

# Exercices : concepts de probabilités

Probabilités et statistique pour la biologie : éléments (STAT1)

*Jacques van Helden*

*2018-11-19*

## Contents

Exercice 03.1: tirage de tétranucléotides dans un génome . . . . .	1
Exercice 03.2 : absence d'un résidu . . . . .	1
Exercice 03.3 : probabilité d'une purine . . . . .	1
Exercice 03.4 : motif dégénéré, nucléotides équiprobables . . . . .	2
Exercice 03.5 : motif dégénéré, nucléotides non-équiprobables . . . . .	2

### Exercice 03.1: tirage de tétranucléotides dans un génome

On tire aléatoirement une position génomique et on analyse le tétranucléotide qui y démarre. Quelles sont les probabilités respectives d'observer chacun aux critères suivants ?

- Etre composé uniquement de  $A$ .
- Ne comporter aucun  $A$ .
- Etre composé de 4 résidus distincts.

Formulez explicitement le raisonnement qui vous amène à la formule de calcul. Indiquez ensuite la formule générale (avec des symboles), puis la formule particulière avec les valeurs numériques. Il n'est pas nécessaire de calculer le résultat final.

### Exercice 03.2 : absence d'un résidu

Quelle est la probabilités de n'observer aucun  $A$  dans un oligonucléotide de taille 12 ?

- En supposant des nucléotides équiprobables.
- Selon un modèle de Bernoulli généralisé dont les paramètres sont estimés à partir des fréquences génomiques suivantes.

Résidu	Occurrences génomiques	Fréquence génomique
A	3766191	0.3098064564636
C	2320522	0.1908858838986
G	2316991	0.1905954242278
T	3752889	0.3087122354100

### Exercice 03.3 : probabilité d'une purine

On tire un nucléotide au hasard dans le génome de la levure. En se basant sur base des fréquences génomiques ci-dessous, quelle est la probabilité de tirer une purine ( $A$  ou  $G$ ) ? Justifiez vos choix.

Résidu	Occurrences génomiques	Fréquence génomique
A	3766191	0.3098064564636
C	2320522	0.1908858838986
G	2316991	0.1905954242278
T	3752889	0.3087122354100

**Exercice 03.4 : motif dégénéré, nucléotides équiprobables**

En supposant qu'une séquence est composée de résidus équiprobables, quelle est la probabilité du motif *GATWNA* (*W* signifie "*T* ou *A*", et *N* correspond à n'importe quel nucléotide) ?

**Exercice 03.5 : motif dégénéré, nucléotides non-équiprobables**

Même question en supposant les probabilités distinctes pour les nucléotides:  $P(A) = P(T) = 0.3$ ,  $P(C) = P(G)$  (à vous de calculer ces dernières).