
Tests d'adéquation basés sur le maximum d'entropie : le package MaxEntGOFTest

Justine Lequesne*¹ and Philippe Regnault²

¹Laboratoire de Mathématiques Nicolas Oresme (LMNO) – CNRS : UMR6139, Université de Caen
Basse-Normandie – BP 5186 14032 Caen Cedex, France

²Laboratoire de Mathématiques (LM-Reims) – CNRS : UMR6056, Université de Reims - Champagne
Ardenne – France

Résumé

Dans la littérature, des tests d'adéquation utilisant l'entropie de Shannon ou la divergence de Kullback-Leibler, concepts fondamentaux de la théorie de l'information, ont été construits pour la plupart des lois les plus usuelles. Le premier est le test de normalité de Vasicek [1] qui utilise la propriété de maximum d'entropie de la loi normale parmi les lois à espérance et variance fixées pour proposer comme statistique de test une différence d'entropies. Plus récemment, Song [2] utilise directement la divergence de Kullback-Leibler pour des tests d'adéquation à des lois réelles très générales.

Le lien entre ces deux techniques est démontré dans Lequesne [3] grâce à une égalité de Pythagore appliquée aux lois à maximum d'entropie, impliquant une unique procédure de tests d'adéquation à de nombreuses familles de lois.

Dans ce poster, nous présentons le package MaxEntGOFTest destiné à appliquer ces tests. En particulier, ce package contient une fonction estimant l'entropie de Shannon d'un échantillon par l'estimateur introduit par Vasicek [1] et une fonction permettant de tester l'adéquation d'un échantillon aux familles de lois uniformes, normales, exponentielles, Gamma, Weibull, Pareto, Fisher, Laplace et Beta. Une application à des données réelles issues d'une étude portant sur la réplication d'ADN est également présentée. [1] Vasicek, O. (1976) A test of normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* V.38(1), p 54-59.

Song, K. S. (2002) Goodness-of-fit tests based on Kullback-Leibler discrimination information. *Information Theory, IEEE Transactions on*, V.48(5), p 1103-1117.

Lequesne, J. (2015) Tests statistiques basés sur la théorie de l'information, applications en biologie et en démographie. Mémoire de thèse de l'Université de Caen.

*Intervenant