

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

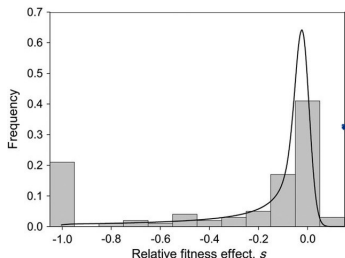
Jérémy Andréoletti et Nathanaël Boutillon

Encadrantes : Marie Doumic et Lydia Robert

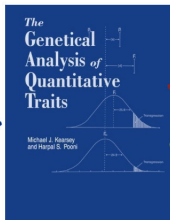
# Objectifs

- 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  $\phi$ 1



*Peris et al., Genetics, 2010*



On pose

$W_t$

fitness (taux de croissance) au temps  $t$

$$s_i = \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}}$$

effet *relatif* de la mutation  $i$

$N_t$

nombre de mutations avant le temps  $t$

On pose

$W_t$	fitness (taux de croissance) au temps $t$
$s_i = \frac{W_{t_{i-1}} - W_{t_i}}{W_{t_{i-1}}}$	effet <i>relatif</i> de la mutation $i$
$N_t$	nombre de mutations avant le temps $t$

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

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

## 1 Simulations

## 2 Problème des moments

- Fonction caractéristique
- Calculs d'erreurs

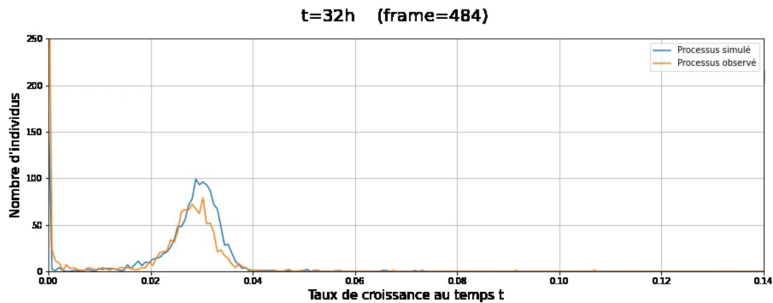
## 3 EDP

- Nouveau point de vue
- Applications

## Section 1

# Simulations

# Simulations : image



# Simulations



## Section 2

### Problème des moments

# Estimation des moments

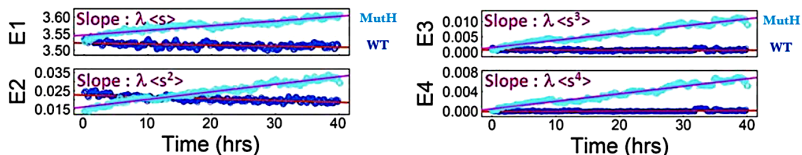
- estimation des **premiers moments** de la loi des effets des mutations ;

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

# 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  $X$ ,  $f$  :

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

# 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  $X$ ,  $f$  :

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

Or on a seulement les  $N$  premiers moments empiriques  $m_k$  :

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

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

# Calculs d'erreurs

## Proposition

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

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

# Calculs d'erreurs

## Proposition

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

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

## Proposition

Supposons que :

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

Alors on a :

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



## Section 3

EDP

# 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 :  
 $u(t, x) dx =$  proportion de cellules telles que  $\ln W_t = x$
- $f \in C^\infty(\mathbb{R})$  : densité de la loi de  $\ln(1 - S)$  ;
- $\lambda$  : taux de mutation ;
- $\mu$  : proportion de mutations létales.

# Explication des termes

Donc (\*) :

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

Traduction :

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

# Explication des termes

Donc (\*) :

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

Traduction :

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

Conservation de la masse :

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

# Solution explicite

On applique la transformée de Fourier sur :

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

donc

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

# Solution explicite

On applique la transformée de Fourier sur :

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

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  $x \in \mathbb{R}$  :

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

# Solution explicite

On applique la transformée de Fourier sur :

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

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 :

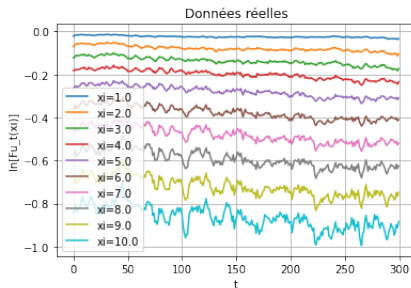
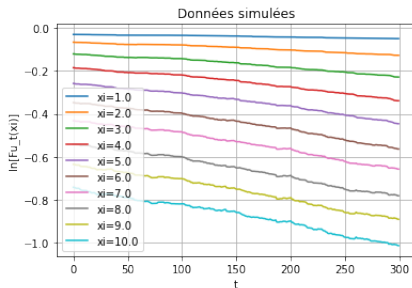
## Proposition

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

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

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

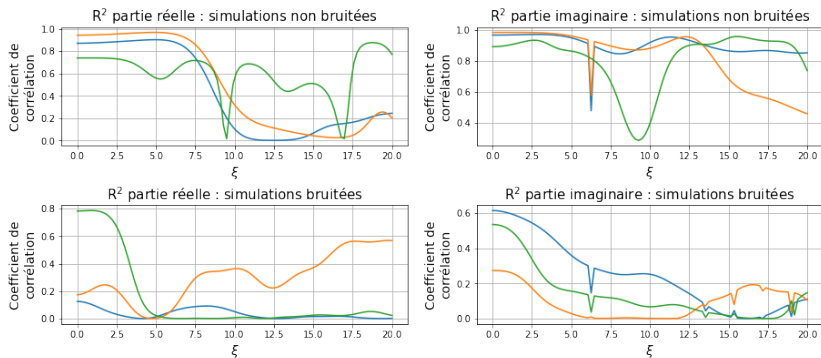
# Vérification



- fonctionne bien pour les petits  $\xi$  ;
- fonctionne très mal pour les grands  $\xi$  ( $\rightarrow$  overfitting?).

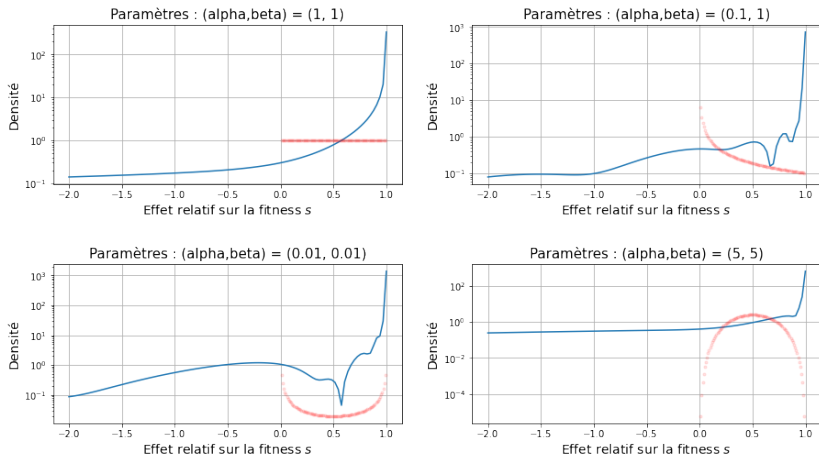


# Inférence des $a_\xi$



**Figure** – Comparaison, selon  $\xi$ , des qualité d'ajustement d'un modèle affine pour les parties réelles (gauche) ou imaginaires (droite) de  $a_\xi$ , sur des simulations bruitées ou non.

# Mauvais résultats



**Figure** – Comparaison entre les DFE inférées (en bleu) et les lois Beta choisies pour effectuer les simulations (en rouge).

# Simulation avec grand nombre de canaux I

On enlève ?

- \* droites magnifiques
- \* stabilité pour  $\xi_{max}$  petit
- \* non-stabilité pour  $\xi_{max}$  grand

# Conclusion et perspectives

- ça ne marche pas

# Références

Merci pour votre attention !



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