

Éléments d'analyse combinatoire

Probabilités et statistique pour la biologie (STAT1)

Jacques van Helden

2017-09-05

Problème: dénombrement d'oligomères

L'ADN est composée de 4 nucléotides distincts dénotés par les lettres A, C, G, T, et les protéines de 20 acides aminés.

Pour chacun de ces deux types de polymères, combien d'oligomères distincts peut-on former en polymérisant 20 résidus ("20-mères") ?

Approche suggérée: simplifiez le problème au maximum, en commençant par des polymères beaucoup plus courts (1 résidu, 2 résidus).

Questions subsidiaires:

- Généralisez la formule pour les oligomères d'une longueur arbitraire k ("k-mères").
- Quel est le nom de la fonction donnant le résultat ?
- Dans ce processus, quel est le mode de sélection des résidus: **avec ou sans remise** ?

Problème: oligomères sans résidus répétés

Combien d'oligomères peut-on former (ADN ou peptides) en utilisant chaque résidu une et une seule fois ?

Approche suggérée: agrégez progressivement les résidus, en vous demandant à chaque étape combien d'entre eux n'ont pas encore été incorporés.

Questions subsidiaires:

- Généralisez la formule pour des séquences d'objets tirés dans un ensemble de taille arbitraire (n).
- Quel est le nom de la fonction donnant le résultat ?
- Dans ce processus, quel est le mode de sélection des résidus: **avec ou sans remise** ?

Problème: sélection d'une liste (ordonnée) de gènes

Lors d'une expérience de transcriptome indiquant le niveau d'expression de tous les gènes de la levure. Sachant que le génome comporte 6000 gènes, combien de possibilités existe-t-il pour sélectionner les 15 gènes les plus fortement exprimés (**en tenant compte** de l'ordre relatif de ces 15 gènes) ?

Approche suggérée: comme précédemment, simplifiez le problème en partant de la sélection minimale, et augmentez progressivement le nombre de gènes (1 gène, 2 gènes, ...).

Questions subsidiaires:

- Trouvez un exemple familier de jeu de pari apparenté à ce problème.
- Généralisez la formule pour la sélection d'une liste de x gènes dans un génome qui en comporte n .

Problème: sélection d'un ensemble (non-ordonné) de gènes

Lors d'une expérience de transcriptome indiquant le niveau d'expression de tous les gènes de la levure. Sachant que le génome comporte 6000 gènes, combien de possibilité existe-t-il pour sélectionner les 15 gènes les plus fortement exprimés (**sans tenir compte** de l'ordre relatif de ces 15 gènes) ?

Approche suggérée: comme précédemment, simplifiez le problème en partant de sélections minimales (1 gène, 2 gènes, ...) et généralisez la formule.

Questions subsidiaires:

- Trouvez un exemple familier de jeu de pari apparenté à ce problème.
- Généralisez la formule pour la sélection d'un ensemble de x gènes dans un génome qui en comporte n .
- Connaissez-vous le nom de la formule ainsi trouvée ?

Exercice : mutagénèse

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

Exercice : oligopeptides d'une composition donnée

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