Estimer l'effet des mutations sur la fitness d'une population de bactéries

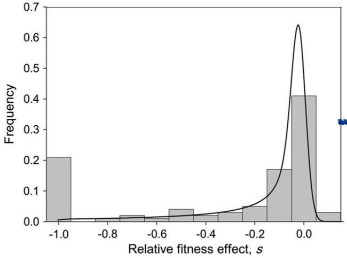
Jérémy Andréoletti et Nathanaël Boutillon

Encadrantes : Marie Doumic et Lydia Robert

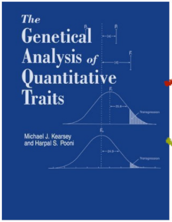
Objectifs

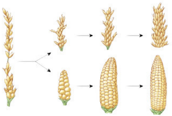
- Comprendre la dynamique d'apparition des mutations, chez *E. coli*
- Estimer la **DFE = Distribution des Effets des mutations sur la Fitness**

Distribution of fitness effects caused by single-nucleotide substitutions in bacteriophage f1



Peris et al., Genetics, 2010





Modèle

On pose

$$\begin{aligned} &W_t && \text{fitness (taux de croissance) au temps } t \\ s_i &= \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}} && \text{effet \textit{relatif} de la mutation } i \\ &N_t && \text{nombre de mutations avant le temps } t \end{aligned}$$

Énoncé du problème : estimer la loi des s_i (qui sont iid) sachant que l'on sait estimer la loi des W_t et N_t ($t \geqslant 0$), et sachant que

$$\frac{W_t}{W_0} = \prod_{i=1}^{N_t} (1 - s_i)$$

Estimation des moments

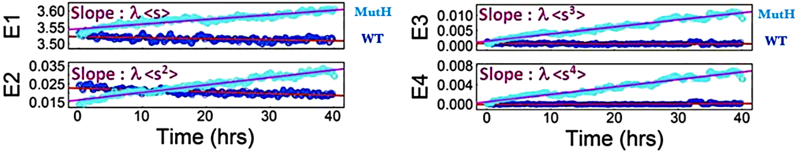
- estimation des **premiers moments** de la loi des effets des mutations ;

$$E_n(t) := \sum_{k=1}^{n} \binom{n}{k} (-1)^k \ln\left(\mathbb{E}\left[W_t^k\right]\right) = (\lambda \mathbb{E}[s^n])\, t$$

Estimation des moments

- estimation des **premiers moments** de la loi des effets des mutations ;

$$E_n(t) := \sum_{k=1}^{n} \binom{n}{k} (-1)^k \ln\left(\mathbb{E}\left[W_t^k\right]\right) = (\lambda \mathbb{E}[s^n])\, t$$



→ On peut facilement estimer les moments.

Méthode par la fonction caractéristique

À partir de tous les moments, on peut calculer la fonction caractéristique :

$$\varphi_X(\xi) := \mathbb{E}\left[e^{i\xi X}\right] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}\left[X^k\right]$$

Méthode par la fonction caractéristique

À partir de tous les moments, on peut calculer la fonction caractéristique :

$$\varphi_X(\xi) := \mathbb{E}\left[e^{i\xi X}\right] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!}\mathbb{E}\left[X^k\right]$$

On peut alors en déduire la distribution de X, f :

$$f(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \varphi_X(\xi)e^{-ix\xi}\,\mathrm{d}\xi$$

Méthode par la fonction caractéristique

À partir de tous les moments, on peut calculer la fonction caractéristique :

$$\varphi_X(\xi) := \mathbb{E}\left[e^{i\xi X}\right] = \sum_{k=0}^{+\infty} \frac{(i\xi)^k}{k!} \mathbb{E}\left[X^k\right]$$

On peut alors en déduire la distribution de X, f :

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \varphi_X(\xi) e^{-ix\xi} \,\mathrm{d}\xi$$

Or on a seulement les N premiers moments empiriques m_k :

$$\hat{\varphi}_X(\xi) = \sum_{k=0}^{N} \frac{(i\xi)^k}{k!} m_k$$
$$\hat{f}(x) = \frac{1}{2\pi} \int_{|\xi| \leqslant A} \hat{\varphi}_X(\xi) e^{-ix\xi} \,\mathrm{d}\xi$$

Espoir : $\varphi_X(\xi) \simeq \hat{\varphi}_X(\xi)$ donc $f(x) \simeq \hat{f}(x)$.

Calculs d'erreurs

Proposition

Pour tous paramètres $A > 0$, $k \geqslant 2$, $N \geqslant 1$ on a : $\forall x \in \mathbb{R}$

$$\left|\hat{f}(x) - f(x)\right| \leqslant \frac{\|f^{(k)}\|_1}{2\pi^2(k-1)A^{k-1}} + \frac{A^{N+1}}{\pi(N+1)!}\mathbb{E}\left[X^N(e^{AX-1})\right] + \frac{\|\varepsilon(N)\|_\infty(e^A-1)}{\pi}$$

Calculs d'erreurs

Proposition

Pour tous paramètres $A > 0$, $k \geqslant 2$, $N \geqslant 1$ on a : $\forall x \in \mathbb{R}$

$$\left|\hat{f}(x) - f(x)\right| \leqslant \frac{\|f^{(k)}\|_1}{2\pi^2(k-1)A^{k-1}} + \frac{A^{N+1}}{\pi(N+1)!}\mathbb{E}\left[X^N(e^{AX-1})\right] + \frac{\|\varepsilon(N)\|_\infty(e^A-1)}{\pi}$$

Proposition

Supposons que :

- l'on soit capable de calculer un nombre arbitrairement grand de moments de f avec une erreur bornée par ε ;
- $\|f^{(2)}\|_1 < +\infty$.

Alors on a :

$$\|f - \hat{f}\|_\infty = O\left(\left|\frac{1}{\pi \ln(\varepsilon)}\right|\right) \qquad \text{quand } \varepsilon \to 0$$

Un autre point de vue : une EDP

On a

$$\ln W_t = \sum_{i=1}^{N_t} \ln(1 - s_i)$$

Soient :

- $u(t, \cdot) \in C^\infty(\mathbb{R})$: « densité de $\ln W_t$ » *ie* :
  $u(t, x)\,\mathrm{d}x =$ proportion de cellules telles que $\ln W_t = x$
- $f \in C^\infty(\mathbb{R})$: densité de la loi de $\ln(1 - S)$;
- λ : taux de mutation ;
- μ : proportion de mutations létales.

Explication des termes

Donc $(*)$:

$$\partial_t u(t,x) = \lambda \left(\int_{\mathbb{R}} f(x-y)u(t,y)\,\mathrm{d}y - u(t,x) \int_{\mathbb{R}} f(y)\,\mathrm{d}y \right) - \lambda\mu u(t,x)$$

Traduction :

- $\partial_t u$: changement de densité de fitness entre t et $t + \mathrm{d}t$;
- $\int f(x-y)u(t,y)\,\mathrm{d}y$: arrivées sur la fitness x ;
- $-u(t,x)\int f(y)\,\mathrm{d}y$: départs de la fitness x ;
- $-\lambda\mu u(t,x)$: morts.

Conservation de la masse :

$$\int_{\mathbb{R}} u(t,x)\,\mathrm{d}x = N(t) = e^{-\lambda\mu t} N(0)$$

Solution explicite

On applique la transformée de Fourier sur :

$$\partial_t u(t,x) = \lambda \left(\int_{\mathbb{R}} f(x-y)u(t,y)\,\mathrm{d}y - u(t,x) \int_{\mathbb{R}} f(y)\,\mathrm{d}y \right) - \lambda\mu u(t,x)$$

donc

$$\partial_t \mathcal{F} u_t(\xi) = \lambda \mathcal{F} f(\xi) \mathcal{F} u_t(\xi) - \lambda \mathcal{F} u_t(\xi)$$

d'où :

Proposition

Pour tout $x \in \mathbb{R}$:

$$f(x) = \mathcal{F}^{-1} \left(\frac{\partial_t \left(\mathcal{F} u_t(\xi) \right)}{\lambda \mathcal{F} u_t(\xi)} + 1 \right)$$

Solution explicite

On applique la transformée de Fourier sur :

$$\partial_t u(t,x) = \lambda \left(\int_{\mathbb{R}} f(x-y)u(t,y)\,\mathrm{d}y - u(t,x) \int_{\mathbb{R}} f(y)\,\mathrm{d}y \right) - \lambda\mu u(t,x)$$

donc

$$\partial_t \mathcal{F}u_t(\xi) = \lambda \mathcal{F}f(\xi)\mathcal{F}u_t(\xi) - \lambda \mathcal{F}u_t(\xi)$$

d'où :

Proposition

Pour tout $x \in \mathbb{R}$:

$$f(x) = \mathcal{F}^{-1}\left(\frac{\partial_t \left(\mathcal{F}u_t(\xi)\right)}{\lambda \mathcal{F}u_t(\xi)} + 1 \right)$$

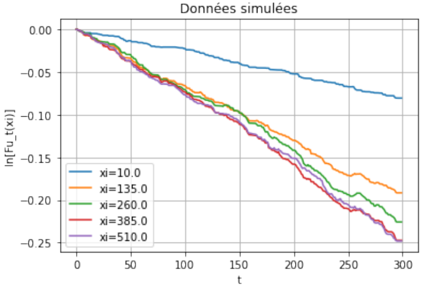
Le terme de droite est indépendant du temps : $\partial_t(\mathcal{F}u_t(\xi)) = c_\xi \mathcal{F}u_t(\xi)$.

Vérification

Transformation :

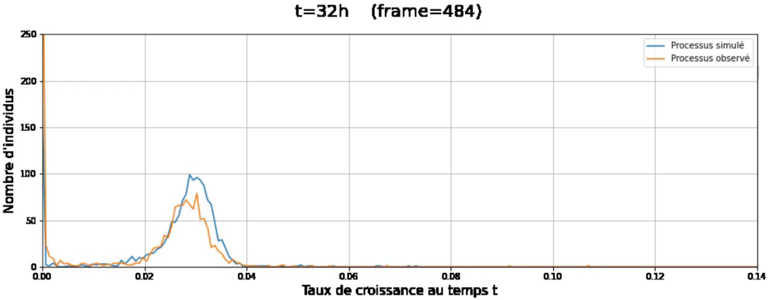
$$\tilde{u}(t,x) = \text{ distribution de } \ln(\mathrm{W}_t/W_0) \qquad \to \qquad u_{t=0} = \delta_0$$

$$\mathcal{F}\tilde{u}_t(\xi) = e^{a_\xi t} \qquad \to \qquad \ln|\mathcal{F}u_t(\xi)| = \Re(a_\xi)t$$



- fonctionne bien pour les données simulées ;
- fonctionne mal pour les données réelles : bruit de la première mesure ?

Simulations : image



Simulations

Idées

- Version multiplicative de l'EDP : si $v(t,\cdot)$ est la densité de W_t on a :

$$\partial_t v(t,x) = \int_{\mathbb{R}_+} \frac{1}{y} g(y) v\left(t, \frac{x}{y}\right) \mathrm{d}y$$

$\rightarrow$ Étudiée dans des problèmes de fragmentation par notre encadrante ;
$\rightarrow$ Comportement en temps long de la solution de l'EDP ;

- Résolution numérique de l'EDP pour comparer aux simulations stochastiques. « Jouer » avec la fonction f en entrée.

Références

- Robert et al., *Mutation dynamics and fitness effects followed in single cells*, Science 359, 1283–1286, 16 March 2018
- Doumic, Escobedo, *Time asymptotics for a critical case in fragmentation and growth-fragmentation equations*, submitted 2015
- Beal et al., *The Division of Amyloid Fibrils : Systematic Comparison of Fibril Fragmentation Stability by Linking Theory with Experiments*, iScience, 25 September 2020