

Contrôle continu de MB61 n° 1

(Documents et calculatrices interdits)

- Le barème est donné à titre indicatif. L'enseignant se réserve le droit de l'adapter.
- Vous aurez au moins 0 et au plus 20.
- Une bonne rédaction est un élément d'appréciation important pour les copies.
- Il est permis d'admettre un résultat et de répondre aux questions suivantes en l'utilisant.
- Les trois parties sont indépendantes.

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires définies sur le même univers Ω à valeurs dans l'alphabet des nucléotides $\mathcal{A} = \{a, c, g, t\}$ indépendantes et équidistribuées suivant la loi $(p(x))_{x \in \mathcal{A}}$ donnée par

$$p(a) = p(g) = \frac{1}{5}, \quad p(c) = p(t) = \frac{3}{10}.$$

On définit l'ensemble R des purines par $R = \{a, g\}$ et l'ensemble des pyrimidines Y par $Y = \{c, t\}$.

Partie 1

1. (0.5pt) Montrer que $\mathbb{P}(X_1 \in Y) = 3/5$.

On définit la suite de variables aléatoires $(W_n(Y))_{n \geq 1}$ par

$$W_n(Y) = \mathbf{1}\{X_n \in Y\}.$$

2. (1pt) Quelle est la loi de $W_n(Y)$? Préciser son paramètre.
3. (1.5pt) Montrer que les variables aléatoires $W_1(Y)$ et $W_2(Y)$ sont indépendantes.

On admet dans la suite que les variables aléatoires $(W_n(Y))_{n \geq 1}$ sont indépendantes.
On définit $N_n(Y)$ et $R_n(Y)$ pour tout $n \geq 1$ par

$$N_n(Y) = \sum_{k=1}^n W_k(Y) \quad \text{et} \quad R_n(Y) = \frac{N_n(Y)}{n}.$$

4. (1pt) Que représentent les variables aléatoires $N_n(Y)$ et $R_n(Y)$?
5. (2pt) Calculer $\mathbb{E}(R_n(Y))$ et $\text{Var}(R_n(Y))$.

6. (1.5pt) En déduire que $R_n(Y)$ converge en probabilité vers $3/5$.

Partie 2

On définit la variable aléatoire T_Y par

$$T_Y = \inf\{n \geq 1 : X_n \in Y\},$$

qui représente la position aléatoire de la première pyrimidine dans la séquence. Par exemple, l'évènement $[T_Y = 1]$ s'écrit $[X_1 \in Y]$ et l'évènement $[T_Y = 2]$ s'écrit $[X_1 \notin Y, X_2 \in Y]$.

7. (2pt) Montrer que pour tout $n \geq 1$

$$\mathbb{P}(T_Y = n) = \frac{3}{5} \left(\frac{2}{5}\right)^{n-1}.$$

8. (1pt) En remarquant que $[T_Y \geq n+1] = [X_1 \notin Y, \dots, X_n \notin Y]$ pour tout $n \geq 1$, calculer $\mathbb{P}(T_Y \geq n+1)$.
9. (2pt) Calculer $\mathbb{P}(T_Y \geq m+n+1 | T_Y \geq n+1)$ pour tous $m, n \geq 1$.

Partie 3

On définit la suite de variables aléatoires $(W_n(\mathbf{aa}))_{n \geq 1}$ par

$$W_n(\mathbf{aa}) = \mathbf{1}\{X_n = \mathbf{a}, X_{n+1} = \mathbf{a}\}.$$

10. (1pt) Quelle est la loi de $W_n(\mathbf{aa})$? Préciser son paramètre.
11. (1.5pt) Les variables aléatoires $W_1(\mathbf{aa})$ et $W_2(\mathbf{aa})$ sont-elles indépendantes ? Justifier.

On admet dans la suite que si $|i - j| \geq 2$ alors les variables aléatoires $W_i(\mathbf{aa})$ et $W_j(\mathbf{aa})$ sont indépendantes. On définit $N_n(\mathbf{aa})$ et $R_n(\mathbf{aa})$ pour tout $n \geq 2$ par

$$N_n(\mathbf{aa}) = \sum_{k=1}^{n-1} W_k(\mathbf{aa}) \quad \text{et} \quad R_n(\mathbf{aa}) = \frac{N_n(\mathbf{aa})}{n-1}.$$

12. (1pt) Calculer $\mathbb{E}(R_n(\mathbf{aa}))$.
13. (2pt) Calculer $\mathbb{E}(W_1^2(\mathbf{aa}))$, $\mathbb{E}(W_1(\mathbf{aa})W_2(\mathbf{aa}))$ et $\mathbb{E}(W_1(\mathbf{aa})W_3(\mathbf{aa}))$.
14. (2.5pt) En déduire $\text{Var}(R_n(\mathbf{aa}))$.
15. (1.5pt) Déduire de ce qui précède que $R_n(\mathbf{aa})$ converge en probabilité vers $1/25$.

Fin