

Proposition de thèse en statistique mathématique en co-tutelle France-Canada

Titre : Tests de détection de rupture dans la dépendance multivariée.

Financement : Co-financement franco-canadien.

Laboratoires d'accueil : Laboratoire de mathématiques et de leurs applications, université de Pau et des Pays de l'Adour, France (<http://lma-umr5142.univ-pau.fr>) et département de mathématiques et d'informatique, université du Québec à Trois-Rivières, Canada (<http://dmi.uqtr.ca>).

Durée et montant de l'allocation : 3 ans (2011-2014), environ 1400 Euros net par mois, plus possibilité de monitorat et de missions industrielles.

Déroulement probable : 1 an en France, 1 an au Québec et 1 an en France.

Pré-requis : Master recherche en mathématiques appliquées, statistique ou similaire, grande école d'ingénieur.

Mots-clés : détection de rupture, dépendance, tests non paramétriques, ré-échantillonnage, processus empiriques, statistique computationnelle, package R.

Contexte scientifique du projet de thèse : Dans de nombreux domaines comme la fiabilité, l'hydrologie ou la finance, l'objectif est de modéliser un phénomène aléatoire à partir de l'observation répétée de ce phénomène. Nous nous intéressons ici au cas standard où le phénomène aléatoire en question est représenté par un vecteur aléatoire X dont les caractéristiques sont inconnues. Un problème classique en statistique est alors d'estimer la fonction de répartition (f.d.r.) F inconnue de X à partir d'un échantillon indépendant et identiquement distribué (i.i.d.) X_1, \dots, X_n de même f.d.r. que X . En pratique, il est souvent nécessaire de vérifier que les données disponibles pour la modélisation sont bien la réalisation d'un échantillon i.i.d., en particulier lorsque les caractéristiques du phénomène aléatoire observé peuvent évoluer au cours du temps, comme en hydrologie, fiabilité ou en finance. L'indépendance entre observations peut être testée à l'aide de tests d'indépendance sérielle, les plus connus étant fondés sur la fonction d'autocorrélation. S'il n'y a pas d'indications de violation de l'indépendance sérielle, la prochaine étape consiste à "vérifier" que les observations proviennent bien de la même distribution. Cette "vérification" peut être effectuée à l'aide de tests de détection de rupture, qui sont au coeur de ce projet de thèse. Dans leur forme classique, ces techniques statistiques permettent de tester l'hypothèse d' "absence de rupture"

$$H_0 : X_1, \dots, X_n \text{ ont la même f.d.r. } F$$

contre l'hypothèse de "présence d'un point de rupture"

$$H_1 : \text{il existe } k \text{ dans } \{1, \dots, n-1\} \text{ tel que } X_1, \dots, X_k \text{ ont pour f.d.r. } F \\ \text{et } X_{k+1}, \dots, X_n \text{ ont pour f.d.r. } G \neq F.$$

Objectifs du projet : Dans le cadre de ce projet, nous souhaitons développer des tests de détection de rupture essentiellement *non paramétriques*, pour données *multivariées*, permettant en particulier de détecter des changements dans la dépendance. Pour cela, nos travaux graviteront naturellement autour de la notion de *copule*, concept fondamental permettant d’isoler la dépendance des comportements marginaux. Nous chercherons non seulement à développer rigoureusement la théorie des tests mentionnés, mais également à étudier leur comportement empirique dans le cadre d’applications réelles issues du monde de la fiabilité, de l’hydrologie ou de la finance.

Plus précisément, soit $Y = (Y_1, \dots, Y_d)$ un vecteur aléatoire d -dimensionnel de f.d.r. F et de f.d.r. marginales continues F_1, \dots, F_d . Il existe alors une unique f.d.r. C sur $[0, 1]^d$ dont les marges sont uniformes, appelée *copule*, telle que

$$F(y_1, \dots, y_d) = C\{F_1(y_1), \dots, F_d(y_d)\}, \quad y_1, \dots, y_d \in \mathbb{R}.$$

La représentation précédente permet d’isoler la dépendance entre les composantes Y_1, \dots, Y_d de Y , caractérisée par la copule C , des comportements marginaux caractérisés par les f.d.r. marginales F_1, \dots, F_d . Dans le contexte de la détection de rupture fondée sur un échantillon d -dimensionnel X_1, \dots, X_n dont les composantes sont des variables aléatoires continues, il est ainsi possible d’associer une unique copule C_i à chaque X_i . Les tests que nous développerons seront ainsi particulièrement sensibles à l’hypothèse

$$H'_1 : \text{il existe } k \text{ dans } \{1, \dots, n-1\} \text{ tel que } X_1, \dots, X_k \text{ ont pour copule } C \\ \text{et } X_{k+1}, \dots, X_n \text{ ont pour copule } C' \neq C.$$

Ils seront essentiellement fondés sur la comparaison de copules empiriques (estimateurs non paramétriques des copules inconnues) et sur la comparaison de mesures de dépendance, comme le tau de Kendall ou le rho de Spearman. Afin de mettre ces tests en œuvre, nous serons amenés à utiliser des techniques de ré-échantillonnage. Ces travaux seront accompagnés d’études de Monte Carlo poussées permettant de comparer les diverses approches proposées et de fournir des recommandations aux utilisateurs.

Encadrants : Ivan Kojadinovic, professeur au laboratoire de mathématiques et de leurs applications, université de Pau et des Pays de l’Adour, France, et Jean-François Quessy, professeur au département de mathématiques et d’informatique, université du Québec à Trois-Rivières, Canada.

Contact : Ivan Kojadinovic. Email : ivan.kojadinovic@univ-pau.fr. Tél : + 33 (0)5 59 40 71 33.