Eléments d'analyse combinatoire - solutions des exercices

Probabilités et statistique pour la biologie (STAT1)

Jacques van Helden

2017-12-05

Contents

Résumé des concepts et formules

Tirages avec / sans remise

Il existe deux types classiques de tirage d'éléments au sein d'un ensemble: avec ou sans remise.

1. **Tirage sans remise**: chaque élément peut être tiré au plus une fois. Exemples:
   - Jeu de loto (ou lotto).
   - Sélection aléatoire d'un ensemble de gènes dans un génome.
2. Tirage **avec remise**: chaque élément peut être tiré zéro, une ou plusieurs fois. Exemples:
   - Jeu de dés. A chaque lancer on dispose des mêmes possibilités (6 faces).
   - Génération d'une séquence aléatoire, par sélection itérative d'un élément dans l'ensemble des résidus (4 nucléotides pour l'ADN, 20 acides aminés pour les protéines).

Formules

Remise	Ordre	Formule	Description
Oui	Oui	n^x	**Exponentielle**: séquences de x éléments tirés dans un ensemble de taille n, avec remise.
Non	Oui	$n!$	**Factorielle**: toutes les permutations d'un ensemble de taille n
Non	Oui	$A_n^x = \frac{n!}{x!}$	**Arrangements** : listes (ordonnée) de x éléments tirés dans un ensemble de taille n
Non	Non	$C_n^x = \binom{n}{x} = \frac{n!}{x!(n-x)!}$	**Combinaisons** : ensembles (non ordonnés) de x éléments tirés dans un ensemble de taille n

Solutions des exercices

Solution exercice 1: mutagenèse

On soumet un fragment d'ADN de 1 kilobase à un traitement mutagène qui provoque des mutations ponctuelles (substitutions) à 5 positions distinctes indépendantes. Combien de séquences possibles existe-t-il pour le fragment muté ?

Il s'agit de choisir au hasard 5 positions mutantes parmi les 1000 nucléotides du fragment d'ADN. Il s'agit d'un choix sans remise (chaque position ne peut être tirée qu'une fois), on choisit donc le coefficient binomial.

$$\binom{n}{x} = \binom{1000}{5} = C^5_{1000} = \frac{1000!}{5!995!} = 8.2502913 \times 10^{12}$$

Solution de l'exercice 2 : oligopeptides 3×20

Combien d'oligopeptides de taille 60 peut-on former en utilisant exactement 3 fois chaque acide aminé ?

Commençons par générer une séquence particulière qui remplit ces conditions, en concaténant 3 copies de chaque acide aminé, dans l'ordre alphabétique.

Table 2: Symboles des acides aminés et codons

Aminoacid	Symbol3	Symbol	Codons
Alanine	Ala	A	GCA, GCC, GCG, GCT
Arginine	Arg	R	CGA, CGC, CGG, CGT, AGA, AGG
Aspartic acid	Asp	D	GAC, GAT
Asparagine	Asn	N	AAC, AAT
Cysteine	Cys	C	TGC, TGT
Glutamic acid	Glu	E	GAA, GAG
Glutamine	Gln	Q	CAA, CAG
Glycine	Gly	G	GGA, GGC, GGG, GGT
Histidine	His	H	CAC, CAT
Isoleucine	Ile	I	ATA, ATC, ATT
Leucine	Leu	L	CTA, CTC, CTG, CTT, TTA, TTG
Lysine	Lys	K	AAA, AAG
Methionine	Met	M	ATG
Phenylalanine	Phe	F	TTC, TTT
Proline	Pro	P	CCA, CCC, CCG, CCT
Serine	Ser	S	TCA, TCC, TCG, TCT, AGC, AGT
Threonine	Thr	T	ACT, ACC, ACG, ACT
Tryptophan	Trp	W	TGG
Tyrosine	Tyr	Y	TAC, TAT
Valine	Val	V	GTA, GTC, GTG, GTT
STOP	-	-	TAG, TAA, TGA

`AAACCCDDDEEEFFFGGGHHHIIIKKKLLLMMMNNNPPPQQQRRRSSSTTTVVVWWWYYY`

Les permutations de ces 60 lettres sont des solutions valudes. En voici trois exemples.

`VFHNLPGVSFQSEATKVDYHEHGYWAKCPCPIQDEQMATNFNYKRSCTDMIRGLIWWMRL`

`IGGPRMANVAPSLIWYCTAVQNSQFFHNTWVPYIHDQTRRFECDLCEHYLEWGMDKKMKS`

`RYVCVMFWKTVRWDCEIIMCRHGLKQESAIFYQENPQDWPFPKMSDTGNLHANSALGTYH`

...

`VQPPMGCTQWYMAVRKFDHPSQTINESCYGESDHDRGAFEKNAILMICNVLYWWHLTFKR`

Le nombre total de permutations possibles parmi 60 éléments est $60! = 8.3209871 \times 10^{81}$. Cependant, cenombre dépasse de loin le nombre de séquences distinctes. En effet, dans chacune des séquences ci-dessus, chaque lettre apparaît 3 fois. Or, des permutations entre les trois positions occupées par des A ne changeront pas la séquence. Il en va de même pour les permutations entre les positions occupées par chacun des acides aminées : A, C, D, ...

Il faut donc diviser le nombre total de permutations (60!) par le nombre de permutations qui ne modifient pas la séquence: 3! pour A, 3! pour D, et ainsi de suite pour chacun des 20 acides aminés.

La formule finale est donc:

$$N = \frac{\overbrace{60!}^{\text{60 lettres}}}{\underbrace{3! \cdot 3! \cdot \ldots 3!}_{\text{20 acides aminés présents 3 fois}}} = \frac{60!}{(3!)^{20}} = 2.2758825 \times 10^{66}$$