

flan : Un package pour l'analyse de fluctuations des modèles de mutations

Adrien Mazoyer

Laboratoire Jean Kuntzmann, UGA GRENoble

Rencontres R 2016

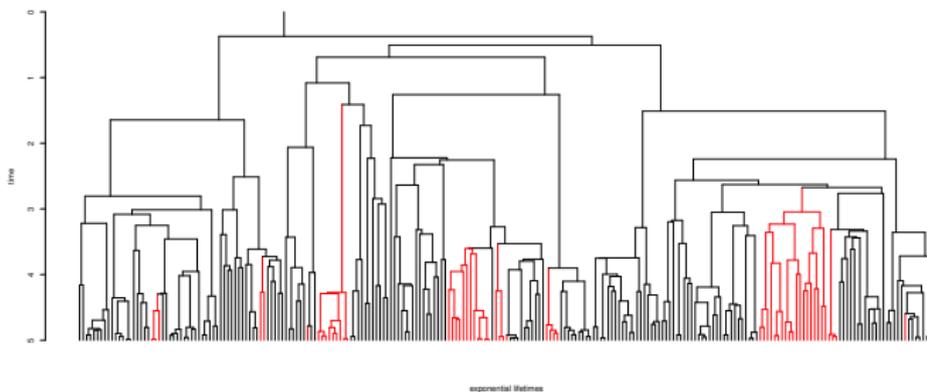


Plan

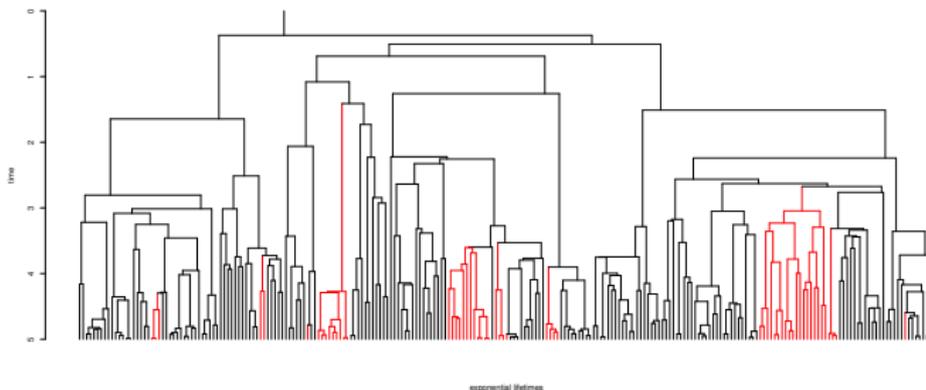
- 1 Modèle de mutations
- 2 Estimation paramétrique
- 3 Fonctionnalités de flan
- 4 Exemples

Modèle de mutations

Modèle de mutations



Modèle de mutations



Principe

- Croissance d'une population de cellules avec mutations ;
- Nombreuses divisions et faible probabilité de mutations ;
- Une mutation \Rightarrow développement d'un clone ;
- Observation du nombre final de cellules (total et mutantes).

Motivations

Observations

 N_{mut}

2

3

0

0 $N_f = 1e9$

5

6

110

1

0

Motivations

Observations

⇒

Estimation

N_{mut}

2

3

0

0

5

6

110

1

0

$N_f = 1e9$

⇒

Paramètres d'intérêt :

→ π : Probabilité de mutation ;

→ α : Nombre moyen de mutations ;

...

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;
- Les durées de vie des cellules normales sont i.i.d. ;

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;
- Les durées de vie des cellules normales sont i.i.d. ;
- Lorsqu'elle arrive en fin de vie, une cellule normale produit :
 - une cellule normale et une mutante avec probabilité π ;
 - deux cellules normales avec probabilité $1 - \pi$;

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;
- Les durées de vie des cellules normales sont i.i.d. ;
- Lorsqu'elle arrive en fin de vie, une cellule normale produit :
 - une cellule normale et une mutante avec probabilité π ;
 - deux cellules normales avec probabilité $1 - \pi$;
- Les durées de vie des mutantes sont exponentiellement i.i.d. ;

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;
- Les durées de vie des cellules normales sont i.i.d. ;
- Lorsqu'elle arrive en fin de vie, une cellule normale produit :
 - une cellule normale et une mutante avec probabilité π ;
 - deux cellules normales avec probabilité $1 - \pi$;
- Les durées de vie des mutantes sont exponentiellement i.i.d. ;
- Lorsqu'une mutante arrive en fin de vie, elle produit deux mutantes :

Modèle de Luria-Delbrück

Hypothèses de modélisations

- n cellules normales à $t = 0$;
- Les durées de vie des cellules normales sont i.i.d. ;
- Lorsqu'elle arrive en fin de vie, une cellule normale produit :
 - une cellule normale et une mutante avec probabilité π ;
 - deux cellules normales avec probabilité $1 - \pi$;
- Les durées de vie des mutantes sont exponentiellement i.i.d. ;
- Lorsqu'une mutante arrive en fin de vie, elle produit deux mutantes :
- Toutes les variables et tous les événements (divisions et mutations) sont mutuellement indépendants.

Nombre final de mutants

Caractéristique de la distribution

- Loi (asymptotiquement) **explicite** \Rightarrow inférence paramétrique ;

Nombre final de mutantes

Caractéristique de la distribution

- Loi (asymptotiquement) **explicite** \Rightarrow inférence paramétrique ;
- Deux paramètres :
 - α : le nombre moyen de mutations ;
 - ρ : paramètre de "fitness" (\Leftrightarrow rapport des taux de croissance) ;

Nombre final de mutantes

Caractéristique de la distribution

- Loi (asymptotiquement) **explicite** \Rightarrow inférence paramétrique ;
- Deux paramètres :
 - α : le nombre moyen de mutations ;
 - ρ : paramètre de “fitness” (\Leftrightarrow rapport des taux de croissance) ;
- Loi à queue lourde de paramètre de queue ρ (\Rightarrow “jackpots”) ;

Estimation paramétrique

Estimation paramétrique

3 méthodes pour estimer α , π et ρ

- Lien entre $\mathbb{P}[0 \text{ mutantes}]$ et $\mathbb{P}[0 \text{ mutations}] \Rightarrow p_0$;
- Loi explicite \Rightarrow Maximum de Vraisemblance ;
- Composée Poissonnienne \Rightarrow Méthode basée sur la Fonction Génératrice ;

Estimation paramétrique

3 méthodes pour estimer α , π et ρ

- Lien entre $\mathbb{P}[0 \text{ mutantes}]$ et $\mathbb{P}[0 \text{ mutations}] \Rightarrow p_0$;
- Loi explicite \Rightarrow Maximum de Vraisemblance ;
- Composée Poissonnienne \Rightarrow Méthode basée sur la Fonction Génératrice ;

Remarques

- Calcul de $\hat{\pi}$ en divisant $\hat{\alpha}$ par le nombre total de cellules ;
- Estimateurs asymptotiquement normaux \Rightarrow test d'hypothèse.

Estimation paramétrique

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

Estimation paramétrique

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

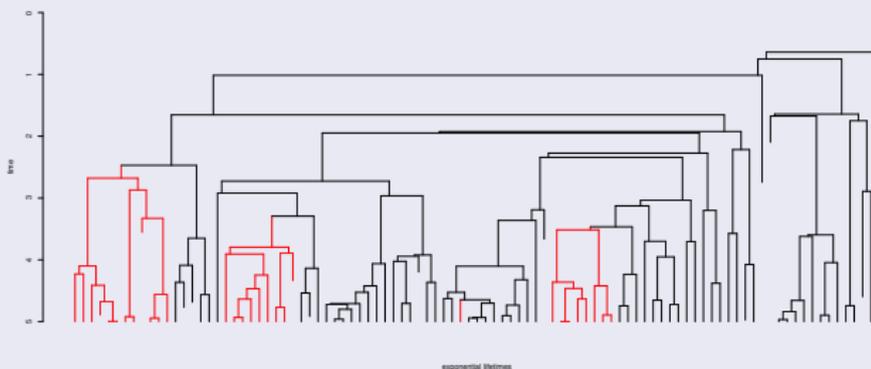
Sources de biais

- Certaines cellules meurent au lieu de se diviser ;
- Les durées de vies ne sont pas exponentielles ;
- Le nombre final de cellules N_f n'est pas constant.

Morts de cellules

Modélisation

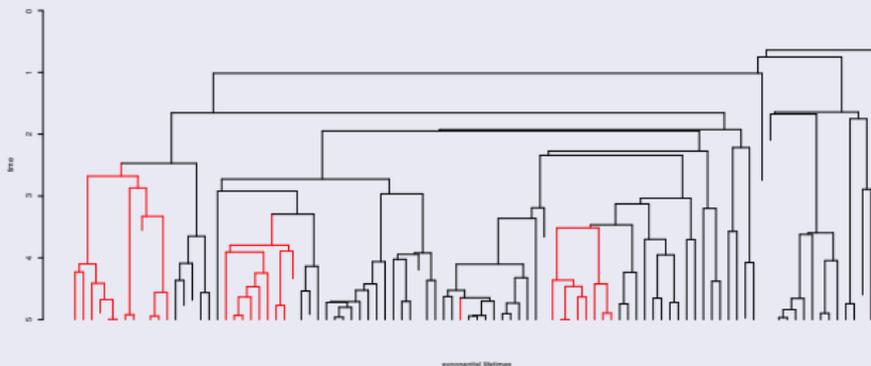
Une mutante peut mourir avec probabilité δ au lieu de se diviser.



Morts de cellules

Modélisation

Une mutante peut mourir avec probabilité δ au lieu de se diviser.

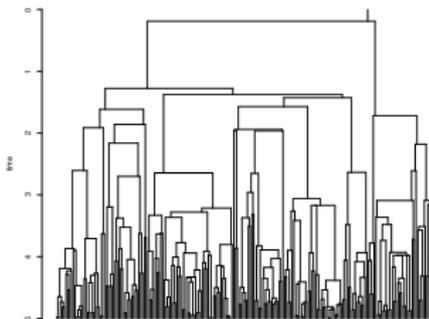


Effets sur l'estimation

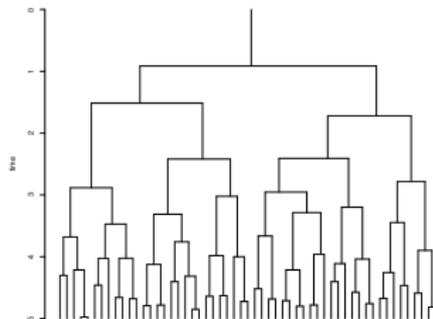
- Même méthode d'estimation ;
- Complications pour estimer δ .

Durées de vie des mutants

Pas forcément exponentielles !



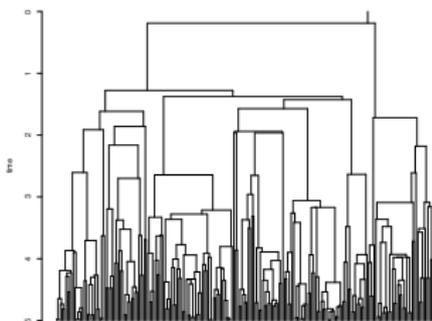
exponential lifetimes



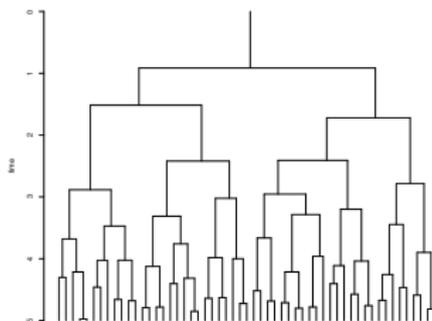
log-normal lifetimes

Durées de vie des mutants

Pas forcément exponentielles !



exponential lifetimes



log-normal lifetimes

Effets sur l'estimation

- Le vrai modèle est inconnu ;
- 2 cas explicites : exponentielles (Luria-Delbrück), constantes (Haldane).

Fluctuations de N_f

Au lieu de

 N_{mut}

2

3

0

0

5

 $N_f = 1e9$

6

110

1

0

Fluctuations de N_f

Au lieu de

 N_{mut}

2

3

0

0

5

6

110

1

0

 $N_f = 1e9$

On a

 N_{mut}

2

3

0

0

5

6

110

1

0

$$\mathbb{E}[N_f] = 1e9$$

$$\frac{\mathbb{E}[N_f]}{\text{sd}[N_f]} = 0.5$$

Fluctuations de N_f

Au lieu de

 N_{mut}

2

3

0

0

5

6

110

1

0

 $N_f = 1e9$

On a

 N_{mut}

2

3

0

0

5

6

110

1

0

$$\mathbb{E}[N_f] = 1e9$$

$$\frac{\mathbb{E}[N_f]}{\text{sd}[N_f]} = 0.5$$

Ou encore

 N_{mut}

2

3

0

0

5

6

110

1

0

 N_f

1.36e9

1.05e9

4.28e8

6.24e8

7.36e8

4.90e8

1.36e9

9.56e8

6.82e8

Fluctuations de N_f

Effets sur l'estimation

- N_f constant $\Rightarrow \pi = \frac{\alpha}{N_f}$;
- $N_f \Rightarrow \pi$ et α liés par transformée de Laplace de N_f ;

Fluctuations de N_f

Effets sur l'estimation

- N_f constant $\Rightarrow \pi = \frac{\alpha}{N_f}$;
- $N_f \Rightarrow \pi$ et α liés par transformée de Laplace de N_f ;
- Loi de N_f inconnue \Rightarrow réduction du biais par approximation ;

Fluctuations de N_f

Effets sur l'estimation

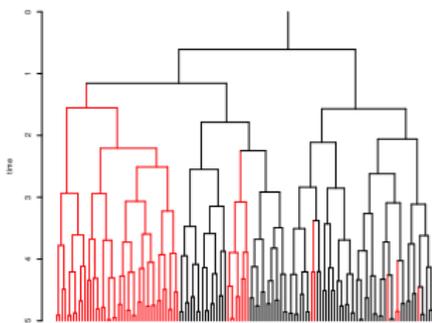
- N_f constant $\Rightarrow \pi = \frac{\alpha}{N_f}$;
- $N_f \Rightarrow \pi$ et α liés par transformée de Laplace de N_f ;
- Loi de N_f inconnue \Rightarrow réduction du biais par approximation ;
- Cas de ML sur données $(N_{\text{mut}} ; N_f)$: calcul direct de $\hat{\pi}$.

Fonctionnalités de **flan**

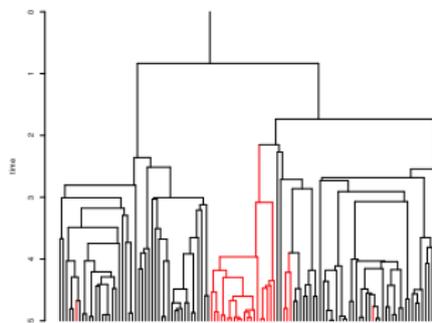
Fonctionnalités de flan

Loi du nombre de mutants

- dflan, pflan, qflan, rflan (cf. *norm) ;
- Différentes hypothèses de modélisations ;
- Une fonction graphique : draw.clone.



log-normal (fltime)



exponential (fltime)

Fonctionnalités de **flan**

Estimation paramétrique : `mutestim`

- Application des 3 méthodes ;
- Différentes hypothèses de modélisation ;

Fonctionnalités de **flan**

Estimation paramétrique : `mutestim`

- Application des 3 méthodes ;
- Différentes hypothèses de modélisation ;
- Estimation de (α, ρ) ou (π, ρ) selon les données :

| Données | Sorties |
|------------------|------------------------------|
| N_{mut} | $(\hat{\alpha}, \hat{\rho})$ |

Fonctionnalités de flan

Estimation paramétrique : mutestim

- Application des 3 méthodes ;
- Différentes hypothèses de modélisation ;
- Estimation de (α, ρ) ou (π, ρ) selon les données :

| Données | Sorties |
|--|--|
| N_{mut} | $(\hat{\alpha}, \hat{\rho})$ |
| N_{mut} et $\left(\mathbb{E}[N_f], \frac{\mathbb{E}[N_f]}{\text{sd}[N_f]} \right)$ | $(\hat{\pi}, \hat{\rho})$ déduit de $(\hat{\alpha}, \hat{\rho})$ |

Fonctionnalités de flan

Estimation paramétrique : mutestim

- Application des 3 méthodes ;
- Différentes hypothèses de modélisation ;
- Estimation de (α, ρ) ou (π, ρ) selon les données :

| Données | Sorties |
|--|--|
| N_{mut} | $(\hat{\alpha}, \hat{\rho})$ |
| N_{mut} et $\left(\mathbb{E}[N_f], \frac{\mathbb{E}[N_f]}{\text{sd}[N_f]} \right)$ | $(\hat{\pi}, \hat{\rho})$ déduit de $(\hat{\alpha}, \hat{\rho})$ |
| $(N_{\text{mut}} ; N_f)$ | <ul style="list-style-type: none"> • ML : $(\hat{\pi}, \hat{\rho})$ direct ; • p_0, GF : $(\hat{\pi}, \hat{\rho})$ déduit avec correction de $(\hat{\alpha}, \hat{\rho})$. |

Fonctionnalités de flan

Estimation paramétrique : mutestim

- Application des 3 méthodes ;
- Différentes hypothèses de modélisation ;
- Estimation de (α, ρ) ou (π, ρ) selon les données :

| Données | Sorties |
|--|--|
| N_{mut} | $(\hat{\alpha}, \hat{\rho})$ |
| N_{mut} et $\left(\mathbb{E}[N_f], \frac{\mathbb{E}[N_f]}{\text{sd}[N_f]} \right)$ | $(\hat{\pi}, \hat{\rho})$ déduit de $(\hat{\alpha}, \hat{\rho})$ |
| $(N_{\text{mut}} ; N_f)$ | <ul style="list-style-type: none"> • ML : $(\hat{\pi}, \hat{\rho})$ direct ; • p_0, GF : $(\hat{\pi}, \hat{\rho})$ déduit avec correction de $(\hat{\alpha}, \hat{\rho})$. |

+ estimation des écart-types.

Fonctionnalités de **flan**

Test d'hypothèses : `flan.test`

- Wrapper de `mutestim` ;
- Hypothèses nulles:
 - Sur 1 échantillon : valeur théorique de α , π et ρ ;
 - Sur 2 échantillons : différence des valeurs théoriques de α , π et ρ ;
- Affichage similaire à `t.test`.

Exemple 1

```
> b1 <- unlist(boeal[1:10])
> b2 <- unlist(boeal[11:20])
> b <- list(b1,b2)

# is the difference in mean number of mutation equal to 0, and in fitness equal to 0?

> flan.test(b,mutations0=0,fitness0=0,method="GF")

      Two sample GF-test (LD model)

----- Data -----
data:  b1 and b2
Sample 1 parameters : death = 0
Sample 2 parameters : death = 0
----- Statistics -----
Tstat = (0.40619 , 0.044766)
p-value for mutations difference = 0.6828
p-value for fitness difference = 0.9623
Alternative hypotheses: true mutations difference is not equal to 0
                       true fitness difference is not equal to 0
95 percent confidence interval for mutations difference:
-0.09797254   0.14958420
95 percent confidence interval for fitness difference:
-0.1625458   0.1705751
Sample 1 estimates :
mutations  fitness
0.7123002 0.7878078
Sample 2 estimates :
mutations  fitness
0.6864943 0.7837932
```

Exemple 2

```
> x <- rflan(1e3,mutprob=1e-10,fitness=2.5,dist=list("exp",rate=1),mfn=1e9,cvfn=0.2)

# according to GF method, is the mutation probability smaller than 2e-10,
# and the fitness greater than 2?
> fln.test(mc=x$mc,fn=x$fn,alternative=c("less","greater"),mutprob0=2e-10,fitness0=2,method="GF")
```

One sample GF-test (LD model)

```
----- Data -----
data: x$mc with x$fn
Sample parameters : death = 0, mfn = 1002100000, cvfn = 0.20032
----- Statistics -----
Tstat = (-8.2612 , 1.8311)
p-value for mutation probability < 2.2e-16
p-value for fitness = 0.03354
Alternative hypotheses: true mutation probability is less than 2e-10
                       true fitness is greater than 2
95 percent confidence interval for mutation probability:
      -Inf  1.284488e-10
95 percent confidence interval for fitness:
2.100942      Inf
Sample estimates :
      mutprob      fitness
1.106608e-10 2.992389e+00
```

Fin !

Merci !

Disponible sur :

▸ Github (ID : AdriMaz)

▸ CRAN