Stage de M2 en Biostatistiques :

Estimation de fonction de survie à partir de données familiales.

Durée du stage : 6 mois.

Lieu du stage :

Laboratoire MAP5 - UMR CNRS 8145
Université Paris Cité
45 rue des Saints-Pères
75006 Paris

Encadrement :

Flora Alarcon,
flora.alarcon@u-pariscite.fr
MCF, MAP5 UMR CNRS 8145
Université Paris Cité

Lucas Ducrot
lucas.ducrot@univ-rouen.fr
MCF, Inserm UMR1245 - Cancer and Brain Genomics (CBG)
Université de Rouen Normandie

Contexte :

L'estimation de la pénétrance, c'est-à-dire du risque de développer une maladie en fonction de l'âge chez les porteurs d'une mutation prédisposante, est essentielle pour le conseil génétique et la prise en charge des personnes à risque. Elle repose notamment sur des données familiales, souvent caractérisées par des génotypes partiellement observés, des données censurées et des biais de recrutement liés aux critères de recensement des familles.

Plusieurs méthodes d'estimation ont été développées dans ce contexte. La Genotype Restricted Likelihood (GRL) est une méthode paramétrique adaptée aux familles recensées selon des critères familiaux [1]. Elle est notamment implémentée dans le package R `generisk` [2] et a été utilisée pour estimer les risques de cancer associés au syndrome de Peutz-Jeghers [3]. La Proband's phenotype Exclusion Likelihood (PEL) est une approche non paramétrique qui corrige le biais de recensement en excluant l'information phénotypique du proposant [4]. Une application à l'estimation de la pénétrance des maladies pulmonaires interstitielles et du cancer bronchopulmonaire chez les porteurs de variants pathogènes de *SFTPA1* ou *SFTPA2* est présentée dans [5]. Enfin, le package R `penetrance` propose une approche bayésienne permettant notamment d'intégrer des connaissances a priori issues de la littérature [6].

Les performances relatives de ces différentes approches dépendent potentiellement des caractéristiques des familles, du mécanisme de recensement et de la qualité des données disponibles. Une comparaison systématique par simulation permettrait de mieux caractériser leurs propriétés et d'identifier les situations dans lesquelles chacune est la plus adaptée.

Objectifs du stage :

L'objectif du stage est de mettre en place une étude de simulation comparant différentes méthodes d'estimation de la pénétrance, puis d'évaluer leur sensibilité aux hypothèses du modèle et aux caractéristiques des données.

Le stage s'organisera autour de trois objectifs :

1. Prise en main du sujet et des méthodes existantes. Comprendre les principes de l'estimation de la pénétrance à partir de pedigrees, notamment l'estimation de la fonction de survie, la prise en compte de la censure et la gestion des génotypes manquants. Étudier les différentes méthodes disponibles, leurs hypothèses et leurs modalités de correction du biais de recensement, en particulier à travers les packages R `generisk` et `penetrance`.
2. Mise en place de simulations réalistes. Développer un cadre de simulation de pedigrees reproduisant différents scénarios de recensement des familles. Les simulations pourront faire varier les critères conduisant au recrutement des familles.
3. Comparaison des méthodes et études de sensibilité. Évaluer les performances des méthodes dans les différents scénarios simulés.

Profil recherché :

Le candidat devra être étudiant en Master 2 ou en école d'ingénieur, avec une formation en statistiques ou biostatistique. Des connaissances en analyse de survie sont souhaitées. Des connaissances en génétique statistique, en algorithmes EM ou en méthodes Monte-Carlo par chaînes de Markov (MCMC) seraient appréciées.

Contact : Si vous êtes intéressé(e), envoyez un CV détaillé et une lettre de motivation à `flora.alarcon@u-pariscite.fr` et/ou `lucas.ducrot@univ-rouen.fr`

Références

[1] Bonaïti B., Bonadona V., Perdry H., Andrieu N., Bonaïti-Pellié C. (2011). Estimating penetrance from multiple case families with predisposing mutations : extension of the 'genotype-restricted likelihood' (GRL) method. *European Journal of Human Genetics*, 19, 173–179. doi:10.1038/ejhg.2010.158.

[2] Drouet Y. `generisk` : package R pour l'estimation du risque génétique. `https://github.com/youenndrouet/generisk`.

[3] Rémond M., Drouet Y. et al. (2026). Follow-up, cancer risk and mortality in Peutz-Jeghers syndrome : Data from the PRED-IdF network. *European Journal of Human Genetics*. doi:10.1038/s41431-026-02116-8.

[4] Alarcon F., Planté-Bordeneuve V., Olsson M., Nuel G. (2018). Non-parametric estimation of survival in age-dependent genetic disease and application to the transthyretin-related hereditary amyloidosis. *PLOS ONE*.

[5] Ducrot L., Nathan N., Benusiglio P. R., Borie R., Nuel G., Legendre M. (2025). Penetrance of interstitial lung disease and lung cancer in carriers of *SFTPA1* or *SFTPA2* pathogenic variants. *ERJ Open Research*, 11(3), 01348-2024. doi:10.1183/23120541.01348-2024.

[6] Kubista N., Braun D., Parmigiani G. (2025). The penetrance R package for estimation of age specific risk in family-based studies. *Bioinformatics Advances*.