Tests de sélection pour des populations surcritiques

Benoît Henry

École de printemps de la Chaire MMB, 31 Mai 2017

Idées et but

Construire un test statistique pour comparer les taux de croissances de différentes populations dans un modèle de branchement.

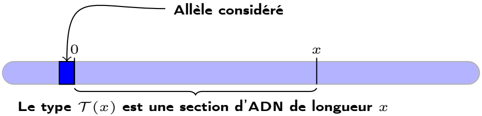
Exemple:

- Une population résidente et une population mutante, tester si la population mutante est en cours de sélection

Problème:

- On ignore l'âge des deux populations

Idées et but

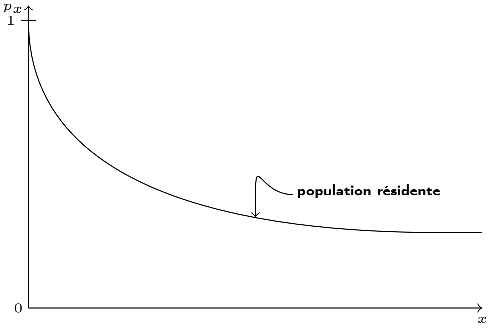


- x joue le rôle d'un taux de mutation
- p_x **probabilité** de tirer **uniformément** au hasard **deux individus de même type** (Homozygotie, e.g. **Sabetti et al.**, 2002)

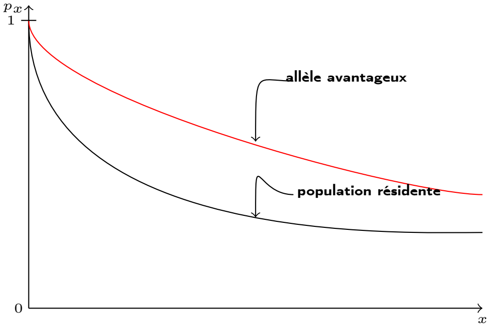
Hypothèse:

- Mutations neutres sur la section d'ADN considérée

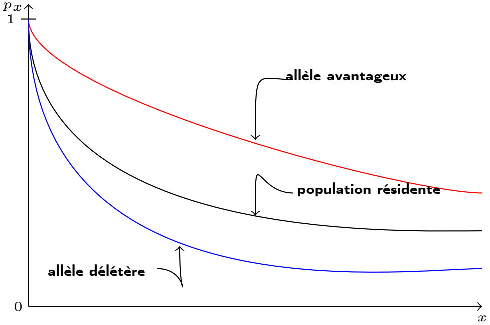
Idées et but



Idées et but



Idées et but

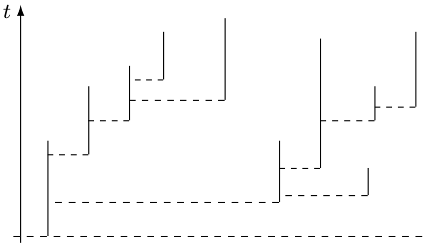


p_x
1
allèle avantageux
population résidente
allèle délétère
0
x

Modèle

On considère un arbre aléatoire $\mathbb{T}$ (splitting trees) (Geiger-Kersting, IMA Math. Appl. 1997) dont la loi peut être décrite informellement sous la forme d'un modèle de dynamique des populations.

- les durées de vies de individus sont **i.i.d.** de distribution $\mathbb{P}_V$.
- conditionnellement aux durées de vies, **nouvelles naissances à taux** $b \in \mathbb{R}_+^*$ (indépendamment les uns des autres).
- la population démarre d'un unique individu (la *racine*).

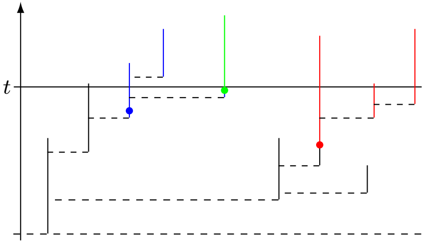


$(N_t,\ t \in \mathbb{R}_+)$ est un processus de branchement (généralement non-Markovien).

Mutations

On considère un arbre aléatoire $\mathbb{T}$ (splitting trees) (Geiger-Kersting, IMA Math. Appl. 1997) dont la loi peut être décrite informellement sous la forme d'un modèle de dynamique des populations.

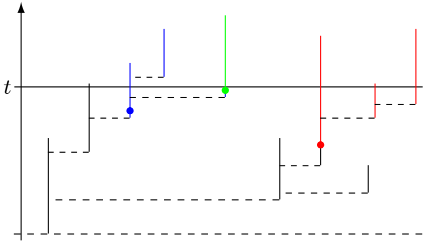
- $\mathbb{T}$ peut être vu comme un espace métrique muni d'une mesure $\mathcal{M}$.
- mutations selon une mesure de Poisson $\mathcal{N}$ sur $\mathbb{T}$ d'intensité $\mathcal{M}$.



Mutations

On considère un arbre aléatoire $\mathbb{T}$ (splitting trees) (Geiger-Kersting, IMA Math. Appl. 1997) dont la loi peut être décrite informellement sous la forme d'un modèle de dynamique des populations.

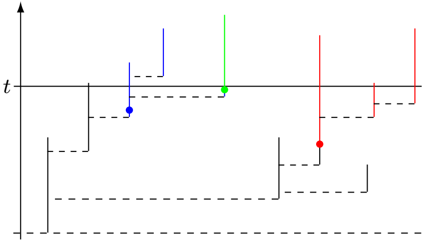
- $\mathbb{T}$ peut être vu comme un espace métrique muni d'une mesure $\mathcal{M}$.
- mutations selon une mesure de Poisson $\mathcal{N}$ sur $\mathbb{T} \times \mathbb{R}_+$ d'intensité $\mathcal{M} \otimes \lambda$.
- À taux θ, les mutations sont données par $\mathcal{N}^\theta := \mathcal{N}(\cdot \times [0, \theta])$.



Spectre de fréquence

On considère un arbre aléatoire $\mathbb{T}$ (splitting trees) (Geiger-Kersting, IMA Math. Appl. 1997) dont la loi peut être décrite informellement sous la forme d'un modèle de dynamique des populations.

- $\mathbb{T}$ peut être vu comme un espace métrique muni d'une mesure $\mathcal{M}$.
- mutations selon une mesure de Poisson $\mathcal{N}$ sur $\mathbb{T} \times \mathbb{R}_+$ d'intensité $\mathcal{M} \otimes \lambda$.
- À taux θ, les mutations sont données par $\mathcal{N}^\theta := \mathcal{N}(\cdot \times [0, \theta])$.



$$A(1,t) = 2,\ A(2,t) = 0,\ A(3,t) = 1, \dots$$

Le **spectre de fréquence** est la suite d'entier $A(k,t)$ comptant le **nombre de familles** *non clonales* **de taille** k **dans la population au temps** t.

Représentation du spectre

Formellement, $\mathbb{T}$ s'identifie à un sous-ensemble de $\cup_{n\geq 0}\mathbb{N}^n \times \mathbb{R}_+$.

- $(\sigma, t) \in \mathbb{T}$ *ssi* l'individu σ est vivant au temps t.

La descendance $C^\theta((\sigma,t))$ clonale d'un point (σ,t) de $\mathbb{T}$ au temps t est définie par

$$C^\theta((\sigma,s),t) = \text{Card}\left\{(\nu,v)\in\mathbb{T} \mid \sigma \preceq \nu,\ \ v=t,\ \ \mathcal{N}^\theta((\sigma,s),(\nu,v)] = 0)\right\}$$

On peut alors écrire

$$A^\theta(k,t) = \int_{\mathbb{T}} \mathbb{1}_{C^\theta(x,t)=k}\, \mathcal{N}^\theta(dx)$$

Le calcul de moments du type

$$\mathbb{E}\left[A^{\theta_1}(k,t)A^{\theta_2}(k,t) \mid N_t > 0\right]$$

repose essentiellement sur l'obtention de la loi du couple

$$\left(Z_0^{\theta_1}(t), Z_0^{\theta_2}(t)\right),$$

avec

$$Z_0^{\theta_1}(t) := C^{\theta_1}((\emptyset,0),t).$$

Famille ancestrale

La **famille ancestrale** $Z_0(t)$ au temps t est l'ensemble des individus vivant au temps t portant le type que portait la branche ancestrale à l'instant 0.

$Z_0(t)$ peut être étudié grâce à l'arbre de nouvelle loi de durée de vie donnée par:

$$V_\theta \overset{\mathcal{L}}{=} V \wedge \mathcal{E}(\theta)$$