

Modèles de mutations : étude probabiliste et estimation paramétrique
Package R `flan` (FLuctuation ANalysis)

A. Mazoyer^a, R. Drouilhet^b, S. Despréaux^c et B. Ycart^d

Laboratoire Jean Kuntzmann
Université Grenoble-Alpes
51 rue des Mathématiques
38041 Grenoble cedex , FRANCE

^aadrien.mazoyer@imag.fr, ^bremy.drouilhet@imag.fr, ^cstephane.despreaux@imag.fr, ^dbernard.ycart@imag.fr

Mots clefs : Cinétique cellulaire, modèle de mutation, analyse de fluctuation, estimation paramétrique.

1 Contexte

L'estimation des probabilités de mutation est d'une importance cruciale dans plusieurs domaines de la médecine et de biologie: cancer, tuberculose, microbiologie, etc. Une construction générale des modèles de mutation a été proposée dans [1]. Elle se décompose en trois niveaux : l'apparition de mutations aléatoires au cours d'un processus de croissance, la durée de développement de chaque clone issu des cellules mutantes et le nombre de cellules produites par un clone durant un temps de développement donné.

En pratique, les divisions cellulaires sont très nombreuses, et la probabilité de mutation est très faible, ce qui mène à une modélisation asymptotique. Le modèle classique de Luria-Delbrück représente le nombre final de mutantes observées comme la composée poissonnienne (nombres de mutations) d'un mélange exponentiel (durée de développement des clones) de lois géométriques (taille finale du clone). Cette loi dépend alors de deux paramètres, la probabilité individuelle de mutation π , et le rapport des taux de croissance des cellules normales et mutantes ρ , aussi appelé fitness. Le problème statistique consiste à estimer π et ρ au vu d'un échantillon de décomptes finaux de cellules mutantes. Il peut être résolu par la méthode du maximum de vraisemblance, par une méthode de moment exponentiel [1], ou par la méthode dite " p_0 " [4]. Cependant, les hypothèses de modélisation sous laquelle la distribution du nombre final de mutantes est explicite sont irréalistes : absence de morts cellulaires, durées de vie exponentielles, nombre final de cellules constant, etc. On obtient des estimations biaisées par rapport aux valeurs réelles.

Le package `flan` que nous développons permet d'effectuer des simulations et de l'inférence statistique dans le cas du modèle classique de Luria-Delbrück, mais également sous des hypothèses de modélisation plus générales : durées de vie non-exponentielles [2], morts cellulaires [3], variabilité du nombre final de cellules [4].

2 Fonctionnalités du package

Notre package est codé via le package Rcpp. Le package est composé de deux groupes principaux de fonctions :

1. **Distribution** : `rflan`, `pflan`, `dflan`, `qflan`

Ces fonctions permettent de : simuler (`rflan`) un échantillon de nombre finaux de mutants (couplé à un échantillon de nombres finaux de cellules, simulé selon une loi log-normale), calculer la fonction de répartition (`pflan`), la fonction de masse (`dflan`) et les quantiles (`qflan`) du nombre final de mutants. Chacune de ces fonctions fait appel aux paramètres suivants :

- `n` : taille de l'échantillon.
- `mutations` : le nombre moyen de mutation. Par défaut $1e-9$.
- `mutprob` : probabilité de mutation π . Par défaut $1e-9$.
- `fitness` : paramètre de fitness ρ . Par défaut 1.
- `mfn` : espérance du nombre final de cellules. Par défaut $1e9$.
- `cvfn` : coefficient de variation du nombre final de cellules. Par défaut 0.
- `dist` : loi des durées de vie des clones. Au choix parmi les 4 lois : Dirac ("`dir`"), exponentielle ("`exp`"), gamma ("`gamma`"), log-normale ("`lnorm`"). Par défaut "`lnorm`".
- `death` : probabilité de mort. Par défaut 0.

2. **Inférence** : `mutestim`, `flan.test`

La fonction `mutestim` permet d'estimer les paramètres d'intérêt, à savoir la probabilité de mutation π et le paramètre de fitness ρ . Elle renvoie également les estimations des écart-types (et les intervalles de confiance correspondant). On dispose de différentes méthodes d'estimation, selon différents types de données :

- `mc` : nombres finaux de mutantes.
- `fn` : nombres finaux de cellules. On peut également entrer `mfn` et `cvfn`, l'espérance et le coefficient de variation du nombre final de cellules.
- `method` : méthode d'estimation employée. Au choix parmi les 3 méthodes : max. de vraisemblance ("`ML`"), méthode p_0 ("`P0`"), méthode de moment exponentiel ("`GF`"). Par défaut "`ML`".
- `model` : hypothèse de la loi des durées de vie sous laquelle est effectuée l'estimation. Au choix parmi : modèle exponentiel ("`LD`") ou modèle Dirac ("`H`"). Par défaut "`LD`".
- `death` : probabilité de mort. Par défaut 0.

Les 3 méthodes citées donnent des estimateurs asymptotiquement normaux, ce qui permet de construire des tests d'hypothèses via la fonction `flan.test`.

Références

- [1] A. HAMON AND B. YCART, *Statistics for the Luria-Delbrück distribution*, Elect. J. Statist., 6 (2012), pp. 1251–1272.
- [2] B. YCART, *Fluctuation analysis: can estimates be trusted?*, PLoS One, 8 (2013).
- [3] ———, *Fluctuation analysis with cell deaths*, J. Appl. Probab. Statist., 9 (2014), pp. 13–29.
- [4] B. YCART AND N. VEZIRIS, *Unbiased estimates of mutation rates under fluctuating final counts*, PLoS One, 9 (2014).