceux qui souhaitent du citron !

On note C le nombre d'enfant préférant le citron. On sait que

$$C \sim \mathcal{B}(200; 0,4).$$

Comme avant

$$200 \geq 30, \quad 200 \times 0,4 = 80 \geq 5, \quad 200 \times 0,6 = 120 \geq 5.$$

d'où

$$C \sim \mathcal{B}(200; 0,4) \simeq \mathcal{N}(80, 48)$$

On note encore N le nombre de bonbons à acheter. On souhaite encore que

$$P(C \leq N) \geq 0,90 \quad \text{ou} \quad P(C \leq N) \geq 0,99$$

- Pour 90% :

On trouve, par correction de continuité :

$$N + 0.5 \simeq 88.87$$

D'où

$$N \simeq 88.37$$

on prend donc

$$89 \text{ bonbons}$$

- Pour 99% :

On trouve, par correction de continuité :

$$N + 0.5 \simeq 96.12$$

D'où

$$N \simeq 95.72$$

on prend donc

$$96 \text{ bonbons pour } 90\%.$$

Ainsi, il faut acheter au moins

96 bonbons au chocolat pour avoir 99% de chances de contenter les amateurs de citron.

4. Au final, dans chaque cas, il faut acheter

$129 + 89 = 218$ bonbons au total pour avoir 90% de chances de contenter tout le monde.
 $136 + 96 = 232$ bonbons au total pour avoir 99% de chances de contenter tout le monde.

ce qui est plus que 200 dans chaque cas ... (mais qui reste raisonnable !)

Exercice 9

Soit n le nombre de gateaux à acheter et X_n le nombre de messages positifs parmi les n . On sait alors que

$$X_n \sim \mathcal{B}(n, \frac{1}{10})$$

n n'étant pas connu, il est difficile de faire le calcul à l'aide de la loi binomiale. Supposons donc dans un premier temps que n soit assez grand pour pouvoir faire une approximation par la loi normale :

$$B(n, \frac{1}{10}) \simeq \mathcal{N}(E(X_n), V(X_n)) = \mathcal{N}(\underbrace{\frac{n}{10}}_{\mu}, \underbrace{\frac{9n}{100}}_{\sigma^2})$$

On cherche n tel que $P(X_n \geq 5) \geq 0.9$. Or,

• *Sans correction de continuité :*

$$\begin{aligned} P(X_n \geq 5) &= P\left(\frac{X_n - \mu}{\sigma} \geq \frac{5 - \mu}{\sigma}\right) \\ &= P\left(\frac{X_n - \mu}{\sigma} \geq \frac{5 - \frac{n}{10}}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{5 - \frac{n}{10}}{\sigma}\right) = 1 - \Phi\left(\frac{5 - \frac{n}{10}}{\sqrt{\frac{9n}{100}}}\right) = 1 - \Phi\left(\frac{5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}}\right) \end{aligned}$$

D'où

$$\Phi\left(\frac{5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}}\right) \leq 0.1$$

Par la table de la loi normale inversée, on obtient

$$\frac{5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}} \leq -1.29$$

Or,

$$\begin{aligned} \frac{5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}} \leq -1.29 &\Leftrightarrow \frac{50 - n}{3\sqrt{n}} \leq -1.29 \\ &\Leftrightarrow 50 - n \leq -1.29 \times 3\sqrt{n} \leq 0 \\ &\Leftrightarrow n \geq 50, \quad (50 - n)^2 \geq (-1.29 \times 3)^2 n \\ &\Leftrightarrow n \geq 50, \quad 2500 + n^2 - 100n \geq (-1.29 \times 3)^2 n \\ &\Leftrightarrow n \geq 50, \quad n^2 - (100 + 1.29^2 \times 9)n + 2500 \geq 0 \end{aligned}$$

$$\Delta = (100 + 1.29^2 \times 9)^2 - 4 \times 2500 \simeq 3219.69, \sqrt{\Delta} \simeq 56.74.$$

$$x_1 = \frac{100 + 1.29^2 \times 9 - \sqrt{\Delta}}{2} \simeq 29.12 \quad x_2 = \frac{100 + 1.29^2 \times 9 + \sqrt{\Delta}}{2} \simeq 85.86$$

Les solutions sont donc

$$\boxed{n \geq 86}$$

• *Avec correction de continuité :*

$$\begin{aligned} P(X_n \geq 5) &= P(X_n \geq 4.5) \\ &= 1 - \Phi\left(\frac{4.5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}}\right) \end{aligned}$$

D'où

$$\Phi\left(\frac{4.5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}}\right) \leq 0.1$$

Par la table de la loi normale inversée, on obtient

$$\frac{4.5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}} \leq -1.29$$

Or,

$$\begin{aligned} \frac{4.5 - \frac{n}{10}}{\frac{3}{10}\sqrt{n}} \leq -1.29 &\Leftrightarrow \frac{45 - n}{3\sqrt{n}} \leq -1.29 \\ &\Leftrightarrow 45 - n \leq -1.29 \times 3\sqrt{n} \leq 0 \\ &\Leftrightarrow n \geq 45, \quad (45 - n)^2 \geq (-1.29 \times 3)^2 n \\ &\Leftrightarrow n \geq 45, \quad 2025 + n^2 - 90n \geq (-1.29 \times 3)^2 n \\ &\Leftrightarrow n \geq 45, \quad n^2 - (90 + 1.29^2 \times 9)n + 2025 \geq 0 \end{aligned}$$

$$\Delta = (90 + 1.29^2 \times 9)^2 - 4 \times 2025 \simeq 2920.1495, \sqrt{\Delta} \simeq 54.0384.$$

$$x_1 = \frac{90 + 1.29^2 \times 9 - \sqrt{\Delta}}{2} \simeq 25.47 \quad x_2 = \frac{90 + 1.29^2 \times 9 + \sqrt{\Delta}}{2} \simeq 79.51$$

Les solutions sont donc

$$\boxed{n \geq 80}$$

En conclusion, Pour $n = 80$, comme n est grand, l'approximation était raisonnable et on conclut qu'il faut acheter au minimum 80 gateaux pour avoir 90% de chances de tomber sur au moins 5 messages positifs.

En remarque, le calcul exact donne 78 gateaux...

Exercice 10

- Traduisons l'exercice en terme de probabilités : L'expression dans la somme nous fait penser à des des sommes de probabilités de type $P(X = k)$, où X est une variable aléatoire de loi binomiale. Posons donc X_n une variable aléatoire telle que

$$X_n \sim \mathcal{B}(n, p)$$

On note au passage que $E(X_n) = np$ et $\sigma(X_n) = \sqrt{np(1-p)}$.

On a alors

$$\begin{aligned} \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} p^k (1-p)^{n-k} &= \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n P(X_n = k) \\ &= P\left(\lfloor \frac{n}{2} \rfloor + 1 \leq X_n \leq n\right) \end{aligned}$$

D'après le TCL, on sait que pour n assez grand, X_n^* (loi centrée réduite de X_n) va suivre environ une loi normale centrée réduite. Supposons donc n assez grand. Appliquons ceci à l'égalité ci-dessus :

$$\begin{aligned} \sum_{k=\lfloor \frac{n}{2} \rfloor + 1}^n \binom{n}{k} p^k (1-p)^{n-k} &= P\left(\frac{\frac{\lfloor \frac{n}{2} \rfloor + 1}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq \frac{X_n - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq \frac{\frac{n}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}}\right) \\ &= P\left(\underbrace{\frac{\frac{\lfloor \frac{n}{2} \rfloor + 1}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}}}_{\alpha_n} \leq X_n^* \leq \underbrace{\frac{\frac{n}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}}}_{\beta_n}\right) \\ &\simeq \Phi(\beta_n) - \Phi(\alpha_n) \end{aligned}$$

Comme $1-p \neq 0$, on a que

$$\lim_{n \rightarrow +\infty} \beta_n = \lim_{n \rightarrow +\infty} \sqrt{n} \sqrt{\frac{1-p}{p}} = +\infty$$

De plus, comme

$$\frac{n}{2} \leq \lfloor \frac{n}{2} \rfloor < \frac{n}{2} + 1$$

on obtient

$$\frac{\frac{\lfloor \frac{n}{2} \rfloor + 1}{n} - p}{\sqrt{p(1-p)}} \xrightarrow{n \rightarrow +\infty} \frac{1}{2} - p$$

Or, par hypothèse, $p > \frac{1}{2}$. On en déduit que

$$\lim_{n \rightarrow +\infty} \alpha_n = \lim_{n \rightarrow +\infty} \sqrt{n} \frac{\frac{\lfloor \frac{n}{2} \rfloor + 1}{n} - p}{\sqrt{p(1-p)}} = -\infty$$

Donc, si n est assez grand,

$$\Phi(\beta_n) \simeq 1 \quad ; \quad \Phi(\alpha_n) = 0$$

Finalement, on trouve que, pour n assez grand,

$$P\left(\sqrt{n} \frac{\frac{\lfloor \frac{n}{2} \rfloor + 1}{n} - p}{\sqrt{p(1-p)}} \leq X_n^* \leq \sqrt{n} \sqrt{\frac{1-p}{p}}\right) = P(-\infty \leq X \leq +\infty) \simeq 1$$

- On peut interpréter X_n comme le nombre de voix recueillies par un candidat à des élections pour n électeurs. Le résultat montre que si $p > 1/2$ (les voix peuvent être très légèrement en faveur du candidat.) alors la probabilité pour le candidat d'être élu à la majorité d'une seule voix est de 1 lorsque le nombre de votants tend vers l'infini.

Exercice 11

- Les variables X_i étant mutuellement indépendantes, on sait que $X_1 + \dots + X_n$ suit une loi de Poisson de paramètre n . On remarque donc immédiatement que $P_n(n) = P(S_n \leq n)$.
- S_n étant une somme, on peut écrire tout ceci comme une variable centrée réduite issue d'une moyenne empirique.

$$P_n(n) = P(S_n - n \leq 0) = P(S_n - E(S_n) \leq 0) = P\left(\frac{S_n - 1}{\frac{1}{\sqrt{n}}} \leq 0\right)$$

On termine en invoquant le théorème de la limite centrée, qui nous dit que la limite existe et vaut

$$\lim_{n \rightarrow +\infty} P_n(n) = \Phi(0) = \frac{1}{2}$$

où Φ est la fonction de répartition d'une variable centrée réduite.

Exercice 12

1. • Support de T_n :

$X_i \in [0; 1]$ pour tout $i \in \mathbb{N}^*$. Donc,

$$\min_{1 \leq i \leq N_n} X_i \in [0; 1]$$

puis

$$T_n \in [0; n]$$

• Fonction de répartition de T_n :

★ Si $t < 0$, d'après le support de T_n , il est clair que : $F_{T_n}(t) = 0$.

★ Si $t > n$, d'après le support de T_n , il est clair que : $F_{T_n}(t) = 1$.

★ Si $t \in [0; n]$,

$$\begin{aligned} F_{T_n}(t) &= P(T_n \leq t) \\ &= P(n \min_{1 \leq i \leq N_n} X_i \leq t) \\ &= P(\min_{1 \leq i \leq N_n} X_i \leq \frac{t}{n}) \\ &= 1 - P(\min_{1 \leq i \leq N_n} X_i > \frac{t}{n}) \\ &= 1 - P\left(\bigcap_{1 \leq i \leq N_n} (X_i > \frac{t}{n})\right) \end{aligned}$$

Attention, N_n est une variable aléatoire, on ne peut donc pas transformer en un produit sans prendre de précautions. Comme

$$N_n \rightsquigarrow \mathcal{B}(n, p)$$

on a un système complet d'événements

$$(N_n = k)_{k \in \llbracket 0; n \rrbracket}$$

D'où

$$\begin{aligned} P\left(\bigcap_{1 \leq i \leq N_n} (X_i > \frac{t}{n})\right) &= \sum_{k=0}^n P((N_n = k) \cap \bigcap_{1 \leq i \leq N_n} (X_i > \frac{t}{n})) \\ &= \sum_{k=0}^n P((N_n = k) \cap \bigcap_{1 \leq i \leq k} (X_i > \frac{t}{n})) \\ &= \sum_{k=0}^n P(N_n = k) P\left(\bigcap_{1 \leq i \leq k} (X_i > \frac{t}{n})\right) \quad \text{ind. des } N_n \\ &\quad \text{par rapport aux } X_i. \\ &= \sum_{k=0}^n P(N_n = k) \prod_{1 \leq i \leq k} P(X_i > \frac{t}{n}) \quad \text{ind. des } X_i \\ &= \sum_{k=0}^n P(N_n = k) \prod_{1 \leq i \leq k} \left(1 - P(X_i \leq \frac{t}{n})\right) \\ &= \sum_{k=0}^n P(N_n = k) \prod_{1 \leq i \leq k} \left(1 - F_{X_i}\left(\frac{t}{n}\right)\right) \\ &= \sum_{k=0}^n P(N_n = k) \left(1 - F_{X_1}\left(\frac{t}{n}\right)\right)^k \quad \text{Les } X_i \text{ ont même loi.} \end{aligned}$$

Application numérique :

$$\begin{aligned} P\left(\bigcap_{1 \leq i \leq N_n} (X_i > \frac{t}{n})\right) &= \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} \left(1 - \frac{t}{n}\right)^k \quad \text{car } t \in [0; n] \text{ et donc } \frac{t}{n} \in [0; 1] \\ &= \sum_{k=0}^n \binom{n}{k} \left(p\left(1 - \frac{t}{n}\right)\right)^k (1-p)^{n-k} \\ &= \left(p\left(1 - \frac{t}{n}\right) + 1 - p\right)^n \quad \text{Formule du binôme} \\ &= \left(1 - \frac{pt}{n}\right)^n \end{aligned}$$

En bref,

$$F_{T_n}(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - \left(1 - \frac{pt}{n}\right)^n & \text{si } t \in [0; n] \\ 1 & \text{si } t > n \end{cases}$$

2. Pour $t \in [0; +\infty[$, pour $n \geq t$, on a

$$\left(1 - \frac{pt}{n}\right)^n = e^{n \ln(1 - \frac{pt}{n})} \xrightarrow{n \rightarrow +\infty} e^{-pt}$$

D'où

$$\lim_{n \rightarrow +\infty} F_{T_n}(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 - e^{-pt} & \text{si } t \geq 0 \end{cases}$$

On reconnaît une loi exponentielle de paramètre p .

Exercice 13

1. $(Y_n \leq t)$ signifie : il existe un jour pour lequel le bateau arrive entre minuit et l'heure t .
 $(k \leq X_k \leq k + t)$ signifie : le bateau arrive le jour k entre minuit et l'heure t .
D'où la réponse.

2. • Support de Y_n : Il est clair que $0 \leq Y_n \leq 1$.

• Fonction de répartition :

★ Soit $t < 0$. De par le support, on a $P(Y_n \leq t) = 0$.

★ Soit $t > 1$. De par le support, on a $P(Y_n \leq t) = 1$.

★ Soit $t \in [0; 1]$. D'après la première question, on a

$$\begin{aligned} P(Y_n \leq t) &= P\left(\bigcup_{k=0}^{+\infty} X_n \in [k; k+t]\right) \\ &= \sum_{k=0}^{+\infty} P(X_n \in [k; k+t]) \\ &= \sum_{k=0}^{+\infty} \frac{1}{n} \int_k^{k+t} e^{-\frac{1}{n}t} dt \\ &= \sum_{k=0}^{+\infty} \left(-e^{-\frac{1}{n}(k+t)} + e^{-\frac{1}{n}k} \right) \\ &= \left(1 - e^{-\frac{t}{n}}\right) \sum_{k=0}^{+\infty} e^{-\frac{k}{n}} \\ &= \left(1 - e^{-\frac{t}{n}}\right) \frac{1}{1 - e^{-\frac{1}{n}}} \end{aligned}$$

D'où finalement,

$$\forall n \in \mathbb{N}^*, \quad F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ \frac{1 - e^{-\frac{t}{n}}}{1 - e^{-\frac{1}{n}}} & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases}$$

3. On a

$$1 - e^{-\frac{1}{n}} \underset{n \rightarrow +\infty}{\sim} \frac{1}{n}, \quad 1 - e^{-\frac{t}{n}} \underset{n \rightarrow +\infty}{\sim} \frac{t}{n}$$

D'où

$$\lim_{n \rightarrow +\infty} F_n(t) = \begin{cases} 0 & \text{si } t < 0 \\ t & \text{si } 0 \leq t \leq 1 \\ 1 & \text{si } t > 1 \end{cases}$$

On reconnait bien une loi uniforme sur $[0; 1]$.

Exercice 14

1. On désigne par X_i le volume contenu dans la tasse i (qui sont toutes indépendantes). Alors on sait que

$$V_n = \frac{X_1 + \dots + X_n}{n}$$

Si la machine n'est pas dérégulée, comme n est "grand", on sait alors par TCL (2ème forme) qu'on peut supposer :

$$\frac{V_n - \mathbb{E}[X_1]}{\sqrt{S_n}/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

i.e.

$$\frac{V_n - 40}{\sqrt{S_n}/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

On sait alors que

$$P\left(\left|\frac{V_n - 40}{\sqrt{S_n}/\sqrt{n}}\right| \leq x\right) \simeq 2\Phi(x) - 1$$

Avec la probabilité cherchée, on a donc

$$2\Phi(x) - 1 = 0.95$$

et donc

$$x = \Phi^{-1}\left(\frac{0.95 + 1}{2}\right) \simeq 1.96$$

2. Il suffit de développer l'expression ci-dessus.

$$\left|\frac{V_n - 40}{\sqrt{2}/\sqrt{n}}\right| \leq x \Leftrightarrow 40 - \frac{x\sqrt{2}}{\sqrt{n}} \leq V_n \leq 40 + \frac{x\sqrt{2}}{\sqrt{n}}$$

On trouve environ

$$I \simeq [39, 608; 40, 391]$$

3. D'après les données précédentes, on sait que

$$0.95 \simeq P(V_n \in [39, 608; 40, 391])$$

Or ici, on a

$$V_n = 39 \notin I$$

On peut donc rejeter l'hypothèse de travail $\mathcal{H}_0$.

On affirme que la machine est dérégulée.

Exercice 15

1. Si la pièce n'est pas truquée, alors le nombre X de piles va suivre une loi binomiale

$$X \hookrightarrow \mathcal{B}(50, \frac{1}{2})$$

Or, $F_n = \frac{X}{50}$. C'est une moyenne du type $F_n = \frac{X_1 + \dots + X_n}{n}$ avec $X_i \hookrightarrow \mathcal{B}(\frac{1}{2})$ d'après le théorème de Moivre Laplace, on sait alors que

$$\frac{X - n\frac{1}{2}}{\sqrt{\frac{1}{4}n}} \hookrightarrow \mathcal{N}(0; 1)$$

i.e.

$$\frac{F_n - \frac{1}{2}}{\sqrt{\frac{1}{4}/\sqrt{n}}} \hookrightarrow \mathcal{N}(0; 1)$$

On sait alors que

$$P\left(\left|\frac{F_n - \frac{1}{2}}{\sqrt{\frac{1}{4}/\sqrt{n}}}\right| \leq x\right) \simeq 2\Phi(x) - 1$$

Avec la probabilité cherchée, on a donc

$$2\Phi(x) - 1 = 0.95$$

et donc

$$x = \Phi^{-1}\left(\frac{0.95 + 1}{2}\right) \simeq 1.96$$

2. Il suffit de développer l'expression ci-dessus.

$$\left|\frac{F_n - \frac{1}{2}}{\sqrt{\frac{1}{4}/\sqrt{n}}}\right| \leq x \Leftrightarrow \frac{1}{2} - \frac{x}{2\sqrt{n}} \leq F_n \leq \frac{1}{2} + \frac{x}{2\sqrt{n}}$$

On trouve environ

$$a \simeq 0.139$$

3. a) D'après les données précédentes, on sait que

$$0.95 \simeq P\left(F_n \in \left[\frac{1}{2} - a, \frac{1}{2} + a\right]\right) \simeq P(F_n \in \underbrace{[0.3614, 0.6386]}_I)$$

Or ici, on a

$$F_n = \frac{29}{50} \simeq 0.58 \in I$$

On ne peut pas rejeter l'hypothèse de travail $\mathcal{H}_0$.

On ne peut donc pas être certain que la pièce soit truquée.

- b) On sait de même que

$$0.95 \simeq P(F_n \in \underbrace{[0.3614, 0.6386]}_I)$$

Or ici, on a

$$F_n = \frac{32}{50} = 0.64 \notin I$$

On rejette donc l'hypothèse de travail $\mathcal{H}_0$ et on annonce que la pièce est truquée.

Exercice 16

1. Ainsi, la famille $(X_1, \dots, X_{100})$ est un 100-échantillon de X_1 . Comme $p = \mathbb{E}[X]$, d'après la loi faible des grands nombres, la moyenne empirique

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$$

avec $n = 100$ approche p .

2. • Moyenne empirique :

Avec $n = 100$, on sait que

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n} = \frac{38}{100} = 0,38$$

(donc une approximation de p serait 0,38)

- Variance empirique :

On sait que la variance empirique S_n^2 se calcule par

$$S_n^2 = \overline{X^2} - \overline{X}^2$$

Or, comme X est une variable de Bernoulli,

$$X^2 = X$$

D'où

$$S_n^2 = \overline{X_n} - \overline{X_n}^2 = \overline{X_n^2}(1 - \overline{X_n^2})$$

(On constate au passage que c'est la même formule que pour la loi de Bernoulli avec p ...)

On obtient alors

$$S_n^2 = 0,38(1 - 0,38) = 0,38 \times 0,62 = 0,2356$$

3. a) Remarquons que

$$\overline{X_n} - u \frac{S_n}{\sqrt{n}} \leq p \leq \overline{X_n} + u \frac{S_n}{\sqrt{n}} \Leftrightarrow \left| \frac{\overline{X_n} - p}{S_n/\sqrt{n}} \right| \leq u$$

On rappelle que, d'après la deuxième forme du TCL, on peut supposer que

$$\frac{\overline{X_n} - p}{S_n/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

D'où

$$\begin{aligned} P\left(\left|\frac{\overline{X_n} - p}{S_n/\sqrt{n}}\right| \leq u\right) &= 0,99 \Leftrightarrow 2\Phi(u) - 1 = 0,99 \\ &\Leftrightarrow \Phi(u) = \frac{1 + 0,99}{2} \\ &\Leftrightarrow u = \Phi^{-1}\left(\frac{1 + 0,99}{2}\right) = 0,995 \end{aligned}$$

Par la calculatrice, Python ou la table de la loi normale, on trouve

$$u \simeq 2,576$$

b) Le calcul avec les valeurs trouvées donne alors

$$I_{0,99} = \left[\overline{X_n} - u \frac{S_n}{\sqrt{n}} \leq p \leq \overline{X_n} + u \frac{S_n}{\sqrt{n}} \right]$$

et finalement

$$I_{0,99} = [0,255; 0,505]$$

c) Pour un risque de 5%, on trouve

$$u \simeq 1,960$$

et donc, par les mêmes calculs,

$$I_{0,95} = [0,285; 0,475]$$

4. On peut affirmer avec confiance de 95% que la probabilité pour un client de laisser un pourboire n'est pas de 50% car on est certain que $p \in [0,285; 0,475]$ avec une confiance de 95%.

En revanche, si on raisonne avec une confiance de 99%, on ne peut pas exclure cette possibilité car $0,5 \in I_{0,99} = [0,255; 0,505]$.

Dans tous les cas, on ne peut jamais conclure que la proportion de client sera de 50%. On ne peut au mieux que l'exclure.

5. a) On sait que $X \hookrightarrow \mathcal{B}(n, p)$ où p est la probabilité pour un client de laisser un pourboire.
On a ici $n \geq 30$.

- Voyons si on a également $np \geq 5$ et $np(1-p) \geq 5$:
Avec une confiance de 95%, on a

$$0,285 \leq p \leq 0,475$$

D'où

$$\underbrace{120 \times 0,285}_{34,2} \leq np \leq \underbrace{120 \times 0,475}_{57}$$

En particulier,

$$np \geq 5.$$

De plus,

$$\underbrace{1 - 0,475}_{0,525} \leq 1 - p \leq \underbrace{1 - 0,285}_{0,715},$$

d'où

$$\underbrace{0,525 \times 34,2}_{17,955} \leq np(1-p) \leq \underbrace{0,715 \times 57}_{40,755}$$

En particulier,

$$np(1-p) \geq 5$$

- D'après le théorème central limite (plus précisément le théorème de Moivre-Laplace), comme $n \geq 30$, $n \cdot p \geq n \cdot 0,285 \geq 5$ et $n \cdot p \cdot (1-p) \geq 5$, on peut admettre que

$$X^* = \frac{X - np}{\sqrt{np(1-p)}} \hookrightarrow \mathcal{N}(0, 1)$$

D'où

$$\begin{aligned} P(X \leq 100) &= P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{100 - np}{\sqrt{np(1-p)}}\right) \\ &= \Phi\left(\frac{100 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

Commençons par encadrer $\frac{100 - np}{\sqrt{np(1-p)}}$:
Avec les calculs effectués plus haut, on peut dire que

$$\underbrace{100 - 57}_{43} \leq 100 - np \leq \underbrace{100 - 34,2}_{65,8}$$

et également que

$$\underbrace{\frac{1}{\sqrt{40,755}}}_{0,1566} \leq \frac{1}{\sqrt{np(1-p)}} \leq \underbrace{\frac{1}{\sqrt{17,955}}}_{0,236}$$

D'où

$$\underbrace{0,1566 \times 43}_{6,736} \leq \frac{100 - np}{\sqrt{np(1-p)}} \leq \underbrace{0,236 \times 65,8}_{15,58}$$

Par croissance de Φ , ceci signifie que

$$\Phi(6,736) \leq \Phi\left(\frac{100 - np}{\sqrt{np(1-p)}}\right) \leq \Phi(15,58)$$

En prenant la calculatrice (ou en se souvenant que pour la loi normale $\mathcal{N}(0, 1)$, tout se passe entre -4 et 4 , on obtient

$$1 \leq \Phi\left(\frac{100 - np}{\sqrt{np(1-p)}}\right) \leq 1$$

$$\boxed{P(X \leq 100) \simeq 1}$$

- b) Il suffit de remplacer 100 par 60 dans la réponse précédente. Ceci donne

$$\underbrace{60 - 57}_3 \leq 100 - np \leq \underbrace{60 - 34,2}_{25,8}$$

$$\underbrace{0,1566 \times 3}_{0,47} \leq \frac{60 - np}{\sqrt{np(1-p)}} \leq \underbrace{0,236 \times 25,8}_{6,09}$$

d'où au final

$$\underbrace{\Phi(0,47)}_{0,68} \leq \Phi\left(\frac{60 - np}{\sqrt{np(1-p)}}\right) \leq \underbrace{\Phi(6,09)}_{=1}$$

D'où

$$\boxed{P(X \leq 60) \in [0,68; 1]}$$

1. a) X_A est la taille d'un seul individu. On sait que

$$\mathbb{E}[X_A] = \mu_A \quad \sigma(X_A) = \sigma_A$$

Or, d'après le TCL première forme, on sait que l'on peut supposer que

$$T = \frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \hookrightarrow \mathcal{N}(0, 1)$$

D'où

$$P\left(\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \geq a\right) \simeq 1 - \phi(a)$$

d'où, en notant $z = 550$:

$$\begin{aligned} P(M_n \geq z) &= P\left(\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \geq \frac{z - \mu_A}{\sigma_A/\sqrt{n}}\right) \\ &= P\left(\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \geq \frac{550 - 570}{152/\sqrt{250}}\right) \\ &\simeq P\left(\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \geq -2,080\right) \\ &\simeq 1 - \phi(-2,080) \simeq 0,98 \end{aligned}$$

D'où

$$\boxed{P(M_n \geq 550) \simeq 0,98}$$

- b) D'après le même raisonnement que tout à l'heure, on a

$$0,95 = P[M_n > h] = P\left(\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}} \geq \frac{h - \mu_A}{\sigma_A/\sqrt{n}}\right) = 1 - \phi\left(\frac{h - \mu_A}{\sigma_A/\sqrt{n}}\right)$$

Ainsi,

$$\frac{h - \mu_A}{\sigma_A/\sqrt{n}} = \phi^{-1}(1 - 0,95) \simeq -1,645$$

Ainsi,

$$h \simeq -1,645 \times \frac{\sigma_A}{\sqrt{n}} + \mu_A \simeq 554,2$$

i.e.

$$\boxed{h \simeq 554}$$

- c) D'après le TCL (1^{ère} forme), on a

$$P\left(\underbrace{\left|\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}}\right| < u}_{\phi(u) - \phi(-u) = 2\phi(u) - 1}\right) \simeq \underbrace{0,95}_{\beta} \quad (1)$$

avec

$$\phi(u) = \frac{1 + \beta}{2} = \frac{1 + 0,95}{2} = 0,975$$

et donc

$$u = \phi^{-1}(0,975) \simeq 1,96$$

Or, d'après 1, on a

$$P\left(|M_n - \mu_A| < \frac{\sigma_A}{\sqrt{n}}u\right) \simeq 0,95$$

Ainsi,

$$a = \frac{\sigma_A}{\sqrt{n}}u$$

donc

$$\boxed{a \simeq 18,8}$$

d'où

$$\boxed{I_A \simeq [551; 589]}$$

- d) D'après le TCL (1^{ère} forme), on a

$$P\left(\left|\frac{M_n - \mu_A}{\sigma_A/\sqrt{n}}\right| < u_{1-\frac{0,1}{2}}\right) \simeq 0,9$$

avec

$$\phi(u_{1-\frac{0,1}{2}}) = 0,9$$

D'où

$$P\left(|M_n - \mu_A| < \frac{\sigma_A}{\sqrt{n}}u\right) \simeq 0,9$$

Ainsi,

$$a = \frac{\sigma_A}{\sqrt{n}}u$$

avec

$$u = \phi^{-1}(0,95) \simeq 1,645$$

d'où

$$\boxed{I_{A'} \simeq [554; 585]}$$

- e) On souhaite connaître n tel que M_n (moyenne empirique observée) satisfasse

$$P(|\mu_A - M_n| \leq 10) = 0,99$$

Or, d'après le TCL (1^{ère} forme), on a

$$\begin{aligned} 0,99 = P(|\mu_A - M_n| \leq 10) &= P\left(\frac{|\mu_A - M_n|}{\sigma_A/\sqrt{n}} \leq \frac{10}{\sigma_A/\sqrt{n}}\right) \\ &= 2\phi\left(\frac{10}{\sigma_A/\sqrt{n}}\right) - 1 \end{aligned}$$

Ainsi,

$$\phi\left(\frac{10}{\sigma_A/\sqrt{n}}\right) = \frac{1}{2}(0,99 + 1) = 0,995$$

mais

$$\phi^{-1}(0,995) \simeq 2,579$$

D'où

$$\frac{10\sqrt{n}}{\sigma_A} \simeq 2,579$$

et donc

$$n \simeq \left(2,579 \frac{\sigma_A}{10}\right)^2 \simeq 1532,9$$

On prendrait donc

$$\boxed{\text{au moins 1 533 individus}}$$

2. Dans la première question, on a trouvé que la moyenne empirique de la population doit se trouver dans I_A ou dans $I_{A'}$ (suivant la confiance que l'on met dans l'intervalle.), plus précisément, on a montré que :

$$P(M_n \in [551; 589]) \simeq 0,95, \quad P(M_n \in [554; 585]) \simeq 0,9$$

S'il y a une différence de taille entre les populations A et B , alors

$$\underbrace{M_n}_{\text{moyenne empirique pour } A} \neq \underbrace{m_B}_{\text{moyenne empirique pour } B}$$

et donc dans ce cas Mais on a

$$m_B = 555 \in [551; 589] \quad ; \quad m_B = 555 \in [554; 585]$$

On ne peut donc pas exclure le fait que $M_n = m_B$ (dans aucun des deux cas)
A priori,

on ne peut donc pas affirmer que la différence soit notable, ni à 90%, ni à 95%.

3. a) En faisant le total du nombre d'individu pour chaque classe, on trouve

$$\boxed{150 \text{ individus}}$$

- b) Pour calculer la moyenne et l'écart-type, on ramène chaque classe à son centre. De ce fait, on calcule finalement la moyenne et l'écart-type de

x	200	350	450	550	650	800
n_i	10	19	30	38	31	22

Calcul de la moyenne :

$$m_C = \frac{10 \times 200 + \dots + 22 \times 800}{150} = 538,7$$

Calcul de l'écart-type :

Le calcul peut se faire de plusieurs manières :

- par la formule de Huygens Koenig :

$$s_C^2 = s_{C,C} - m_C^2$$

où

$$s_{C,C} = \frac{1}{n} \sum n_i x_i^2 = \frac{1}{150} (10 \times 200^2 + \dots + 22 \times 800^2)$$

- par la formule de la différence à la moyenne :

$$s_C^2 = \frac{1}{n} \sum n_i (x_i - m_C)^2 = \frac{1}{150} (10 \times (200 - m_C)^2 + \dots + 22 \times (800 - m_C)^2)$$

Résumé :

On obtient :

$$\boxed{m_C = 538,7 \quad s_C \simeq 162,3}$$

- c) Si la question avait été " donner une estimation de μ_C " sans parler de confiance, on aurait répondu :

par la loi faible des grands nombres, on sait que $\mu_C \simeq m_C$

mais ici, on nous parle de confiance à 95%. Il faut donc préciser un peu. On veut trouver un intervalle I tel que

$$P(\mu_C \in I) = 0.95$$

Il prend donc ici l'intervalle de confiance de μ_C donné par la formule de cours :

$$\left[m_C - u_\beta \frac{s_n^2}{\sqrt{n}} < \mu < m_C + u_\beta \frac{s_n^2}{\sqrt{n}} \right]$$

avec ici $n = 150$ et par le cours, u_β déterminé par $\Phi(u_\beta) = \frac{1+0.95}{2}$ car la confiance est de 0.95. On obtient

$$\mu_C \in I_C = [512, 7; 564, 6] \quad \text{avec une confiance de 95\%}$$

- d) Étant donné que

$$\mu_A = 570 \notin I_C$$

et que

$$\mu_C \in I_C \text{ avec probabilité } 0,95$$

on peut affirmer avec une confiance de 95% que

$$\mu_A \neq \mu_C$$

et qu'il y a donc une différence significative de taille entre A et C .

On pourrait même dire que la taille moyenne dans la population A est supérieure à celle de C car μ_A est supérieur à toutes les valeurs de I_C .

4. On a, avec probabilité 0,95,

$$I_D \cap I_C \neq \emptyset$$

On ne peut donc pas exclure le fait que $\mu_C = \mu_D$, autrement dit,

On ne peut donc pas conclure à une différence de taille moyenne entre C et D .