DS 03 - 29 novembre

Durée : 3h. Les calculatrices ne sont pas autorisées. Les exercices sont indépendants.

Questions de cours

1. Soit E de cardinal n. Quel est le cardinal de l'ensemble des parties de E ? (sans justification)
2. Quelle est l'équation de la droite formant l'ajustement linéaire de deux séries (X, Y) ?

Exercice 1

On donne les séries statistiques suivantes :

X	3	8	5	6	1	7	2	4	9
Y	7	17	11	13	3	15	5	9	19

1. Calculer la médiane de chacune des séries.
2. Calculer la moyenne et la variance de chaque série.
3. Calculer leur covariance. Les variables sont-elles corrélées linéairement ? Était-ce prévisible ?

Exercice 2

1. Soit $n \in \mathbb{N}$. Montrer que la suite $((n-k)!(n+k)!)_{k \in [\![0,n]\!]}$ est croissante.
2. En écrivant $k = n - (n-k)$, en déduire que $\forall k \in [\![0, 2n]\!]$, $\binom{2n}{k} \leq \binom{2n}{n}$.
3. Montrer que $1,001^{20}$ est majoré par la somme des 21 premiers termes d'une suite géométrique, dont on précisera la raison.
4. À l'aide d'une majoration grossière, montrer que $\binom{20}{10} \leq 64 \times 10^5$. En déduire que $1,001^{20} \leq \frac{64}{999} 10^8$.
5. On pose $A = \binom{20}{0} + \binom{20}{1} 0,001 + \binom{20}{2} 0,001^2$ et on prend A comme approximation de $1,001^{20}$. Majorer la différence entre ces deux quantités, et en déduire que A est une approximation à 10^{-2} près.
6. Que pensez-vous de la majoration trouvée à la question 4 ?

Exercice 3

Soit $n \in \mathbb{N}^*$. On appelle

— *codon* une succession de trois lettres choisies parmi A, C, G, T ;

— *brin* un assemblage ordonné de n codons ;

— *codon d'initiation* (ou codon start) le codon ATG ;
— *codon de terminaison* (ou codon stop) l'un des trois codons TAG, TGA ou TAA ;
— *séquence codante* un morceau contigu d'un brin qui commence par une codon d'initiation et finit par un codon de terminaison (on lit de gauche à droite).

1. a) Justifier qu'il existe 64 codons différents.
   b) Combien y aurait-il de codons si ceux-ci n'étaient constitués que de 2 lettres parmi A, C, G, T ? Commenter, sachant que ces codons servent à coder 20 acides aminés.
2. Combien y a-t-il de brins différents ?
3. a) Combien de brins possèdent au moins un codon d'initiation ?
   b) Combien de brins possèdent au moins un codon de terminaison ?
   c) Combien de brins satisfont les deux conditions précédentes ?
4. Combien de brins possèdent exactement un codon d'initiation, un codon de terminaison, et une séquence codante ?
5. Lorsqu'un brin contient au moins un codon d'initiation, on note D le numéro du codon d'initiation le plus à gauche.
   a) Soit $k \in [\![1, n-1]\!]$. Déterminer le nombre de brins contenant au moins un codon d'initiation avec $D = k$.
   b) Soit $k \in [\![1, n-1]\!]$. Déterminer le nombre de brins contenant au moins un codon d'initation avec $D = k$ et aucune séquence codante.
   c) Soit $k \in [\![1, n-1]\!]$. Déterminer le nombre de brins contenant au moins un codon d'initiation avec $D = k$ et au moins une séquence codante.
   d) En déduire que le nombre N de brins contenant au moins une séquence codante est
$$N = 64^n - \frac{189}{2} 63^{n-1} + \frac{1}{2} 61^n$$

Exercice 4

On s'intéresse aux mots (concaténation de lettres) qu'il est possible de former avec les lettres A, B, C et obéissant aux contraintes suivantes :

— le mot est de longueur n (il contient n lettres) ;
— il commence et finit par la lettre A ;
— deux lettres adjacentes sont toujours différentes.

Un tel mot sera dit *terne*. On désigne par t_n le nombre de mots ternes de longueur n.

1. Déterminer tous les mots ternes pour $n = 1, 2, 3, 4, 5, 6$.
2. Montrer que pour tout entier $n \geq 3$, $t_n = t_{n-1} + 2t_{n-2}$. *(On justifiera le raisonnement avec soin, en raisonnant par exemple sur la valeur de la $(n-2)$-ème lettre d'un mot terne de longueur n.)*
3. Exprimer t_n en fonction de n.

Questions de cours

1. $\text{card}\mathcal{P}(E) = 2^n$.
2. $y = \frac{\text{Cox}(X,Y)}{V(X)}(x - \overline{X}) + \overline{Y}$.

Correction de l'exercice 1

1. Pour calculer la médiane de ces séries statistiques, il faut tout d'abord ordonner les valeurs. Puis, comme il y a 9 valeurs, la médiane est la valeur du milieu, c'est-à-dire la 5ème, puisqu'il y a alors 4 valeurs avant et 4 après. On trouve que la médiane de X est 5 et la médiane de Y est 11.
2. En ordonnant les valeurs pour la question précédente, on s'aperçoit que la série X est composée de tous les entiers de 1 à 9, et la série Y est arithmétique de premier terme 3 et de raison 2. Les calculs sont alors simplifiés ; on a :

$$\begin{aligned}
\overline{X} &= \frac{1}{9}\sum_{k=1}^{9} k = \frac{9 \times 10}{2} = 5 \\
V(X) &= \left(\frac{1}{9}\sum_{k=1}^{9} k^2\right) - \overline{X} \qquad \text{par la formule de König} \\
&= \frac{1}{9} \times \frac{9 \times 10 \times 19}{6} - 5^2 \\
&= \frac{95}{3} - 25 = \frac{20}{3}
\end{aligned}$$

et pour Y :

$$\begin{aligned}
\overline{Y} &= \frac{1}{9}\sum_{k=1}^{9}(2k+1) = \frac{1}{9} \times \left(9 \times \frac{3+19}{2}\right) = 11 \\
V(Y) &= \frac{1}{9}\sum_{k=1}^{9}(2k+1-11)^2 \\
&= \frac{1}{9}\left(64+36+16+4+0+4+16+64+36\right) \\
&= \frac{100+20+20+100}{9} = \frac{240}{9} = \frac{80}{3}
\end{aligned}$$

3. On applique la formule pour la covariance :

$$\begin{aligned}
\text{Cov}(X,Y) &= \frac{1}{9}\sum_{k=1}^{9}(k-5)(2k+1-11) \\
&= \frac{1}{9}\sum_{k=1}^{9}\left(2k^2 - 10k - 10k + 50\right) \\
&= \frac{1}{9}\left(2 \times \frac{9 \times 10 \times 19}{6} - 20\frac{9 \times 10}{2} + 50 \times 9\right) \\
&= \frac{190}{3} - 100 + 50 = \frac{40}{3}
\end{aligned}$$

Pour savoir si ces variables sont corrélées linéairement on calcule le coefficient de corrélation :

$$\begin{aligned}\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(y)}} &= \frac{40/3}{\sqrt{\frac{20}{3}\times\frac{80}{3}}}\\ &= \frac{40/3}{\sqrt{\frac{1600}{9}}} = 1\end{aligned}$$

Les deux variables sont donc très fortement corrélées positivement. C'était prévisible, car on a $Y = 2X + 1$, qui est bien une corrélation linéaire.

Correction de l'exercice 2

1. Si $k \geq n$, on a $(n-k)!(n+k)! = 1$ car la factorielle d'un nombre négatif est égale à 1 (car c'est un produit vide). Si $k < n$, on regarde le rapport entre deux termes consécutifs :

$$\begin{aligned}\frac{(n-k-1)!(n+k+1)!}{(n-k)!(n+k)!} &= \frac{(n-k-1)!}{(n-k)!}\times\frac{(n+k+1)!}{(n+k)!}\\ &= \frac{1}{n-k}\frac{n+k+1}{1}\\ &= 1 + \frac{2k+1}{n-k} > 0\end{aligned}$$

car $k < n$. Ainsi la suite est croissante.

2. On a

$$\begin{aligned}\binom{2n}{k} \leq \binom{2n}{n} &\Leftrightarrow \frac{(2n)!}{k!(2n-k)!} \leq \frac{(2n)!}{n!(2n-n)!}\\ &\Leftrightarrow \frac{1}{k!(2n-k)!} \leq \frac{1}{n!n!}\\ &\Leftrightarrow k!(2n-k)! \geq (n-0)!(n+0)!\\ &\Leftrightarrow (n-(n-k))!(n+(n-k))! \geq (n+0)!(n-0)!\end{aligned}$$

Cette inégalité est vraie car la suite $((n+k)!(n-k)!)_{k\in\mathbb{N}}$ est croissante.

3. On écrit $1,001 = 1 + 10^{-3}$ et on applique le binôme de Newton :

$$\begin{aligned}1,001^{20} &= (1+10^{-3})^{20}\\ &= \sum_{k=0}^{20}\binom{20}{k}10^{-3k}1^{20-k}\\ &= \sum_{k=0}^{20}\binom{20}{k}10^{-3k}\\ &\leq \binom{20}{10}\sum_{k=0}^{20}10^{-3k}\end{aligned}$$

Ainsi $1,001^{20}$ est majoré par la somme des 21 premiers termes de la suite géométrique de premier terme $\binom{20}{10}$ et de raison 10^{-3}.

4. On écrit

$$\begin{aligned}\binom{20}{10} &= \frac{20\times 19\times 18\times 17\times 16\times 15\times 14\times 13\times 12\times 11}{1\times 2\times 3\times 4\times 5\times 6\times 7\times 8\times 9}\\ &= 19\times 17\times 2\times 5\times 13\times 4\times 11\\ &= 19\times 11\times 17\times 13\times 40\\ &\leq 20\times 20\times 20\times 20\times 40\\ &\leq 64\times 10^5\end{aligned}$$

En reprenant la majoration de la question précédente :

$$\begin{aligned}1,001^{20} &\leq 64\times 10^5 \sum_{k=0}^{20} \frac{1}{1000^k}\\ &\leq 64\times 10^5 \times \frac{1-\frac{1}{10^{63}}}{1-\frac{1}{10^3}}\\ &\leq 64\times 10^5 \frac{1}{1-\frac{1}{10^3}}\\ &\leq \frac{64}{999}\times 10^8\end{aligned}$$

5. En reprenant un raisonnement similaire on a

$$\begin{aligned}1,001^{20} - A &= \sum_{k=3}^{20}\binom{20}{k}10^{-3k}\\ &\leq 64\times 10^5 \times \sum_{k=3}^{20} 10^{-3k}\\ &\leq 64\times 10^5\times \sum_{k=0}^{17} 10^{-3k-9}\\ &\leq 64\times 10^{-4}\frac{1-10^{-54}}{1-10^{-3}}\\ &\leq \frac{64\times 10^{-1}}{999}\end{aligned}$$

Or $\frac{64}{999}\leq\frac{1}{10}$ car $640<999$, donc $1,001^{20}-A\leq 10^{-2}$.

6. Cette approximation était très grossière, puisque $1,001^{20}\simeq 1+0,020+0,00019\simeq 1,02019$, alors que l'approximation nous donnait 64×10^5.

Correction de l'exercice 3

1. a) Un codon est une liste de 3 lettres choisies parmi A, C, G, T ; l'ordre des lettres a une importance, et on a droit aux répétitions, c'est donc une 3-liste avec répétitions sur un ensemble à 4 éléments. Le cardinal est donc $4^3 = 64$.

   b) Si un codon n'était fait que de 2 lettres, on aurait une 2-liste avec répétition, et donc $4^2 = 16$ possibilités, ce qui n'est pas assez pour coder 20 acides aminés.

2. Un brin est un assemblage ordonné, donc une liste, de n codons ; on peut les répéter, c'est donc une n-liste sur un ensemble de 64 codons. Il y a donc 64^n brins différents.

3. a) Il est plus simple (puisqu'on nous dit « au moins un ») de passer par l'ensemble complémentaire, c'est-à-dire de déterminer le nombre de brins où il n'y a pas de codon d'initiation. Un tel brin est une liste de n codons, et chaque codon peut être un des 63 codons qui ne sont pas le codon ATG ; d'après le cours (nombre de p-listes avec répétition), on a 63^n brins sans codons d'initiation. Ainsi, le cardinal recherché est $64^n - 63^n$.

   b) On utilise le même raisonnement que précédemment ; il y a cette fois 3 codons interdits si le brin ne comporte pas de codon de terminaison, et on a donc une n-liste d'éléments choisis (avec répétition) dans un ensemble à 61 codons. Ainsi, on a $64^n - 61^n$ brins qui ont au moins un codon de terminaison.

   c) On cherche le nombre de brins qui ont au moins un codon d'initiation et au moins un codon de terminaison. Passons par l'ensemble complémentaire, et déterminons le nombre de brins qui n'ont pas de codon d'initation ou pas de codon de terminaison. Puisque c'est un « ou », il s'agit de déterminer le cardinal d'une union ; on utilise donc la formule

$$\operatorname{card}(E \cup F) = \operatorname{card} E + \operatorname{card} F - \operatorname{card} E \cap F$$

   Il y a 63^n brins sans codon d'initiation, et 61^n sans codon de terminaison ; quant au nombre de brins qui n'ont ni codon d'initiation, ni codon de terminaison, il est de 60^n puisqu'un tel brin est une n-liste (avec répétition) de codons choisis parmi 60 codons (on ne peut choisir ni ATG, ni TAG, ni TGA, ni TAA). Ainsi, le nombre de brins qui n'ont pas de codon d'initiation ou pas de codon de terminaison est $63^n + 61^n - 60^n$; le cardinal que l'on cherche est donc $64^n - 63^n - 61^n + 60^n$.

4. Il y a 3 types de codons dans ce brin : un codon d'initiation, un codon de terminaison, et les $n-2$ autres codons, qui ne sont ni d'initiation ni de terminaison. Construire un tel brin revient tout d'abord à choisir deux places dans le brin (la première étant pour le codon d'initiation, la deuxième pour le codon de terminaison) ; une fois ces deux emplacements choisis, on a 1 possibilité pour le codon d'initiation et 3 possibilités pour le codon de terminaison, et 60 possibilités pour chacun des $n - 2$ codons restants. Ainsi on a $\binom{n}{2} \times 3 \times 60^{n-2}$.

5. a) Si le codon d'initiation le plus à gauche est en position k, ceci veut dire que les codons à sa gauche ne sont pas des codons d'initiation. Ainsi, un tel brin est constitué d'une liste de $k-1$ codons qui ne sont pas des codons d'initiation (ce qui donne 63^{k-1} possibilités), d'un codon d'initiation (1 possibilité), et du reste de la séquence, c'est-à-dire une liste $n-k$ codons sans aucune contrainte (64^{n-k}). Le cardinal recherché est donc $63^{k-1}64^{n-k}$.

   b) De même que précédemment, un tel brin est constitué de $k - 1$ codons qui ne sont pas des codons d'initiation, un k-ème codon ATG, et $n - k$ codons qui ne peuvent pas être des codons de terminaison (sinon, on aurait une séquence codante). Ainsi le cardinal recherché est $63^{k-1} \times 1 \times 61^{n-k}$.

   c) Un tel brin est constitué d'une liste de $k - 1$ codons qui ne sont pas ATG, un codon ATG, et une liste de $n-k$ codons contenant au moins un codon de terminaison ; d'après ce qui précède, le nombre de listes de $n - k$ codons qui ne contiennent pas de codon de terminaison est 61^{n-k}, donc le nombre de listes de $n - k$ codons en contenant au moins un est $64^{n-k} - 61^{n-k}$. Au final on a $63^{k-1}(64^{n-k} - 61^{n-k})$ brins avec un codon d'initialisation à la position k et au moins une séquence codante.

   d) Les résultats précédents permettent de faire une disjonction de cas sur D : on a $D = 1$ ou $D = 2$ ou … ou $D = n-1$ (D ne peut pas être égal à n car il y a forcément au moins un codon de terminaison après la position D). Dans tous les cas le cardinal correspondant est celui de la question précédente ; et comme c'est une union disjointe de cas, on ajoute les cardinaux. Ainsi le cardinal $\mathcal{C}$ recherché est

$$\begin{aligned}
C &= \sum_{k=1}^{n-1} 63^{k-1}(64^{n-k} - 61^{n-k}) \\
&= \sum_{k=1}^{n-1} 63^{k-1}64^{n-k} - \sum_{k=1}^{n-1} 63^{k-1}61^{n-k} \\
&= \sum_{k=0}^{n-2} 63^{k}64^{n-k-1} - \sum_{k=0}^{n-2} 63^{k}61^{n-k-1} \\
&\qquad \text{par changement d'indice } k \leftarrow k-1 \\
&= 64^{n-1}\sum_{k=0}^{n-2}\left(\frac{63}{64}\right)^k - 61^{n-1}\sum_{k=0}^{n-2}\left(\frac{63}{61}\right) \\
&= 64^{n-1}\frac{\frac{63^{n-1}}{64^{n-1}} - 1}{\frac{63}{64} - 1} - 61^{n-1}\frac{\frac{63^{n-1}}{61^{n-1}} - 1}{\frac{63}{61} - 1} \\
&= 64^{n-1}\frac{64 \times (\frac{63^{n-1}}{64^{n-1}} - 1)}{63 - 64} - 61^{n-1}\frac{61 \times (\frac{63^{n-1}}{61^{n-1}} - 1)}{63 - 61} \\
&= 64^{n}\frac{\frac{63^{n-1}}{64^{n-1}} - 1}{-1} - 61^{n}\frac{\frac{63^{n-1}}{61^{n-1}} - 1}{2} \\
&= \frac{\frac{63^{n-1}64^{n}}{64^{n-1}} - 64^{n}}{-1} - \frac{\frac{63^{n-1}61^{n}}{61^{n-1}} - 61^{n}}{2} \\
&= 64^n - 64 \times 63^{n-1} - 63^{n-1} \times \frac{61}{\times 2} + \frac{1}{2}61^n \\
&= 64^n + \frac{1}{2}61^n - 63^{n-1}(64 + \frac{61}{2}) \\
&= 64^n + \frac{1}{2}61^n - 63^{n-1}\frac{128+61}{2} \\
&= 64^n + \frac{1}{2}61^n - 63^{n-1}\frac{189}{2}
\end{aligned}$$

Correction de l'exercice 4

1. On a
   - Pour $n = 1$: juste "A" ;
   - Pour $n = 2$: il n'y en n'a pas ("AA" a deux lettres adjacentes) ;
   - Pour $n = 3$: "ABA" et "ACA" : il y en a 2 ;
   - Pour $n = 4$: "ABCA" et "ACBA" : il y en a 2 ;
   - Pour $n = 5$: il y a les mots dont la deuxième lettre est B ("ABCBA", "ABABA", "ABACA") et ceux dont la troisième lettre est C ("ACBCA", "ACACA", "ACABA") : il y en a 6 ;
   - Pour $n = 6$: il y a les mots dont la deuxième lettre est B ("ABABCA", "ABACBA", "ABCACA", "ABCABA", "ABCBCA") et ceux dont la deuxième lettre est C ("ACACBA", "ACABCA", "ACBACA", "ACBCBA", "ACBABA") : il y en a 10.
2. Soit un mot terne de longueur n. On peut distinguer deux cas selon la lettre numéro $n - 2$:

— Si c'est un A : la lettre suivante est un B ou un C, ce qui est cohérent avec le fait que la lettre d'après (la n-ème) est un A ; le mot est donc composé d'un mot terne de longueur $n-2$ (puisque la $(n-2)-$ème lettre est un A), puis on a 2 choix pour le reste du mot (BA ou CA). Ainsi il y a $2t_{n-2}$ possibilités.

— Si c'est un B ou un C : la lettre suivante ne peut pas être un A (sinon on aurait 2 A), et ne peut pas être la même que la $(n-2)-$ème, donc il n'y a qu'une seule possibilité. Vu qu'il n'y a qu'une seule possibilité, cette lettre est redondante (et peut se déduire du reste) ; ainsi, choisir un mot terne à n lettres avec un B ou un C en $(n-2)-$ème position revient à choisir un mot terne de longueur $n-1$ (quitte à rajouter une lettre en plus entre les deux dernières, pour laquelle on n'a pas de choix). On obtient donc t_{n-1} possibilités.

On a opéré par disjonction de cas, donc il faut additionner les cardinaux trouvés (car on calcule ainsi le cardinal d'une union disjointe) ; ainsi

$$t_n = t_{n-1} + 2t_{n-2}$$

3. On considère le polynôme caractéristique associé à la relation de récurrence, qui est X^2-X-2 ; on calcule ses racines :

$$\Delta = (-1)^2 - 4 \times (-2) = 9$$
$$x_1 = \frac{1+3}{2} = 2, \qquad x_2 = \frac{1-3}{2} = -1$$

Ainsi $X^2 - X - 2 = (X-2)(X+1)$; on a ainsi

$$t_n = \lambda 2^n + \mu(-1)^n$$

Pour déterminer les coefficients, on utilise la question 1 : $t_1 = 2\lambda - \mu = 1$ et $t_2 = 4\lambda + \mu = 0$, ainsi

$$\begin{cases} 2\lambda - \mu = 1 \\ 4\lambda + \mu = 0 \end{cases} \Leftrightarrow \begin{cases} 2\lambda - \mu = 1 \\ 6\lambda \quad = 1 \end{cases} \Leftrightarrow \begin{cases} \mu = -\frac{4}{6} \\ \lambda = \frac{1}{6} \end{cases}$$

Ainsi $t_n = \frac{1}{6}2^n - \frac{2}{3}(-1)^n$. On peut vérifier que la formule donne bien les résultats de la première question :

$$t_3 = \frac{4}{3} + \frac{2}{3} = 2$$
$$t_4 = \frac{8}{3} - \frac{2}{3} = 2$$
$$t_5 = \frac{16}{3} + \frac{2}{3} = 6$$
$$t_6 = \frac{32}{2} - \frac{2}{3} = 10$$

(Si vous ne trouvez pas les mêmes cardinaux, c'est peut-être que vous avez oublié des mots !)