

NORMALITÉ ASYMPTOTIQUE DES STATISTIQUES DE TESTS DES INDICES RELATIFS DE DISPERSION ET DE VARIATION

Aboubacar Y. TOURÉ¹, Simplicie DOSSOU-GBÉTÉ² & Célestin C. KOKONENDJI³

¹ *Université Bourgogne Franche-Comté, Laboratoire de Mathématiques de Besançon, France.
aboubacar_yacouba.toure@univ-fcomte.fr*

² *Université de Pau et des Pays de l'Adour, Laboratoire de Mathématiques appliquées, France.
simplice.dossou-gbete@univ-pau.fr*

³ *Université Bourgogne Franche-Comté, Laboratoire de Mathématiques de Besançon, France.
celestin.kokonendji@univ-fcomte.fr*

Résumé. À partir des indices de dispersion relatifs aux lois de Poisson et binomiale pour les données de comptage et, récemment, de l'indice de variation exponentielle pour les données continues positives, nous introduisons d'abord la définition unifiée de l'indice de variabilité relative à une famille exponentielle naturelle positive à travers sa fonction variance. Ensuite, nous montrons la normalité asymptotique des statistiques de tests correspondantes et donnons des exemples applicables. Des études de simulations ont mis en évidence de bons comportements de ces statistiques de tests asymptotiques. Des remarques finales sont faites avec de possibles extensions.

Mots-clés. Loi continue, loi de comptage, famille exponentielle, fonction variance.

Abstract. From the classical dispersion indexes with respect to the Poisson and binomial distributions for count data and, recently, the exponential variation index for positive continuous data, we first introduce the unified definition of the relative variability to a non-negative natural exponential family through its variance function. Then, we show the asymptotic normality of the corresponding test statistics and give some applicable examples. Simulations have pointed out good behaviours of these asymptotic test statistics. Concluding remarks are made with possible extensions.

Keywords. Continuous law, counting law, exponential family, variance function.

1 Introduction

Le phénomène de dispersion poissonnienne est bien connu et très largement utilisé en pratique ; voir Jørgensen et Kokonendji (2016), Kokonendji (2014) et Kokonendji et al. (2004, 2007, 2008, 2011). Il existe de nombreux mécanismes menant à ce phénomène qui permettent de faire des inférences ; voir Böhning (1994) et Mizère et al. (2006) pour des tests statistiques. Introduit depuis Fisher (1934), l'indice de dispersion de Poisson d'une variable aléatoire réelle (v.a.r.) de comptage X sur $\mathbb{N}$ peut être défini comme $DI(X) = \text{Var}X/\mathbb{E}X$: le rapport de la variance sur la moyenne. En fait, DI est le rapport de deux variances car $\mathbb{E}X$ est la variance espérée sous la loi de Poisson de référence.

Ainsi, on en déduit facilement le concept d'indice relatif de dispersion (noté par RDI) en choisissant une loi de référence autre que Poisson. En effet, si X et Y sont deux v.a.r. de comptage de même support $\mathcal{S} \subseteq \mathbb{N}$ et avec $\mathbb{E}X = \mathbb{E}Y$ alors

$$\text{RDI}_Y(X) := \text{Var}X/\text{Var}Y = \text{DI}(X)/\text{DI}(Y). \quad (1.1)$$

Voir Weiss (2018, page 15) et Engel et te Brake (1993) pour la référence binomiale.

De manière similaire, Abid et al. (2019b) viennent de proposer l'indice de variation exponentielle (VI) pour une v.a.r. continue X sur $]0, +\infty[$ comme étant $\text{VI}(X) = \text{Var}X/(\mathbb{E}X)^2$. On peut le voir comme le carré du coefficient de variation. Il est parfois utilisé en fiabilité pour discriminer des lois en fonction du taux moyen de défaillance croissante/décroissante ; voir, par exemple, Barlow et Proschan (1981). Cependant, il n'est pas simple de construire un test statistique exact pour VI ; on peut se référer à Lam (1980). L'indice relatif de variation (RVI) est défini, pour deux v.a.r. continues de même support $\mathcal{S} =]0, +\infty[$ avec $\mathbb{E}X = \mathbb{E}Y$, par

$$\text{RVI}_Y(X) := \text{Var}X/\text{Var}Y = \text{VI}(X)/\text{VI}(Y). \quad (1.2)$$

Clairement, la seule différence entre (1.1) et (1.2) se trouve au niveau du support $\mathcal{S}$ des deux variables comparables X et Y .

Les deux objectifs de cette communication sont, *primo*, d'unifier les définitions de RDI et de RVI en distinguant le support $\mathcal{S}$ de leurs lois correspondantes et, *secundo*, de proposer un test asymptotique commun. Pour ce faire, nous considérerons le cadre de famille exponentielle naturelle (FEN) avec leurs fonctions variance. Le reste du document est organisé comme suit. La section 2 rappelle brièvement la FEN avant de proposer le nouvel indice relatif commun dit de variabilité. La section 3 présente les résultats asymptotiques des statistiques de tests. La section 4 étudie des exemples applicables. Enfin, la section 5 conclut sur quelques remarques et extensions possibles.

2 FEN et indice relatif de variabilité

Pour les notations de FEN, on peut se référer à Jørgensen (1997, Chap.2). Étant donné $\lambda \in \Lambda \subseteq]0, +\infty[$, la densité par rapport à la mesure de Lebesgue ou la mesure de comptage ν d'une FEN $F_\lambda := \{\mathcal{F}_\nu(\theta; \lambda); \theta \in \Theta\}$ sur $\mathcal{S} \subseteq \mathbb{R}$ peut être écrite comme

$$f(x; \theta, \lambda) = a(x; \lambda) \exp\{\theta x - \lambda K(\theta)\}, \quad \theta \in \Theta \subseteq \mathbb{R}, \quad (2.1)$$

où θ est le paramètre canonique et $K(\cdot)$ est la fonction cumulante sur Θ . Le paramètre fixé λ est la puissance de convolution et, donc, $\Lambda \supseteq \mathbb{N}^*$. Si $\lambda = 1$ et $F := F_1$ alors la paramétrisation par la moyenne $\mu := K'(\theta)$ dans (2.1) conduit à la caractérisation de F par sa fonction variance

$$V : M \rightarrow]0, +\infty[, \quad \mu \mapsto V(\mu) := K''(\psi(\mu)) = 1/\psi'(\mu), \quad (2.2)$$

où $M := K'(\Theta)$ est le domaine des moyennes et $\psi(\cdot)$ est la fonction inverse de $K'(\cdot)$. Pour $\lambda > 0$ la fonction variance $V_\lambda(\cdot)$ de $F_\lambda := \{\mathcal{F}_V(\mu; \lambda); \mu \in M_\lambda\}$ est obtenue, via (2.2), par

$$V_\lambda(\mu) = \lambda V(\mu/\lambda) \text{ sur } M_\lambda := \lambda M. \quad (2.3)$$

À affinité près, Morris (1982) a classifié les FENs quadratiques de la Table 1 ; voir Letac et Mora (1990) pour les strictement cubiques. On peut se référer à Abid et al. (2019abc), Bonat et Kokonendji (2017) et Bonat et al. (2018) pour d'autres modèles.

Table 1: Résumé des six types de FEN quadratiques (Morris, 1982) sur M avec les ensembles $\mathcal{S}$ et Λ respectifs ; GHS représente la sécante hyperbolique généralisée.

Type	$V(\mu)$	M	$\mathcal{S}$	Λ
Gaussien	1	$\mathbb{R}$	$\mathbb{R}$	$]0, +\infty[$
Poisson	μ	$]0, +\infty[$	$\mathbb{N}$	$]0, +\infty[$
Gamma	μ^2	$]0, +\infty[$	$]0, +\infty[$	$]0, +\infty[$
Binomial	$\mu(1 - \mu)$	$]0, 1[$	$\{0, 1\}$	$\mathbb{N}^*$
Binomial négatif	$\mu(1 + \mu)$	$]0, +\infty[$	$\mathbb{N}$	$]0, +\infty[$
GHS	$\mu^2 + 1$	$\mathbb{R}$	$\mathbb{R}$	$]0, +\infty[$

Proposition 2.1 (Jørgensen et al., 1994) *Soit F une FEN de fonction variance V sur M et de support $\mathcal{S}$. Si $\inf \mathcal{S} = 0$ et notons par $\delta := \inf\{\mathcal{S} \setminus \{0\}\}$ la distance entre 0 et le plus petit point du support, alors V et M satisfont: (1) $\inf M = 0$, (2) $\lim_{\mu \rightarrow 0} V(\mu) = 0$, (3) $\lim_{\mu \rightarrow 0} V(\mu)/\mu = \delta$, (4) si $\mathbb{P}\{0\} > 0$ alors $\lim_{\mu \rightarrow 0} V'(\mu) = \delta$ et (5) si $\mathbb{P}\{0\} > 0$ alors $\lim_{\mu \rightarrow 0} V'(\mu)/\mu^2 = +\infty$.*

Dans la Proposition 2.1, $\delta = 0$ est pour les lois continues (p.ex., gamma) et $\delta = 1$ est pour les lois de comptage (p.ex., binomiale négative). Une loi ayant $\delta > 0$ doit avoir une masse en zéro (p.ex., gamma décentrée). En général, la dérivée à droite de zéro $V'(0^+)$ de V est positive pour les lois discrètes et égale à zéro pour les lois continues.

Voici la définition unifiée des indices relatifs de dispersion et de variation :

Definition 2.2 *Soient X et Y deux v.a.r. à valeurs dans $\mathcal{S} \subseteq [0, +\infty[$. Soit $\mu = \mathbb{E}X = \mathbb{E}Y$, $\sigma^2 = \text{Var}X$ et $V_\lambda(\mu) = \text{Var}Y$, pour $V_\lambda > 0$ fixé, l'indice relatif de variabilité de X par rapport à Y est la quantité positive*

$$\text{RWI}_Y(X) = \text{RWI}_{V_\lambda}(X) := \sigma^2 / V_\lambda(\mu) \geq 1;$$

*autrement dit, la **sur-** (équi- et **sous-variabilité**) de X comparé à Y , et notée par $X > Y$ ($X \asymp Y$ et $X < Y$), est réalisée si $\sigma^2 > V_\lambda(\mu)$ ($\sigma^2 = V_\lambda(\mu)$) et $\sigma^2 < V_\lambda(\mu)$, respectivement).*

Si $\mathcal{S} \subseteq \mathbb{N}$ alors RWI n'est autre que RDI de (1.1), et si $\mathcal{S} =]0, +\infty[$ on a RWI = RVI de (1.2). On peut dire que RWI est le rapport de deux variabilités comparables et l'équi-variabilité RWI = 1 signifie qu'il n'y a pas de différence entre les deux variabilités.

3 Normalité asymptotique des statistiques de tests

Soit $X_1, \dots, X_n$ un n -échantillon aléatoire i.i.d. de même loi que X sur $\mathbb{S} \subseteq [0, +\infty[$ avec $\mathbb{E}X^4 < +\infty$. Notons $\bar{X}_n := n^{-1} \sum_{i=1}^n X_i$, $S_n^2 := n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, $\mu := \mathbb{E}X$ et $\sigma^2 := \text{Var}X$. En prenant $\lambda = 1$ dans (2.3), l'estimateur $S_n^2/V(\bar{X}_n)$ de l'indice relatif de variabilité $\sigma^2/V(\mu)$ est asymptotiquement sans biais : $\lim_{n \rightarrow \infty} \mathbb{E}\{S_n^2/V(\bar{X}_n)\} = \sigma^2/V(\mu)$. Nous établissons les résultats suivants dont les démonstrations sont données dans Touré et al. (2019).

Theorem 3.1 *Sous l'hypothèse nulle $H_{0,V} : \sigma^2 = V(\mu)$ et en supposant que $V''(\mu)$ existe, on a :*

$$\sqrt{n}[S_n^2/V(\bar{X}_n) - 1] \rightsquigarrow \mathcal{N}\left(0; 2 + V''(\mu) + 4\mu^2/V(\mu)\right), \text{ quand } n \rightarrow +\infty,$$

où $\rightsquigarrow$ désigne la convergence en loi et $\mathcal{N}(0; \tau^2)$ est la loi normale centrée de variance $\tau^2 > 0$.

À partir du Théorème 3.1, on a le comportement suivant de la statistique du test :

$$T_{n,V} = \sqrt{\frac{n}{2 + V''(\bar{X}_n) + 4\bar{X}_n^2/V(\bar{X}_n)}} \left(\frac{S_n^2}{V(\bar{X}_n)} - 1 \right) \rightsquigarrow \mathcal{N}(0, 1), \text{ quand } n \rightarrow +\infty.$$

Proposition 3.2 *Supposons seulement que $\mathbb{E}X^2 < +\infty$, alors $S_n^2/V(\bar{X}_n) \xrightarrow{p.s.} \sigma^2/V(\mu)$, quand $n \rightarrow +\infty$ et où $\xrightarrow{p.s.}$ désigne la convergence presque sûre.*

4 Simulations des exemples applicables

Les simulations des cinq exemples intéressants pour la pratique ci-dessous ont fourni de très bons résultats, même pour des échantillons de tailles modérées, et selon des valeurs fixées de $\mu > 0$, $N \in \mathbb{N}^*$ et $\lambda > 0$. En notant u_p le quantile d'ordre p de $\mathcal{N}(0, 1)$:

- (i) $H_{0,G} : \sigma^2 = \mu^2/\lambda$ contre $H_{1,G} : \sigma^2 \neq \mu^2/\lambda$ sert à tester l'équi-variation de RVI de gamma $\mathcal{G}(\lambda; \lambda/\mu)$, où $\lambda = 1$ correspond au cas VI de la loi exponentielle $\mathcal{E}(1/\mu)$ et $H_{0,G}$ est rejetée si

$$T_{n,G} = \left| \sqrt{\frac{n}{2 + 4\lambda + 2/\lambda}} \left(\frac{S_n^2}{\bar{X}_n^2/\lambda} - 1 \right) \right| > u_{1-\alpha/2};$$

- (ii) $H_{0,GI} : \sigma^2 = \mu^3/\lambda^2$ contre $H_{1,GI} : \sigma^2 > \mu^3/\lambda^2$ sert à tester la sur-variation de RVI par rapport à la gaussienne inverse $\mathcal{G}\mathcal{I}(\lambda^2; \mu)$ et $H_{0,GI}$ est rejetée si

$$T_{n,GI} = \sqrt{\frac{n\bar{X}_n}{4\lambda^2 + 2\bar{X}_n + 6\bar{X}_n^2/\lambda^2}} \left(\frac{S_n^2}{\bar{X}_n^3/\lambda^2} - 1 \right) > u_{1-\alpha};$$

- (iii) $H_{0,P} : \sigma^2 = \mu$ contre $H_{1,P} : \sigma^2 \neq \mu$ sert à tester le DI de Poisson $\mathcal{P}(\mu)$ et $H_{0,P}$ est rejetée si

$$T_{n,P} = \left| \sqrt{\frac{n}{2 + 4\bar{X}_n}} \left(\frac{S_n^2}{\bar{X}_n} - 1 \right) \right| > u_{1-\alpha/2};$$

- (iv) $H_{0,B} : \sigma^2 \geq \mu(1 - \mu/N)$ contre $H_{1,B} : \sigma^2 < \mu(1 - \mu/N)$ sert à tester la sous-dispersion de RDI par rapport à la binomiale $\mathcal{B}(N; \mu/N)$ et $H_{0,B}$ est rejetée si

$$T_{n,B} = \sqrt{\frac{n(1 - \bar{X}_n/N)}{(2 - 2/N) + (4 - 2/N + 2/N^2)\bar{X}_n}} \left(\frac{S_n^2}{\bar{X}_n - \bar{X}_n^2/N} - 1 \right) < u_\alpha;$$

- (v) $H_{0,BN} : \sigma^2 \leq \mu(1 + \mu/\lambda)$ contre $H_{1,BN} : \sigma^2 > \mu(1 + \mu/\lambda)$ pour tester la sur-dispersion de RDI à binomiale négative $\mathcal{BN}(\lambda; \lambda/(\lambda + \mu))$ et $H_{0,BN}$ est rejetée si

$$T_{n,BN} = \sqrt{\frac{n(1 + \bar{X}_n/\lambda)}{(2 + 2/\lambda) + (4 + 2/\lambda + 2/\lambda^2)\bar{X}_n}} \left(\frac{S_n^2}{\bar{X}_n + \bar{X}_n^2/\lambda} - 1 \right) > u_{1-\alpha}.$$

5 Conclusion

On a introduit un nouvel indice relatif de variabilité RWI qui inclut le classique RDI et le plus récent RVI. On peut l'étendre au cas où $\mathcal{S} = \mathbb{R}$ afin de comparer par exemple la loi normale centrée à la loi de Student, ainsi qu'au cas multivarié. Pour ce dernier, on peut se référer à Kokonendji et Puig (2018) pour RDI. Les tests asymptotiques se comportent bien et on aura besoin d'une version bootstrapée pour de petits échantillons.

Bibliographie

- Abid, R., Kokonendji, C.C. and Masmoudi, A. (2019a), Geometric dispersion models with real quadratic v-functions, *Statistics and Probability Letters*, 145, 197–204.
- Abid, R., Kokonendji, C.C. and Masmoudi, A. (2019b), Geometric Tweedie regression models for continuous and semicontinuous data with variation phenomenon, *AStA Advances in Statistical Analysis*, DOI : 10.1007/s10182-019-00350-8.
- Abid, R., Kokonendji, C.C. and Masmoudi, A. (2019c), Poisson-exponential-Tweedie regression models for ultra-overdispersed count data and applications, Submitted.
- Böhning, D.(1994), A note on a test for Poisson overdispersion, *Biometrika*, 81, 418–419.
- Barlow, R.A. et Proschan, F. (1981), *Statistical Theory of Reliability and Life Testing: Probability Models*, To Begin With, Silver Springs, MD.
- Bonat, W.H. and Kokonendji, C.C. (2017), Flexible Tweedie regression models for continuous data, *Journal of Statistical Computation and Simulation*, 87, 2138–2152.

- Bonat, W.H., Jørgensen, B., Kokonendji, C.C., Hinde, J. and Demétrio, C.G.B. (2018), Extended Poisson-Tweedie : Properties and regression models for count data, *Statistical Modelling*, 18, 24–49.
- Engel, B. and te Brake, J. (1993), Analysis of embryonic development with a model for under- or overdispersion relative to binomial variation, *Biometrics*, 49, 269–279.
- Fisher, R.A. (1934), The effects of methods of ascertainment upon the estimation of frequencies, *Annals of Eugenics*, 6, 13–25.
- Jørgensen, B. (1997), *The Theory of Dispersion Models*, Chapman & Hall, London.
- Jørgensen, B. and Kokonendji, C.C. (2016), Discrete dispersion models and their Tweedie asymptotics, *AStA Advances in Statistical Analysis*, 100, 43–78.
- Jørgensen, B., Martínez, J.R. and Tsao, M. (1994), Asymptotic behaviour of the variance function, *Scandinavian Journal of Statistics*, 21, 223–243.
- Kokonendji, C.C. (2014). Over- and underdispersion models. In: N. Balakrishnan (Ed.) *The Wiley Encyclopedia of Clinical Trials - Methods and Applications of Statistics in Clinical Trials*, Vol. 2 (Chap. 30), Wiley, New York, pp. 506–526.
- Kokonendji, C.C., Dossou-Gbété, S. and Demétrio, C.G.B. (2004), Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes, *Statistics and Operations Research Transactions*, 28, 201–214.
- Kokonendji, C.C., Demétrio, C.G.B. and Zocchi, S.S. (2007), On Hinde-Demétrio regression models for overdispersed count data, *Statistical Methodology*, 4, 271–291.
- Kokonendji, C.C., Mizère, D. and Balakrishnan, N. (2008), Connections of the Poisson weight function to overdispersion and underdispersion, *Journal of Statistical Planning and Inference*, 138, 1287–1296.
- Kokonendji, C.C., Demétrio, C.G.B., Dossou-Gbété, S. and Zocchi, S.S. (2011), On an extension of the negative binomial distribution, *Journal of Statistical Theory and Applications*, 10, 23–35.
- Kokonendji, C.C. and Puig, P. (2018), Fisher dispersion index for multivariate count distributions: A review and a new proposal, *Journal of Multivariate Analysis*, 165, 180–193.
- Lam, H.-K. (1980), Remarks on the distribution of the sample variance in exponential sampling, *Communications in Statistics-Simulation and Computation*, 9, 639–647.
- Letac, G. and Mora, M. (1990), Natural real exponential families with cubic variance functions, *The Annals of Statistics*, 18, 1–37.
- Mizère, D., Kokonendji, C.C. and Dossou-Gbété, S. (2006), Quelques tests de la loi de Poisson contre des alternatives générales basées sur l'indice de dispersion de Fisher, *Revue de Statistique Appliquée*, 54(4), 61–84.
- Morris, C.N. (1982), Natural exponential families with quadratic variance functions, *The Annals of Statistics*, 10, 65–80.
- Touré, A.Y., Dossou-Gbété, S. and Kokonendji, C.C. (2019), Asymptotic normality of the test statistics for the unified relative dispersion and relative variation indexes, Submitted.
- Weiss, C.H. (2018), *An Introduction to Discrete-Valued Time Series*, Wiley, Hoboken NJ.