

# Modèles de mutation : étude probabiliste et estimation paramétrique

Package R `flan` (FLuctuation ANalysis)

A.Mazoyer, S.Desprésaux, B.Ycart

Rencontres R 2015

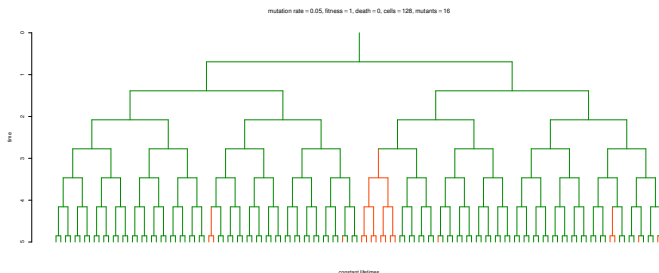
# Plan

- 1 Modèles de mutations
- 2 Estimation paramétrique
- 3 Fonctionnalités du package `flan`
- 4 À venir...

# Modèles de mutations

## Principes

- Croissance d'une population de cellules avec mutations.
- Nombreuses divisions et peu de mutations.
- Une mutation = développement d'un clone.
- Observation du nombre final de cellules (total et mutantes).



# Motivations

$N_{\text{mut}}$	$N_f$
2	1.36e9
3	1.05e9
0	4.28e8
0	6.24e8
5	7.36e8
6	4.90e8
110	1.36e9
1	9.56e8
0	6.82e8

Paramètres d'intérêt :

→  $\pi$  : Probabilité de mutation

→  $\alpha$  : Nombre moyen de mutations

→  $\rho$  : "Fitness"

...

# Méthodes d'estimation

Expressions explicites des probabilités  $\Rightarrow$  Maximum de  
Vraisemblance (ML).

## Obstacle numérique

Présence de "jackpots" (influencé par  $\rho$ ) :

1	2	131	4	1	8480	3	23900	10	8	338	2	5	71	0	25	2	23	7		
11	46	1	0	11	13	4	2324	0	4	6	6	4	1	1	5	65	14	5	4	1

## Winsorization

1	3	16	3	7	2	4	5	1	8	1	1	16	1216	1	1432	5	12	4	0

# Méthodes d'estimation

## Méthode $p_0$ (P0)

Principe :

- 1 (0 *mutantes*) relié à (0 *mutations*).
- 2 Décompte du nombre de 0.
- 3 Estimation de  $\mathbb{P}[0 \text{ mutantes}]$ .
- 4 Estimation de  $\pi$ .

## Inconvénients

- Nécessite présence de 0.
- Estimation de  $\rho$  par Max de Vraisemblance.

# Méthodes d'estimation

Mélange poissonnien  $\Rightarrow$  estimation à l'aide de la fonction génératrice (GF).

## Intérêts

- Temps de calcul très faible.
- Plus stable que ML.
- Initialisation de ML (aide à la stabilité).

# Sources de biais

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

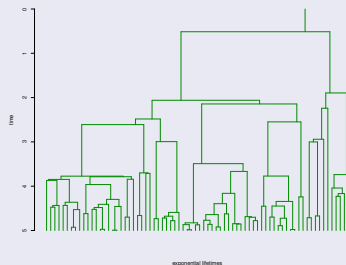
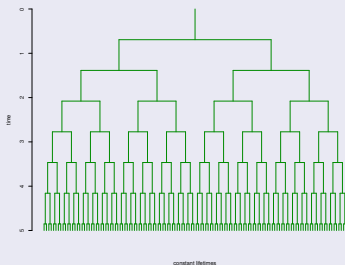
## Sources de biais

- Modèle de croissance des clones.
- Présence de morts cellulaires.
- Fluctuations de  $N_f$ .



# Sources de biais

## Modèle de croissance des clones



- Influence les estimations.
- Problème : le vrai modèle est inconnu
- Hypothèse d'estimation : le modèle est soit exponentiel (E), soit Dirac (D).

# Source de biais : morts cellulaires

## Impact

- À chaque division : probabilité  $\delta$  de mourir sans se diviser.
- Mêmes méthodes d'estimation si  $\delta$  est connu.
- Complications si  $\delta$  est inconnu.

# Source de biais : fluctuation de $N_f$

## Impact

- Si  $N_f$  constant :  $\hat{\pi} = \frac{\hat{\alpha}}{N_f}$ .
- Si  $N_f$  aléatoire : influence de  $\frac{\mathbb{E}[N_f]}{sd[N_f]}$  sur  $\hat{\pi}$ .

# Fonctionnalités du package `flan`

## Simulation : `rmut`, `pmut`, `qmut`

- Mêmes fonctionnalités que `rnorm`, `pnorm`, `qnorm`.
- Hypothèses possibles de construction :
  - Durées de vie non-exponentielles.
  - Morts cellulaires.
  - Fluctuation de  $N_f$ .

# Fonctionnalités du package `flan`

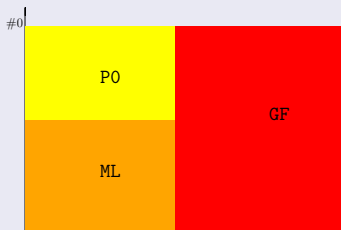
## Estimation : `mutestim`, `mutintestim`

- Renvoi :
  - Estimations de  $\pi$  et  $\rho$ .
  - Estimations des écart-types des estimateurs.
  - Intervalles de confiance.
- Hypothèses d'estimation disponibles :
  - Durées de vie exponentielles ou constantes.
  - Morts cellulaires.
  - Fluctuations de  $N_f$ .

# À venir...

## Futurs ajouts/améliorations

- Ajout de fonctionnalités avec l'avancement de la thèse (simulation/estimation).
- Estimation du paramètre de mort (?).
- Amélioration du débiaisage lorsque  $N_f$  est aléatoire.
- Ajout d'une fonction de test sur un échantillon.
- Déduire directement la méthode d'estimation à employer selon l'échantillon :



# Voilà !

Merci ! Des questions ?