



D_{zz}^2 ou W pour tester la normalité des données d'un échantillon

Louis Laurencelle ^a,

^aUniversité du Québec à Trois-Rivières

Abstract ■ D_{zz}^2 , a new measure of non-normality in a statistical series, is computed as the total squared difference between the order statistics of the series and values of the corresponding normal series. Following the procedure used by Shapiro et al. (1968), we set up an extensive comparative study implying Shapiro and Wilk's W measure, shape indices g_1 , g_2 and their combination $C_{g_1g_2}$, and D_{zz}^2 . The Monte Carlo experimentation surveyed ten non-normal statistical distributions, each with a large sample of their parametric variants, and data series thereof of sizes from 10 to 1000. Notwithstanding the occasional pertinence and efficacy of the three other statistics, D_{zz}^2 and W only can claim to be qualified as omnibus non-normality tests, with a grossly equivalent performance and divided merits, plus a somewhat greater sensitivity of D_{zz}^2 for the subtler non-normal conditions. **Résumé** ■ Une nouvelle mesure, le D_{zz}^2 , est proposée pour vérifier la normalité d'un échantillon de données statistiques, mesure basée sur l'écart global entre les statistiques d'ordre de la série testée et les valeurs normales correspondantes. Une étude comparative extensive, semblable à l'étude princeps de SHAPIRO et coll. (1968), met en jeu la statistique W de SHAPIRO et WILK (1965), les indices de forme g_1 et g_2 , $C_{g_1g_2}$, une combinaison de g_1 et g_2 , et D_{zz}^2 . L'expérimentation Monte Carlo comprend dix lois non-normales, symétriques et asymétriques, un large jeu de variantes de ces lois et des tailles d'échantillons allant de 10 à 1000. Nonobstant certains mérites des trois autres statistiques, seuls D_{zz}^2 et W se qualifient comme tests universels de non-normalité, avec une performance globalement équivalente et des avantages partagés, si ce n'est une tendance à une plus grande sensibilité précoce de D_{zz}^2 .

Keywords ■ Non-normality tests; Shapiro-Wilk W ; Non-normal densities; Comparative Study..

louis.laurencelle@gmail.com

LL: na

[10.20982/tqmp.14.1.p055](https://doi.org/10.20982/tqmp.14.1.p055)

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ No reviewer.

Introduction

La "loi normale", qu'on retrouve partout dans les sciences et les applications scientifiques (démographie, épidémiologie, médecine, psychologie, etc.), et qui sert abondamment dans d'autres sphères de la vie, a une longue histoire : STIGLER (1990), dans un récit passionnant, en retrace les étapes. Dès 1733, de Moivre, qui cherchait alors à exprimer la forme limite de la distribution de probabilité binomiale, a le premier donné une expression mathématique complète de la loi. Laplace, en 1777, puis Gauss en 1810, ont conçu la loi par un raisonnement direct (et non pas comme une approximation d'une autre loi), de sorte qu'on y réfère parfois sous le nom de "loi

de Laplace-Gauss". Quant à l'appellation "loi normale" ou "distribution normale", elle émane de plusieurs sources. Quételet en 1855 rapporte que cette loi décrit excellemment la répartition des mesures biométriques humaines et allègue, avec K. Pearson, qu'on peut la qualifier d'universelle. C'est Galton qui, en 1877, la baptise officiellement "normale", suivi ou imité en cela par Poincaré et K. Pearson en 1893. Sous sa forme moderne, la loi normale de densité de probabilité est :

$$\phi(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2/\sigma^2}, \quad (1)$$

et on peut caractériser une variable normale par la notation $X \sim N(\mu, \sigma^2)$.



Universelle, non pas, mais très commode comme référence première et souvent comme approximation, la loi normale joue un rôle clé dans des domaines comme la statistique, la psychométrie, la prédiction de toute sorte (médecine, actuariat, météorologie, etc.), et ce n'est qu'à contrecœur qu'on renonce à son utilisation. Or, le chercheur, le technicien, l'utilisateur se trouve parfois, voire régulièrement dans certains contextes, placé devant une mesure nouvelle, une variable dont il n'est pas certain de la "normalité" ou dont certains indices laissent croire que sa distribution est "non normale". S'il s'avère que la variable est vraiment non normale, alors il se voit confronté à trois options : - fermer les yeux et faire 'comme si', en espérant que la technique de calcul appliquée (la moyenne arithmétique¹, l'analyse de variance, la corrélation, la prédiction linéaire) sera assez robuste pour résister à l'anomalie de distribution et fournir néanmoins des résultats représentatifs et valides ; - tenter de transformer la variable mathématiquement, par une fonction non-linéaire, afin que la distribution résultante se rapproche de la "normale", puis opérer ses calculs sur la variable transformée ; - renoncer aux techniques d'analyse classiques, quand c'est possible, et recourir à des méthodes ou plus laborieuses, ou moins statistiquement puissantes.

Le test de normalité se montre donc souvent d'une utilité cruciale.

Contexte et objectifs

Le chercheur qui a à sa disposition un échantillon de bonne taille, par exemple $n \geq 200$ ou 300, peut en produire un histogramme et lui-même constater de visu la non-normalité si celle-ci est patente. Toutefois, la normalité, même approximative, est une caractéristique complexe : la répartition descendante et symétrique de part et d'autre d'un sommet modal n'en est que l'aspect le plus apparent. La longueur et la minceur des ailes de la distribution, la plus ou moins forte étroitesse du renflement modal, la position du point d'inflexion dans la chute bilatérale de densité sont autant d'aspects qui, étant plus subtils, peuvent échapper à l'œil et compromettre éventuellement la validité des analyses.

Parmi les techniques que propose la littérature pour tester la normalité d'une série statistique, la plus connue et respectée est celle proposée par Shapiro et Wilk en 1965 (voir aussi SHAPIRO et coll., 1968 ; SHAPIRO & WILK, 1968 ;

SHAPIRO & FRANCIA, 1972 ; SARKADY, 1975 ; ROYSTON, 1982). Sous le symbole W , la statistique de Shapiro et Wilk est essentiellement basée sur un calcul de moments, et elle apparaît comme un estimateur de variance obtenu par la somme des données de l'échantillon pondérées par les statistiques d'ordre de la loi normale. La littérature classique, à la fois mathématique et pédagogique, a proposé d'autres techniques, d'aucunes graphiques, la plupart par calcul. Les plus connus de ces autres tests de normalité par calcul sont :

- le Kolmogorov-Smirnov. La forme la plus courante du test consiste en deux étapes : 1) transformer d'abord les n données de l'échantillon en écarts-réduits (ou scores z) et les mettre en ordre croissant, puis 2) pour ces données $z_{[i]}$ ($i = 1$ à n), repérer la différence absolue maximale entre i/n ou, préférablement, $i/(n + 1)$, et $\Phi(z_{[i]})$, $\Phi(z)$ dénotant l'intégrale normale standard à $Z = z$. Le test indique une non-normalité significative si cette différence d_{\max} déborde la valeur critique.²
- le Khi-deux sur un tableau de fréquences. Pour ce test, les n données de l'échantillon doivent d'abord être classées en k intervalles et dénombrées dans un tableau de fréquences f_1, f_2, \dots, f_k . Ces fréquences sont ensuite comparées aux fréquences "normales" équivalentes, obtenues pour chaque classe à partir des bornes de classe standardisées (via la moyenne et l'écart-type de l'échantillon). Une valeur de khi-deux élevée suggère la non-normalité.
- les moments g_1 et g_2 (aussi nommés "indices de forme")³, qu'on peut définir simplement par $g_1 = \sum z_i^3/n$ et $g_2 = \sum z_i^4/n - 3$, se rapportant respectivement à l'asymétrie et la voussure (ou aplatissement) de la distribution. Par nature et par définition, la normalité implique $g_1 \approx 0$ et $g_2 \approx 0$.

Il existe bien sûr d'autres tests, certains étant des variantes des tests énumérés ci-dessus, d'autres des versions "ordinales", c'est-à-dire basées sur des quantiles,⁴ d'autres encore.

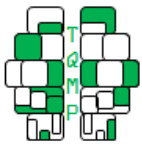
Shapiro, Wilk et Chen, en 1968, ont procédé à une étude comparative extensive de 9 tests différents, incluant leur test W , dit "test Shapiro-Wilk", en étudiant leurs efficacités diagnostiques versus 12 lois de distribution non normales et leurs variantes paramétriques, ce pour des séries (ou échantillons) de tailles $n = 10, 15, 20, 35$ et 50 : une grande abondance de résultats en découle. La conclusion

1. La moyenne arithmétique peut bien sûr être calculée sur n'importe quelle espèce de variable numérique, mais son interprétation reste délicate. En cas d'asymétrie, par exemple, la moyenne ne représente plus la valeur de l'"individu moyen" (concept si cher à Quételet), puisqu'elle est attirée par les données situées dans l'aile plus étendue de la distribution, aboutissant à une sur- ou une sous-estimation de la valeur cherchée, laquelle devrait se rapprocher de la médiane.

2. Voir cependant LILLIFORS (1967) et LAURENCELLE (2001) pour un calcul de valeur critique plus rigoureux, tenant compte de l'estimation (robuste ou non) des paramètres de moyenne et d'écart-type.

3. Les indices de forme sont parfois rapportés comme $\sqrt{b_1} = g_1$ et $b_2 = g_2 + 3$.

4. Pour mesurer le degré d'asymétrie, par exemple, le moment empirique g_1 est remplacé par un calcul de forme $[(C_{90} - C_{50}) - (C_{50} - C_{10})]/(C_{75} - C_{25})$, C_P dénotant le P ème centile empirique de la série statistique.



de l'étude favorise globalement le test W proposé. Nous citons (op. cit., p. 1366, traduction libre) :

La statistique W se montre sensible à la non-normalité pour un grand registre de distributions différentes. Dans la plupart des cas, elle a une puissance aussi bonne ou meilleure que les huit autres procédures de test comparées. Pour les variables continues, le W est le seul test qui n'affiche jamais une puissance vraiment basse dans les cas où un autre test s'avère prédominant.

Depuis l'étude citée ci-dessus, à notre connaissance, il n'est pas paru d'autres études ni publié d'autres résultats d'envergure concernant l'efficacité des tests de normalité. Pour cette raison sans doute, et parce que le test W 'fait l'affaire' - un test de normalité global, c'est-à-dire sensible à la normalité globale plutôt qu'à l'un de ses aspects seulement -, il est devenu le test de référence, et son résultat, une 'garantie' de quasi-normalité.⁵

L'étude des trois auteurs (SHAPIRO et coll., 1968) présente toutefois plusieurs bémols, le moindre étant que certaines des lois ou variantes de loi explorées se recoupent. On ne peut leur reprocher leur répertoire de tests de normalité appliqués, sinon peut-être qu'il pêche par excès. Ainsi, le test de David, Hartley et Pearson (1954; voir Shapiro, Wilk et Chen, 1968, p. 1344), basé sur deux données seulement, x_{max} et x_{min} ,⁶ ne fait pas sérieux; ou bien le test des fréquences par le khi-deux, encore présent dans les manuels de statistique appliquée, dépend et varie selon un classement arbitraire des données de la série dans un ensemble d'intervalles,⁷ en plus de manquer de puissance. Les deux grandes faiblesses de l'étude réfèrent aux tailles échantillonnelles utilisées et aux champs paramétriques des lois explorées. Par exemple, pour la loi lognormale, la seule variante étudiée correspond à une variance (de la normale sous-jacente) de $\sigma^2 = 1$, résultant en une asymétrie de $\gamma_1 \approx 6,185$, une distribution dont la non-normalité saute déjà aux yeux dès l'examen de quelques données, tandis que par exemple plusieurs variables biométriques, foncièrement lognormales, présentent des valeurs γ_1 situées plutôt dans l'intervalle $0 < \gamma_1 < 1$. Il en est de même pour la loi χ^2 (khi-deux), la valeur maximale du paramètre ν (degrés de liberté) explorée étant 10. Les lois Bêta et celle du Double khi-deux (reprise ici dans la loi dite de Subbotin, voir plus loin) tombent sous la même critique. Enfin, les auteurs n'ont pas considéré d'éventuels échantillons composites ou hétérogènes, c'est-à-dire des échantillons constitués

d'éléments issus du mélange de deux ou plusieurs populations. Quant aux tailles d'échantillon que Shapiro et son équipe ont regardées, soit de $n = 10$ à $n = 50$, elles sont faibles, notamment celles de 10 ou 15, et généralement insuffisantes pour hasarder une sanction de normalité ou de non-normalité. L'examen de sensibilité aurait aussi gagné avec des tailles raisonnablement élevées, comme $n = 100, 500$, voire 1000, en permettant de faire ressortir ou d'atténuer les différences de sensibilité des différents tests. Une dernière faiblesse méthodologique, affectant cette fois la précision et la fiabilité des résultats, concerne les assises de l'expérimentation Monte Carlo mise en œuvre. D'une part, les valeurs critiques de sept des neuf tests sont produites et estimées à partir de seulement $N = 500$ échantillons (normaux) aléatoires; celles du test W le sont avec N allant de 5000 à 100, selon la taille n croissante de la série produite; seules, les centiles du χ^2 sont exacts, étant repris d'une table publiée. Quant aux échantillons des lois non-normales étudiées, le nombre d'échantillons ayant servi à produire les indices de puissance et à fonder les conclusions de l'étude a été de 200!

L'étude proposée ici vise à combler les lacunes constatées dans la preuve donnée par SHAPIRO et coll. (1968) en faveur de la supériorité générale du test W , et elle prend pour prétexte la proposition de deux nouvelles statistiques, désignées D_{zz}^2 et $C_{g_1g_2}$, lesquelles mesurent aussi, à leur façon, l'aspect global de la normalité. Nos objectifs sont donc de :

- remettre à l'épreuve la statistique W , dite de Shapiro-Wilk, en la confrontant à quatre autres statistiques concurrentes, le g_1 , le g_2 , la combinaison $C_{g_1g_2}$, et notre nouvelle D_{zz}^2 ;
- confronter les différentes statistiques à un palmarès de lois non-normales et de variantes à tous les niveaux mesurables de leur non-normalité;
- étendre la taille des échantillons testés depuis $n = 20$ jusqu'à $n = 1000$;

De plus, les valeurs critiques appliquées seront basées sur des estimations Monte Carlo sérieuses, et les mesures de puissance des tests (leur significativité) basées sur 2 000 000 échantillons ou plus. Par ces objectifs, nous comptons à la fois moderniser la preuve d'efficacité des tests examinés et nuancer leurs avantages respectifs selon le contexte concerné.

5. Tout comme un rond n'est jamais tout à fait un cercle, une série statistique ne peut jamais s'inscrire parfaitement dans le modèle normal, la quasi-normalité étant une condition suffisante à toutes fins pratiques.

6. La statistique u se présente comme $u = (x_{max} - x_{min})/s$, s étant l'écart-type. J. W. Tukey (voir p. ex. HOWELL, 2008) a proposé cette forme (sous le nom "Studentized range") pour tester les différences paires de ($k > 2$) moyennes sous contrôle du seuil de significativité alpha.

7. Sur ce sujet, voir les commentaires et recommandations de KENDALL et STUART (1979), p. 455 seq.



Méthodologie de l'étude

Les tests comparés et leurs valeurs critiques

Dans leur étude de 1968, Shapiro et collègues incluaient des tests dits "de distance", ou ce qu'on pourrait appeler des variantes du test de Kolmogorov-Smirnov : ces tests sont le Cramer-Von Mises, le Cramer-Von Mises pondéré et le Durbin (op. cit., p. 1344). Leur conclusion est nette : ces tests démontrent une "puissance étonnamment faible" (op. cit., p. 1367). Diagnostic semblable pour le khi-deux sur fréquences, qui pêche déjà par l'arbitraire du groupement des données en intervalles. Quant au test u de David, il n'est pas sérieux, la significativité du test reposant seulement sur les deux valeurs extrêmes de la série statistique. De l'étude de 1968, nous retenons donc les statistiques g_1 , g_2 (en valeurs absolues) et W , à quoi nous ajoutons le composé $C_{g_1 g_2}$ et D_{zz}^2 .

La statistique $C_{g_1 g_2}$ est une variante d'une famille de tests exploitant les indices g_1 et g_2 , telle la combinaison $z(g_1)^2 + z(g_2)^2$ qui forme une quasi variable khi-deux, les deux indices étant fortement interdépendants (D'AGOSTINO, BÉLANGER & D'AGOSTINO, 1990). Quant à D_{zz}^2 , il s'agit d'une mesure descriptive de type fonction de perte : elle quantifie la distance (au carré) entre les données ordonnées de la série observée et les valeurs correspondantes de la série normale équivalente, ajustée d'après la moyenne et l'écart-type de la série donnée : il s'agit donc de l'équivalent d'une comparaison de graphes, D_{zz}^2 fournissant une mesure globale d'écart entre le graphe observé et le graphe de référence.

Les cinq tests étudiés sont définis ci-après (le symbole s dénote l'écart-type de la série statistique, le symbole $\hat{\mu}_r$ réfère à l'estimateur du r ième moment central de la série).

— g_1 , indice d'asymétrie :

$$|g_1| = \left| \hat{\mu}_3 / \hat{\mu}_2^{1,5} \right| = \left| \sqrt{n} \sum (x - \bar{x})^3 / \left(\sum (x - \bar{x})^2 \right)^{1,5} \right|$$

— g_2 , indice de voussure, ou aplatissement :

$$|g_2| = \left| \hat{\mu}_4 / \hat{\mu}_2^2 - 3 \right| = \left| n \sum (x - \bar{x})^4 / \left(\sum (x - \bar{x})^2 \right)^2 - 3 \right|$$

— $C_{g_1 g_2}$, indice composite de normalité :

$$C_{g_1 g_2} = 2 \times |g_1| + |g_2|,$$

8. Un estimateur général de l'espérance $E(z_{[i]})$ a pour forme générale $F^{-1}(u)$, où $u = (i - \alpha) / (n + 1 - 2\alpha)$, où $F(x)$ est la fonction de répartition de la variable ; par exemple, pour la loi uniforme, on a exactement $E(z_{[i]}) = i / (n + 1)$, appliquant $\alpha = 0$. Pour la loi normale, l'estimation est approximative, et deux valeurs se concurrencent, soit $\alpha = 3/8$ et $\alpha = 2/5$. Les données de HARTER (1961) donnent la préférence à la première valeur ($\alpha = 3/8$) pour les petites séries, environ $n \leq 20$, la seconde ($\alpha = 2/5$) l'emportant pour les séries plus longues, ce que confirment aussi nos expérimentations.

la pondération appliquée ayant pour but d'équilibrer les erreurs-types respectives des deux composants (voir Kendall et Stuart, 1979, p. 258).

— W , indice global de Shapiro-Wilk (méthode de Royston 1982) :

$$W = \left(\sum a_i x_{(i)} \right)^2 / ((n - 1) \times s^2)$$

Le coefficient a_i dérive de l'espérance d'une statistique d'ordre de la loi normale standard, $z_{[i;n]}$ ou $z_{[i]}$, espérance basée ici sur l'approximation de HARTER (1961), soit $c_i = E(z_{[i]}) \approx \Phi^{-1}((i - 0, 4) / (n + 0, 2))$;⁸ la correction de Royston (1982, éq. 2.2) est appliquée aux deux valeurs de rangs extrêmes, a_1 et a_n . Le lecteur est référé à Royston pour le détail des calculs ; l'annexe 4 en présente la séquence.

— D_{zz}^2 , indice global descriptif :

$$D_{zz}^2 = \sum (z_{[i]} - c_i^*)^2 \times n$$

Les données $z_{[i]}$ sont les statistiques d'ordre standardisées de l'échantillon observé, soit $z_{[i]} = (x_{[i]} - \bar{x}) / s_x$, alors que les c_i^* , calculés comme $(c_i - \bar{c}) / s_c$ (où $\bar{c} = 0$ et s_c est l'écart-type des c_i), les coefficients c_i étant les mêmes que ceux exploités dans le calcul de l'indice W , voir ci-dessus. Il est intéressant de noter que, comme le montre l'équivalence

$$1 - \frac{D_{zz}^2}{2n(n - 1)} = r(z_{[i]}, c_i^*),$$

l'indice reflète directement la conformité linéaire entre les statistiques d'ordre observées et leurs correspondants de loi normale, la forme "r" ramenant l'indice à une métrique semblable à celle utilisée pour le W . La forme D_{zz}^2 a cependant été retenue, d'abord parce qu'elle exprime directement le concept d'écart entre les formes distributionnelles en jeu et aussi pour la commodité de ses valeurs critiques,

Valeurs critiques. Les valeurs critiques des cinq statistiques utilisées dans l'étude ont été établies par échantillonnage Monte Carlo, ce pour une grille de 27 tailles n , soit $n = 10(5)30(10)100(25)200(50)500(100)1000$, à raison de 5×10^6 séries normales standards par taille, chaque valeur normale étant obtenue par la méthode de Box-Muller (DEVROYE, 1986 ; LAURENCELLE, 2001). Dans la partie de l'étude concernant l'efficacité des



tests, les valeurs critiques correspondant aux tailles n intermédiaires (de $n = 20$ à $n = 1000$) sont obtenues par interpolation dans la grille : l'interpolation est linéaire (exposant $e = 1$) pour D_{zz}^2 et quasi harmonique pour les autres tests, soit selon un exposant de 0,6 pour W et de $-0,1$ pour g_1, g_2 et $C_{g_1g_2}$.⁹

Nous avons tenté de modéliser la valeur critique (VC) du test D_{zz}^2 , la valeur augmentant quasi linéairement avec n . La fonction prédictive suivante se montre toutefois plus précise (pour $n \geq 20$), soit :

$$VC_{n,\alpha} \approx n^{(3/2)} \times a_\alpha (n + b_\alpha)^{-c_\alpha} .$$

Le tableau ci-dessous caractérise cette fonction : la précision affichée apparaît globalement satisfaisante. Pour le lecteur intéressé, le tableau 7 en annexe 1 présente la grille ajourée de valeurs critiques préparée pour cette étude et décrite ci-dessus.

Seuil α	a	b	c	Erreur absolue relative	
				moyenne	maximale
0,10	1,649	19	0,4313	0,0010	0,0029
0,05	2,047	17	0,4390	0,0012	0,0048
0,01	3,244	16	0,4629	0,0009	0,0041
0,001	5,262	16	0,4858	0,0024	0,0064

Ces statistique appliquées, à une série donnée, suggèrent la non-normalité quand leur valeur déborde par le haut la valeur critique, sauf dans le cas du W de Shapiro-Wilk, qui est dite significative au seuil α si elle se trouve en-deçà, soit $W \leq W(\alpha)$.

Les lois non-normales utilisées et leurs variantes

Le répertoire exploré comporte 5 familles de lois symétriques et 5 asymétriques, avec leurs variantes. Nous présentons chaque loi ci-dessous, dans un court paragraphe, les tableaux 1 et 2 plus bas énumérant toutes les variantes étudiées.

Lois symétriques ($\gamma_1 = 0$)

1. La loi de Subbotin. La forme standard de la loi de SUBBOTIN (1923), à paramètre δ , est

$$f_S(X) = \frac{\exp[-\frac{1}{2}|X|^{2/\delta}]}{2^{\delta/2}\Gamma(\frac{1}{2}\delta + 1)}, -\infty < X < +\infty,$$

les paramètres de position (= 0) et d'échelle (= 1) étant ici à leur valeur standard. Les densités de cette loi s'étendent d'une forme aplatie ($\delta \rightarrow 0$) vers une Normale ($\delta = 1$), puis traversent une Double-exponentielle ou loi de Laplace ($\delta = 2$) et au-delà ($\delta > 2$). La loi dite du Double-khi-deux, dérivée de celle de Subbotin

9. La formule générale d'interpolation de la valeur V pour n , entre (V_-, n_-) et (V_+, n_+) , est $\hat{V} = V_- + (V_+ - V_-) \times (n^e - n_-^e) / ((n_+^e - n_-^e))$, e étant l'exposant.

10. Isométrique signifie de dimensions ou tailles égales, hétérométrique signifiant le contraire. Dans le cas présent, la densité binormale (d'intégrale 1) est composée de deux demi-densités égales (d'intégrales respectives 0,5).

par BOX et TIAO (1964) et exploitée par SHAPIRO et coll. (1968), utilise le paramètre β , égal à $\delta - 1$.

2. La loi t de Student. Son paramètre est ν ("nu") ≥ 1 , aussi désigné "degrés de liberté". Cette loi coïncide avec la loi de Cauchy pour $\nu = 1$.
3. La loi Bêta symétrique, ou Bêta(a, a). Sa densité, de paramètre a , est simplement

$$f_{B_S} = \frac{\Gamma(2a)}{\Gamma^2(a)} \times (x(1-x))^{a-1},$$

pour $0 \leq x \leq 1$ et $a > 0$; la loi Bêta(1, 1) coïncide avec la loi Uniforme (ou Rectangulaire).

4. La loi binormale isométrique¹⁰ homogène centrée (désignée ici Bi-Nor centrée). Il s'agit d'une mixture (des densités) de deux normales, $\approx 0,5 \times N(\hat{\mu}, \sigma) + 0,5 \times N(\hat{\mu}, \varepsilon \times \sigma)$, le champ paramétrique exploré étant ici $\varepsilon > 1$.
5. La loi binormale isométrique homogène décentrée (désignée ici Bi-Nor décentrée). Une mixture encore de deux normales, $\approx 0,5 \times N(\hat{\mu}, \sigma) + 0,5 \times N(\hat{\mu} + \delta, \sigma)$, le champ paramétrique exploré étant $\delta > 0$.

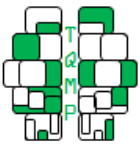
Lois asymétriques ($\gamma_1 \neq 0$)

6. La loi lognormale. La loi lognormale, telle que $X = a + \exp(\hat{\mu} + \sigma \times z)$ où $z = [\ln(X - a) - \hat{\mu}] / \sigma \approx N(0, 1)$, $X > a$, voit sa forme changer selon σ , les indices d'asymétrie et de voussure croissant avec lui. Le paramètre appliqué est donc $\sigma > 0$.
7. La loi khi-deux (χ^2). Loi positive, telle que la loi lognormale, son paramètre étant $\nu \geq 1$, aussi désigné "degrés de liberté". Cette loi coïncide avec la loi Exponentielle pour $\nu = 2$.
8. La loi Bêta asymétrique, ou Bêta(a, b). La densité de cette loi s'exprime par

$$f_{B_A} = \frac{\Gamma(a+b)}{\Gamma(a) \times \Gamma(b)} \times x^{a-1} (1-x)^{b-1},$$

pour $0 \leq x \leq 1$ et $a > 0, b > 0$. Elle est asymétrique si $a \neq b$. Le paramètre contrôlé est ici b (où $b \geq a = 2$).

9. La loi binormale isométrique inhomogène (désignée ici Bi-Nor inhomogène). Une mixture de normales équilibrées, aux deux paramètres inégaux, soit $\approx 0,5 \times N(\hat{\mu}, \sigma) + 0,5 \times N(\hat{\mu} + \delta, \varepsilon \times \sigma)$, $\delta \neq 0, \varepsilon \neq 1$. Les paramètres δ et ε sont variés.
10. La loi binormale hétérométrique inhomogène (désignée ici Bi-Nor "ghettoisée"), représentant une sous-population abritée dans une plus grande. Ici, une mixture de populations inégales, selon le ratio $P : 1 - P$, où $1/2 < P < 1$ et à dispersions différentes,



soit $\approx 0,8 \times N_1(\mu, 1) + 0,2 \times N_2(\mu + \delta, 1/2)$, où δ constitue le paramètre contrôlé, la loi étant donc décentrée. La sous-population de taille plus petite représente par exemple un “ghetto”.

Noter que nous avons fixé les domaines paramétriques des lois, c’est-à-dire l’échantillon de variantes $(\delta, \varepsilon, a, b, \hat{\mu}, \sigma, \nu)$ de chacune, de façon à couvrir la zone de sensibilité des cinq tests, depuis la zone à peu près non-sensible jusqu’à la détection (quasi) parfaite de la non-normalité.

Les variables des lois de Subbotin, lognormale et toutes celles dites binormales ont été produites par la méthode de rejet (DEVROYE, 1986; LAURENCELLE, 2001), à partir d’un modèle normal approprié.¹¹ La technique de BAILEY (1994) a été appliquée pour la variable du t de Student, celle de CHENG et FEAST (1979) pour loi khi-deux, celle de CHENG (1978) pour la loi Bêta asymétrique, celle d’Ulrich (1984) pour la loi Bêta symétrique. Ces techniques spécifiques se retrouvent aussi dans DEVROYE (1986) et LAURENCELLE (2001).

Échantillonnage Monte Carlo et analyses

L’étude s’est faite en deux temps. En premier lieu, nous avons ‘ausculté’ la puissance respective des cinq tests pour chacune des 10 lois considérées, ce en fonction du ou des paramètres associé(s) à cette loi : par exemple, il est connu que les lois t et khi-deux tendent vers une normale à mesure que leur paramètre (ν) augmente, de sorte que, pour une taille d’échantillon n donnée, nous comparons la sensibilité des cinq tests selon l’évolution de ce paramètre. Les puissances (ou proportions de dépassement de la valeur critique) rapportées sont basées sur 2×10^6 séries Monte Carlo de tailles $n = 20$ et $n = 100$.

En second lieu, et parce que cette approche nous paraît plus parlante pour le chercheur, nous avons repéré, pour chaque loi et variante, la taille n requise pour atteindre une puissance P donnée; les valeurs cibles de puissance étaient de $P = 0,50$ (équivalant au seuil de significativité du test¹²) et $P = 0,80$: il s’agit ici d’un repérage algorithmique assez laborieux. Pour ce faire, nous avons restreint le champ d’exploration à des tailles allant de 20 à 1000. Quant à la base échantillonnale de ces calculs, le repérage débutait avec 10 000 séries Monte Carlo puis, à mesure que le champ d’exploration allait se précisant, le nombre de séries évoluait vers 100 000. Quant aux niveaux de significativité (α) étudiés, notre étude comportait les seuils de

0,10, 0,05, 0,01 et 0,001. Nous ne présentons ici qu’une partie de ce devis d’étude, retenant seul le seuil $\alpha = 0,05$ et les puissances $P = 0,50$ et $P = 0,80$.

D’abord, un exemple !

Un exemple

La série statistique du Tableau 3 comporte $n = 139$ données. La moyenne calculée est 211,106, l’écart-type (échantillonnal) 23,418.

L’histogramme affiché à la Figure 1 fait deviner de l’asymétrie positive. Les descripteurs de la série sont : $n = 139$, $\bar{X} = 211,106$, $s = 23,418$, $g_1 = 0,888$, $g_2 = 1,281$: l’asymétrie semble confirmée, de même qu’une bonne leptokurtose (voussure étranglée) de la distribution.

Les résultats de nos cinq tests confirment la non-normalité de la série de 139 données, soit :

D_{zz}^2	= 885,09	(> 743,78, $\alpha = 0,001$),	$P \approx 0,0003$;
W	= 0,9493	(< 0,9538, $\alpha = 0,001$),	$P \approx 0,0003$;
$ g_1 $	= 0,888	(> 0,717, $\alpha = 0,001$),	$P \approx 0,0001$;
$ g_2 $	= 1,281	(> 1,171, $\alpha = 0,01$),	$P \approx 0,0071$;
$C_{g_1g_2}$	= 3,056	(> 1,972, $\alpha = 0,01$),	$P \approx 0,0011$.

Les probabilités approximatives fournies ci-dessus se basent sur 2×10^6 échantillons normaux de taille $n = 139$. L’examen des données suggère qu’un modèle lognormal à trois paramètres (incluant une origine positive) conviendrait sans doute à cette série.

Résultats de puissance et d’efficacité

Deux ensembles de figures sont fournis en annexe, présentant l’ensemble des données retenues sous forme de graphiques. Le premier ensemble, intitulé Puissances (Annexe 2), fournit pour chaque loi la puissance atteinte par des séries de $n = 20$ et $n = 100$ données. Dans le second ensemble (Annexe 3), on trouve les graphiques d’Efficacités, rapportant pour chaque loi et la puissance cible $P = 0,80$, la taille n requise pour l’atteindre. Le lecteur notera que, pour fins de clarté, le domaine de la loi de Subbotin a été rompu en deux segments, soit $0 \leq \delta < 1$ (voussure plus évasée que la normale) et $1 < \delta \leq 3$ (voussure plus pointue). De même, pour la loi dite inhomogène, les deux variantes à ratio $\varepsilon = \sigma_2/\sigma_1 = 2$ et 3 sont présentées séparément. Dans la lecture des tableaux comme l’interprétation des graphiques, le lecteur est prié de porter attention au contexte paramétrique spécifique de chaque loi. Ainsi, la distribution obtenue se “dénormalise”

11. Rappelons que la méthode de rejet consiste à générer une variable à partir d’une loi ‘facile’ (c.-à-d. pour laquelle la variable est facile à produire) qui domine la loi cible, puis à rejeter la variable candidate si elle n’est pas incluse dans la densité sous-jacente de la loi cible. Soit $f_C(x)$ la densité de la loi cible, et $f_D(x)$, celle de la dominante. Ici, nous avons choisi pour dominante une loi normale, $f_D(x) = p \times N(x : \mu, \sigma)$ telle que $f_D(x) \geq f_C(x)$, $p \geq 1$, tout x , après quoi l’algorithme suivant est appliqué : Répéter $x \leftarrow N(x : \mu, \sigma)$, $U \leftarrow \text{gen}[U(0, 1)]$ Jusqu’à $U \leq f_C(x)/f_D(x) \times p$. Les paramètres suffisants de la loi dominante (ici, μ, σ et p) sont à trouver dans chaque cas.

12. La valeur de la variable telle que le test appliqué atteint tout juste sa valeur critique (au seuil α choisi) a une puissance de 0,50, c.-à-d. une chance sur deux d’être significative ou non (KENDALL & STUART, 1979).



Tableau 1 ■ Variantes et caractérisation des 5 lois symétriques retenues

1. Subbotin ($\gamma_1 = 0$)†													
δ	0	0,2	0,5	0,8	(1)	1,25	1,5	1,75	2	2,25	2,5	2,75	3
γ_2	-1,200	-1,116	-0,812	-0,369	0	0,553	1,222	2,029	3	4,167	5,565	7,244	9,257
2. <i>t</i> de Student ($\gamma_1 = 0$)													
ν	1	2	3	4	5	6	7	8	9	10	12		
γ_2	-	-	-	-	6	3	2	1,5	1,2	1,0	0,75		
3. Bêta symétrique ($\gamma_1 = 0$)													
<i>a</i>	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7
γ_2	-1,200	-1,000	-0,857	-0,750	-0,667	-0,600	-0,545	-0,500	-0,462	-0,429	-0,400	-0,375	-0,353
4. Bi-Nor centrée ($\gamma_1 = 0$) : $0,5 \times N_1(0, 1) + 0,5 \times N_2(0, \varepsilon)$													
ε	1,5	2	2,5	3	3,5	4							
γ_2	0,444	1,080	1,573	1,920	2,163	2,336							
5. Bi-Nor décentrée ($\gamma_1 = 0$) : $0,5 \times N_1(0, 1) + 0,5 \times N_2(\delta, 1)$													
δ	1,5	2	2,5	3	3,5	4							
γ_2	-0,256	-0,500	-0,744	-0,956	-1,137	-1,280							

Note. † : La variante $\delta = 0$ de la loi de Subbotin est un cas limite, représenté ici par une variable uniforme, de domaine [0..1].

■ pour la loi de Subbotin lorsque δ s'éloigne de 1, de part et d'autre ; ■ pour les lois de Student, Khi-deux (χ^2) et Bêta symétrique lorsque le paramètre décroît vers 1 ; ■ pour toutes les autres lois, lorsque le paramètre croît.

Puissances

Les tableaux 4 et 5 donnent une lecture chiffrée des graphiques de puissance présentés en annexe 2 : la donnée au tableau, obtenue par interpolation grossière, indique la valeur approximative du paramètre de la loi (p. ex. δ , pour la loi Subbotin) à laquelle le test concerné atteint la puissance prescrite.

Les graphiques de puissance tout comme leur interprétation chiffrée au tableau 4 montrent d'abord que, généralement, la taille $n = 20$ fournit une base bien précaire, voire inadéquate pour juger de la non-normalité d'une série, la puissance accessible y étant bien faible. Il faut rappeler que, dans le présent contexte, les séries non-normales testées sont propres, voire « pures », étant générées par ordinateur, et ne contiennent rien d'hétérogène, alors que ces défauts sont communs dans les mesures réelles et susceptibles d'introduire une ou quelques données déviantes : le poids relatif de ces coquilles numériques dans une série empirique de taille $n \leq 20$ peut ainsi peser lourd dans la décision de non-normalité. Pour le lecteur intéressé, nous avons préparé un second tableau, le tableau 6 (basé sur nos données

dites d'efficacité), lequel reproduit les estimations du tableau 4, cette fois pour des séries de $n = 50$ données, la taille maximale considérée dans l'étude de SHAPIRO et coll. (1968). Le lecteur constatera que, encore ici, la taille de 50 laisse à désirer, ne permettant de détecter que des cas de non-normalité grave (par exemple, un *t* à 2 degrés de liberté, une lognormale dont les indices $\gamma_1(\sim 1,4)$ et $\gamma_2(\sim 3,6)$ sont très élevés, etc., et échappant aussi plusieurs cas d'autres lois.

Considérant maintenant le tableau 5, on peut constater que, globalement, les tests *W* et D_{zz}^2 réagissent correctement à la non-normalité, se livrant une bonne concurrence. De plus, pour les lois symétriques, le test g_2 et, pour les lois asymétriques, le test g_1 sont généralement fiables, un constat qui fait d'eux de bons tests spécifiques mais qui les disqualifie au titre de tests universels.

Efficacités

L'examen détaillé et comparatif des graphiques d'efficacité (Annexe 3) nous donne une image plus claire du mérite comparatif des procédures de test : la capacité de détection d'un test est exprimée ici par la taille n (entre 20 et 1000) à laquelle il atteint la puissance $P = 0,80$ (au seuil $\alpha = 0,05$), pour chaque palier de valeur paramétrique de la loi considérée : plus basse est la taille requise, plus le test est efficace.

Cet examen nous amène aux constatations suivantes :



Tableau 2 ■ Variantes et caractérisation des 5 lois asymétriques retenues

6. Lognormale (voir texte)												
σ	0,0665	0,1320	0,1956	0,2564	0,3143	0,3686	0,4194	0,4667	0,5106	0,5514	1	
γ_1	0,2	0,4	0,6	0,8	1	1,2	1,4	1,6	1,8	2	6,185	
γ_2	0,071	0,286	0,647	1,158	1,830	2,665	3,676	4,872	6,263	7,864	110,9	
7. Khi-deux (χ^2)												
ν	1	2	3	5	7	10	20	50	100	150	200	250
γ_1	2,828	2	1,633	1,265	1,069	0,894	0,632	0,400	0,283	0,231	0,200	0,179
γ_2	12	6	4	2,4	1,714	1,200	0,600	0,240	0,120	0,080	0,060	0,048
8. Bêta asymétrique, $B(a = 2, b)$												
b	2	3	4	5	6	7	8	9	10			
γ_1	0	0,286	0,468	0,596	0,693	0,768	0,829	0,879	0,921			
γ_2	-0,857	-0,643	-0,375	-0,120	0,109	0,312	0,490	0,648	0,789			
9. Bi-Nor ($\sigma_1/\sigma_2 \neq 1$) inhomogène : $0,5 \times N_1(0, 1) + 0,5 \times N_2(\delta, \varepsilon)$												
$\delta; \varepsilon$	0,5; 2	1; 2	2; 2	3; 2	4; 2	5; 2	0,5; 3	1; 3	2; 3	3; 3	4; 3	5; 3
γ_1	0,274	0,493	0,687	0,652	0,543	0,435	0,263	0,499	0,816	0,922	0,899	0,795
γ_2	1,027	0,876	0,388	-0,150	-0,598	-0,932	1,873	1,737	1,278	0,721	0,198	-0,238
10. Bi-Nor "ghettoisée" : $0,8 \times N_1(0, 1) + 0,2 \times N_2(\delta, \frac{1}{2})$												
δ	0,5	1	1,5	2	2,5	3	3,5	4				
γ_1	-0,200	-0,260	-0,162	0,026	0,238	0,436	0,606	0,747				
γ_2	0,205	-0,153	-0,457	-0,611	-0,637	-0,591	-0,514	-0,430				

Subbotin ($\delta < 1$). D_{zz}^2 et W dominant,¹³ avec avantage pour W ; g_2 les accompagne. $C_{g_1g_2}$ suit de loin. g_1 est hors-jeu.

Subbotin ($\delta > 1$). D_{zz}^2 domine, et g_2 l'accompagne. $C_{g_1g_2}$ suit, et W suit de loin. g_1 est hors-jeu.

t de Student. D_{zz}^2 domine, accompagné par g_2 et $C_{g_1g_2}$. W suit de loin. g_1 est marginal.

Bêta(a, a). W domine. g_2 , puis D_{zz}^2 suivent, comme $C_{g_1g_2}$ de loin. g_1 est hors-jeu.

Bi-Nor centrée. D_{zz}^2 domine, accompagné de g_2 et $C_{g_1g_2}$. W suit de loin. g_1 est hors-jeu.

Bi-Nor décentrée. D_{zz}^2 et W dominant, avec avantage pour W . g_2 les accompagne. $C_{g_1g_2}$ suit de loin. g_1 est hors-jeu.

Lognormale. D_{zz}^2 et W dominant, avec un avantage précoce pour D_{zz}^2 . g_1 et $C_{g_1g_2}$ les accompagnent, suivis de loin par g_2 .

Khi-deux. D_{zz}^2 et W dominant, W perdant la course pour ν plus élevés. g_1 épouse D_{zz}^2 , $C_{g_1g_2}$ suit et, plus loin, g_2 est hors-jeu.

Bêta (2, b). D_{zz}^2 et W dominant, avec avantage pour W . $C_{g_1g_2}$ et g_1 suivent.

Bi-Nor inhomogène. D_{zz}^2 et W dominant dans les deux cas ($\sigma_2/\sigma_1 = 2$ et 3), avec avantage précoce pour D_{zz}^2 et tardif pour W . $C_{g_1g_2}$ les accompagne et g_1 suit. g_2 est hors-jeu.

Bi-Nor « ghettoisée ». D_{zz}^2 et W dominant, avec avantage précoce pour D_{zz}^2 et tardif pour W . $C_{g_1g_2}$ les accompagne et g_1 suit. g_2 est hors-jeu.

Pour l'ensemble des tests comparés, il ressort avec évidence que seuls les tests D_{zz}^2 et W peuvent être qualifiés d'universels, puisqu'ils réagissent adéquatement à toutes les formes de non-normalité envisagées. Ce n'est pas le cas des autres tests, g_1 , g_2 et $C_{g_1g_2}$, dont le mérite varie selon le type de non-normalité rencontrée. On remarque notamment le simple test g_1 , lequel remporte la palme pour les lois Khi-deux et Lognormale, faisant la barbe à D_{zz}^2 dans ces deux cas. À signaler aussi, l'excellent comportement du test g_2 pour les lois symétriques, faisant aussi la barbe à D_{zz}^2 pour le t de Student.

Le recours à diverses formes de lois binormales, une primeur dans ce type d'étude, a permis d'établir qu'il est possible de diagnostiquer une série statistique en vérifiant l'hétérogénéité : l'échantillon observé émane-t-il d'une seule population, ou reflète-t-il une mixture de deux

13. Le mot « domine » signifie d'abord que l'un des tests D_{zz}^2 et W , ou les deux, fait ou font partie du peloton de tête, sans en exclure explicitement les autres tests.



Tableau 3 ■ Série statistique de $n = 139$ données

186,39	231,63	199,60	211,35	205,00	194,18	194,06
190,00	188,61	194,00	208,00	182,85	204,00	180,16
199,72	197,20	206,31	215,11	212,93	190,00	231,00
234,31	176,90	241,85	223,04	229,68	225,00	218,00
196,00	196,25	194,73	216,79	204,00	200,11	204,07
235,02	221,00	204,01	192,66	195,20	201,70	201,79
198,18	173,00	206,02	206,01	205,00	220,05	223,23
191,97	268,00	202,00	194,07	197,65	216,15	244,57
199,62	176,59	184,78	199,50	184,00	227,48	233,45
224,44	189,26	211,99	202,50	194,31	213,00	212,76
249,26	167,00	223,43	180,00	186,00	222,89	179,15
219,72	210,52	237,00	216,82	175,63	214,00	194,00
222,00	211,00	232,14	190,74	176,08	204,63	293,74
214,00	200,00	264,14	242,59	220,22	198,33	199,92
222,21	229,00	223,75	175,00	179,00	182,80	223,00
232,00	222,00	185,10	239,76	296,39	184,00	195,27
198,00	259,04	230,41	201,63	209,00	209,18	203,55
225,00	217,07	213,65	185,45	221,76	259,56	214,08
213,53	246,85	182,51	202,00	264,00	248,14	214,46
219,35	223,29	243,40	245,08	196,72	228,98	

ou plusieurs populations à paramètres distincts ? Ici aussi, le test D_{zz}^2 détecte plus vite la mixture, W le rejoignant ou le dépassant pour une hétérogénéité plus grande.

Entre D_{zz}^2 et W , l'avantage de détection revient à W pour la loi Bêta, une loi doublement bornée ; c'est aussi le cas pour la loi Binormale décentrée. Dans tous les autres cas, les deux tests s'équivalent à peu près, montrant cependant un avantage précoce de D_{zz}^2 (à l'orée de la non-normalité) et parfois un avantage tardif de W (dans la non-normalité plus prononcée). Globalement, au titre de test universel ou test omnibus de non-normalité et comme compétiteur du test W de SHAPIRO et WILK (1965), le nouveau test D_{zz}^2 s'en tire bien, ayant plus souvent l'avantage sur le W et se montrant précocement plus sensible dans plusieurs cas.

Conclusion

Le test D_{zz}^2 se présente comme une mesure descriptive de non-normalité, son élément étant la différence entre chaque statistique d'ordre de la série testée ($z_{[i]}$) et la valeur probable d'une série normale de taille égale (c_i), soit ($z_{[i]} - c_i$). Cet élément admettra tout aussi facilement un autre modèle de référence, par exemple un Khi-deux, une Bêta, un F , le modèle choisi fournissant alors sa série c_i et ses valeurs critiques. Le test reste intuitivement simple, très performant, et ses valeurs critiques, soit modélisées, soit interpolées, sont commodément calculables.

Il reste bien sûr d'autres formes de non-normalité à considérer, d'autres tests à mettre en compétition, de

même que l'usage qui, à la longue, fera apparaître les avantages et sans doute les faiblesses comparatives de ce test. Néanmoins, au risque de chatouiller leurs droits d'auteurs, nous pouvons paraphraser Shapiro, Wilk et Chen (1968, p. 1371) dans la première conclusion de leur étude, en affirmant :

Les statistiques D_{zz}^2 et W fournissent un test omnibus supérieur de la non-normalité d'une série, tel que jugé à partir de diverses densités symétriques, asymétriques, bornées, semi-bornées ou non bornées et pour toutes les tailles de série considérées, le D_{zz}^2 montrant généralement une plus grande sensibilité.

Que demander de mieux ?

Références

BAILEY, R. W. (1994). Polar generation of random variates with the t-distribution. *Mathematics of Computation*, 62, 779–781.

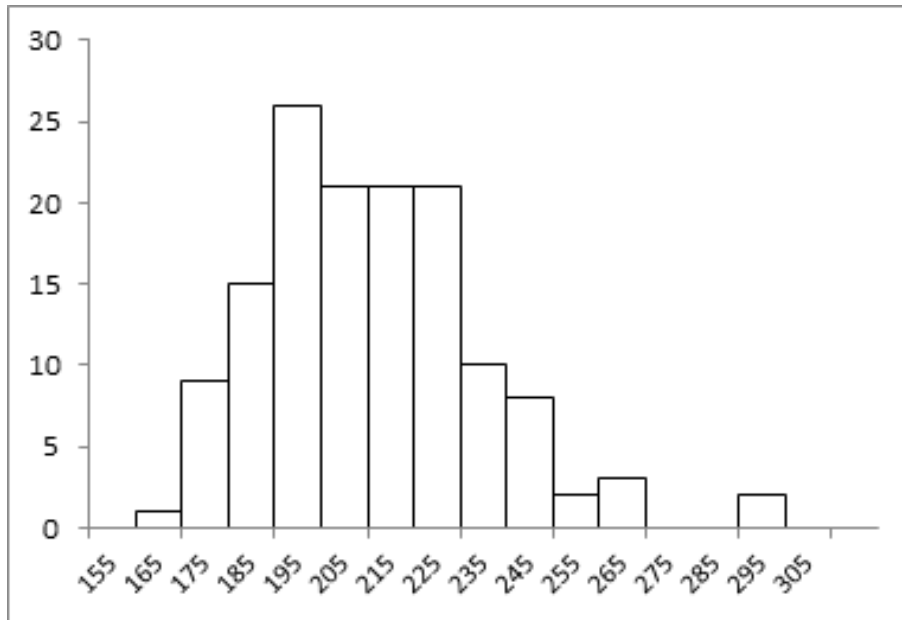
BOX, G. E. P. & TIAO, G. C. (1964). A Bayesian approach to the importance of assumptions applied to the comparison of variances. *Biometrika*, 51, 155–167.

CHENG, R. C. H. (1978). Generating Beta variates with nonintegral shape parameters. *Communications of the ACM*, 21, 317–322.

CHENG, R. C. H. & FEAST, G. M. (1979). Some simple Gamma variate generators. *Applied Statistics*, 28, 290–295.



FIGURE 1 ■ Histogramme des 139 données énumérées au Tableau 3



D'AGOSTINO, R. B., BÉLANGER, A. & D'AGOSTINO, R. B. J. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44, 316–321.

DAVID, H. A., HARTLEY, H. O. & PEARSON, E. S. (1954). The distribution of the ratio in a normal sample, of range to standard deviation. *Biometrika*, 41, 482–493.

DEVROYE, L. (1986). *Non-uniform random variate generation*. New York : Springer-Verlag.

HARTER, H. L. (1961). Expected values of normal order statistics. *Biometrika*, 48, 151–165.

HOWELL, D. C. (2008). *Méthodes statistiques en sciences humaines*. Bruxelles : De Boeck.

KENDALL, M. & STUART, A. (1979). *The advanced theory of statistics, vol. 2 : Inference and relationship (4e édition)*.

LAURENCELLE, L. (2001). *Hasard, nombres aléatoires et méthode Monte Carlo*. Québec : Presses de l'Université du Québec.

LILLIFORS, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399–402.

ROYSTON, J. P. (1982). An extension of Shapiro-Wilk's W test for normality to large samples. *Applied Statistics*, 31, 115–124.

SARKADY, K. (1975). The consistency of the Shapiro-Francia test. *Biometrika*, 62, 445–450.

SHAPIRO, S. S. & FRANCIA, R. S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association*, 67, 215–216.

SHAPIRO, S. S. & WILK, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.

SHAPIRO, S. S. & WILK, M. B. (1968). Approximations to the null distribution of the W statistic. *Technometrics*, 10, 861–866.

SHAPIRO, S. S., WILK, M. B. & CHEN, H. J. (1968). A comparative study of various tests for normality. *Journal of the American Statistical Association*, 63, 1343–1372.

STIGLER, S. M. (1990). *The History of Statistics : The Measurement of Uncertainty before 1900*. Harvard : Belknap Press of Harvard University Press.

SUBBOTIN, M. T. (1923). On the law of frequency of errors. *Mathematicheskii Sbornik*, 31, 296–301.

ULRICH, G. (1984). Computer generation of distributions on the m-sphere. *Applied Statistics*, 33, 158–163.



Tableau 4 ■ Niveau paramétrique approximatif de détection des seuils de puissance 0,50 et 0,80 pour un seuil de significativité $\alpha = 0,05$, dans des séries de $n = 20$ données

Loi	Paramètre	$P = 0,50$					$P = 0,80$				
		D_{zz}^2	W	g_1	g_2	$C_{g_1g_2}$	D_{zz}^2	W	g_1	g_2	$C_{g_1g_2}$
Subbotin <1	δ	-	-	-	-	-	-	-	-	-	-
Subbotin >1	δ	2,65	2,93	-	2,89	2,87	>3	>3	>3	>3	>3
t de Student	ν	2,3	2,1	2,0	2,3	2,3	1,2	1,3	-	1,2	1,2
B(a, a)	a	<1	<1	<1	<1	<1	<1	<1	<1	<1	<1
B-N centrée	σ_2/σ_1	>4	>4	>4	>4	>4	>4	>4	>4	>4	>4
B-N décentrée	$\mu_2 - \mu_1$	>4	>4	>4	>4	>4	>4	>4	>4	>4	>4
Lognormale	σ normal	,47	,47	,48	0,49	,49	,79	,76	,85	>1	>1
χ^2	ν	4	4	4	2	2	2	2	2	-	1
B(2, b)	b	>10	>10	>10	>10	>10	>10	>10	>10	>10	>10
B-N inhom 2	$\mu_2 - \mu_1$	>5	>5	>5	>5	>5	>5	>5	>5	>5	>5
B-N inhom 3	$\mu_2 - \mu_1$	4,38	3,21	>5	>5	>5	>5	>5	>5	>5	>5
B-N ghetto	$\mu_2 - \mu_1$	3,81	3,61	>4	>4	>4	>4	>4	>4	>4	>4

Note. La valeur fournie est obtenue par interpolation (linéaire) simple. Le demi-tiret (-) indique que le seuil de puissance n'est pas accessible dans le domaine paramétrique considéré (voir Tableau 1).

Tableau 5 ■ Niveau paramétrique de détection des seuils de puissance 0,50 et 0,80 pour un seuil de significativité $\alpha = 0,05$, dans des séries de $n = 100$ données

Loi	Paramètre	$P = 0,50$					$P = 0,80$				
		D_{zz}^2	W	g_1	g_2	$C_{g_1g_2}$	D_{zz}^2	W	g_1	g_2	$C_{g_1g_2}$
Subbotin <1	δ	0,36	0,57	-	0,49	-	0,21	0,35	-	0,29	-
Subbotin >1	δ	1,57	1,95	2,30	1,54	1,59	1,94	2,41	>3	1,94	2,04
Student	ν	6,1	4,0	4,0	6,7	6,5	3,7	2,7	2,1	4,0	3,8
B(a, a)	a	1,6	2,6	-	2,1	-	1,2	1,9	-	1,6	-
B-N centrée	σ_2/σ_1	2,2	2,9	>4	2,1	2,2	2,8	3,8	>4	3,0	3,5
B-N décentrée	$\mu_2 - \mu_1$	2,9	2,5	-	2,7	3,9	3,3	2,9	-	3,1	>4
Lognormale	σ normal	0,19	0,20	0,17	0,32	0,22	0,29	0,28	0,31	0,49	0,33
χ^2	ν	26	25	32	7	18	12	12	13	3	8
B(2, b)	b	3,5	2,0	4,2	-	5,2	4,9	3,5	5,8	-	8,4
B-N inhom 2	$\mu_2 - \mu_1$	0,64	1,30	0,98	~	0,91	1,92	2,11	2,25	~	3,71
B-N inhom 3	$\mu_2 - \mu_1$	<0,50	<0,50	0,89	<0,50	<0,50	<0,50	1,53	1,86	↓	1,24
B-N ghetto	$\mu_2 - \mu_1$	2,50	2,13	3,11	>4	3,00	2,87	2,63	3,40	~	3,43

Note. La valeur fournie est obtenue par interpolation simple. Le demi-tiret (-) indique que le seuil de puissance n'est pas accessible dans le domaine paramétrique considéré (voir Tableau 1); le tilde (~) signale une variation incohérente de la statistique g_2 et la flèche descendante (↓) reflète une variation contraire de g_2 .



Annexe 1 : Tableaux de références (tableaux 6 et 7)

Tableau 6 ■ Niveau paramétrique de détection du seuil de puissance 0,80 au seuil de significativité $\alpha = 0,05$, dans des séries de $n = 50$ données (La valeur fournie est obtenue par interpolation)

Loi	Paramètre	$P = 0,80$				
		D_{zz}^2	W	g_1	g_2	C_{g_1, g_2}
Subbotin <1	δ	0,08	-	-	-	-
Subbotin >1	δ	2,50	2,94	-	2,61	2,79
Student	ν	2,4	2,0	2,5	3,5	3,4
B(a, a)	a	-	1,10	-	-	-
B-N centrée	σ_2/σ_1	>4	>4	-	>4	>4
B-N décentrée	$\mu_2 - \mu_1$	4,00	3,61	-	3,89	>4
Lognormale	σ normal	0,419	0,406	0,414	0,797	0,527
χ^2	ν	5,8	5,6	5,5	1,6	3,2
B(2, b)	b	>10	9,5	>10	-	>10
B-N inhomo 2	$\mu_2 - \mu_1$	4,84	3,80	-	-	>5
B-N inhomo 3	$\mu_2 - \mu_1$	2,35	2,71	-	-	>5
B-N ghetto	$\mu_2 - \mu_1$	3,40	3,20	>4	-	>4

Tableau 7 ■ Grille de valeurs critiques de la statistique D_{zz}^2 (estimations Monte Carlo par centiles basées sur 5×10^6 séries normales standard re-standardisées, pour chaque taille n retenue)

$n \backslash \alpha$	0.10	0.05	0.01	0.001
10	11.86	14.75	21.78	31.89
15	20.84	25.81	38.14	56.86
20	30.39	37.52	55.19	82.35
25	40.39	49.69	72.98	108.77
30	50.64	62.19	90.90	135.48
40	71.89	87.88	127.44	189.49
50	93.84	114.35	165.14	244.59
60	116.31	141.39	202.85	299.27
70	139.15	168.89	241.81	356.01
80	162.49	196.74	280.70	410.39
90	185.97	224.93	319.53	467.36
100	209.46	252.97	359.46	523.72
125	269.82	324.75	458.08	663.09
150	331.42	397.60	559.43	807.18
175	393.22	471.25	659.92	948.18
200	456.54	546.16	762.05	1088.52
250	584.11	697.06	968.07	1378.36
300	713.69	849.24	1174.21	1671.64
350	844.53	1003.50	1382.22	1956.74
400	976.88	1158.77	1592.63	2247.36
450	1110.00	1314.37	1802.80	2535.25
500	1244.02	1471.93	2013.35	2825.34
600	1514.92	1789.11	2436.85	3412.81
700	1789.11	2110.66	2863.19	4002.37
800	2065.16	2430.96	3289.39	4584.26
900	2341.88	2755.36	3723.51	5150.78
1000	2621.52	3081.81	4159.86	5758.69



Annexe 2 : Graphiques de puissance (figures 2, 3 et 4)

FIGURE 2 ■ Puissance pour les quatre premières distributions testées

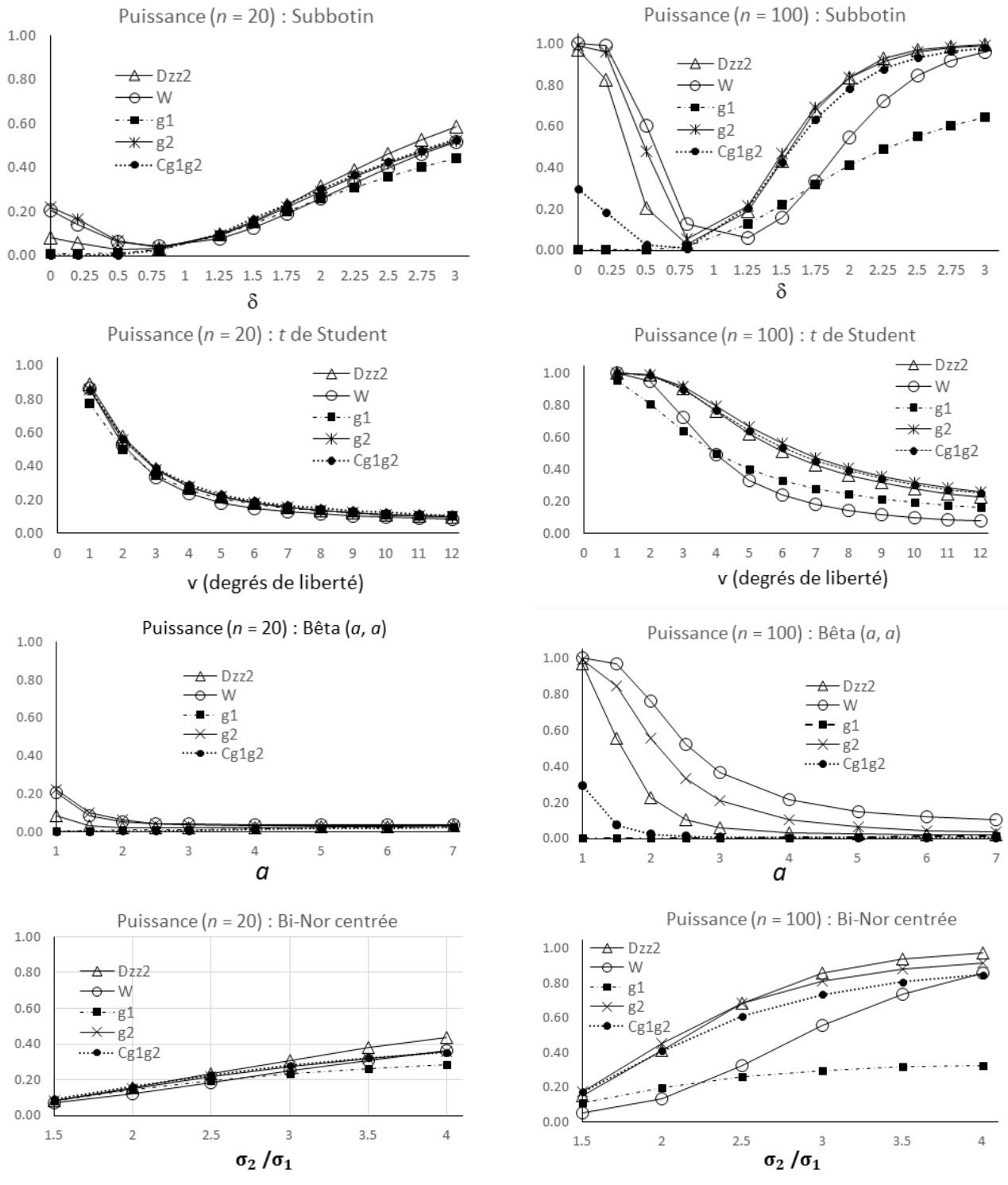
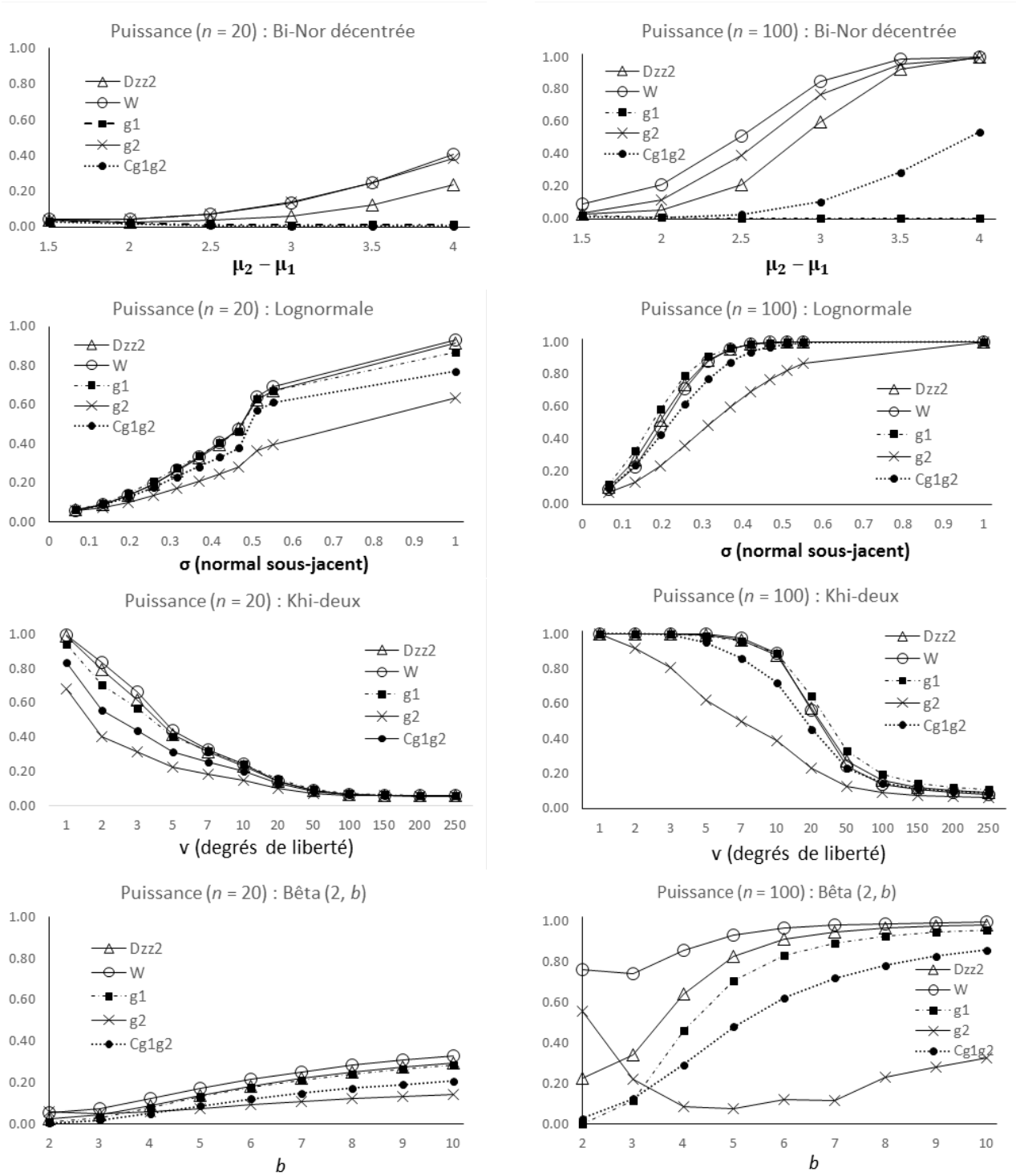




FIGURE 3 ■ Puissance pour les quatre distributions testées suivantes



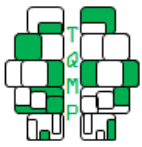
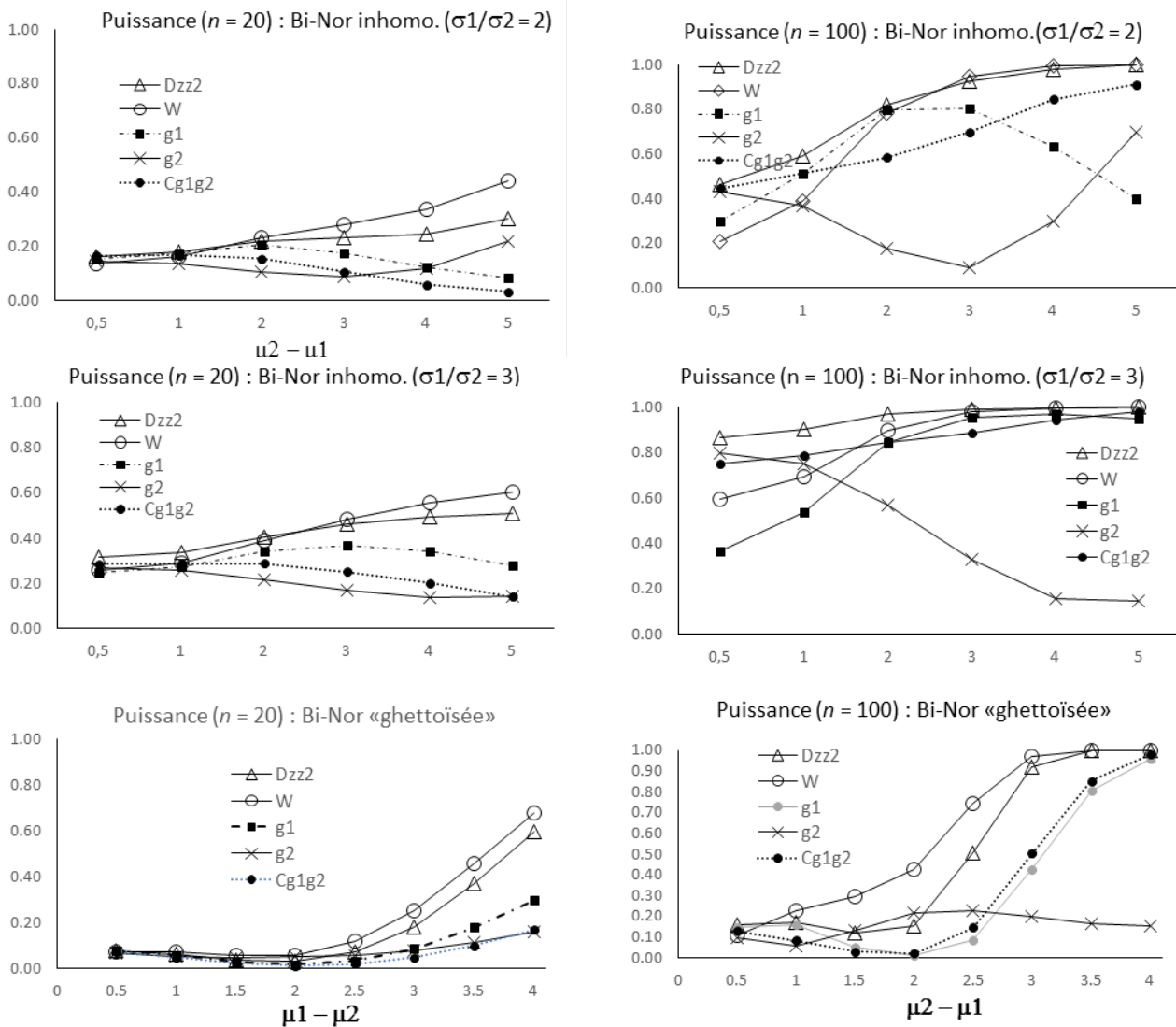
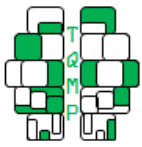


FIGURE 4 ■ Puissance pour les trois dernières distributions testées





Annexe 3 : Graphiques d'efficacité (figures 5 et 6)

FIGURE 5 ■ Efficacité pour les six premières distributions testées

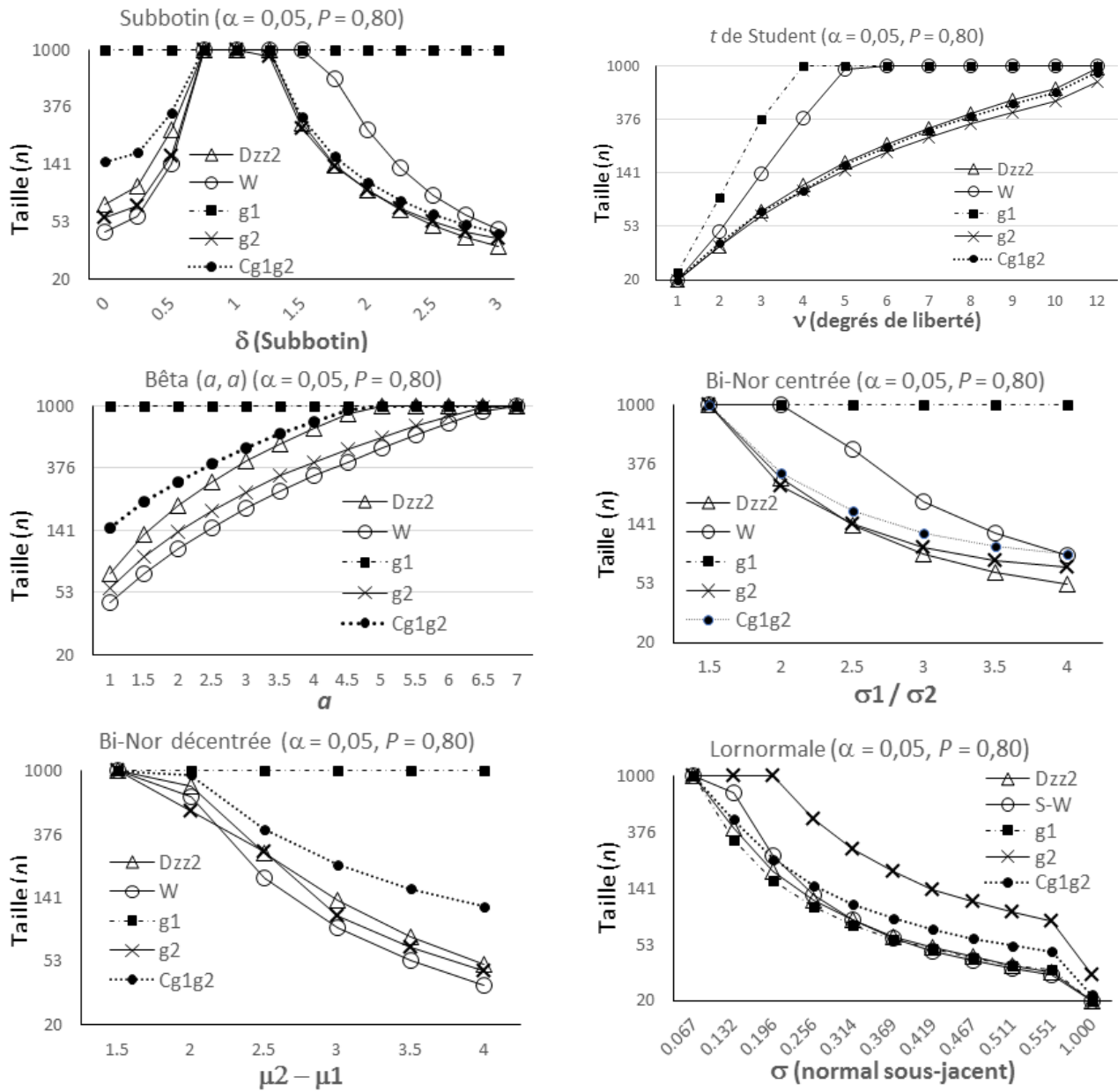
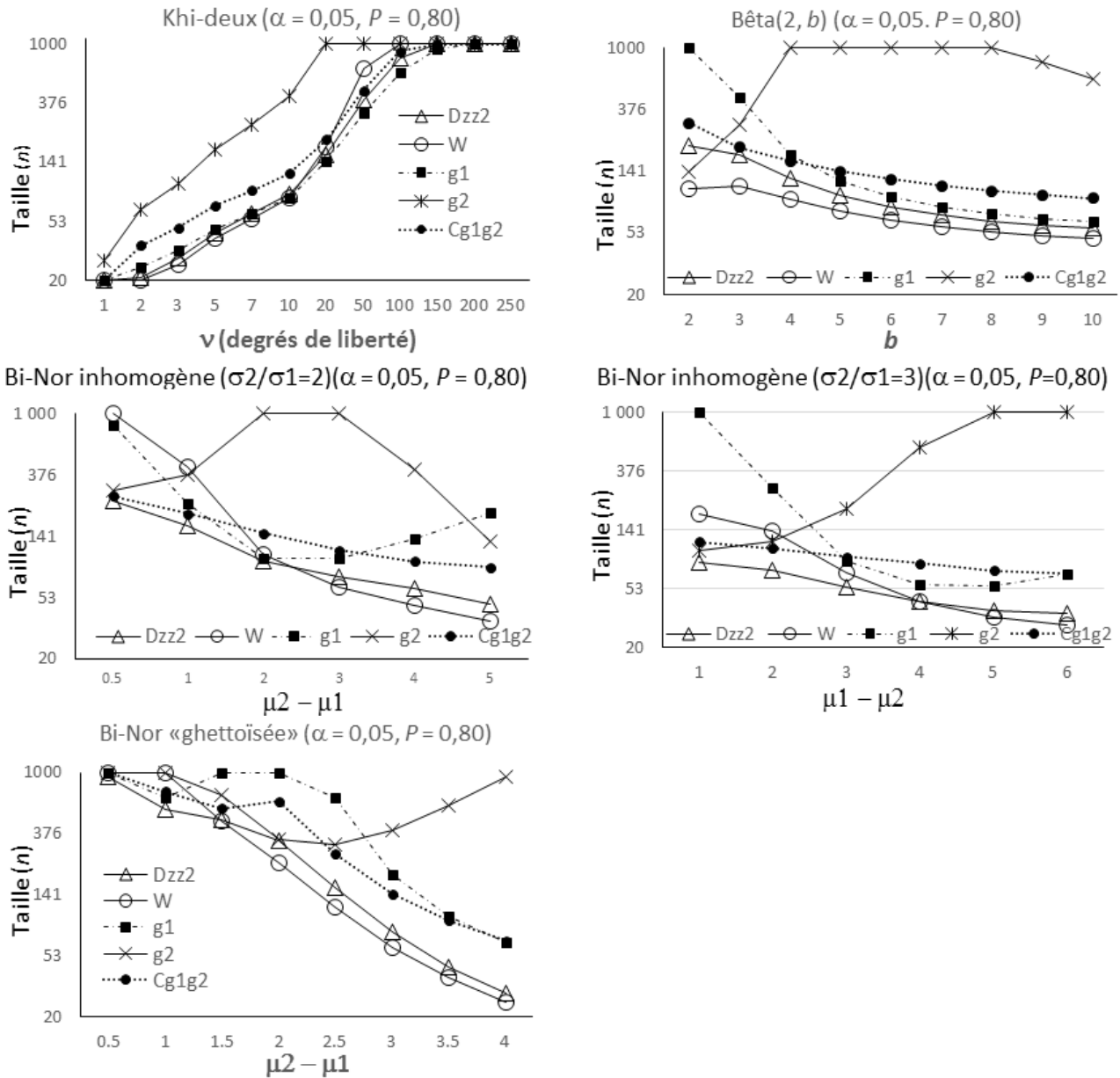
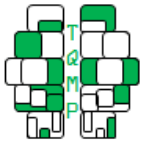




FIGURE 6 ■ Efficacité pour les cinq dernières distributions testées





Annexe 4 : Détails des calculs pour les coefficients de Royston

Les coefficients a_i incorporés dans la formule de calcul du W selon la solution de Royston (1982) découlent d'une approximation de la valeur d'une statistique d'ordre normale tirée d'une série de taille n , puis leur calcul procède par les étapes suivantes :

1. Soit $c_i = \varepsilon z_{[i:n]} = f(i; n)$, pour $i = 2$ à $n - 1$. La valeur originale de c_i (dans SHAPIRO & WILK, 1965) était l'espérance (exacte) de $z_{[i:n]}$ dans le modèle normal standard, à quoi nous proposons de substituer l'approximation citée de HARTER (1961), soit $c_i \approx \Phi^{-1}((i - 0,4)/(n + 0,2))$, au lieu de l'approximation de ROYSTON (1982), plus complexe et globalement équivalente. Le recours à l'espérance de $z_{[i:n]}$ pour le test Shapiro-Wilk ou ses avatars suppose que cette valeur est, globalement, la plus représentative des distributions respectives des statistiques d'ordre de la loi normale. Or, la densité des statistiques d'ordre est elle-même non-normale et, notamment, asymétrique (sauf pour $i = \frac{1}{2}(n + 1)$, n impair), de sorte que la médiane (plutôt que l'espérance), par exemple, en constituerait une meilleure valeur typique. Nos expérimentations Monte Carlo sur la distance entre une série normale aléatoire et un estimateur fixe de $z_{[i:n]}$, distance moyennée pour toutes les positions i et des milliers de séries, est la plus réduite en recourant à l'approximation Harter appliquée ici.
2. Ensuite, pour $i = 2$ à $n - 1$, on obtient $b_i = 2 \times c_i$, et $S = \sum_2^{n-1} c_i^2$.
3. Enfin, $b_1 = -b_n = \sqrt{t(n')/(1 - 2t(n')) \times S}$, où $t(n') = \Gamma(1/2(n' + 1))/(\sqrt{2}\Gamma(1/2n' + 1))$, avec $n' = n$ si $n > 20$ ou $n' = n - 1$ sinon.
4. Obtenir finalement $a_i = b_i / \sqrt{(\sum_1^n b_i^2)/(n - 1)}$ pour tous i .

Citation

LAURENCELLE, L. (2018). Dzz2 ou W pour tester la normalité des données d'un échantillon. *The Quantitative Methods for Psychology*, 14(1), 55–72. doi :[10.20982/tqmp.14.1.p055](https://doi.org/10.20982/tqmp.14.1.p055)

Copyright © 2018, LAURENCELLE. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Received: 06/10/2017 ~ Accepted: 18/12/2017