

AVIS DE SOUTENANCE DE THÈSE

Madame Margaux THOREZ est autorisée à présenter ses travaux en vue de l'obtention du diplôme national de DOCTORAT délivré par l'école CENTRALE de MARSEILLE

Le mercredi 11 décembre 2024 à 10h00

Lieu : Salle Frumam - séminaire 3ème étage, Aix-Marseille Université - Site de Saint Charles (Bâtiment 7)
3 place Victor Hugo 13003 Marseille

Titre : **Développement et étude de nouveaux tests statistiques d'adéquation et d'homogénéité pour des données incomplètes en actuariat.**

Ecole doctorale : ED 184 - Mathématiques et Informatique de Marseille

Spécialité : Mathématiques

Composition du jury :

M. CHRISTOPHE POUET	Centrale Méditerranée, Institut de Mathématiques de Marseille UMR 7373	Directeur de thèse
Mme Ghislaine GAYRAUD	Université Technologique de Compiègne, Laboratoire de Mathématiques Appliquées de Compiègne EA 2222	Rapporteuse
M. Christophe DENIS	Paris 1 Panthéon-Sorbonne, Laboratoire Statistique, Analyse et Modélisation Multidisciplinaire UMR 4543	Rapporteur
M. Denys POMMERET	Aix Marseille Université - Institut de Mathématiques de Marseille UMR 7373	Examineur
M. Eric MATZNER-LOBER	Rennes 2 et ENSAE-ENSAI, Centre de recherche en économie et statistique UMR 9194	Examineur
Mme Maud THOMAS	Université Lyon 1, Institut de Sciences Financière et d'Assurances, Laboratoire de Sciences Actuarielle et Financière	Examinatrice

Résumé (FR)

Nous considérons un problème en sciences actuarielles, dans lequel nous avons M compagnies d'assurance (échantillons) et K groupes d'assurés. Nous voulons modéliser les arrêts de travail pour chaque groupe. Cependant, le label du groupe de chaque observation n'est pas défini avec certitude pour des raisons de confidentialité. Nous supposons que chaque compagnie d'assurance peut fournir la proportion de chaque groupe dans sa base de données. Dans ce cas, nous allons comparer deux estimateurs de Kaplan-Meier différents pour les modèles de mélange. Nous mesurons la perte d'information lors de la modélisation des arrêts de travail avec des modèles de mélange à poids variables dans des modèles de durée censurés, en comparant les méthodes de Ryzhov et de Maiboroda. Nous utilisons à la fois des simulations suivant différentes lois et des données empiriques. Ces deux estimateurs

sont identiques dans le cas sans censure, mais peuvent être différents dans le cas avec censure. Le choix des poids a par ailleurs un impact sur la qualité de l'estimation. Nous avons également appliqué nos modèles de mélange à poids variables au calcul de différents quantiles et aux M-estimateurs. Nous nous sommes intéressés ensuite aux tests d'homogénéité et d'adéquation, comme le test de Kolmogorov-Smirnov, qui peuvent être effectués avec ces modèles. L'objectif est d'avoir des outils permettant de mieux maîtriser les enjeux liés aux arrêts de travail. En particulier, nous utilisons les M-estimateurs pour réaliser des tests d'homogénéité sur les moyennes et médianes des différents groupes d'assurés, et pour réaliser un test de conformité de la moyenne.

Mots-clés : modèles de mélange, modèles de durée, données censurées, poids variables, arrêts de travail, tests statistiques

Abstract (EN)

We consider a problem in actuarial sciences, in which we have M insurance companies (samples) and K groups of insured persons. We want to model sick leaves for each group. However, the group label of each observation is not defined with certainty for confidential reasons. We assume that each insurance company can provide the proportion of each group in its database. In this case, we are going to compare two different Kaplan-Meier estimators for mixture models. We measure the loss of information when modeling sick leaves with variable-weight mixture models in censored duration models, comparing Ryzhov's and Maiboroda's methods. We use both simulations following different laws and empirical data. Both estimators are identical in the uncensored case, but may be different in the censored case. The choice of weights also has an impact on the quality of the estimate. We have also applied our varying weight mixture models to the calculation of different quantiles and to M-estimators. We then looked at homogeneity and adequacy tests, such as the Kolmogorov-Smirnov test, which can be performed with these models. The aim is to provide tools for better management of the issues associated with sick leaves. In particular, we use M-estimators to perform homogeneity tests on the means and medians of the different groups of insureds, and to perform a mean conformity test.

Keywords: mixture models, duration models, censored data, varying weights, sick leaves, statistical testing