

# Mathilde Sautreuil

35 avenue du général de Gaulle  
Bâtiment A1  
91140 Villebon sur Yvette  
☎ 06 19 97 91 24  
✉ [mathilde.sautreuil@gmail.com](mailto:mathilde.sautreuil@gmail.com)  
📄 [mathildesautreuil.github.io](https://github.com/mathildesautreuil)

## Formation

- 2017-2021 **Doctorat en Mathématiques Appliquées**, MICS, CentraleSupélec, Université Paris-Saclay.  
2015-2017 **Master 2 Bioinformatique en alternance**, Mention Très Bien, Normandie Université, Université de Rouen Normandie.  
2014-2015 **Master 1 Bioinformatique**, Normandie Université, Université de Rouen Normandie.  
2011-2014 **Licence de Mathématiques**, Normandie Université, Université de Rouen Normandie.

## Expériences professionnelles

- 2020- **Ingénieure de recherche en Biostatistiques**, sous la direction de Guillem Rigaill et Etienne Delannoy  
*LaMME, Équipe Stat et Génome - IPS2, Équipe Gnet*, Université Paris-Saclay.  
Le sujet concerne la détection d'empreintes PPR à partir de méthodes de Machine Learning chez l'*Arabidopsis Thaliana*.
- 2017-2020 **Doctante en Biostatistiques**, sous la direction du Pr Paul-Henry Cournède et du Dr Sarah Lemler  
*Biomathematics - Laboratoire MICS*, CentraleSupélec.  
Le sujet concerne l'analyse de survie et la détection de marqueurs en oncologie dont les principaux objectifs sont :
  - l'étude des méthodes de sélection de variables et de régularisation pour détecter les biomarqueurs dans le cas du cancer du rein;
  - l'étude et l'adaptation pour la grande dimension de réseaux de neurones pour la prédiction de la durée de survie.
- 2015-2017 **Apprentie en Biostatistiques**, sous la direction du Dr Caroline Bérard et du Dr Nicolas Vergne  
21 mois *LITIS EA 4108, Équipe TIBS - LMRS UMR 6085 CNRS, Équipe Statistique*, Rouen.  
Les objectifs de mon apprentissage consistait en le développement :
  - d'une nouvelle méthode statistique pour l'analyse de données RNA-seq;
  - d'un package R pour les modèles de Markov et semi-Markov;
  - d'une interface web automatisée Shiny pour le traitement de données qRT-PCR.
- 2015 **Stagiaire en Biostatistiques**, sous la direction du Dr Caroline Bérard  
4 mois *Équipe TIBS, Laboratoire d'Informatique, du Traitements de l'Information et des Systèmes (LITIS) - EA 4108*, Rouen.  
Développement d'une nouvelle méthode statistique pour l'analyse de données RNA-seq.

## Activités de recherche

### Publications

- Mathilde Sautreuil, Sarah Lemler, Paul-Henry Cournède, *Benchmarking the Stability of Variable Selection Methods in the Cox Model*, en préparation.
- Mathilde Sautreuil, Sarah Lemler, Paul-Henry Cournède, *Neural networks for survival analysis*, en préparation.
- Diana Tronik-Le Roux, Mathilde Sautreuil, Mahmoud Bentrion, et al., *Comprehensive landscape of immune-checkpoints uncovered in clear cell renal cell carcinoma reveals new and emerging therapeutic targets*, Cancer Immunology, Immunotherapy, 2020.
- Vlad Stefan Barbu, Caroline Bérard, Dominique Cellier, Mathilde Sautreuil, Nicolas Vergne. *SMM: An R Package for Estimation and Simulation of Discrete-time semi-Markov Models*, The R Journal, Norman Matloff, 2018.

### Acte de conférences

- Mathilde Sautreuil, Sarah Lemler, Paul-Henry Cournède. *Study of neural networks to predict the survival in oncology*, StatMod, 2020, Bucharest.
- Mathilde Sautreuil, Sarah Lemler, Paul-Henry Cournède. *Réseaux de neurones pour l'analyse de survie en grande dimension*, JDS, 2019, Nancy.
- Mathilde Sautreuil, Caroline Bérard, Gaëlle Chagny, Antoine Channarond, Angelina Roche, Nicolas Vergne. *Modèle de mélange binomial négatif bivarié pour l'analyse de données RNA-Seq*, JDS, 2018, Paris.
- Vlad Stefan Barbu, Caroline Bérard, Dominique Cellier, Mathilde Sautreuil, Nicolas Vergne. *Estimation paramétrique des chaînes semi-markoviennes pour des données censurées*, JDS, 2017, Avignon.

### Communications orales

- 2020 StatMod (Statistical Modeling with Applications), Bucharest (Romania).  
2019 50èmes Journées de Statistique (JDS2019), Nancy.

- 2018 49èmes Journées de Statistique (JDS2018), Paris-Saclay.
- 2017 48èmes Journées de Statistique (JDS2017), Avignon.
- 2017 Journée NormaSTIC (axe Données, Apprentissage, Connaissance), Rouen.

### Posters

- 2018 Journée de l'école doctorale (ED 573), Versailles.
- 2017 Journées Ouvertes Biologie, Informatique et Mathématiques (JOBIM2017), Lille.

## Responsabilités

- 2018- Membre du bureau élargi du groupe "Jeunes" de la SFdS
- 2018-2020 Responsable doctorants du laboratoire MICS
- 2018-2019 Projet BOUM - Séminaire IA (Intelligence Artificielle)
- 2019 Congrès MATH.en.JEANS 2019 - Saclay

## Développement d'outils

- Package R **survMS: an R package for simulation of survival data from Cox, AFT and AH models**, <https://gitlab-research.centralesupelec.fr/2017sautreuim/survms>.  
Ce package permet de simuler des données à partir de différents modèles de survie (modèle de Cox, de durée de vie accélérée (AFT) et des risques accélérés (AH)). L'utilisation de différents modèles de survie pour la simulation des données est intéressante car les hypothèses associées à ces modèles ne sont pas les mêmes. En effet, le modèle de Cox est un modèle de risque proportionnel. Dans le modèle AFT, les variables auront un effet d'accélération ou de décélération sur la survie des individus. Malgré cela, les courbes des fonctions de survie d'un modèle AFT ne se croisent jamais comme dans le modèle Cox. Le modèle AH est plus flexible que les deux autres modèles mentionnés, les variables vont accélérer ou décélérer le risque instantané de décès. Les courbes de survie du modèle AH peuvent donc se croiser. La génération des temps de survie est effectuée à partir des différents modèles mentionnés ci-dessus, où la fonction de risque de base des modèles est supposée être connue et suit une certaine distribution de probabilité.
- Package R **SMM: An R Package for Estimation and Simulation of Discrete-time semi-Markov Models**, <https://cran.r-project.org/web/packages/SMM/>.  
Ce package performe l'estimation et la simulation paramétrique et non paramétrique pour les processus de semi-Markov en temps discret à plusieurs états. Pour l'estimation paramétrique, plusieurs distributions discrètes sont prises en compte pour les temps de séjour : Uniforme, géométrique, de Poisson, Weibull discrète et binomiale négative. L'estimation non-paramétrique concerne les distributions des temps de séjour, où aucune hypothèse n'est faite sur la forme des distributions. De plus, l'estimation peut être faite sur la base d'un ou plusieurs échantillons, avec ou sans censure au début et/ou à la fin des échantillons. Les méthodes mises en œuvre sont décrites dans les publications suivantes doi:10.1007/978-0-387-73173-5, doi:10.1080/10485250701261913 et doi:10.1080/10485252.2011.555543. L'estimation et la simulation de chaînes de Markov d'ordre k en temps discret sont également prises en compte.
- Application web Shiny **PROqPCR: PROcessing of qRT-PCR data**, <https://qpcrapp.shinyapps.io/proqpcr/>.  
Cette application permet aux biologistes d'effectuer facilement le traitement des données de la qRT-PCR. L'utilisateur fournit uniquement le plan d'expérience et les fichiers créés par le logiciel de l'instrument de PCR quantitative et trois graphiques différents sont proposés et peuvent être téléchargés en format png (Barplot de toutes les conditions pour un gène donné, Barplot de tous les gènes pour une condition donnée, Comparaison des barplot pour des gènes sélectionnés dans toutes les conditions). De plus, deux graphiques sont également fournis afin de faciliter la comparaison entre les résultats de la qRT-PCR et ceux d'une expérience de séquençage de l'ARN.

## Enseignements

- 2018-2021 **Enseignante vacataire**, M2 Bioinformatique, Université de Rouen, Normandie Université.  
Création de supports :
  - Cours Introduction à Shiny (Framework d'applications web R)/TP de Shiny (R)
  - Cours et TP : Rappels des tests statistiques dans le cadre de GSEA (*Gene Set Enrichment Analysis*)
- 2019 **Enseignante vacataire**, CentraleSupélec, Université Paris-Saclay.
  - TP de Machine learning en python
- 2018-2019 **Enseignante vacataire**, DUT GEA (*Gestion des Entreprises et des Administrations*), IUT de Sceaux, Université Paris-Sud.
  - TD de mathématiques (Étude de fonctions, matrices)
  - TD/TP de statistiques descriptives
  - TD de probabilités (Modèle probabiliste, variables aléatoires)
- 2017-2018 **Enseignante vacataire**, CentraleSupélec, Université Paris-Saclay.
  - TD/TP de Modélisation du Vivant en troisième année
  - Encadrements de projets en première année
- 2017 **Enseignante vacataire**, Université de Rouen, Normandie Université.
  - TD de Biosstatistiques en L2 SVT (Sciences de la Vie et de la Terre)
  - TP de Statistiques en L3 EBO (Écologie et Biologie des Organismes)