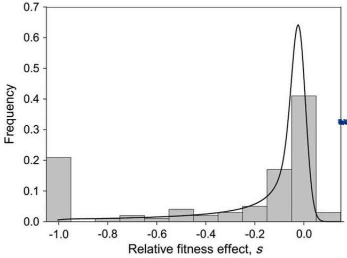
Estimer les effets des mutations sur la valeur sélective d'une bactérie

Jérémy Andréoletti et Nathanaël Boutillon

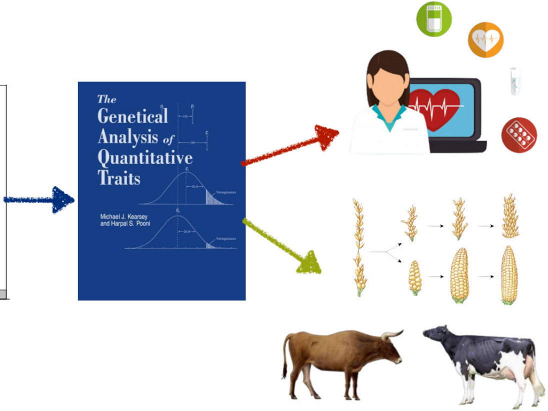
Encadrantes : Marie Doumic et Lydia Robert

- 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

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

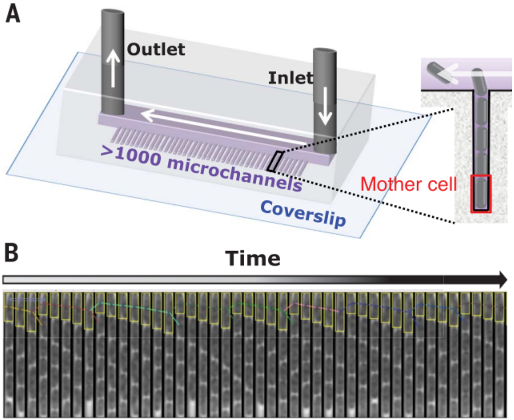
Plan

1. Expérience & simulations
2. Problème des moments
   - Fonction caractéristique
   - Calculs d'erreurs
3. EDP
   - Nouveau point de vue
   - Applications

Section 1

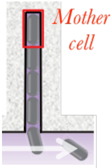
Expérience & simulations

Expérience : accumulation de mutations



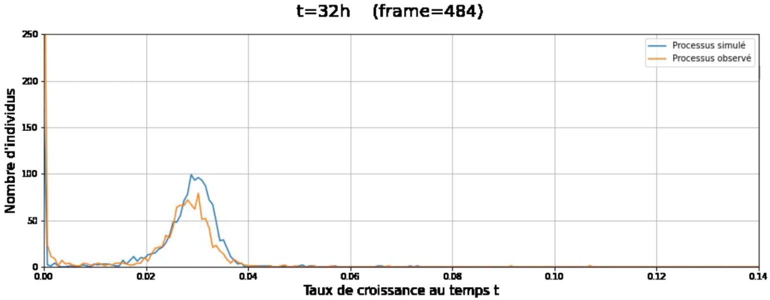
Robert et al., 2018

Expérience : accumulation de mutations



Robert et al., 2018

Simulations : image



Simulations

Section 2

Problème des moments

Estimation des moments

Analyse de *Robert et al.*, 2018

Moments de la loi des effets des mutations ;

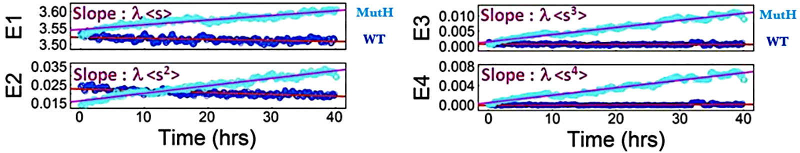
$$E_n(t) := \sum_{k=1}^{n} \binom{n}{k} (-1)^k \ln\left(\mathbb{E}\left[W_t^k\right]\right) = \left(\lambda \mathbb{E}\left[S^n\right]\right) t$$

Estimation des moments

Analyse de *Robert et al.*, 2018

Moments de la loi des effets des mutations ;

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



$\rightarrow$ estimation facile des $\mathbb{E}[S^n]$.

Méthode par la fonction caractéristique

- Moments $\to$ fonction caractéristique φ_S :

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

Méthode par la fonction caractéristique

- Moments $\to$ fonction caractéristique φ_S :

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

- Fonction caractéristique $\varphi_S \to$ densité f de la loi de S.

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

Méthode par la fonction caractéristique

- Moments $\to$ fonction caractéristique φ_S :

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

- Fonction caractéristique $\varphi_S \to$ densité f de la loi de S.

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

- Avec les N premiers moments estimés m_k :

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

- **Espoirs :** $\varphi_S(\xi) \simeq \hat{\varphi}_S(\xi)$ donc $f(x) \simeq \hat{f}(x)$.

Calculs d'erreurs I

Proposition

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

$$\left|\hat{f}(x) - f(x)\right| \leqslant \underbrace{\frac{\|f^{(k)}\|_1}{2\pi^2(k-1)A^{k-1}}}_{\text{erreur de régularisation}} + \underbrace{\frac{A^{N+1}}{\pi(N+1)!}\mathbb{E}\left[S^N(e^{AS-1})\right]}_{\text{nb fini de moments}} + \underbrace{\frac{\|\varepsilon(N)\|_\infty(e^A - 1)}{\pi}}_{\text{erreur sur les moments}}$$

Calculs d'erreurs II

Proposition

Si :

- erreur sur les $\mathbb{E}[S^n]$ bornée par ε (pour tout $n \in \mathbb{N}$) ;
- $\|f^{(2)}\|_1 < +\infty$,

alors :

$$\|f - \hat{f}\|_\infty = O\left(\left|\frac{1}{\ln^2(\varepsilon)}\right|\right) \qquad \text{quand } \varepsilon \to 0$$

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 $\mathcal{L}(\ln W_t)$;
- $m(t)$: proportion de cellules mortes en t
  $\rightarrow$ Loi de $\ln W_t$:

$$m(t)\delta_{-\infty} + u(t, \cdot)$$

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 $\mathcal{L}(\ln W_t)$;
- $m(t)$: proportion de cellules mortes en t
  $\rightarrow$ Loi de $\ln W_t$:

$$m(t)\delta_{-\infty} + u(t, \cdot)$$

- $f(\cdot) \in C^{\infty}(\mathbb{R})$: densité de $\mathcal{L}(\ln(1 - S))$;
- μ : proportion de mutations létales
  $\rightarrow$ Loi de $\ln(1 - S)$:

$$\mu\delta_{-\infty} + f(\cdot)$$

- λ : taux de mutation

EDP et explication des termes

Donc $(*)$:

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

Traduction :

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

Proposition

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

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

Terme de droite indépendant du temps

Solution explicite

Proposition

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

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

Terme de droite indépendant du temps : $\partial_t(\mathcal{F}u_t(\xi)) = a_\xi \mathcal{F}u_t(\xi)$.

D'où :

$$\begin{aligned}
\mathcal{F}u_t(\xi) &= e^{a_\xi t}\mathcal{F}u_0(\xi) \\
\ln|\mathcal{F}u_t(\xi)| &= \ln|\mathcal{F}u_0(\xi)| + t \times \Re(a_\xi) \\
\arg\left(\mathcal{F}u_t(\xi)\right) &= \arg \mathcal{F}u_0(\xi) + t \times \Im(a_\xi)
\end{aligned}$$

Vérification

$$\ln|\mathcal{F}u_t(\xi)| = \ln|\mathcal{F}u_0(\xi)| + t \times \Re(a_\xi)$$
$$\arg(\mathcal{F}u_t(\xi)) = \arg\mathcal{F}u_0(\xi) + t \times \Im(a_\xi)$$

Vérification

$$\ln|\mathcal{F}u_t(\xi)| = \ln|\mathcal{F}u_0(\xi)| + t \times \Re(a_\xi)$$
$$\arg(\mathcal{F}u_t(\xi)) = \arg\mathcal{F}u_0(\xi) + t \times \Im(a_\xi)$$

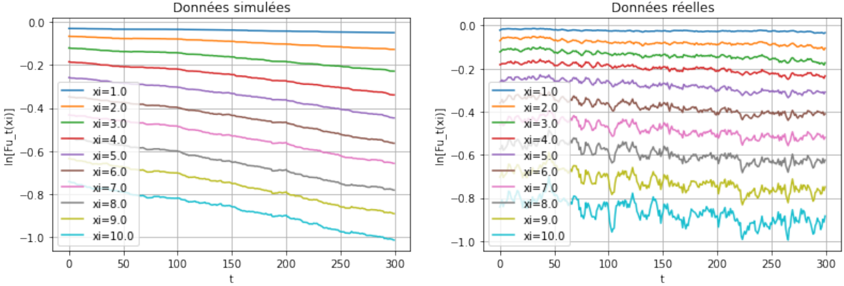


Figure – Graphes de $t \mapsto \ln|\mathcal{F}u_t(\xi)|$ pour données simulées et expérimentales.

Mauvais résultats

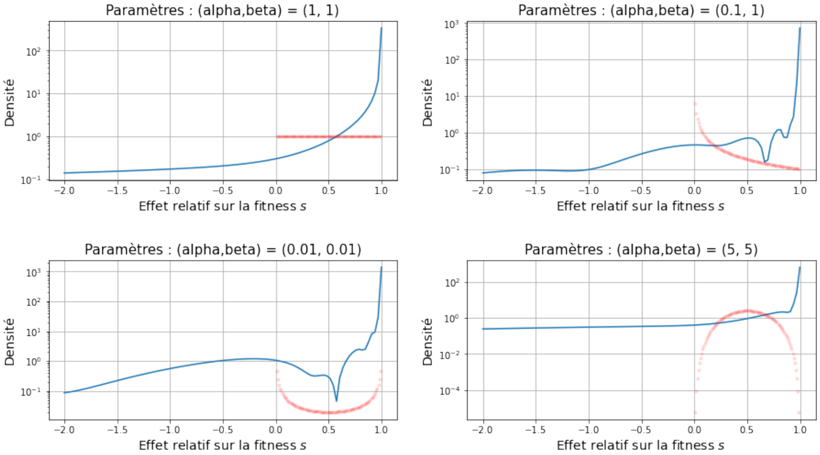


Figure – Comparaison entre les DFE inférées (en bleu) et les lois Beta choisies pour effectuer les simulations (en rouge). Utilisation de $f(x) = \mathcal{F}^{-1}\left(\frac{a_\xi}{\lambda} + 1\right)$

Conclusion et perspectives

- Problème inverse sévèrement mal posé ;
- 3 approches : Analyses naïves, Problème des moments, EDP
  $\rightarrow$ méthodes insuffisantes en pratique ;

Conclusion et perspectives

- Problème inverse sévèrement mal posé ;
- 3 approches : Analyses naïves, Problème des moments, EDP
  $\rightarrow$ méthodes insuffisantes en pratique ;
- Notebooks annotés : analyses reproductibles
  $\rightarrow$ en attente de données plus nombreuses ou moins bruitées ;

Conclusion et perspectives

- Problème inverse sévèrement mal posé ;
- 3 approches : Analyses naïves, Problème des moments, EDP
  $\rightarrow$ méthodes insuffisantes en pratique ;
- Notebooks annotés : analyses reproductibles
  $\rightarrow$ en attente de données plus nombreuses ou moins bruitées ;
- Améliorations :
  - $\star$ Combinaison EDP + estimation des moments ;
  - $\star$ Similitudes avec les problèmes de fragmentation ;
  - $\star$ Propriétés de la DFE ;
  - $\star$ Résultats théoriques sur l'erreur ;

  $\rightarrow$ simulations efficaces pour tester ces approches.

Références

Merci pour votre attention !

- Robert et al., *Mutation dynamics and fitness effects followed in single cells*, Science 359, 1283–1286, 16 March 2018
- Trudy F. C. Mackay, Eric A. Stone, Julien F. Ayroles, *The genetics of quantitative traits : challenges and prospects*, Nature Reviews Genetics, 565–577, 2009
- 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