

---

## USING EXTREME VALUE THEORY TO TEST FOR OUTLIERS

Nathaniel GBENRO

*École Nationale Supérieure de Statistique et d'Économie Appliquée, Abidjan, Côte d'Ivoire*  
*Université Cergy-Pontoise*

[ngbenro@gmail.com](mailto:ngbenro@gmail.com)

**Mots-clés** : théorie des valeurs extrêmes, outliers, test, hypothèses

---

### Résumé

L'importance de l'identification des valeurs aberrantes dans un échantillon de données a été longuement abordée dans la littérature. Plusieurs tests d'identification des outliers à l'image du célèbre test de Grubb (ou encore de Pearson) sont proposés dans la littérature. Ces tests, pour la plupart, reposent sur l'hypothèse de la normalité du processus générateur des données (DGP). Toutefois, cette hypothèse de normalité des données n'est pas toujours vérifiée en pratique, d'où l'importance de pouvoir identifier les valeurs aberrantes dans les cas de distributions anormales des données. Cet article teste l'identification des valeurs aberrantes en utilisant une approche novatrice basée sur la théorie des valeurs extrêmes. L'objectif de ce papier est de proposer un algorithme permettant d'identifier des valeurs aberrantes dans des échantillons de grandes tailles issus d'une distribution non normale ou même normale. Pour ce faire, nous nous basons sur une statistique semblable à celle proposée par Pearson et montrons à partir de la théorie des valeurs extrêmes la possibilité d'obtenir sa loi.

Notre démarche a consisté à montrer dans un premier temps l'échec du test de Grubb lorsque l'hypothèse de normalité des observations est violée en utilisant des données simulées à partir de plusieurs familles de lois usuelles. Dans ce dernier cas, le test de Grubb a tendance à rejeter l'hypothèse nulle systématiquement pour certaines familles de lois et pour d'autres familles, l'erreur de 1<sup>ère</sup> espèce estimée est loin de la valeur théorique de 5% retenue. Dans un second temps, nous avons révisé l'approche proposée par (Olmo, 2009) dans le cas de test unilatéral. Nous proposons une version modifiée de son algorithme et évaluons l'algorithme proposé par deux tests empiriques. Les deux tests utilisent des données simulées et du bootstrap. D'une part, nous simulons des données à partir d'une loi normale et comparons l'approche basée sur la théorie des valeurs extrêmes et celle du test de référence de Grubb en utilisant les erreurs de 1<sup>ère</sup> et de 2<sup>ème</sup> espèce. D'autre part, nous simulons des données à partir de diverses familles de loi afin d'analyser la puissance de l'approche proposée. Nos premiers résultats nous indiquent que l'approche de Grubbs a une meilleure performance que notre approche, toutefois avec une puissance comparable des deux tests. Dans le second test qui le plus important au vu de l'objectif de notre papier, l'approche que nous proposons rejette l'hypothèse nulle dans une proportion raisonnable (de l'ordre de l'erreur de première espèce), ce qui correspond bien aux prédictions théoriques.

### Bibliographie

- [1] Beirlant, J., Goegebeur, Y., Teugels, J. & Segers, J. (2004), 'Front matter', *Statistics of Extremes: Theory and Applications* pp. i–xiii.
- [2] Dey, D., Roy, D. & Yan, J. (2016), *Extreme Value Modeling and Risk Analysis: Methods and Applications*, CRC Press.
- [3] Dixon, W. J. (1950), 'Analysis of extreme values', *The Annals of Mathematical Statistics* 21(4), 488–506.

- [4] Embrechts, P., Klüppelberg, C. & Mikosch, T. (2013), Modelling extremal events: for insurance and finance, Vol. 33, Springer Science & Business Media.
- [5] Gonzalo, J. & Olmo, J. (2004), 'Which extreme values are really extreme?', Journal of Financial Econometrics 2(3), 349–369.
- [6] Grubbs, F. E. (1950), 'Sample criteria for testing outlying observations', The Annals of Mathematical Statistics pp. 27–58.
- [7] Olmo, J. (2009), 'Extreme value theory filtering techniques for outlier detection'.
- [8] Pearson, E. S. & Sekar, C. C. (1936), 'The efficiency of statistical tools and a criterion for the rejection of outlying observations', Biometrika 28(3/4), 308–320.