Rapport sur l'épreuve de "Modélisation mathématique et informatique" 2017

Contexte scientifique du sujet de l'épreuve

L'épreuve de modélisation mathématique et informatique visait à guider les candidat·e·sdans la mise en place de différentes méthodes de modélisation mathématique issues du domaine des statistiques et de l'analyse numérique pour aborder un problème de recherche actuel en écophysiologie végétale.

Dans ce domaine, la recherche s'intéresse à l'évolution du nombre de feuilles de plantes en fonction du "temps thermique". Ce dernier correspond à la somme cumulée des températures au cours des différents jours de l'expérience. De telles mesures permettent de mettre en évidence trois phases de développement :

- une phase dite "de rosette" pendant laquelle le nombre de feuilles augmente lentement en fonction du temps thermique ;
- une phase dite "d'élongation" qui correspond à la croissance en hauteur de la tige de la plante. Pendant cette phase, le nombre de feuilles augmente plus vite en fonction du temps thermique que pendant la phase précédente ;
- une phase dite "de floraison" pendant laquelle la plante arrête de produire des feuilles et où le nombre de feuilles est à peu près constant en fonction du temps thermique.

La modélisation la plus souvent utilisée consiste à supposer que l'évolution du nombre de feuilles en fonction du temps thermique se comporte comme une fonction affine au cours de chacune des trois phases précédentes (cf. [2, 6, 1, 7, 3, 5, 4] pour plus de détails sur cette modélisation).

Le problème qui se pose alors consiste à mettre en place des méthodes automatiques permettant de détecter les instants séparant d'une part les phases de rosette et d'élongation et d'autre part les phases d'élongation et de floraison. Il est, en effet, intéressant de voir si ces instants changent de façon significative en fonction des conditions expérimentales auxquelles les plantes pourraient être soumises.

Dans cette épreuve, différentes notions du programme de mathématiques de classe préparatoires ont été abordées : probabilités, algèbre linéaire et analyse. Cette épreuve contient également une partie importante d'informatique. Il est à noter que l'épreuve contenait un certain nombre de questions ouvertes permettant d'évaluer les capacités d'initiative des candidat·e·s en matière de modélisation et d'interprétation des résultats.

Nous reviendrons sur la manière dont les différentes questions ont été traitées, les principales erreurs et les points positifs relevés par l'équipe de correction. Nous proposerons à ce propos quelques pistes pour le traitement des questions les plus délicates.

Commentaires généraux sur le traitement du sujet

Le sujet abordait des thématiques variées : statistiques bivariées, probabilités, algèbre linéaire, notions d'analyse numérique, algorithmique et programmation. La majorité des candidat·e·s ont réparti leur travail de façon plutôt équilibrée entre les questions théoriques et l'informatique.
Parmi les points positifs l'équipe de correction a pu noter la quasi absence de copies vides, l'informatique en particulier permettant même aux candidat·e·s les plus fragiles d'assurer un minimum de points. Les connaissances et les méthodes classiques de programmation sont à présent bien maîtrisés, les candidat·e·s n'hésitant pas à aller chercher jusqu'au bout de la partie 4 des questions à traiter.
Toujours dans le domaine des satisfactions, le sérieux des candidat·e·s et la qualité de la préparation transparaissent dans les questions classiques d'algèbre linéaire ou de probabilité. Cependant cette maitrise, très assurée dans un contexte familier et numérique, se fait beaucoup plus problématique lorsque le contexte devient plus théorique avec des sommes, des espérances, des matrices comportant des paramètres et des variables. Cette difficulté a amené parfois les candidat·e·s à effectuer des manipulations de sommes erronées ou à écrire des formules du type $\mathbb{E}(X/Y) = \mathbb{E}(X)/\mathbb{E}(Y)$, qui ne seraient sans doute pas intervenues dans un contexte plus familier. De même, énoncer une formule de Taylor sans l'appliquer directement à une fonction précise a semblé représenter une difficulté insurmontable pour la majorité des candidat·e·s. Le jury a conscience que ce type de contexte constitue une source de difficultés pour les candidat·e·s, mais la possibilité de raisonner avec des variables et des paramètres constitue un des points incontournables de la démarche de modélisation. S'entraîner à de telles manipulations dans des cas simples est donc nécessaire pour ne pas être déstabilisé le jour de l'épreuve.

Commentaires détaillés et pistes de résolution

PARTIE 1

1. Nous avons constaté beaucoup de manipulations erronées dans les sommations. La gestion du symbole Σ est en effet parfois problématique. Les écritures du type $\sum_{i=1}^{n}\sum_{i=1}^{n} x_i$ conduisent souvent à des erreurs. Il convient de penser à introduire des indices de sommations différents pour des sommes différentes dans un même calcul.
2. Dans une part non négligeable des copies, le calcul de la dérivée partielle est faux dès le départ. Celles qui passent ce cap, après avoir traversé avec succès celui de la première question, arrivent au résultat final.
   On note des problèmes dans la gestion de l'équivalence : le système $\begin{cases} \frac{\partial F}{\partial a}(a,b) = 0 \\ \frac{\partial F}{\partial b}(a,b) = 0 \end{cases}$ n'équivaut pas à $\frac{\partial F}{\partial a}(a,b) = \frac{\partial F}{\partial b}(a,b)$. Il convenait ici d'exprimer a en fonction de b puis de remplacer dans la seconde équation par exemple.
   Remplacer les valeurs proposer par l'énoncé dans les expressions des dérivées partielles pour vérifier la nullité est une bonne idée mais ne permet pas de s'assurer de l'unicité.
3. Cette question a été rarement réussie avec des interprétations parfois très étranges.
4. Il est étonnant de constater que nombreux sont les résultats erronés. Le rôle de a et b, classique en démarche de modélisation mais inversé par rapport aux habitudes sco-

laires concernant les fonctions affines, a pu dérouter les candidat·e·s dans ces premières questions.

5. Cette partie est globalement bien réussie. Il y a moins d'erreurs grossières liées à Python (indentation, range, initialisation) que les dernières années même si certains candidats écrivent encore ab au lieu de $a*b$ ou bien a^2 au lieu de $a**2$. Par contre, on peut lire des algorithmes impeccables, même dans des copies très faibles sur les questions précédentes. En 5 c), il est possible dans une même boucle **for** de faire plusieurs calculs plutôt que d'enchaîner des boucles avec le même indice. Au niveau du nombre de calculs, il est préférable de calculer la valeur **m=moy(x)** avant la boucle et d'utiliser **m** plutôt que le faire calculer à chaque tour de boucles. Par ailleurs la modularité des écritures est appréciée : il est plus lisible d'utiliser **moy(x)** que de ré-écrire des instructions dans la fonction **Bchap(x,y)** qui reprennent les instructions écrites en 5(a)

PARTIE 2

Dans cette partie, on note comme mentionné dans le commentaire général des erreurs assez grossières sur les notions d'espérance et de variance : $\mathbb{E}(a) = 0$, $\mathbb{E}(x_i) = \bar{x}$, $\mathbb{E}(X/Y) = \mathbb{E}(X)/\mathbb{E}(Y)$, $\text{Var}(X/a) = \text{Var}(X)/\text{Var}(a)$, lorsque a est une constante et X et Y sont des variables aléatoires.

8. a) Nombreux sont les candidat·e·s qui connaissent la valeur de la somme des n premiers carrés d'entiers. Il est dommage que la gestion des calculs mène ensuite souvent à une réponse erronée ou incomplète.

8. b) L'inégalité de Bienaymé-Tchebychev est très mal utilisée, voire mal connue.

PARTIE 3

1. Cette question a été très rarement traitée. Le fait qu'une combinaison linéaire de variables aléatoires INDÉPENDANTES qui suivent une loi normale suit une loi normale est peu connu.

2. Le calcul des valeurs propres de la matrice Γ a souvent manqué de rigueur.
   En 2 e) : Être de rang 1 ne signifie pas qu'il n'y a qu'une seule valeur propre.
   En 2 g) : Un certain nombre ont diagonalisé explicitement en trouvant une matrice de passage P. Ils invoquent ensuite le théorème spectral pour prouver que $P \in \mathcal{O}(3)$ ce qui est faux. Le théorème spectral assure qu'il existe une matrice de passage orthogonale mais pas que toutes les matrices de passages sont orthogonales.

3. et 4. Ces questions ont été peu ou mal traitées. Les critères attendus étaient :

$$\widehat{n}_1 = \text{Argmin}_{n_1} \left[\min_{(a_1,b_1,a_2,b_2,a_3,b_3)} \left\{ \sum_{i=1}^{n_1} (Y_i - a_1 - b_1 x_i)^2 + \sum_{i=n_1+1}^{n} (Y_i - a_2 - b_2 x_i)^2 \right\} \right]$$

$$\widehat{T}_1 = x_{\widehat{n}_1}.$$

d'une part, et :

$$(\widehat{n}_1, \widehat{n}_2) = \text{Argmin}_{(n_1,n_2)} \left[\min_{(a_1,b_1,a_2,b_2,a_3,b_3)} \left\{ \sum_{i=1}^{n_1} (Y_i - a_1 - b_1 x_i)^2 + \sum_{i=n_1+1}^{n_2} (Y_i - a_2 - b_2 x_i)^2 + \sum_{i=n_2+1}^{n} (Y_i - a_3 - b_3 x_i)^2 \right\} \right]$$

$$\widehat{T}_1 = x_{\widehat{n}_1} \text{ et } \widehat{T}_2 = x_{\widehat{n}_2},$$

d'autre part. Nous n'attendions pas que les candidats écrivent des critères aussi précis mais qu'il en aient au moins une intuition.

PARTIE 4

1. De façon surprenante, la formule de Taylor-Young est apparue comme mal maîtrisée dans ce contexte.
2. Cette question a été globalement mal traitée ou n'a pas été abordée.
3. a) Cette question a été bien réussie même dans les mauvaises copies.
3. b) Les justifications sont souvent douteuses. La fonction renvoie bien une erreur lorsque la liste est vide car **m=L[0]** pose problème mais pas à cause du **range(1,0)**. La boucle ne s'exécute pas simplement mais ne cause pas l'erreur.
3. c) et d) Les réponses ont été décevantes.
   3(c) : Il fallait remplacer **m<L[k]** par **L[p]<L[k]**. Il s'agissait d'un problème de programmation pas d'un problème d'algorithmique. On ne demandait pas de changer l'algorithme. D'ailleurs souvent une comparaison avec le terme précédent dans la liste a été proposée : **m<L[k]** par **L[p]<L[k+1]** qui ne répond pas à la question.
   3(d)-i : L'objet de la question n'est pas de proposer un algorithme de tri.

La suite du problème n'a été que peu abordée, en dehors du dernier algorithme en 5.c) qui contient souvent de bons éléments lorsque la question est traitée.
Pour les questions plus ouvertes 5.d) et 5.e), nous attendions les réponses suivantes :

5. d) On applique plusieurs itérations de lissage. On calcule les D_i. On cherche les instants correspondants aux deux plus grands D_i en valeur absolue.
5. e) Cette méthode n'est pas très satisfaisante car le lissage va lisser la rupture de pente et donc l'instant trouvé sera approximatif.

Nous avons par ailleurs constaté que nombreux sont les candidats qui rendent des copies mal présentées, mal écrites ou contenant beaucoup de fautes d'orthographe. À partir de l'année prochaine, ces candidats seront sanctionnés.

Références

[1] J. Baker, L. Allen, K. Boote, P. Jones, and J. Jones. Developmental responses of rice to photoperiod and carbon dioxide concentration. *Agricultural and Forest Meteorology*, 50 :201–210, 1990.

[2] R. Bonhomme. Bases and limited to using 'degree.day' units. *European Journal of Agronomy*, 13(1) :1–10, 2000.

[3] M. de Raissac, A. A., R. S., and B. J. Competition between plants affects phenology in rice cultivars. In N. Turner, J. Angus, L. Mc Intyre, M. Robertson, A. Borrell, and D. Lloyd, editors, *New directions for a diverse planet : Proceedings for the 4th International Crop Science Congress*. Gosford : Regional Institute, 26 Sep - 1 Oct 2004.

[4] N. Gomez and D. Miralles. Factors that modify early and late reproductive phases in oilseed rape (brassica napus l .) : Its impact on seed yield and oil content. *Industrial Crops and Products*, 34 :1277–1285, 1990.

[5] S. Lemaire, F. Maupas, P. Cournede, and P. de Reffye. A morphogenetic crop model for sugar-beet (beta vulgaris l.). *International Symposium on Crop Modeling and Decision Support : ISCMDS*, 5 :19–22, 2008.

[6] R. Rickman and B. Klepper. The Phyllochron : where do we go in the future ? *Crop Science*, 35 :44–49, 1995.

[7] F. Tivet. *Etude des facteurs génotypiques et environnementaux déterminant la mise en place de la surface foliaire chez le riz. Incidence particuliere d'un deficit hydrique.* PhD thesis, INA P-G, 2000.