

M2 Mycologie

Notes et compléments

Inférence et test d'hypothèse

Comme dans un essai clinique, on cherche donc à comparer l'effet d'un traitement A à celui d'un traitement B de référence (ou placebo), qui sert de comparateur. La différence observée entre les effets de A et B peut faire l'objet d'un test statistique. Celui-ci permet de confronter la valeur observée à celles pouvant résulter de simples fluctuations d'échantillonnage (c.a.d. des différences dues au "hasard", signifiant ici l'absence d'une réelle différence).

D'un autre côté, il existe un risque β de ne pas être en mesure de rejeter l'hypothèse nulle lorsqu'une réelle différence existe. Le complément $1 - \beta$, appelée puissance du test, représente donc la probabilité de rejeter correctement l'hypothèse nulle en faveur de l'hypothèse alternative.

Recherche de motifs

Consider a sequence of size k and a window of size n (n much smaller than the genome length). If all nucleotides are assumed independent, then the probability of not finding this particular sequence is $(1 - 1/4^k)$ (using the complementary is useful to avoid dealing with the number of times this sequence can appear) times the total number of positions available in the window (n or $n - k + 1$, but if $k \ll n$ this correction factor can be omitted). Hence the answer : $1 - (1 - 1/4^k)^n$.

Seconde loi de Mendel

Le terme $1/4$ vient du fait que la probabilité d'observer un descendant de ce sous-type est uniforme quelque soit le croisement considéré.

On calcule dans un premier temps la probabilité d'observer n Aa Bb descendants après k générations. Il s'agit d'une succession d'événements de Bernoulli jusqu'à la génération k , d'où $P(b, k) = \binom{2^k}{n} (1/4)^n (1 - 1/4)^{2^k - n}$, qui n'est autre que la fonction de densité d'une variable aléatoire suivant une loi Binomiale de paramètre $n = 2k$ (il y a deux enfants à chaque étape) et $p = 1/4$. Ensuite il s'agit de trouver la probabilité qu'au moins N organismes se trouvent à la génération k , qui vaut $1 - \sum_{x=0}^{N-1} P(X = x)$, où X est une variable aléatoire représentant le nombre de descendants, c'est-à-dire 1 moins tous les cas $P(X \neq N)$.

Prolifération bactérienne

Le nombre moyen de bactéries par boîte est 5. On suppose que le nombre de colonies par boîte est le même que le nombre moyen de bactéries par boîte. On a alors $P(X = 0) = \frac{5^0 e^{-5}}{0!} = 0.0067$, soit approximativement 0,67 % de chance. On remarquera que la probabilité qu'il y ait au moins une colonie sur la boîte de Pétri vaut $P(X > 0) = 1 - P(X = 0) = 1 - 0.0067 = 0.9933$ (99,3 % de chance)!

Le modèle de Jukes-Cantor en phylogénie

En somme, si l'on part de A, on attendra $\exp(0.886) = 2.425$ unités de temps, et la probabilité de transitionner vers un autre état sera de 0.214 (pour C), 0.714 (G) et 0.071 (T). La transition la plus probable A→G implique d'attendre $\exp(-0.633) = 0.531$ unité de temps.

PCR et processus de branchement

From a practical point of view, it is often desired to know N_0 in advance, although N_0 cannot be identified uniquely. Note, however, that it is possible to consistently estimate the mathematical expectation of the reproduction m with the estimator $\hat{m}_n = \sum_{k=0}^{\infty} k \hat{p}_{k_n} = \frac{Y_{(n+1)} - 1}{Y_n}$, that is m can be estimated from the observed size of each generation. Hence, the following estimator has been proposed : $\hat{N}_{0n} = \frac{Z_n}{\hat{m}_n}$. As an illustration, if the success rate of the PCR is 80% (i.e., $p = 0.8$), its variance equals $\frac{1-p}{1+p} = 0.11$.