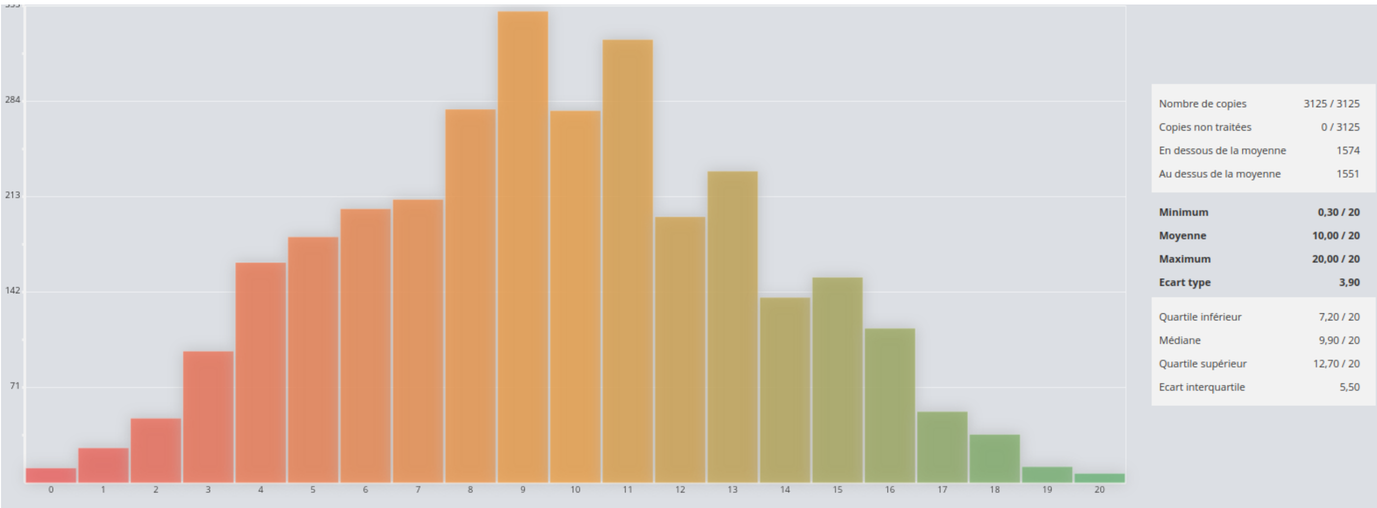
Rapport épreuve Modélisation mathématique et Informatique 2021

1 Statistiques générales de l'épreuve



2 Introduction

Le sujet traitait de la modélisation de l'espérance de vie d'une cohorte d'individus subissant la variole et l'impact de la vaccination.
L'épreuve comportait 4 parties, la partie 3 étant consacrée à l'informatique. La première partie proposait une modélisation classiquede l'espérance de vie à partir d'une équation différentielle linéaire du premier ordre. Les parties 2 et 3 développaient un modèle historique proposé par Bernoulli en 1760 pour étudier les bénéfices d'une vaccination sur la cohorte. La dernière partie proposait une troisième approche probabiliste basée sur une chaîne de Markov.

3 Remarques générales sur l'épreuve

Le sujet abordait des thématiques variées : analyse fonctionnelle, probabilités, algèbre linéaire, algorithmique et programmation. Dans l'ensemble, les résultats ont été satisfaisants. Parmi les points positifs l'équipe de correction a pu noter la quasi absence de copies vides. Les meilleures copies ont abordé la totalité du sujet. La majorité des candidats a traité les 4 parties. Les parties 1, 3 et 4 ont été correctement traitées, la partie 2 a posé plus de problèmes aux candidats. La partie 4 a sans doute été plus discriminante que les autres en raison de questions accessibles en fin de sujet qui ont bénéficié aux candidats qui ont préféré passer du temps sur cette partie. Le jury salue également un effort de rédaction, notamment en informatique où de plus en plus de candidats pensent bien à expliquer leurs codes.
De nombreux candidats s'en sortent bien en informatique. En particulier, certains candidats qui n'ont pas su montrer leur qualités mathématiques ont su montrer de réelles compétences en informatique.

Dans de nombreuses copies le modèle probabiliste de la partie 4 a été bien compris et la rédaction des questions de probabilités a été bien faite.
Beaucoup de candidats n'ont pas reculé devant l'aspect calculatoire de la dernière partie, et notamment le calcul des valeurs propres et de l'inverse d'une matrice, avec succès.
Dans les points négatifs, les interprétations ont souvent posé problème aux candidats et notamment la considération des unités des grandeurs étudiées. La notion de cohorte a été mal comprise dans beaucoup de copies parlant de naissances.

4 Difficultés mathématiques notables

1. Beaucoup de confusions entre le symbole "$\Longrightarrow$", le symbole "$\Longleftrightarrow$" et le "donc".
2. Parfois de très grosses erreurs dans l'intégration par parties (Partie 1, question 2.(c))
3. La notation de la dérivée physique $\frac{d}{dt}$ a posé problème à un certain nombre de candidats. Beaucoup de candidats ont du mal à produire des raisonnements et calculs justes quand le cadre utilisé est moins mathématique mais plus appliqué.
4. Une difficulté de la partie 2 du sujet venait du fait que la variable ne soit pas explicitée dans l'expression $z(t)$ ou $a(t)$ qui était parfois écrite dans le sujet z ou a.
5. Dans beaucoup de copies la dérivée d'un quotient (Partie 2, question 2.(a)) est devenu le quotient des dérivées.
6. Peu de candidats pensent à la constante d'intégration à la question 3.(b) (partie 2).
7. Le rapport entre la quantité $\mathcal{E}_{\Delta_t}$ et l'espérance de vie était difficile à appréhender
8. Le jury constate que la justification de la diagonalisabilité est souvent mal comprise par les candidats et que les propriétés sur le calcul matriciel souvent mal connues (non-commutativité du produit, puissance d'une matrice A qui n'est pas la matrice dont les coefficients sont les puissances des coefficients de A)

5 Difficultés informatiques notables

1. Quelques confusions entre les **print** et les **return**
2. Quelques erreurs sur les bornes fournies par la fonction **range**
3. La méthode des moindres carrés est bien maîtrisée dans l'ensemble, un peu plus de confusion quand il s'est agi de justifier l'utilisation d'une modélisation exponentielle plutôt que linéaire
4. La question 3.(b) était la plus difficile de la partie, elle a été peu abordée et souvent confondue avec la dichotomie classique

6 Éléments de correction

Le jury propose dans cette partie des éléments de réponse à des questions choisies ayant posé des problèmes spécifiques aux candidats.

- **Partie 1, question 3.(b)**
$\mathcal{E}_{\Delta_t}$ est homogène à un temps. On peut l'interpréter comme le temps de vie moyen : cela fournira une valeur approchée de l'espérance de vie à la question 3.(c))

- **Partie 2, question 2.(a)**

$$\frac{dq}{dt} = \frac{x'S - S'x}{S^2}$$

or $x = R + S$ donc d'après le système d'équation initial :

$$\begin{aligned} x' &= R' + S' \\ &= S - aR - (a+b+c)S \\ &= -a(R+S) - cS \\ &= -ax - cS \end{aligned}$$

Par ailleurs $S' = -(a+b+c)S$ d'où finalement :

$$\begin{aligned} \frac{dq}{dt} &= \frac{-axS - cS^2 + (a+b+c)xS}{S^2} \\ &= \frac{(b+c)xS - cS^2}{S^2} \\ &= (b+c)q - c \end{aligned}$$

- **Partie 2, question 3.(b)**
On a $\frac{d\ln G}{dt} = \frac{d}{dt}\left(\ln \frac{z}{x}\right)$ donc il existe une constante C telle que pour tout t :

$$\ln(G(t)) = \ln\left(\frac{z(t)}{x(t)}\right) + C.$$

Ainsi :

$$\frac{z}{x} = e^C G.$$

Or $\frac{z(0}{x(0)} = 1$ (la cohorte fictive est initialement identique à la cohorte réelle) et $G(0) = \frac{1}{8}$ donc $e^C = 8$.

- **Partie 2, question 5**
D'après la question 3. de la partie 1 on peut approcher l'espérance de vie de cette population fictive par

$$\mathcal{E}_{\Delta_t} = \frac{\sum_{i=0}^{+\infty} i\,(y(i) - y(i+1))}{y_0} = \frac{\gamma}{y_0}\sum_{i=0}^{+\infty} i\frac{8e^{i/8}}{1+7e^{i/8}}\,(x(i) - x(i+1))$$

Il reste à fixer le pas de temps Δ_t. Dans la cohorte de Bernoulli, le plus petit pas de temps considéré est l'année.

- **Partie 3, question 4.(a)**
Le "return True" ne doit pas être à l'intérieur de la boucle, sinon le programme s'arrête dès la première itération positive.

- **Partie 4, question 2.(a)**
  On montre que
  $$Z_{n+1} = QZ_n$$
  en utilisant la formule des probabilités totales appliquée au système complet d'évènement $\{S, D, R\}$. On montre la question par récurrence.

- **Partie 4, question 2.(b)**
  Q est une matrice diagonale, ses valeurs propres sont donc les coefficients de sa diagonale, a, $\lambda_1 = 1 - (a + b + c)$ et $\lambda_2 = 1 - a$. De plus $a > 0$ (car les individus remis ont une probabilité non nulle de mourir) donc $\lambda_2 < 1$. De même $b + c > 0$ (car les probabilités d'attraper la variole ou de mourir de la variole sont également non-nulles) donc $\lambda_1 < \lambda_2$. Enfin $\lambda_1 > 0$ car λ_1 est la probabilité qu'un individu susceptible soit encore susceptible l'année suivant.

- **Partie 4, question 3.(a)**
  Q est une matrice carré de taille 3 admettant trois valeurs propres distinctes donc Q est diagonalisable. Il existe donc une matrice P constituée des vecteurs propres de Q.

- **Partie 4, question 3.(c)**
  En suivant la suggestion de l'énoncé on a $\mathcal{E} = \mathbb{E}[T]$. D'où :
  $$\begin{aligned}\mathcal{E} &= \sum_{n=1}^{+\infty} n\mathbb{P}(T = n)\\ &= \sum_{n=1}^{+\infty} n\mathbb{P}(A_n)\end{aligned}$$
  en effet, si A_n est vrai cela signifie que l'individu est vivant jusqu'au temps n et meurt au temps $n + 1$, soit de façon équivalente que $T_n = n$. On termine la question en se servant de la question précédente.