Modélisation mathématique et information : rapport du jury

Introduction

Le sujet traitait de la cinétique d'une réaction catalysée par une enzyme agissant sur un substrat unique et de sa modélisation mathématique. Un outil omniprésent est la méthode proposée par Michaelis et Menten, qui aboutit à l'équation qui porte leur nom. Cette méthode s'avère être un cas particulier de la théorie des perturbations singulières, qui s'intéresse aux situations dans lesquelles l'évolution d'un système a lieu suivant plusieurs dynamiques différentes, certaines rapides et d'autres lentes.

Les parties 1 à 3 utilisaient la réduction de modèle ainsi obtenue et se consacraient aux conséquences et aux prédictions obtenues à l'aide de cette réduction.

La partie 4 présentait une approche statistique simplifiée du traitement des données expérimentales.

Enfin, la partie 5 proposait une justification mathématique de la réduction utilisée jusque là.

1 Remarques générales sur l'épreuve

Le sujet abordait de nombreux sujets : équations différentielles, probabilités et statistiques, analyse et informatique. Dans l'ensemble, les résultats ont été satisfaisants. Les parties 1 et 2 ont été globalement bien réussies. Notamment, alors que l'an dernier les candidats avaient été très nombreux à omettre l'explication de leur code dans la partie informatique, ils ont cette année majoritairement pensé à l'inclure. Les parties 3, 4 et 5, plus difficiles, ont été moins bien réussies. Malgré la longueur du sujet, presque tous les candidats ont abordé les 5 parties, même si la partie 5 a quasiment toujours été survolée.

2 Difficultés mathématiques notables

1. De nombreux étudiants ont des difficultés à appréhender les équations différentielles. Certains mélangent les constantes et les variables. Souvent, les étudiants se raccrochent à l'interprétation physique du problème pour pallier leur incompréhension des équations différentielles. C'est un piège, puisque le coeur du sujet consistait justement à donner des preuves mathématiques de principes chimiques bien connus.
2. L'interprétation du coefficient de corrélation linéaire est souvent incorrecte. De nombreux candidats essayent de relier sa valeur avec le nombre de points par lequel passe la droite de régression linéaire.
3. Peu de candidats savent que la droite de régression linéaire passe par le point de coordonnées (moyenne des x, moyenne des y). Très peu sont capables de donner l'équation de la droite, alors même qu'elle était écrite dans le script fourni.
4. Il y a beaucoup d'erreurs sur le théorème de la bijection. Les hypothèses ne sont pas, ou mal, vérifiées (en particulier la continuité est souvent oubliée, de même que la stricte monotonie). Les limites de la fonction initiale ne sont pas calculées pour déterminer l'ensemble de définition de la fonction réciproque. Le tracé du graphe de la réciproque est régulièrement incorrect : la fonction est représentée décroissante alors qu'elle devrait être croissante, elle est représentée avec une asymptote alors qu'elle tend vers l'infini, etc.
5. De manière générale, peu de soin est accordé aux représentations graphiques, qui devraient pourtant être des questions « cadeaux ».

6. La densité d'une loi continue n'est pas toujours centrée sur sa moyenne. La quasi-totalité des étudiants invoque pourtant cet argument pour sélectionner la bonne courbe d'une loi du χ^2 et se trompe. La positivité de la loi du χ^2 n'est presque jamais utilisée pour discriminer parmi les courbes.
7. De nombreux candidats ne connaissent pas la formule de la covariance. En outre, la formule de la variance est trop souvent incorrecte.

3 Difficultés informatiques notables

1. Même si cela était souvent acceptable au vu des consignes, de nombreux candidats écrasent les données d'entrée pour construire les données de sortie. C'est une mauvaise habitude.
2. De nombreux candidats effectuent des opérations entre des données de types différents, comme une liste et un nombre. La surcharge de l'opérateur + n'y est probablement pas étrangère, mais une réflexion rapide sur le type des objets en jeu dans une expression permettrait souvent d'éviter des erreurs manifestes.
3. Souvent, les candidats pensent pouvoir effectuer des opérations sur les listes comme sur des vecteurs (par exemple 2*L pour multiplier une liste par 2).

4 Eléments de correction

Le jury propose dans cette partie des éléments de réponse à des questions ayant posé des problèmes spécifiques aux candidats. Il est à noter que les méthodes employées n'étaient pas les seules possibles et que toute autre méthode mathématiquement valide était acceptée.

- **Partie 3, question 3**
  Très peu de bonnes réponses ont été données à cette question qui repose pourtant exclusivement sur des manipulations algébriques.
  On définit $y(t) = g\left(\frac{s(t)}{K_M}\right)$. La fonction y est dérivable sur $\mathbb{R}_+^*$ en tant que composée de fonctions dérivables. Pour tout $t > 0$,
  $$\begin{aligned} y'(t) &= \frac{s'(t)}{K_M} g'\left(\frac{s(t)}{K_M}\right) \\ &= \frac{s'(t)}{K_M}\left(1 + \frac{s(t)}{K_M}\right) e^{s(t)/K_M}. \end{aligned}$$
  Or, pour $t \geq \delta$, $s'(t) = -\frac{v_{\max} s(t)}{K_M + s(t)}$ d'après la question 1. Donc, pour $t \geq \delta$,
  $$\begin{aligned} y'(t) &= -\frac{v_{\max} s(t)}{K_M + s(t)} \times \frac{K_M + s(t)}{K_M^2} e^{s(t)/K_M} \\ &= -\frac{v_{\max} s(t)}{K_M^2} e^{s(t)/K_M} \\ &= -\frac{v_{\max}}{K_M} g\left(\frac{s(t)}{K_M}\right) = -\frac{v_{\max}}{K_M} y(t). \end{aligned}$$
  Ainsi, y est solution de l'équation différentielle $y' = -\frac{v_{\max}}{K_M} y$, ayant pour condition initiale $y(\delta) = g\left(\frac{s(\delta)}{K_M}\right) = g\left(\frac{s_0}{K_M}\right) = \frac{s_0}{K_M} e^{s_0/K_M}$.
- **Partie 4, question 4**
  Cette question consistait simplement à lire le sujet et à appliquer la définition d'une loi du χ^2 donnée au préalable ; elle a pourtant été mal réussie. En effet, l'hypothèse d'indépendance a été excessivement rarement invoquée. La plupart du temps, les candidats ont voulu appliquer un résultat de leur cours sur la somme de variables aléatoires gaussiennes indépendantes pour conclure que la

variable Z était elle aussi gaussienne. Cependant, ici, la somme portait sur des variables aléatoires gaussiennes indépendantes **au carré** et ce résultat de cours ne pouvait donc pas s'appliquer.
D'après l'hypothèse (H) :

$$\forall i \in \{1, \ldots, n\}, \quad M_i - p(t_i) = R_i \sim \mathcal{N}(0; \sigma^2)$$

et ces variables sont indépendantes. Donc les $\frac{M_i - p(t_i)}{\sigma} \sim \mathcal{N}(0; 1)$ et sont indépendantes. Par définition de la loi du χ^2,

$$Z \sim \chi^2(n).$$

- Les questions d'interprétation ont reçu des réponses plutôt décevantes. En particulier, peu de candidats sont capables de lier la réalité, le modèle et les expériences.
  Partie 4, question 7
  On a vu que, sous l'hypothèse (H), Z suivait une loi du $\chi^2(9)$ dont la densité est représentée à la figure (c) page 9 de l'énoncé. D'après la question précédente, la réalisation z de Z vaut 17.0914, qui est une valeur peu plausible. Cela amène à remettre en question l'hypothèse (H), c'est-à-dire à remettre en question le fait que nos mesures correspondent aux valeurs théoriques prédites par le modèle à une erreur de mesure près.
  Une cause envisageable est l'hypothèse d'AEQS. En effet, d'après les parties précédentes, elle implique les valeurs théoriques données dans le modèle (H). Si l'AEQS n'est pas vérifiée, le modèle considéré est alors faux.
  Une autre cause envisageable se situe du côté des erreurs de mesures R_i. Peut-être ne sont-elles pas normales, ou pas indépendantes, ou peut-être qu'elles n'ont pas la même variance. Dans tous ces cas, on ne peut plus déterminer la loi de la variable aléatoire Z et les conclusions précédentes ne sont plus valables.
  Dans tous les cas, il faut remettre en cause soit les équation différentielles et/ou l'AEQS, soit le processus expérimental, soit les deux.
- **Partie 5, question 1**
  Un certain nombre de candidats a pris pour acquis l'unité de k_1 en invoquant le fait que k_1 est une constante de réaction, et donc en mol^{-1}.L.s^{-1}. Même si cela s'avère vrai dans ce cas précis, l'unité de k_2 et de k_{-1} est différente, en l'occurrence s^{-1}. On attendait donc que les candidats justifient l'unité de k_1, d'autant plus que, l'unité de t_0 étant donnée, on pouvait déduire de l'énoncé l'unité de k_1.
  e_0 et s_0 ont la même unité, donc ε est sans unité. D'après l'équation (E), on sait que $\frac{\mathrm{d}s}{\mathrm{d}t}$ et $k_1 es$ ont la même unité. L'unité de $\frac{\mathrm{d}s}{\mathrm{d}t}$ est mol.L^{-1}.s^{-1} et celle de es est mol^2.L^{-2}, donc l'unité de k_1 est mol^{-1}.L.s^{-1}. Ainsi, l'unité de $k_1 e_0$ est s^{-1}, et on en déduit que t_0 est bien homogène à un temps.

5 Conseils aux candidats

1. Lors d'une épreuve de concours, certaines questions sont plus difficiles que d'autres. Cependant, le fait qu'une question soit plus simple ne signifie pas pour autant qu'il faille la négliger. Par exemple, il arrive que certains candidats ne soignent pas suffisamment les questions d'applications numériques et perdent ainsi les points associés de manière regrettable.
2. Beaucoup trop souvent, et à tort, les questions d'interprétation ne sont pas traitées. Ces questions sont habituellement des questions plutôt ouvertes, dans lesquelles une réponse précise n'est pas attendue, et où il s'agit, pour le candidat, de montrer qu'il prend du recul sur le sujet. Toutes les pistes proposées par les candidats sont susceptibles d'être valorisées.