

Modèles de mutation : extension à des processus non-homogènes en temps

Adrien Mazoyer
sous la dir. de Bernard Ycart

Laboratoire Jean Kuntzmann, UGA GRENoble

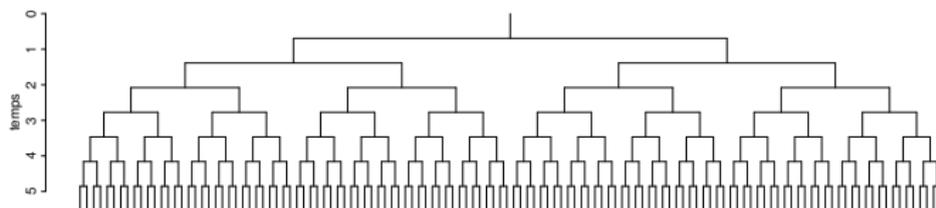
Rencontres des Jeunes Statisticiens 2017



Plan

- 1 Modèles de mutations
- 2 Estimation paramétrique
- 3 Modèles inhomogènes en temps

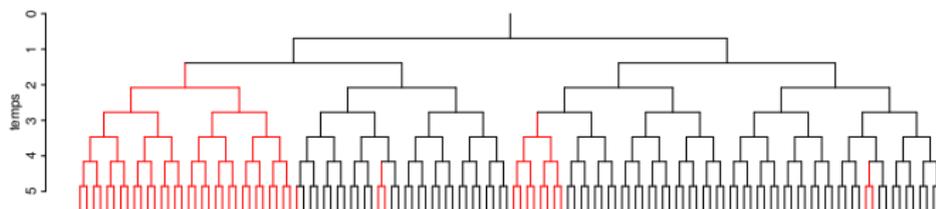
Modèle de mutations



Principe

- Croissance d'une population de cellules ;

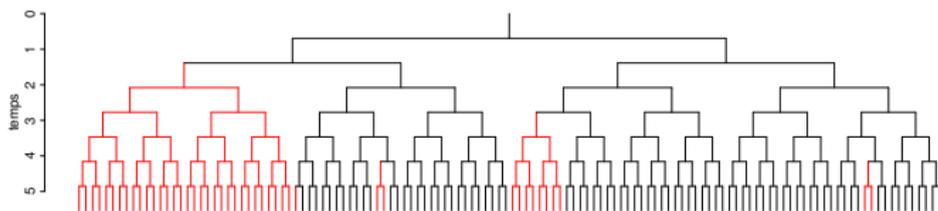
Modèle de mutations



Principe

- Croissance d'une population de cellules **avec mutations** ;
- Une mutation \Rightarrow développement d'un **clone** ;

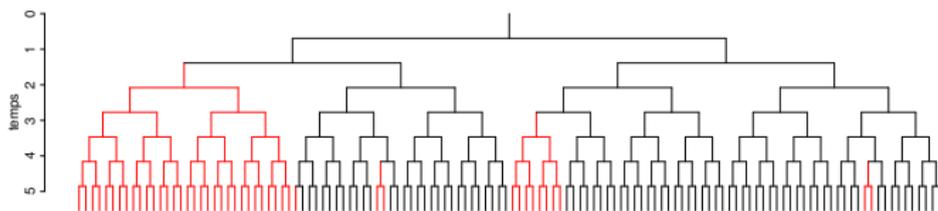
Modèle de mutations



Principe

- Croissance d'une population de cellules **avec mutations** ;
- Une mutation \Rightarrow développement d'un **clone** ;
- Nombreuses divisions et faible probabilité de mutations ;

Modèle de mutations



Principe

- Croissance d'une population de cellules **avec mutations** ;
- Une mutation \Rightarrow développement d'un **clone** ;
- Nombreuses divisions et faible probabilité de mutations ;
- Observation de nombres finaux de cellules.

Motivations

Observations

 M

2

3

0

0

5

 $N_{\text{tot}}=1\text{e}9$

6

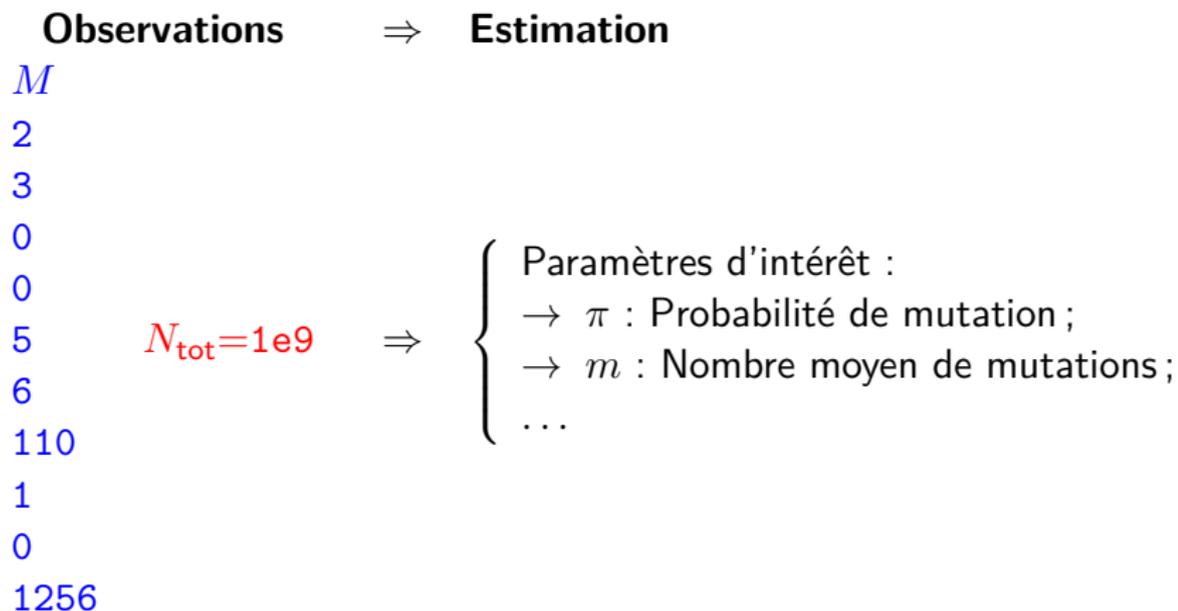
110

1

0

1256

Motivations



Nombre final de mutants

Fonction génératrice (PGF) [Yca13]

Nombre final de mutants

Fonction génératrice (PGF) [Yca13]

$$\phi(z) = \mathbb{E} \left[z^M \right] = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \int_0^\infty \psi(z, u) \rho e^{-\rho u} du,$$

Nombre final de mutantes

Fonction génératrice (PGF) [Yca13]

$$\phi(z) = \mathbb{E} \left[z^M \right] = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \int_0^\infty \psi(z, u) \rho e^{-\rho u} du,$$

et

- m : le nombre moyen de mutations ;
- ρ : paramètre de « fitness » (rapport des taux de croissance) ;

Nombre final de mutantes

Fonction génératrice (PGF) [Yca13]

$$\phi(z) = \mathbb{E} \left[z^M \right] = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \int_0^\infty \psi(z, u) \rho e^{-\rho u} du,$$

et

- m : le nombre moyen de mutations ;
- ρ : paramètre de « fitness » (rapport des taux de croissance) ;
- $\psi(\cdot, t)$: PGF de la taille en t d'un clone démarré en 0 ;
- ψ dépend de la loi des durées de vie des mutantes.

Nombre final de mutantes

Fonction génératrice (PGF) [Yca13]

$$\phi(z) = \mathbb{E} \left[z^M \right] = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \int_0^\infty \psi(z, u) \rho e^{-\rho u} du,$$

et

- m : le nombre moyen de mutations ;
- ρ : paramètre de « fitness » (rapport des taux de croissance) ;
- $\psi(\cdot, t)$: PGF de la taille en t d'un clone démarré en 0 ;
- ψ dépend de la loi des durées de vie des mutantes.
- Loi à queue lourde de paramètre de queue ρ (\Rightarrow « jackpots ») ;

Méthodes d'estimation

De nombreuses méthodes proposées

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombres de mutantes [LC49 ; Fos06] ;

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombres de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombres de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;
- ... [Fos06].

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;
- ... [Fos06].

Propriétés d'un « bon » estimateur

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutants [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;
- ... [Fos06].

Propriétés d'un « bon » estimateur

- 1 Consistance, i.e. $\lim_{n \rightarrow +\infty} \hat{\theta}_n = \theta$;

Méthodes d'estimation

De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;
- ... [Fos06].

Propriétés d'un « bon » estimateur

- 1 Consistance, i.e. $\lim_{n \rightarrow +\infty} \hat{\theta}_n = \theta$;
- 2 Convergence en loi de $\sqrt{n} (\hat{\theta}_n - \theta)$.

Méthodes d'estimation

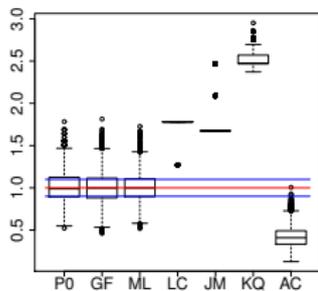
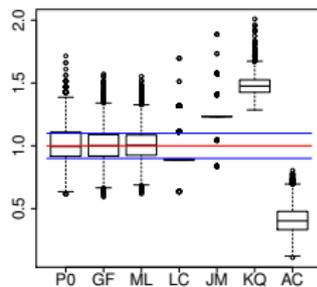
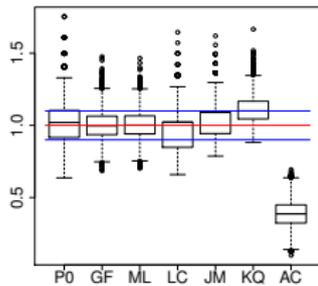
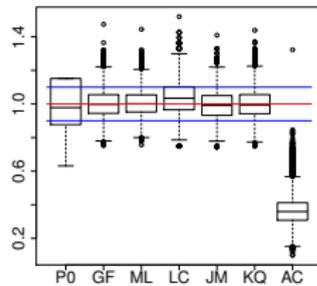
De nombreuses méthodes proposées

- Basées sur des quantiles du nombre de mutantes [LC49 ; Fos06] ;
- Basée sur le nombre de décomptes nuls [LD43] ;
- Par le Maximum de Vraisemblance ;
- Basée sur la fonction génératrice [HY12] ;
- ... [Fos06].

Propriétés d'un « bon » estimateur

- 1 Consistance, i.e. $\lim_{n \rightarrow +\infty} \hat{\theta}_n = \theta$;
- 2 Convergence en loi de $\sqrt{n} (\hat{\theta}_n - \theta)$.

« Comparaison » des méthodes

 $m = 0.5$  $m = 1$  $m = 2$  $m = 4$ 

Méthode GF

Estimateurs [HY12]

Étant donné un échantillon de décomptes de mutantes $(M_i)_{i=1\dots n}$;
Si $0 < z_1 < z_2 < 1$, $z_3 \in]0; 1[$, alors m et ρ peuvent être estimés
par :

$$\hat{m}_{GF}(z_3) = \frac{\log(\hat{\phi}_n(z_3))}{\hat{h}(z_3) - 1} \quad \text{et} \quad \hat{\rho}_{GF}(z_1, z_2) = f_{z_1, z_2}^{-1}(\hat{y}_n),$$

Méthode GF

Estimateurs [HY12]

Étant donné un échantillon de décomptes de mutantes $(M_i)_{i=1\dots n}$;
Si $0 < z_1 < z_2 < 1$, $z_3 \in]0; 1[$, alors m et ρ peuvent être estimés
par :

$$\hat{m}_{GF}(z_3) = \frac{\log(\hat{\phi}_n(z_3))}{\hat{h}(z_3) - 1} \quad \text{et} \quad \hat{\rho}_{GF}(z_1, z_2) = f_{z_1, z_2}^{-1}(\hat{y}_n),$$

où $\hat{\phi}_n(z) = \frac{1}{n} \sum_{i=1}^n z^{M_i}$,

Méthode GF

Estimateurs [HY12]

Étant donné un échantillon de décomptes de mutantes $(M_i)_{i=1\dots n}$;
Si $0 < z_1 < z_2 < 1$, $z_3 \in]0; 1[$, alors m et ρ peuvent être estimés
par :

$$\hat{m}_{GF}(z_3) = \frac{\log(\hat{\phi}_n(z_3))}{\hat{h}(z_3) - 1} \quad \text{et} \quad \hat{\rho}_{GF}(z_1, z_2) = f_{z_1, z_2}^{-1}(\hat{y}_n),$$

où $\hat{\phi}_n(z) = \frac{1}{n} \sum_{i=1}^n z^{M_i}$, \hat{h} est la fonction h avec $\rho = \hat{\rho}_{GF}$

Méthode GF

Estimateurs [HY12]

Étant donné un échantillon de décomptes de mutantes $(M_i)_{i=1\dots n}$;
Si $0 < z_1 < z_2 < 1$, $z_3 \in]0; 1[$, alors m et ρ peuvent être estimés
par :

$$\hat{m}_{GF}(z_3) = \frac{\log(\hat{\phi}_n(z_3))}{\hat{h}(z_3) - 1} \quad \text{et} \quad \hat{\rho}_{GF}(z_1, z_2) = f_{z_1, z_2}^{-1}(\hat{y}_n),$$

où $\hat{\phi}_n(z) = \frac{1}{n} \sum_{i=1}^n z^{M_i}$, \hat{h} est la fonction h avec $\rho = \hat{\rho}_{GF}$ et

$$f_{z_1, z_2}(\rho) = \frac{h(z_1) - 1}{h(z_2) - 1} \quad \text{et} \quad \hat{y}_n(z_1, z_2) = \frac{\log(\hat{\phi}_n(z_1))}{\log(\hat{\phi}_n(z_2))}.$$

Méthode GF

Intérêts

- Temps de calcul très faible ;
 - Précision comparable à celle de ML ;
 - Numériquement plus stable que ML (moins sensible aux jackpots) ;
- ⇒ Initialisation de ML (aide à la stabilité numérique et à la vitesse d'optimisation).

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

- Le nombre final de cellules N_{tot} n'est pas constant ;

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

- Le nombre final de cellules N_{tot} n'est pas constant ;
- Certaines cellules meurent au lieu de se diviser ;

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

- Le nombre final de cellules N_{tot} n'est pas constant ;
- Certaines cellules meurent au lieu de se diviser ;
- Un échantillon peut être extrait d'une population plus grande ;

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

- Le nombre final de cellules N_{tot} n'est pas constant ;
- Certaines cellules meurent au lieu de se diviser ;
- Un échantillon peut être extrait d'une population plus grande ;
- Les durées de vies ne sont pas exponentielles ;

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

- Le nombre final de cellules N_{tot} n'est pas constant ;
- Certaines cellules meurent au lieu de se diviser ;
- Un échantillon peut être extrait d'une population plus grande ;
- Les durées de vies ne sont pas exponentielles ;
- Les durées de vies ne sont pas identiquement distribuées.

Des estimations biaisées

Nécessité d'identifier les sources de biais et de corriger les estimations.

Sources de biais

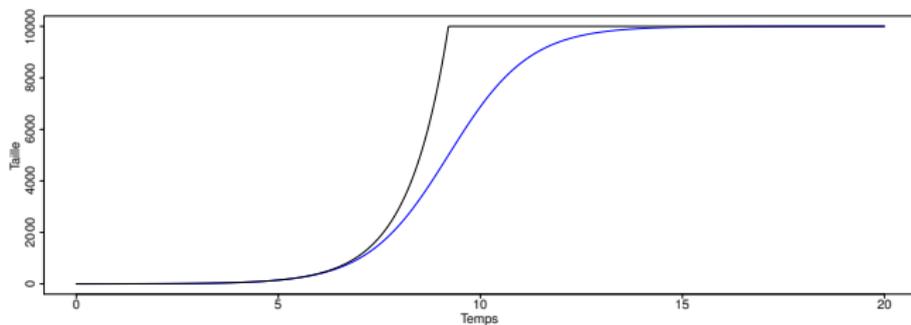
- Le nombre final de cellules N_{tot} n'est pas constant [YV14];
- Certaines cellules meurent au lieu de se diviser [Yca14];
- Un échantillon peut être extrait d'une population plus grande;
- Les durées de vies ne sont pas exponentielles [Yca13];
- Les durées de vies ne sont pas identiquement distribuées.

Source de biais

Les durées de vie ne sont pas identiquement distribuées.

Source de biais

Les durées de vie ne sont pas identiquement distribuées.



Modélisation

Hypothèses

- L'instant de division d'une cellule **normale** dépend de sa date de naissance s ;

Modélisation

Hypothèses

- L'instant de division d'une cellule **normale** dépend de sa date de naissance s ;
- Sa fonction de répartition (CDF) $F_{\nu}(s, t)$ est définie sur $\overline{\mathbb{R}}_+$

Modélisation

Hypothèses

- L'instant de division d'une cellule **normale** dépend de sa date de naissance s ;
- Sa fonction de répartition (CDF) $F_{\nu}(s, t)$ est définie sur $\overline{\mathbb{R}}_+$ à partir d'une CDF $F_{\nu}^*(s, \cdot)$ d'une mesure de sous-prob. sur \mathbb{R}_+ t.q. :
 - $\lim_{t \rightarrow \infty} F_{\nu}^*(\cdot, t) \leq 1$ et $F_{\nu}^*(s, t) = 0$ si $t < s$ et :

Modélisation

Hypothèses

- L'instant de division d'une cellule **normale** dépend de sa date de naissance s ;
- Sa fonction de répartition (CDF) $F_{\nu}(s, t)$ est définie sur $\overline{\mathbb{R}}_+$ à partir d'une CDF $F_{\nu}^*(s, \cdot)$ d'une mesure de sous-prob. sur \mathbb{R}_+ t.q. :
 - $\lim_{t \rightarrow \infty} F_{\nu}^*(\cdot, t) \leq 1$ et $F_{\nu}^*(s, t) = 0$ si $t < s$ et :

$$F_{\nu}(s, t) = F_{\nu}^*(s, t)\mathbf{1}_{t \in [s; \infty)} + \mathbf{1}_{t = \infty} ;$$

Modélisation

Hypothèses

- L'instant de division d'une cellule **normale** dépend de sa date de naissance s ;
- Sa fonction de répartition (CDF) $F_{\nu}(s, t)$ est définie sur $\overline{\mathbb{R}}_+$ à partir d'une CDF $F_{\nu}^*(s, \cdot)$ d'une mesure de sous-prob. sur \mathbb{R}_+ t.q. :

- $\lim_{t \rightarrow \infty} F_{\nu}^*(\cdot, t) \leq 1$ et $F_{\nu}^*(s, t) = 0$ si $t < s$ et :

$$F_{\nu}(s, t) = F_{\nu}^*(s, t) \mathbf{1}_{t \in [s; \infty)} + \mathbf{1}_{t = \infty} ;$$

- la fonction η_{ν} définie par :

$$\eta_{\nu}(s, t) = -\log(1 - F_{\nu}^*(s, t)) ,$$

vérifie

$$\eta_{\nu}(s, t) = \eta_{\nu}(0, t) - \eta_{\nu}(0, s) .$$

Modélisation

Hypothèses

- L'instant de division d'une cellule **mutante** dépend de sa date de naissance s ;
- Sa fonction de répartition (CDF) $F_{\mu}(s, t)$ est définie sur $\overline{\mathbb{R}}_+$ à partir d'une CDF $F_{\mu}^*(s, \cdot)$ d'une mesure de sous-prob. sur \mathbb{R}_+ t.q. :

- $\lim_{t \rightarrow \infty} F_{\mu}^*(\cdot, t) \leq 1$ et $F_{\mu}^*(s, t) = 0$ si $t < s$ et :

$$F_{\mu}(s, t) = F_{\mu}^*(s, t) \mathbf{1}_{t \in [s; \infty)} + \mathbf{1}_{t = \infty} ;$$

- la fonction η_{μ} définie par :

$$\eta_{\mu}(s, t) = -\log(1 - F_{\mu}^*(s, t)) ,$$

vérifie

$$\eta_{\mu}(s, t) = \eta_{\mu}(0, t) - \eta_{\mu}(0, s) .$$

Loi du nombre de mutantes

Fonction génératrice (PGF) (Maz16)

$$\phi(z) = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \lim_{t \rightarrow \infty} \int_0^t \psi(z, u, t) \frac{\lambda_\nu(u) e^{-\eta_\nu(u, t)}}{1 - e^{-\eta_\nu(0, t)}} du,$$

Loi du nombre de mutantes

Fonction génératrice (PGF) (Maz16)

$$\phi(z) = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \lim_{t \rightarrow \infty} \int_0^t \psi(z, u, t) \frac{\lambda_\nu(u) e^{-\eta_\nu(u, t)}}{1 - e^{-\eta_\nu(0, t)}} du,$$

et

- $\psi(\cdot, u, t)$: PGF de la taille en t d'un clone mutant démarré en u ;

Loi du nombre de mutantes

Fonction génératrice (PGF) (Maz16)

$$\phi(z) = \exp(-m(1 - h(z))),$$

avec

$$h(z) = \lim_{t \rightarrow \infty} \int_0^t \psi(z, u, t) \frac{\lambda_\nu(u) e^{-\eta_\nu(u, t)}}{1 - e^{-\eta_\nu(0, t)}} du,$$

et

- $\psi(\cdot, u, t)$: PGF de la taille en t d'un clone mutant démarré en u ;
- λ_ν : taux inst. de division des non-mutantes ;

Conséquences

- Possible de généraliser la notion de « fitness » :

$$\rho = \frac{\lambda_\nu(u)}{\lambda_\mu(u)} ;$$

Conséquences

- Possible de généraliser la notion de « fitness » :

$$\rho = \frac{\lambda_\nu(u)}{\lambda_\mu(u)};$$

- Possible d'ajuster la croissance d'un clone à « n'importe quelle » fonction f :

$$F_\mu^*(s, t) = 1 - \frac{f(s)}{f(t)};$$

Conséquences

- Possible de généraliser la notion de « fitness » :

$$\rho = \frac{\lambda_\nu(u)}{\lambda_\mu(u)};$$

- Possible d'ajuster la croissance d'un clone à « n'importe quelle » fonction f :

$$F_\mu^*(s, t) = 1 - \frac{f(s)}{f(t)};$$

- Dans ces deux cas :

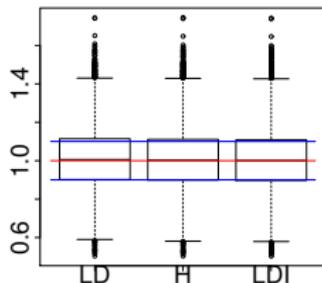
$$\Rightarrow h(z) = \lim_{t \rightarrow \infty} \frac{1}{1 - \left(\frac{f(0)}{f(t)}\right)^\rho} \int_{\frac{f(0)}{f(t)}}^1 \frac{\rho z v^\rho}{1 - z(1-v)} dv;$$

- Méthodes d'estimation toujours applicables (et propriétés conservées)!

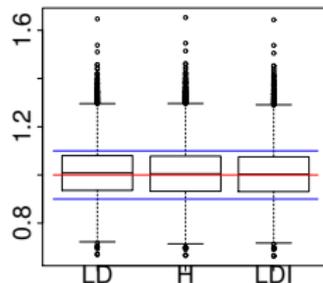
Résultats empiriques

$$\frac{f(\infty)}{f(0)} = 100$$

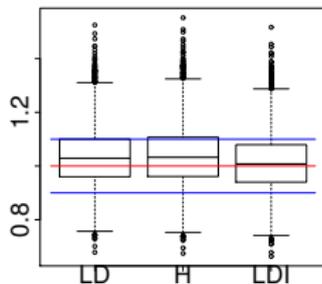
$m = 0.5$



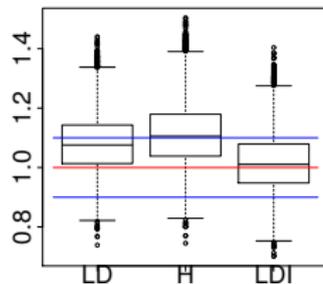
$m = 2$



$m = 4$



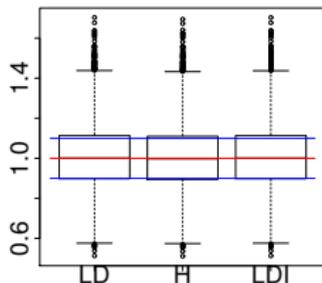
$m = 8$



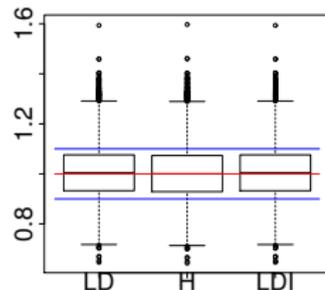
Résultats empiriques

$$\frac{f(\infty)}{f(0)} = 10^4$$

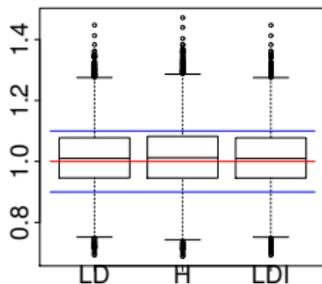
$m = 0.5$



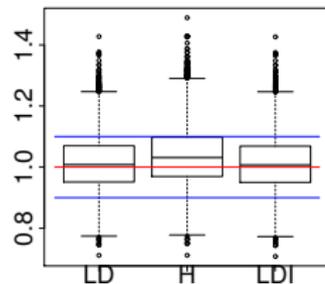
$m = 2$



$m = 4$



$m = 8$



flan

Package R pour l'ANalyse de FLuctuation

Fonctionnalités

- Loi du nombre de mutantes (`rflan`, `dflan`,...);
- Estimation paramétrique (`mutestim`);
- Tests statistiques (`flan.test`).

› Github (ID : AdriMaz)

› CRAN

Fin !

Merci de votre attention !

Références I



P.L. FOSTER. « Methods for Determining Spontaneous Mutation Rates ». In : *Method. Enzymol.* 409 (2006), p. 195–213.



A. HAMON et B. YCART. « Statistics for the Luria-Delbrück distribution ». In : *Elect. J. Statist.* 6 (2012), p. 1251–1272.



D.E. LEA et C.A. COULSON. « The distribution of the number of mutants in bacterial populations ». In : *J. Genet.* 49.3 (1949), p. 264–285.



S.E. LURIA et M. DELBRÜCK. « Mutations of bacteria from virus sensitivity to virus resistance ». In : *Genetics* 28.6 (1943), p. 491–511.



B. YCART. « Fluctuation analysis : can estimates be trusted ? » In : *PLoS One* 8.12 (2013), p. 1–12. URL : <http://dx.doi.org/10.1371/journal.pone.0080958>.



B. YCART. « Fluctuation analysis with cell deaths ». In : *J. Appl. Probab. Statist* 9.1 (2014), p. 13–29.



B. YCART et N. VEZIRIS. « Unbiased estimates of mutation rates under fluctuating final counts ». In : *PLoS One* 9.7 (2014), p. 1–10. URL : <http://dx.doi.org/10.1371/journal.pone.0101434>.