Actuariat et data mining: prise en compte des dépendances

Arthur Charpentier

ENSAE-CREST, France
arthur.charpentier@ensae.fr

Résumé. Cet article vise à présenter un outil central en modélisation et en représentation de la dépendance entre risques: les copules. La corrélation n'ayant pas de motivation en assurance, il convient de pouvoir utiliser des notions plus générales, et nous verrons que les rangs sont en particulier des notions de première importance. Nous évoquerons également les problèmes de corrélations entre évènements extrêmes, particulièrement importants en gestion des risques.

1 Introduction et motivation

Les problèmes de dépendance et de corrélation en assurance ne sont pas nouveaux. On prendra d'ailleurs le terme "*corrélation*" au sens littéraire du terme, c'est à dire un "*rapport de dépendance dû à un lien de cause à effet, ou un lien créé par une cause commune, déterminée ou non*". En particulier on le distinguera de la notion de "*mesure de corrélation au sens de Pearson*", correspodant au coefficient (usuel) de corrélation (dont les propriétés seront évoquées dans la Section. 2).

Frees et Valdez (1995) avait ainsi proposé une revue de la littérature sur les copules, et en particulier les copules archimédiennes, en présentant des applications en assurance, avec en particulier les "*corrélations*" entre les durées de vie dans un couple et l'impact sur la tarification des primes sur deux têtes (assurance au dernier survivant, ou assurance veuvage), ou bien sur la tarification de contrats de réassurance, avec la séparation entre les coûts des sinistres (avec un traité non-proportionnel avec franchise) et les frais associées (partagés *pro rata capita*).

Dans le premier cas, si T_x et T_y désignent les durées de vie résiduelles d'un homme et d'une femme respectivement d'âge x et y à la signature du contrat, on peut ainsi écrire les annuités (c'est à dire les primes pour le versement d'un capital unitaire à terme échu) vie-jointes sous la forme

$$a_{xy:n\urcorner} = \sum_{k=1}^{n} v^k \mathbb{P}(T_x > k \text{ et } T_y > k) = \sum_{k=1}^{n} v^k {}_kp_{xy},$$

en utilisant les notations classiques d'assurance-vie, où ${}_kp_x = \mathbb{P}(T_x > k)$ désigne la probabilité qu'une personne d'âge x aujourd'hui soit encore en vie dans k années. Les annuités au dernier survivant s'écrivent sous la forme

$$a_{\overline{xy}:n\urcorner} = \sum_{k=1}^{n} v^k \mathbb{P}(T_x > k \text{ ou } T_y > k) = \sum_{k=1}^{n} v^k {}_kp_{\overline{xy}},$$