

Master 2 Internship Offer

Nonparametric hazard estimation in cure models and application to intensive care data



Start Date : February/march 2026

Duration : 6 months

Location : Montpellier (France) IDESP

Context

When analyzing time-to-event data, it often happens that a certain fraction of the population will never experience the event of interest. These individuals are considered "cured", and their event times are effectively infinite. In this context, the population is a mixture of two sub-populations, a cured and an uncured one, and standard survival models, which assume that all individuals will eventually experience the event, are not adequate anymore. Instead, specialized models known as *cure models* are used to account for the presence of this cured fraction and enable the estimation of two quantities, the survival function and the cure fraction. Time-to-event data are often analyzed through the hazard function instead of the survival, which corresponds to the instantaneous risk of experiencing the event of interest. However, very few have been made on this aspect in the context of cure models, particularly in a nonparametric setting. Therefore, the main objective of this internship is to study and develop nonparametric estimation methods for the hazard function in the presence of a cure fraction.

Internship work plan

The project will be structured in the following stages. The first step will consist of a comprehensive review of the existing literature on nonparametric cure models. Particular attention will be paid to methods that estimate the survival function and/or the cure fraction without relying on parametric assumptions (see for example Taylor, 1995; López-Cheda et al., 2017a and 2017b; Patilea and Van Keilegom, 2020; López-Cheda et al., 2023). The second part will involve understanding and analyzing projection-based estimators in duration (survival) models. These estimators allow for flexible and data-driven inference and have shown optimality properties in various contexts (see Placade, 2011; Brunel and Comte, 2021).

Based on the previous steps, the goal will be to propose a novel projection-based method to estimate the instantaneous hazard function (or hazard rate) in a cure model framework. The core of the project will involve adapting a suitable *projection contrast* to the cure model setting. The newly developed estimator will be implemented and applied to both simulated and real-world data. Application on intensive care data, where the presence of cured individuals is common – such as when studying the time to re-intubation, with a fraction of patients never being re-intubated – would represent an ideal use case for this method. Its

empirical performance will be assessed and compared to existing approaches from the literature.

This internship will provide the student with a strong foundation in modern survival analysis, hands-on experience with statistical modeling, and practical skills in data analysis and numerical computation. The work may lead to a publication or contribute to ongoing research in the field of nonparametric inference in survival analysis.

Skills/Experience

- You are a MSc or Engineer student in Applied Mathematics or Statistics, currently enrolled in a full-time degree program
- You are motivated to apply recent statistical methods for biomedical data.
- You have experience with R and/or Python programming languages.

Practical Aspect

The internship will take place at IDESP Campus Balard Bâtiment 39 (avenue du Professeur Émile Jeanbrau, Montpellier) under the supervision of **Maïlis AMICO (IDESP)** and **Élodie BRUNEL-PICCININI (IMAG)**.

You can apply for this internship position by sending your CV and Cover letter to the following email addresses: mailis.amico@umontpellier.fr ; elodie.brunel-piccinini@umontpellier.fr

Bibliography:

Amico M. and Van Keilegom I., (2018) [Cure Models in Survival Analysis](#), *Annual Review of Statistics and Its Applications*, 5, 311-342.

Brunel E. and Comte F., (2021) [Hazard regression with non-compactly supported bases](#), *The Canadian Journal of Statistics*, 49, 4, 1273-1297.

López-Cheda A., Cao R., Jácome MA., and Van Keilegom I. (2017a) [Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure model](#), *Computational Statistics and Data Analysis*, 105, 144-165.

López-Cheda A., Jácome MA., and Cao R. (2017b). [Nonparametric latency estimation for mixture cure models](#), *TEST*, 26, 353-376.

López-Cheda A., Peng, Y., and Jácome MA. (2023) Nonparametric estimation in mixture cure models with covariates, *TEST*, 32, 467-495.

Patilea V. and Van Keilegom I. (2020). [A general approach for cure models in survival analysis](#), *The Annals of Statistics*, 48, 4, 2323-2346.

Plancade, S. (2011) Model selection for hazard rate estimation in presence of censoring. *Metrika*, 74, 313-347

Taylor JMG. (1995). [Semi-parametric estimation in failure time mixture cure model](#), *Biometrics*, 51, 899-907.

Offre de stage de Master 2

Estimation non paramétrique de la fonction de hasard dans les modèles de guérison et application aux données de soins intensifs



Début : Février ou mars 2026

Durée : 6 months

Lieu : Montpellier (France) IDESP

Gratification : selon la tarification en vigueur au (environ 600 €/mois)

Contexte

En analyse de survie, les données sont des durées jusqu'à la survenue d'un événement (par exemple le décès), et il arrive souvent qu'une certaine proportion de la population ne connaisse jamais l'événement d'intérêt. Ces individus sont considérés comme « guéris », et leurs durées d'événement sont effectivement infinies. Dans ce contexte, la population est un mélange de deux sous-populations — l'une guérie et l'autre non guérie — et les modèles de survie standards, qui supposent que tous les individus connaîtront éventuellement l'événement, ne sont alors plus adaptés. À la place, on utilise des modèles spécifiques appelés modèles de guérison (*cure models*), qui tiennent compte de la présence d'individus qui ne connaîtront jamais l'événement (guérison) et permettent d'estimer deux quantités : la fonction de survie et la proportion de guérison. Les données de durées de vie sont souvent analysées à travers la fonction de hasard (*hazard function*) plutôt que la fonction de survie, car elle correspond au risque instantané de survenue de l'événement d'intérêt. Cependant, très peu de travaux ont été réalisés sur cet aspect dans le cadre des modèles de guérison, en particulier dans un cadre non paramétrique. Ainsi, l'objectif principal de ce stage est d'étudier et de développer des méthodes d'estimation non paramétriques de la fonction de risque en présence d'une proportion de guérison.

Plan de travail du stage

Le projet sera structuré selon les étapes suivantes. La première étape consistera en une revue de la littérature existante sur les modèles de guérison non paramétriques. Une attention particulière sera portée aux méthodes permettant d'estimer la fonction de survie et/ou la proportion de guérison sans recourir à des hypothèses paramétriques (par exemple Taylor, 1995 ; López-Cheda *et al.*, 2017a et 2017b ; Patilea et Van Keilegom, 2020 ; López-Cheda *et al.*, 2023).

La seconde partie portera sur la compréhension et l'analyse des estimateurs par projection dans les modèles de durée (ou de survie). Ces estimateurs permettent une inférence flexible et guidée par les données, et présentent des propriétés d'optimalité dans divers contextes (Plancade, 2011 ; Brunel et Comte, 2021).

Sur la base de ces travaux préliminaires, l'objectif sera de proposer une nouvelle méthode d'estimation par projection de la fonction de risque instantané (ou de hasard) dans le cadre d'un modèle de guérison. Le cœur du projet consistera à adapter un contraste de projection approprié à ce contexte spécifique. Le nouvel estimateur sera ensuite implémenté et appliqué à la fois sur des données simulées et sur des données réelles. Une application sur des données de soins intensifs, où la présence d'individus guéris est fréquente — par exemple lors de l'étude du temps jusqu'à la ré-intubation, avec une fraction de patients n'étant jamais réintubés — constituerait un cas d'utilisation idéal pour cette méthode. Ses performances empiriques seront évaluées et comparées aux approches existantes issues de la littérature. Ce stage offrira à l'étudiant(e) une solide formation en analyse de survie moderne, une expérience pratique en modélisation statistique, ainsi que des compétences appliquées en analyse de données et implémentation numérique. Le travail pourra donner lieu à une publication scientifique ou contribuer à des recherches en cours dans le domaine de l'inférence non paramétrique en analyse de survie.

Compétences / Profil recherché

- Vous êtes étudiant(e) en Master 2 ou en école d'ingénieur, spécialisé(e) en mathématiques appliquées ou statistiques.
- Vous êtes motivé(e) pour appliquer des méthodes statistiques modernes à des données biomédicales.
- Vous avez une expérience en programmation avec les langages R et/ou Python.

Aspects pratiques

Le stage se déroulera à l'IDESP, Campus Balard – Bâtiment 39 (avenue du Professeur Émile Jeanbrau, Montpellier), sous la supervision de **Mailis AMICO (IDESP)** et **Élodie BRUNEL-PICCININI (IMAG)**.

Vous pouvez postuler à ce stage en envoyant votre CV et votre lettre de motivation aux deux encadrantes : mailis.amico@umontpellier.fr ; elodie.brunel-piccinini@umontpellier.fr

Bibliography:

Amico M. and Van Keilegom I., (2018) [Cure Models in Survival Analysis](#), *Annual Review of Statistics and Its Applications*, 5, 311-342.

Brunel E. and Comte F., (2021) [Hazard regression with non-compactly supported bases](#), *The Canadian Journal of Statistics*, 49, 4, 1273-1297.

López-Cheda A., Cao R., Jácome MA., and Van Keilegom I. (2017a) [Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure model](#), *Computational Statistics and Data Analysis*, 105, 144-165.

López-Cheda A., Jácome MA., and Cao R. (2017b). [Nonparametric latency estimation for mixture cure models](#), *TEST*, 26, 353-376.

López-Cheda A., Peng, Y., and Jácome MA. (2023) Nonparametric estimation in mixture cure models with covariates, *TEST*, 32, 467-495.

Patilea V. and Van Keilegom I. (2020). [A general approach for cure models in survival analysis](#), *The Annals of Statistics*, 48, 4, 2323-2346.

PlanCADE, S. (2011) Model selection for hazard rate estimation in presence of censoring. *Metrika*, 74, 313-347

Taylor JMG. (1995). [Semi-parametric estimation in failure time mixture cure model](#), *Biometrics*, 51, 899-907.