

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR

MÉLINA MAILHOT

PUISSANCES ASYMPTOTIQUES ET À TAILLES FINIES DE
TESTS DE NORMALITÉ SOUS DES ALTERNATIVES
LOCALES

NOVEMBRE 2009

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

REMERCIEMENTS

Premièrement, je tiens à remercier sincèrement mon directeur de recherche, M. Jean-François Quessy, professeur au Département de mathématiques et d'informatique de l'Université du Québec à Trois-Rivières. Il m'a permis d'explorer la statistique à travers un sujet fort intéressant qui a su combiner nos intérêts scientifiques respectifs. Il a su guider mes travaux par ses remarques pertinentes, en plus de m'encourager et me soutenir durant toutes les étapes de réalisation de ce mémoire. La confiance qu'il m'a accordée, ainsi que sa grande disponibilité, ont contribué à la réussite de ma maîtrise. Sa capacité à susciter la motivation des étudiants fut exemplaire pour moi.

Je suis également reconnaissante envers Alain Desgagné, professeur au Département de mathématiques de l'Université du Québec à Montréal. J'ai grandement apprécié travailler avec lui au début de ma formation à la maîtrise, ce qui m'a permis de bénéficier de ses judicieux conseils. Il m'a accordé beaucoup de temps et m'a appris beaucoup en statistique et en actuariat. Il m'a également beaucoup motivé afin que je complète mes études à la maîtrise.

De plus, je désire souligner les encouragements constants des membres de ma famille tout au long de mes études. J'en profite aussi pour remercier mes amis, qui m'ont toujours supportée et permis de me divertir même dans les moments les plus difficiles.

Un remerciement spécial va à mon amoureux, Fabien Girard. Il m'a beaucoup encouragée, soutenue et conseillée. Sa présence a contribué à élaborer un mémoire dont je suis très fière.

Également, je souhaite remercier Mhamed Mesfioui et Sébastien Tremblay, professeurs au *département de mathématiques et d'informatique appliquées* de l'UQTR, pour avoir accepté de lire et d'évaluer mon mémoire. Les commentaires qu'ils m'ont transmis ont permis d'améliorer la version finale.

Une partie des travaux de ce mémoire a été financée par des octrois individuels accordés à M. Jean-François Quessy par le *Conseil de Recherche en Sciences Naturelles et en Génie* du Canada, ainsi que par le *Fonds Québécois de Recherche sur la Nature et les Technologies*. Je remercie également l'*Institut des Sciences Mathématiques du Québec* pour les bourses d'études qu'il m'a accordées tout au long de ma maîtrise.

AVANT-PROPOS ET STRUCTURE DU MÉMOIRE

Ce mémoire de maîtrise est composé de six chapitres. Les cinq premiers sont rédigés en français. Le chapitre 6 comprend l'intégral d'un article scientifique soumis à la revue *Journal of Statistical Planning and Inference*.

Ma contribution principale à ce mémoire se situe au niveau des cinq premiers chapitres. Ceux-ci donnent une introduction au problème de l'adéquation à la distribution normale ainsi qu'une revue de littérature assez complète sur les tests existants. J'ai également conçu et mis en œuvre un plan de simulation pour évaluer la puissance de plusieurs de ces tests.

Spécifiquement, le chapitre 1 introduit le lecteur à la problématique des tests de normalité. On y présente le problème de l'adéquation à la distribution normale et quelques-unes des caractéristiques les plus importantes de cette loi. Dans les chapitres 2, 3 et 4, trois catégories populaires de tests de normalité sont abordées. Il s'agit respectivement des tests basés sur les moments, des tests basés sur la fonction de répartition et les tests de régression et d'entropie. Le chapitre 5 est dévolu à l'évaluation de la puissance à tailles finies de plusieurs tests présentés aux chapitres 2, 3 et 4. Cette étude de puissance se distingue des nombreuses autres offertes dans la littérature par la considération d'alternatives locales.

Le chapitre 6, constitué d'un article soumis, est en quelque sorte une extension aux cinq premiers chapitres. En effet, au lieu de s'intéresser à la puissance des tests pour des tailles échantillonales finies, l'efficacité asymptotique est considérée. Cela nécessite l'obtention de la loi limite des statistiques de test sous des types particuliers de contre-hypothèses, à savoir des alternatives contiguës. Pour constituer l'article, quelques éléments de ma revue de littérature et de mon étude de simulation ont été repris. Cependant, les résultats asymptotiques, qui constituent le cœur de l'article, sont redevables à mon directeur de recherche, Jean-François Quessy.

La réalisation de l'étude de puissance ainsi que les approximations des courbes de puissance locale ont nécessité de la programmation. Celle-ci a été effectuée à l'aide de MATLAB. L'intégralité des codes informatiques utilisés se retrouve à l'annexe A.

Table des matières

Remerciements	ii
Avant-propos et structure du mémoire	iv
Liste des tableaux	x
Liste des figures	xii
Chapitre 1. Introduction	1
Chapitre 2. Tests basés sur des moments	6
2.1 Coefficients d'asymétrie et d'aplatissement	7
2.2 Test basé sur le coefficient d'asymétrie	12
2.3 Test basé sur le coefficient d'aplatissement	12
2.4 Test de Jarque–Bera	13
2.5 Version robuste du test de Jarque–Bera	13
2.6 Test de Gel et al. (2007)	14
Chapitre 3. Tests basés sur la fonction de répartition	15
3.1 Processus empiriques pour l'adéquation	15
3.2 Le test de Cramér–von Mises	18
3.3 Le test de Anderson–Darling	19
3.4 Le test de Kolmogorov–Smirnov	20

TABLE DES MATIÈRES	vii
Chapitre 4. Autres tests	21
4.1 Tests de régression	21
4.1.1 Idée générale	21
4.1.2 Test de Shapiro–Wilk	22
4.1.3 Test de D’Agostino	23
4.2 Test basé sur l’entropie d’une loi de probabilité	24
4.2.1 Notion d’entropie	24
4.2.2 Test de Vasicek	25
Chapitre 5. Évaluation de la puissance pour des tailles finies	27
5.1 Alternatives locales à la loi normale	27
5.1.1 Mélanges de lois normales	27
5.1.2 Alternative à queues lourdes	29
5.1.3 Une alternative asymétrique	30
5.2 Résultats des simulations	31
5.2.1 Puissance sous des mélanges de lois normales	32
5.2.2 Puissance sous l’alternative à queues lourdes	35
5.2.3 Puissance sous une alternative asymétrique	40
Chapitre 6. Asymptotic power of tests of normality under local alternatives	43
6.1 Introduction	43
6.2 Test statistics for non-normality	46
6.2.1 Tests based on sample skewness and kurtosis	46
6.2.2 The test of Gel et al. (2007)	48
6.2.3 Functionals of an empirical goodness-of-fit process	48
6.3 Asymptotic behavior of the test statistics under local sequences	52

6.4	Three kinds of local alternatives	57
6.5	Asymptotic power and efficiency	61
6.6	Investigation of the asymptotic efficiency of the seven tests . .	66
6.6.1	Local power curves	66
6.6.2	Relative efficiency	69
6.7	Discussion and link with empirical results	73
	Conclusion	78
	Annexe A. Programmes en Matlab	80
A.1	Calcul des statistiques de test	80
A.1.1	Statistique S_n	80
A.1.2	Statistique K_n	81
A.1.3	Statistique JB_n	82
A.1.4	Statistique R_n	83
A.1.5	Statistique CVM_n	84
A.1.6	Statistique AD_n	85
A.1.7	Statistique KS_n	86
A.2	Programmes de simulation selon des alternatives locales	87
A.2.1	Mélanges de lois normales	87
A.2.2	Mélanges de lois normale et Student	88
A.2.3	Alternative asymétrique	89
A.3	Estimation des valeurs critiques	90
A.4	Calculs de puissance	91
A.4.1	Puissance à tailles finies sous des mélanges de normales .	91
A.4.2	Puissance à tailles finies sous des mélanges de normale et Student	92

TABLE DES MATIÈRES	ix
A.4.3 Puissance à tailles finies sous l'alternative asymétrique .	93
A.4.4 Courbes de puissance locale	94
Références	95

LISTE DES TABLEAUX

5.1	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.1) avec $\nu = 1$ et $\gamma = 1$	34
5.2	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.2) avec $\nu = 5$	36
5.3	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.2) avec $\nu = 10$	37
5.4	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.2) avec $\nu = 20$	38
5.5	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.2) avec $\nu = 30$	39
5.6	Estimation, basée sur 10 000 répétitions, de la puissance de $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ et KS_n sous l'alternative (5.3) .	42
6.1	Estimation, based on 100 000 replicates, of the 95-th percentile of the distribution of $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ and KS_n under the normality assumption	51
6.2	Asymptotic relative efficiency of $S_n, K_n, R_n, CVM_n, AD_n$ and KS_n with respect to the Jarque–Bera test statistic JB_n	73

- 6.3 Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 50$ 75
- 6.4 Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 100$ 76
- 6.5 Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 250$ 77

LISTE DES FIGURES

1.1	Histogramme de 2 500 observations	2
1.2	Densités normales d'écart-type $\sigma = 1$ et de moyennes $\mu = 0, 1, 2$	3
1.3	Densités normales de moyenne $\mu = 0$ et d'écart-types $\sigma = 1, 2, 3$	4
2.1	Densités asymétriques à gauche ($\beta_1 < 0$) et à droite ($\beta_1 > 0$)	7
2.2	Densités aplatie (figure de gauche) et gonflée (figure de droite)	8
4.1	Entropie de la densité normale en fonction de l'écart-type σ	25
5.1	Mélanges de lois $\mathcal{N}(0, 1)$ et $\mathcal{N}(\nu, 1)$ pour différentes valeurs de δ . À gauche, $\nu = 1$; à droite, $\nu = 2$	28
5.2	Mélanges de lois $\mathcal{N}(0, 1)$ et $\mathcal{N}(0, 4)$ pour différentes valeurs de δ	29
5.3	Alternative à queues lourdes lorsque g est la densité de Student. À gauche, $\nu = 5$; à droite, $\nu = 10$	30
5.4	Densités asymétriques avec diverses valeurs de δ	31
5.5	Coefficients d'asymétrie (en haut) et d'aplatissement (en bas) en fonction de δ dans le modèle (5.1) quand $\nu = 1$ et $\gamma = 1$	33
5.6	Coefficients d'asymétrie (en haut) et d'aplatissement (en bas) en fonction de δ dans le modèle (5.3)	41
6.1	Local power curves of $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ and KS_n under model \mathcal{A}_1 . Left panel : $\mu = 1$; right panel : $\mu = 2$	67
6.2	Local power curves of $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ and KS_n under model \mathcal{A}_2 . Left panel : $\sigma = 1.5$; right panel : $\sigma = 2$	68
6.3	Local power curves of $S_n, K_n, JB_n, R_n, CVM_n, AD_n$ and KS_n under model \mathcal{A}_3 . Left panel : $\nu = 5$; right panel : $\nu = 10$	69

6.4	Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of μ in model \mathcal{A}_1 . Left panel : $\mu \in [0, 1]$; right panel : $\mu \in [1, 3]$.	70
6.5	Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of $\sigma \in [1, 3]$ in model \mathcal{A}_2	71
6.6	Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of $\nu \in [5, 15]$ in model \mathcal{A}_3	72

CHAPITRE 1

INTRODUCTION

Supposons que l'on observe un phénomène qui fait intervenir une variable aléatoire. Par exemple, la crue annuelle sur une certaine rivière, les montants de réclamation en rapport avec une certaine assurance ou l'âge moyen où les gens contractent une certaine maladie. Pour bien comprendre ce phénomène, une façon efficace consiste à considérer la répartition des valeurs observées. Spécifiquement, on considère l'histogramme de ces observations. Un tel outil, basé sur 2 500 observations, est présenté à la figure 1.1.

Une étape de compréhension supplémentaire consiste à ajuster un modèle théorique à la représentation empirique de l'histogramme. Pour le faire d'une façon formelle et rigoureuse, il s'agit d'employer un test d'adéquation. Spécifiquement, notons f la densité de probabilité inconnue d'une certaine population. On désire vérifier si la densité f_0 fournit un bon modèle théorique pour f , c'est-à-dire s'il explique adéquatement ce qui a été observé. Pour ce faire, les hypothèses nulle et alternative à confronter sont \mathcal{H}_0 : *la densité f est de la forme f_0* et \mathcal{H}_1 : *la densité f n'est pas de la forme f_0* . La littérature statistique traitant de ce problème est très volumineuse.

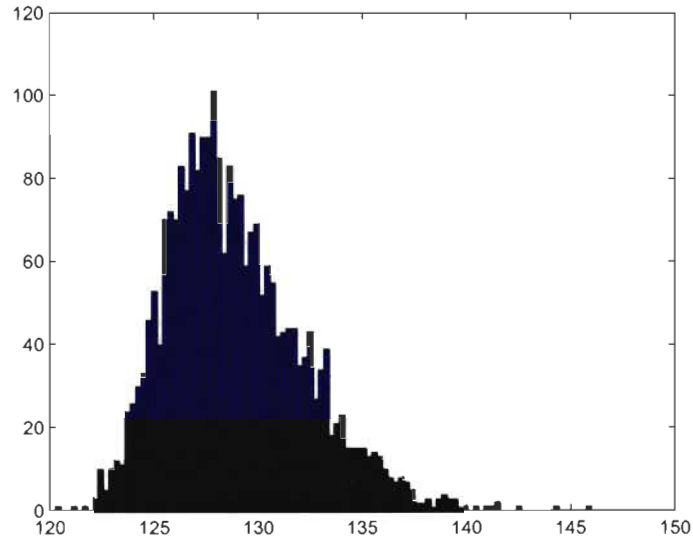


Figure 1.1: Histogramme de 2 500 observations

Un problème d'adéquation qui a retenu particulièrement l'attention concerne la distribution normale. Cette loi de probabilité a en effet connu beaucoup de succès pour la représentation de nombreux phénomènes. Ceci s'explique probablement, en partie du moins, par le théorème de la limite centrale. Celui-ci indique qu'une somme de variables aléatoires dont les variances respectives existent se comporte, asymptotiquement, comme une variable aléatoire de loi normale. Ainsi, comme plusieurs phénomènes que l'on observe sont le résultat d'une sommation, cette distribution est souvent adéquate pour représenter la réalité. À titre d'exemples, on peut citer le nombre de cellules cancéreuses dans un organe ou le montant total des réclamations d'une assurance. La vérification formelle de la normalité est

donc un problème important.

La densité de la loi normale de moyenne $\mu \in \mathbb{R}$ et d'écart-type $\sigma > 0$ est

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\},$$

où $x \in \mathbb{R}$. Des exemples de cette densité sont donnés aux figures 1.2 et 1.3.

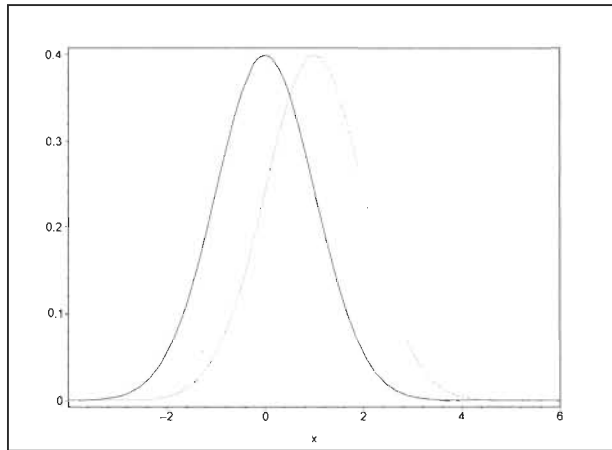


Figure 1.2: Densités normales d'écart-type $\sigma = 1$ et de moyennes $\mu = 0, 1, 2$

Une procédure d'adéquation à la famille de densités $\{\phi_{\mu,\sigma}; \mu \in \mathbb{R}, \sigma > 0\}$ s'appelle un test de normalité. Si f dénote la densité inconnue d'une certaine population, il consiste à confronter les hypothèses

$$\mathcal{H}_0 : f(x) = \phi_{\mu,\sigma}(x) \quad \text{et} \quad \mathcal{H}_1 : f(x) \neq \phi_{\mu,\sigma}(x) \text{ pour au moins un } x \in \mathbb{R}.$$

À noter qu'ici, les paramètres μ et σ seront supposés inconnus. Les hypothèses \mathcal{H}_0 et \mathcal{H}_1 précédentes sont alors qualifiées de *composites*. Celles-ci sont donc beaucoup plus flexibles et réalistes que si on supposait des valeurs

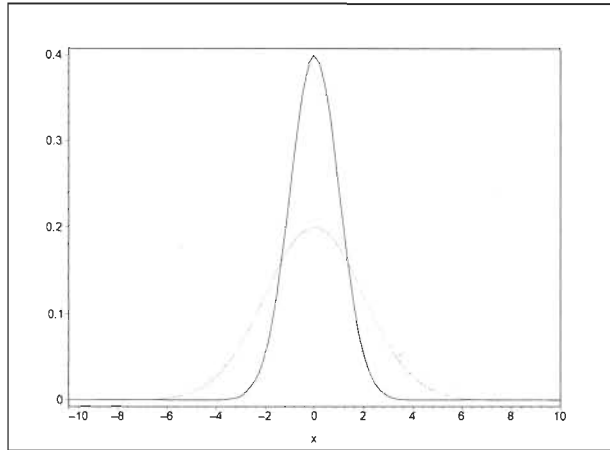


Figure 1.3: Densités normales de moyenne $\mu = 0$ et d'écart-types $\sigma = 1, 2, 3$

μ_0 et σ_0 aux paramètres, puisqu'en pratique ces quantités sont rarement connues. Ainsi, pour toutes les procédures qui seront décrites dans la suite, il faudra estimer μ et σ^2 . Leurs estimateurs naturels sont la moyenne et la variance empiriques, à savoir

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{et} \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2.$$

Les Chapitres 2, 3 et 4 constituent une revue de littérature assez exhaustive sur les tests de normalité. Une attention particulière sera portée aux procédures les plus populaires. De tels recensements se retrouvent, entre autres, dans les travaux de D'Agostino & Stephens (1986) et de Thode Jr. (2002). Dans la suite, on classera les tests de normalité en trois familles :

- (i) les tests basés sur les moments ;
- (ii) les tests basés sur la fonction de répartition ;
- (iii) les tests de régression et d'entropie.

Ces familles constitueront, respectivement, les chapitres 2, 3 et 4.

Afin de choisir un test adéquat parmi la panoplie de procédures disponibles, il convient de sélectionner le plus puissant. On rappelle que la puissance d'un test est sa capacité à rejeter l'hypothèse nulle lorsque celle-ci est fautive. Pour ce faire, on considérera dans ce mémoire des contre-hypothèses dites *locales*. Celles-ci permettent d'évaluer l'efficacité des tests sous de faibles écarts à \mathcal{H}_0 .

Au Chapitre 5, la puissance de plusieurs procédures est évaluée à l'aide de simulations de type Monte-Carlo. Trois alternatives locales sont considérées. Ceci permet d'identifier les tests les plus puissants sous un scénario donné. Au Chapitre 6, qui contient l'intégral d'un article déjà soumis, l'idée de puissance locale est étendue au cas asymptotique, c'est-à-dire lorsque la taille d'échantillon est supposée infinie. Ceci amène à obtenir les lois asymptotiques des statistiques considérées. Cela permet de calculer leurs courbes de puissance asymptotiques locales. Ces dernières renseignent sur les procédures les plus efficaces sous de grandes tailles d'échantillons.

CHAPITRE 2

TESTS BASÉS SUR DES MOMENTS

Soit f , une certaine densité définie sur un domaine $\mathcal{D} \subseteq \mathbb{R}$. C'est donc dire que $f(x) \geq 0$ pour tout $x \in \mathcal{D}$ et

$$\int_{\mathcal{D}} f(x) dx = 1.$$

Pour une variable aléatoire X dont la densité est f , son moment d'ordre k est défini par

$$E(X^k) = \int_{\mathcal{D}} x^k f(x) dx.$$

En particulier, le moment d'ordre 1 est la moyenne, c'est-à-dire que $\mu = E(X)$. Le moment centré d'ordre k de X est

$$\mu_k = E \left\{ (X - \mu)^k \right\}.$$

Le moment centré d'ordre 2 est donc la variance de X .

Dans ce chapitre, on décrira des tests de normalité qui exploitent certaines caractéristiques des moments de la loi normale.

2.1 Coefficients d'asymétrie et d'aplatissement

Soit une variable aléatoire X de densité f , de moyenne μ et de variance σ^2 . Le coefficient d'asymétrie de f est défini par

$$\beta_1 = E \left\{ \left(\frac{X - \mu}{\sigma} \right)^3 \right\}. \quad (2.1)$$

Cet indice correspond au moment standardisé d'ordre trois, c'est-à-dire que $\beta_1 = E(Z^3)$, où

$$Z = \frac{X - \mu}{\sigma}. \quad (2.2)$$

Les valeurs possibles de β_1 sont les réels. Quand $\beta_1 > 0$, on parle d'asymétrie positive. Ceci est la conséquence d'une densité dont la masse de probabilité est plus importante du côté des grandes valeurs de la distribution. De façon similaire, $\beta_1 < 0$ indique une masse de probabilité plus importante pour les petites valeurs de la distribution. La figure 2.1 présente des distributions asymétriques à gauche et à droite.

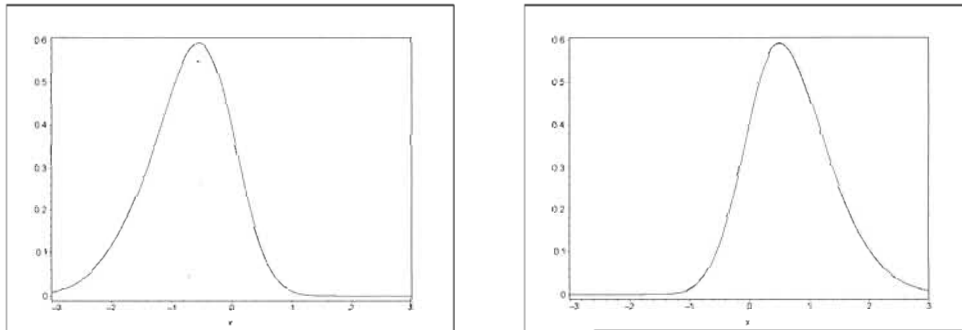


Figure 2.1: Densités asymétriques à gauche ($\beta_1 < 0$) et à droite ($\beta_1 > 0$)

Le coefficient d'aplatissement de f est le moment standardisé d'ordre quatre, c'est-à-dire

$$\beta_2 = \mathbb{E} \left\{ \left(\frac{X - \mu}{\sigma} \right)^4 \right\}. \quad (2.3)$$

Les valeurs possibles de β_2 sont les réels positifs. Quand cet indice est élevé, cela indique que les ailes de la distribution sont lourdes. À l'inverse, une faible valeur de β_2 indique une distribution *gonflée* autour de sa moyenne.

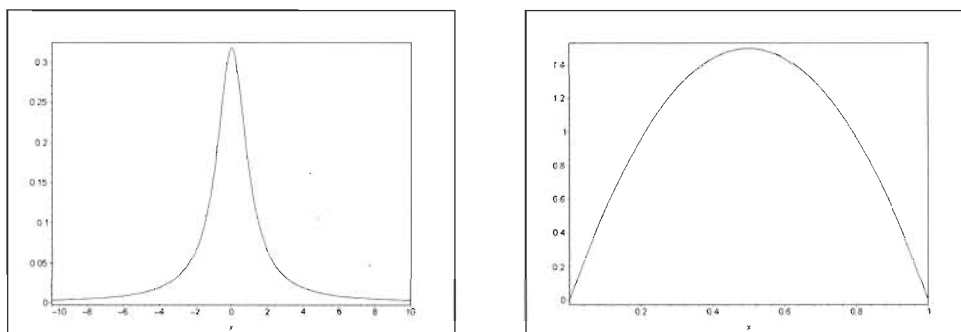


Figure 2.2: Densités aplatie (figure de gauche) et gonflée (figure de droite)

Souvent, on juge de l'asymétrie et de l'aplatissement d'une distribution en rapport avec les valeurs obtenues pour la loi normale. Notons d'abord que si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors sa forme centrée réduite (2.2) est de loi $\mathcal{N}(0, 1)$. On peut donc écrire

$$\beta_1 = \int_{-\infty}^{\infty} x^3 \phi_{0,1}(x) dx.$$

Comme la fonction $g(x) = x^3 \phi_{0,1}(x)$ est impaire, c'est-à-dire que $g(-x) = (-x)^3 \phi_{0,1}(-x) = -x^3 \phi_{0,1}(x) = -g(x)$, on déduit que $\beta_1 = 0$. Pour ce qui est du coefficient d'aplatissement, on note que

$$\beta_2 = \mathbb{E}(Z^4) = \text{var}(Z^2) + \mathbb{E}^2(Z^2).$$

Par un résultat très connu, Z^2 suit une loi khi-carré à 1 degré de liberté. Il s'ensuit que $E(Z^2) = 1$ et $\text{var}(Z^2) = 2$. Ainsi, $\beta_2 = 2 + (1)^2 = 3$.

Dans la suite, les coefficients d'asymétrie et d'aplatissement sont calculés pour les densités uniformes, Beta, exponentielle et double exponentielle. À noter qu'en développant les formules (2.1) et (2.3), on obtient

$$\beta_1 = \frac{E(X^3) - 3\mu E(X^2) + 2\mu^3}{\sigma^3} \quad (2.4)$$

et

$$\beta_2 = \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4}{\sigma^4}. \quad (2.5)$$

Exemple 2.1. Soit une variable aléatoire distribuée uniformément sur l'intervalle (a, b) . Dans ce cas, la densité est donnée par

$$f_{a,b}(x) = \frac{1}{b-a} \mathbb{I}_{[a,b]}(x),$$

où $\mathbb{I}_A(x)$ est la fonction indicatrice d'un ensemble, c'est-à-dire

$$\mathbb{I}_A(x) = \begin{cases} 1, & \text{si } x \in A; \\ 0, & \text{si } x \notin A. \end{cases}$$

Pour une variable aléatoire X distribuée selon cette loi, on a

$$E(X^j) = \frac{b^{j+1} - a^{j+1}}{(j+1)(b-a)}.$$

On en déduit que

$$E(X) = \frac{b^2 - a^2}{2(b-a)}, \quad E(X^2) = \frac{b^3 - a^3}{3(b-a)},$$

$$E(X^3) = \frac{b^4 - a^4}{4(b-a)} \quad \text{et} \quad E(X^4) = \frac{b^5 - a^5}{5(b-a)}.$$

Par une application des formules (2.4) et (2.5), on trouve

$$\beta_1 = 0 \quad \text{et} \quad \beta_2 = -\frac{6}{5}.$$

Exemple 2.2. La densité Beta est de la forme

$$f_{a,b}(x) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} x^{a-1}(1-x)^{b-1} \mathbb{I}_{[0,1]}(x),$$

où

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$$

est la fonction Gamma. La forme de cette densité pour $a = b = 2$ est présentée au graphique droit de la figure 2.2. À noter que l'on retrouve la densité uniforme sur $[0, 1]$ dans le cas particulier où $a = b = 1$. Pour une variable aléatoire $X \sim f_{a,b}$, on a

$$\mathbb{E}(X) = \frac{a}{a+b}, \quad \text{et} \quad \text{var}(X) = \frac{ab}{(a+b)^2(a+b+1)}.$$

Aussi,

$$\mathbb{E}(X^k) = \frac{\Gamma(a+b)\Gamma(a+k)}{\Gamma(a)\Gamma(a+b+k)}.$$

En particulier,

$$\mathbb{E}(X^3) = \frac{a(a+1)(a+2)}{(a+b)(a+b+1)(a+b+2)}$$

et

$$\mathbb{E}(X^4) = \frac{a(a+1)(a+2)(a+3)}{(a+b)(a+b+1)(a+b+2)(a+b+3)}.$$

Finalement, à l'aide des formules (2.4) et (2.5), on trouve

$$\beta_1 = \frac{2(b-a)\sqrt{a+b+1}}{(a+b+2)\sqrt{ab}}$$

et

$$\beta_2 = 6 \left\{ \frac{a^3 - a^2(2b - 1) + b^2(b + 1) - 2ab(b + 2)}{ab(a + b + 2)(a + b + 3)} \right\}.$$

Dans le cas particulier où $a = b$, on déduit

$$\beta_1 = 0 \quad \text{et} \quad \beta_2 = \frac{-6}{2a + 3}.$$

Pour la loi uniforme sur $(0, 1)$, c'est-à-dire quand $a = b = 1$, les formules précédentes donnent $\beta_1 = 0$ et $\beta_2 = -6/5$, en accord avec les calculs de l'exemple 2.1.

Exemple 2.3. Soit une variable X de loi exponentielle de paramètre λ , c'est-à-dire dont la densité est

$$f_\lambda(x) = \frac{1}{\lambda} \exp\left(\frac{-x}{\lambda}\right) \mathbb{I}_{\mathbb{R}^+}(x).$$

Pour cette loi, on a $E(X) = \lambda$, $E(X^2) = 2\lambda^2$, $E(X^3) = 6\lambda^3$ et $E(X^4) = 24\lambda^4$.

Ainsi,

$$E\{(X - \lambda)^3\} = E(X^3) - 3\lambda E(X^2) + 3\lambda^2 E(X) - \lambda^3 = 2\lambda^3$$

et

$$E\{(X - \lambda)^4\} = E(X^4) - 4\lambda E(X^3) + 6\lambda^2 E(X^2) - 4\lambda^3 E(X) + \lambda^4 = 9\lambda^4.$$

Comme $\text{var}(X) = \lambda^2$, on a $\beta_1 = 2$ et $\beta_2 = 9$.

Exemple 2.4. La densité de la loi double exponentielle standard est

$$f(x) = \frac{1}{2} e^{-|x|}.$$

La forme de cette densité est présentée au graphique gauche de la figure 2.2.

On calcule facilement $\mu = 0$ et $\sigma^2 = 2$. Ainsi,

$$\beta_1 = \frac{E(X^3)}{(\sqrt{2})^3} = 0 \quad \text{et} \quad \beta_2 = \frac{E(X^4)}{(\sigma^2)^2} = \frac{24}{4} = 6.$$

2.2 Test basé sur le coefficient d'asymétrie

La version empirique du coefficient d'asymétrie β_1 est donnée par

$$\hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right)^3.$$

Lorsque les observations X_1, \dots, X_n proviennent d'une population de loi normale, le théorème de la limite centrale et la méthode *Delta* (voir van der Vaart & Wellner (1996)) permettent de montrer que $\sqrt{n} \hat{\beta}_1$ converge en loi vers une loi normale de moyenne 0 et de variance 6. Un test de normalité consiste donc à rejeter l'hypothèse nulle lorsque $|S_n| > \Phi^{-1}(1 - \alpha/2)$, où

$$S_n = \frac{\sqrt{n} \hat{\beta}_1}{\sqrt{6}} \quad (2.6)$$

et Φ est la fonction de répartition de la loi normale standard.

2.3 Test basé sur le coefficient d'aplatissement

La version empirique du coefficient d'aplatissement β_2 est donnée par

$$\hat{\beta}_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right)^4.$$

Lorsque les observations X_1, \dots, X_n proviennent d'une population de loi normale, on peut montrer que $\sqrt{n}(\hat{\beta}_2 - 3)$ converge en loi vers une loi normale de moyenne 0 et de variance 24. Un test de normalité consiste donc à rejeter l'hypothèse nulle lorsque $|K_n| > \Phi^{-1}(1 - \alpha/2)$, où

$$K_n = \frac{\sqrt{n}(\hat{\beta}_2 - 3)}{\sqrt{24}}. \quad (2.7)$$

2.4 Test de Jarque–Bera

La statistique du test de Jarque–Bera est

$$JB_n = S_n^2 + K_n^2, \quad (2.8)$$

où S_n et K_n sont définies aux équations (2.6) et (2.7). Un test de normalité basé sur JB_n a été introduit par D’Agostino & Pearson (1973) et popularisé par Jarque & Bera (1980, 1987). Vue sa forme, la statistique de Jarque–Bera est sensible à des écarts des coefficients d’asymétrie et d’aplatissement par rapport aux valeurs associées à la loi normale, à savoir $\beta_1 = 0$ et $\beta_2 = 3$. Sous \mathcal{H}_0 , les statistiques S_n et K_n sont indépendantes et asymptotiquement distribuées selon des normales centrées réduites. Par conséquent, JB_n converge en loi vers une khi-carré à deux degrés de liberté. En pratique, toutefois, il vaut mieux éviter d’utiliser la valeur critique asymptotique, vue la faible vitesse de convergence de cette statistique. On emploie généralement les valeurs critiques estimées par des simulations de type Monte–Carlo.

2.5 Version robuste du test de Jarque–Bera

Une hypothèse implicite dans plusieurs procédures statistiques est l’absence de valeurs aberrantes. En particulier, les conclusions du test de Jarque–Bera peuvent être influencées par la présence de telles valeurs. Ceci a motivé Gel & Gastwirth (2008) à élaborer une version robuste de JB_n , à savoir

$$\widetilde{JB}_n = \frac{n}{C_1} \left(\frac{\hat{\beta}_1}{J_n^3} \right)^2 + \frac{n}{C_2} \left(\frac{\hat{\beta}_2}{J_n^4} - 3 \right)^2,$$

où

$$J_n = \frac{\sqrt{\pi/2}}{n} \sum_{i=1}^n |X_i - M_n|.$$

Ici, C_1, C_2 sont des constantes et M_n est la médiane empirique. Sous \mathcal{H}_0 , la statistique $\widetilde{J}B_n$ a le même comportement asymptotique que JB_n .

2.6 Test de Gel et al. (2007)

L'écart-type σ est habituellement estimé par $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$, où

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2.$$

Néanmoins, il est possible d'estimer σ autrement. En effet, dans le modèle normal, un estimateur sans biais est donné par

$$\tilde{\sigma}^2 = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^n |X_i - \hat{\mu}|.$$

Ainsi, sous l'hypothèse nulle de normalité, le rapport $\hat{\sigma}/\tilde{\sigma}$ devrait être près de 1. Ceci a motivé Gel et al. (2007) à proposer la statistique de test

$$R_n = \sqrt{n} \left(\frac{\hat{\sigma}}{\tilde{\sigma}} - 1 \right). \quad (2.9)$$

Ces auteurs démontrent que R_n converge en loi vers une normale de moyenne nulle et de variance $\sigma_R^2 = (\pi - 3)/2$ sous l'hypothèse de normalité. Ainsi, un test de normalité de niveau de confiance asymptotique $1 - \alpha$ consiste à rejeter l'hypothèse nulle lorsque

$$\frac{|R_n|}{\sigma_R} > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right).$$

CHAPITRE 3

TESTS BASÉS SUR LA FONCTION DE RÉPARTITION

3.1 Processus empiriques pour l'adéquation

La fonction de répartition d'une variable aléatoire X à valeurs dans \mathbb{R} est définie par

$$F(x) = P(X \leq x).$$

Lorsque la loi de X est continue à valeurs dans \mathcal{D} , on a

$$F(x) = \int_{(-\infty, x) \cap \mathcal{D}} f(s) ds,$$

où f est la densité de X . À partir de n copies indépendantes X_1, \dots, X_n de X , on peut estimer F par la fonction de répartition empirique, à savoir

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x).$$

Ainsi, pour tout $x \in \mathcal{D}$, $F_n(x)$ est un estimateur sans biais de $F(x)$ car

$$E\{F_n(x)\} = \frac{1}{n} \sum_{i=1}^n E\{\mathbb{I}(X_i \leq x)\} = \frac{1}{n} \sum_{i=1}^n P(X_i \leq x) = \frac{nF(x)}{n} = F(x).$$

De plus, on note que $nF_n(x)$ est une binomiale de paramètres n et $F(x)$. De là, on déduit que la variance de $F_n(x)$ est $F(x)\{1 - F(x)\}/n$. De plus, si on applique le théorème de la limite centrale, on a que pour tout $x \in \mathcal{D}$ fixé,

$$\mathbb{B}_n^*(x) = \sqrt{n} \{F_n(x) - F(x)\}$$

converge en loi vers la loi normale de moyenne 0 et de variance $F(x)\{1 - F(x)\}$. Pour tester l'adéquation à une fonction de répartition donnée F , une idée consiste à comparer F_n et F pour toutes les valeurs $x \in \mathcal{D}$. Cela amène à considérer $\mathbb{B}_n^*(x)$ en tant que *processus*, c'est-à-dire comme une fonction aléatoire prenant ses valeurs dans un espace de fonctions. Un résultat classique assure que \mathbb{B}_n^* converge en loi vers un processus limite \mathbb{B}^* , où pour $s \in \mathcal{D}$ fixé, la loi de $\mathbb{B}^*(s)$ est normale de moyenne 0 et de variance $F(s)\{1 - F(s)\}$, et $E\{\mathbb{B}^*(s)\mathbb{B}^*(t)\} = \min\{F(s), F(t)\} - F(s)F(t)$.

Plus généralement, supposons que $F(x) = F_\theta(x)$ dépend d'un paramètre inconnu $\theta \in \Theta \subseteq \mathbb{R}^k$. On souhaite tester l'adéquation à la famille paramétrique $\mathcal{F}_\theta = \{F_\theta; \theta \in \Theta\}$ sur la base de F_n . Dans ce cas, Pourvu que l'on dispose d'un estimateur $\hat{\theta}_n$ convergent pour θ , une idée consiste à considérer le processus empirique à paramètres estimés

$$\mathbb{B}_n(x) = \sqrt{n} \{F_n(x) - F_{\hat{\theta}_n}(x)\}.$$

Ce processus a été amplement étudié dans l'élaboration de tests d'adéquation. On pense notamment aux travaux de Stephens (1974, 1976).

Une approche alternative a été suggérée par Durbin (1973). Pour la décrire, soient les pseudo-observations $\hat{T}_{i,n} = F_{\hat{\theta}_n}(X_i)$. Sous l'hypothèse que la loi sous-jacente aux observations appartient à la famille \mathcal{F}_θ , la loi de $\hat{T}_{i,n}$, $i \in$

$\{1, \dots, n\}$ est approximativement uniforme sur $(0, 1)$. Ceci vient du résultat théorique bien connu qui stipule que la loi de $F(X)$ est uniforme sur $(0, 1)$ lorsque $X \sim F$ et que F est continue. Ceci a amené Durbin (1973) à considérer le processus empirique

$$\mathbb{F}_n(x) = \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}(\hat{T}_{i,n} \leq t) - t \right\}. \quad (3.1)$$

Bien que les comportements de \mathbb{B}_n et \mathbb{F}_n soient différents pour des tailles échantillonales finies, leurs distributions asymptotiques sont identiques. Ainsi, quand $n \rightarrow \infty$, \mathbb{B}_n et \mathbb{F}_n convergent vers un certain processus limite \mathbb{F} . Ce résultat est rappelé dans l'article présenté au Chapitre 6, où une technique différente pour le démontrer est offerte.

Dans ce mémoire, l'approche de Durbin (1973) sera privilégiée pour tester la normalité. Pour ce faire, on définit la fonction de répartition d'une variable aléatoire de loi $\mathcal{N}(0, 1)$ par Φ . L'hypothèse de normalité est donc équivalente à l'appartenance de la fonction de répartition à la famille

$$\mathcal{F}_{\mu, \sigma} = \left\{ \Phi \left(\frac{x - \mu}{\sigma} \right); \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \right\}.$$

Les pseudo-observations utilisées seront donc

$$\hat{T}_{i,n} = \Phi \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right), \quad 1 \leq i \leq n.$$

Dans la suite, trois tests d'adéquation sont définis comme fonctionnels du processus (3.1). Ce sont les distances de Cramér-von Mises, Anderson-Darling et Kolmogorov-Smirnov.

3.2 Le test de Cramér–von Mises

Une distance possible entre des fonctions de répartition consiste à prendre la moyenne des différences au carré. Cette idée a été explorée par Cramér (1928), von Mises (1931) et Smirnov & Polejaeff (1934). On réfère maintenant à cette fonctionnelle comme la distance de Cramér–von Mises. Dans notre cas, cela amène la statistique de test

$$\text{CVM}_n = \int_0^1 \{\mathbb{F}_n(t)\}^2 dt. \quad (3.2)$$

En utilisant le fait que

$$\{\mathbb{F}_n(t)\}^2 = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{I}(t \geq \hat{T}_{i,n} \vee \hat{T}_{j,n}) - 2t \sum_{i=1}^n \mathbb{I}(t \geq \hat{T}_{i,n}) + nt^2, \quad (3.3)$$

on arrive à montrer, en intégrant entre 0 et 1, que

$$\text{CVM}_n = \frac{n}{3} + \sum_{i=1}^n \hat{T}_{i,n}^2 - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (\hat{T}_{i,n} \vee \hat{T}_{j,n}). \quad (3.4)$$

Puisque \mathbb{F}_n converge en loi vers la limite \mathbb{F} , on déduit que CVM_n converge en loi vers

$$\text{CVM} = \int_0^1 \{\mathbb{F}(t)\}^2 dt.$$

On peut représenter cette loi limite en terme d'une somme pondérée de variables khi-carrés indépendantes. Toutefois, les valeurs critiques pour des tailles finies seront obtenues par des simulations Monte–Carlo.

3.3 Le test de Anderson–Darling

La statistique de Anderson–Darling est une version pondérée de la fonctionnelle de Cramér–von Mises. En fait, il s’agit de considérer une version pondérée du processus \mathbb{F}_n avec la fonction de poids $t^{-1/2}(1-t)^{-1/2}$. Ceci permet d’accorder plus d’importance aux valeurs de $\mathbb{F}_n(t)$ près de 0 et 1. Cela peut s’avérer utile pour détecter des écarts à la normalité par des lois ayant des queues lourdes. Ainsi, la statistique du test de Anderson–Darling est définie par

$$\text{AD}_n = \int_0^1 \frac{\{\mathbb{F}_n(t)\}^2}{t(1-t)} dt. \quad (3.5)$$

Ainsi, AD_n converge en loi vers

$$\text{AD} = \int_0^1 \frac{\{\mathbb{F}(t)\}^2}{t(1-t)} dt.$$

En utilisant (3.3), on montre que

$$\text{AD}_n = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \log \left(\frac{1}{\hat{T}_{i,n} \vee \hat{T}_{j,n}} - 1 \right) - 2 \sum_{i=1}^n \log (1 - \hat{T}_{i,n}) - n. \quad (3.6)$$

La moyenne et la variance de AD sont données par

$$\begin{aligned} \text{E}(\text{AD}) &= \int_0^1 \frac{\text{E} \{\mathbb{F}(t)\}^2}{t(1-t)} dt = \int_0^1 \frac{t(1-t)}{t(1-t)} dt = 1, \\ \text{var}(\text{AD}) &= \int_0^1 \int_0^1 \text{cov} \left\{ \frac{\mathbb{F}^2(s)}{s(1-s)}, \frac{\mathbb{F}^2(t)}{t(1-t)} \right\} ds dt = \frac{2}{3} (\pi^2 - 9). \end{aligned}$$

3.4 Le test de Kolmogorov–Smirnov

Tel que mentionné par Stephens (1974, 1976), ce test s'inspire de la statistique proposée par Kolmogorov (1933) pour l'ajustement à une distribution. L'idée consiste à calculer la différence maximale, en valeur absolue, entre la distribution empirique et la distribution supposée sous l'hypothèse nulle. Pour tester la normalité, la statistique à utiliser est définie par

$$KS_n = \sqrt{n} \sup_{t \in [0,1]} |F_n(t) - t|.$$

Puisque la fonction $F_n(t)$ présente des *sauts* de hauteur $1/n$ lorsque $t \in \{\hat{T}_{1,n}, \dots, \hat{T}_{n,n}\} = \{1/n, 2/n, \dots, 1\}$, on en déduit que

$$KS_n = \sqrt{n} \max_{1 \leq i \leq n} \left\{ \left| \hat{T}_{(i),n} - \frac{i}{n} \right|, \left| \hat{T}_{(i),n} - \frac{i-1}{n} \right| \right\}, \quad (3.7)$$

où $\hat{T}_{(1),n}, \dots, \hat{T}_{(n),n}$ sont les statistiques d'ordre de l'échantillon $\hat{T}_{1,n}, \dots, \hat{T}_{n,n}$. L'hypothèse de normalité est rejetée pour de grandes valeurs de KS_n . Les valeurs critiques de cette statistique de test s'obtiennent aisément par une méthode Monte–Carlo. À partir du comportement asymptotique du processus \mathbb{F}_n , on déduit facilement celui de KS_n :

$$KS_n = \sup_{t \in [0,1]} |\mathbb{F}_n(t)| \rightsquigarrow KS = \sup_{t \in [0,1]} |\mathbb{F}(t)|,$$

où \rightsquigarrow signifie *convergence en loi*.

CHAPITRE 4

AUTRES TESTS

4.1 Tests de régression

4.1.1 Idée générale

La régression est un modèle mathématique utilisé pour expliquer le lien entre deux ou plusieurs variables aléatoires. Plus spécifiquement, on s'intéresse à l'effet qu'ont des variables *explicatives* X_1, \dots, X_d sur une variable *dépendante* Y . La régression linéaire propose que ce lien soit de la forme

$$Y = \beta_0 + \sum_{j=1}^d \beta_j X_j + \epsilon,$$

où $\epsilon \sim \mathcal{N}(0, \sigma^2)$. À partir d'un échantillon $(Y_i, X_{1i}, \dots, X_{di}), i \in \{1, \dots, n\}$, on cherche les valeurs des paramètres $(\beta_0, \beta_1, \dots, \beta_d)$ qui font en sorte que la droite de régression soit le plus près possible des observations. Un tel critère de proximité est donné par la méthode des moindres carrés, qui consiste à

choisir $(\beta_0, \beta_1, \dots, \beta_d)$ qui minimisent

$$L(\beta_0, \beta_1, \dots, \beta_d) = \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^d \beta_j X_{ji} \right)^2.$$

La solution $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_d)$ constitue les estimateurs de moindres carrés.

4.1.2 Test de Shapiro–Wilk

Soit F_n , la fonction de répartition d'un échantillon X_1, \dots, X_n . On peut définir les statistiques d'ordre par

$$X_{(i)} = F_n^{-1} \left(\frac{i}{n} \right).$$

Sous l'hypothèse de normalité, $F_n^{-1}(u)$ est un bon estimateur de $\Phi_{\mu, \sigma}^{-1}(u) = \sigma \Phi^{-1}(u) + \mu$. Ainsi, la statistique d'ordre $X_{(i)}$ devrait être fortement corrélée avec $\Phi^{-1}(i/n)$. Ainsi, la régression linéaire

$$\Phi^{-1} \left(\frac{i}{n} \right) = \beta_0 + \beta_1 X_{(i)} + \epsilon$$

devrait être significative. La statistique du test de Shapiro–Wilk est

$$W_n = \frac{\left(\sum_{i=1}^n a_i X_{(i)} \right)^2}{(n-1)S_n^2}, \quad \text{où } (a_1, \dots, a_n) = \frac{m' \Sigma^{-1}}{(m' \Sigma^{-1} \Sigma^{-1} m)^{1/2}},$$

avec $m = (m_1, \dots, m_n)$, $m_i = E(X_{(i)})$ et Σ est la matrice de variance-covariance des statistiques d'ordre $X_{(1)}, \dots, X_{(n)}$. La statistique W_n s'interprète comme un coefficient de corrélation quadratique entre les distributions observée et normale. Sous \mathcal{H}_0 , la statistique W_n prend des valeurs près de 1. On rejette l'hypothèse lorsque W_n est inférieure à la valeur critique qui peut être extraite d'une table de Shapiro–Wilk.

4.1.3 Test de D'Agostino

La statistique de D'Agostino est une alternative à la statistique de test de Shapiro–Wilk. Son avantage principal est qu'elle est plus simple et plus rapide à calculer. Spécifiquement, on a

$$D_n = \frac{\left(\sum_{i=1}^n a_i X_{(i)} \right)^2}{(n-1)\hat{\sigma}^2},$$

où

$$a_i = i - \frac{n+1}{2}.$$

On peut montrer que

$$\lim_{n \rightarrow \infty} \mathbb{E}(D_n) = \frac{1}{2\sqrt{\pi}} \quad \text{et} \quad \lim_{n \rightarrow \infty} \text{var}(D_n) \approx \frac{0.03}{\sqrt{n}}.$$

Une version normalisée de D_n est alors

$$\tilde{D}_n = \frac{\sqrt{n}}{0.03} \left(D_n - \frac{1}{2\sqrt{\pi}} \right).$$

On rejette l'hypothèse de normalité quand

$$|\tilde{D}_n| > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right).$$

4.2 Test basé sur l'entropie d'une loi de probabilité

4.2.1 Notion d'entropie

L'entropie d'une fonction de densité f est définie par

$$\mathbb{E}(f) = - \int_{-\infty}^{\infty} f(x) \log f(x) dx.$$

Pour la densité normale de moyenne μ et de variance σ^2 , on calcule

$$\mathbb{E}(\phi_{\mu,\sigma}) = - \int_{-\infty}^{\infty} \phi_{\mu,\sigma}(x) \log \phi_{\mu,\sigma}(x) dx.$$

En appliquant le changement de variable $z = (x - \mu)/\sigma$, on a

$$\begin{aligned} \mathbb{E}(\phi_{\mu,\sigma}) &= - \int_{-\infty}^{\infty} \phi(z) \log \left\{ \frac{1}{\sigma} \phi(z) \right\} dz \\ &= \log \sigma \int_{-\infty}^{\infty} \phi(z) dz - \int_{-\infty}^{\infty} \phi(z) \log \phi(z) dz \\ &= \log \sigma + \log \sqrt{2\pi} \int_{-\infty}^{\infty} \phi(z) dz + \frac{1}{2} \int_{-\infty}^{\infty} z^2 \phi(z) dz \\ &= \log \sigma + \log \sqrt{2\pi} + \frac{1}{2} \\ &= \frac{1}{2} \log (2\pi e \sigma^2). \end{aligned}$$

On remarque que l'entropie de la densité normale ne dépend pas de la moyenne μ . Toutefois, l'entropie augmente comme fonction de l'écart-type. La figure 4.1 présente la courbe de $\mathbb{E}(\phi_{\mu,\sigma})$ en fonction de σ .

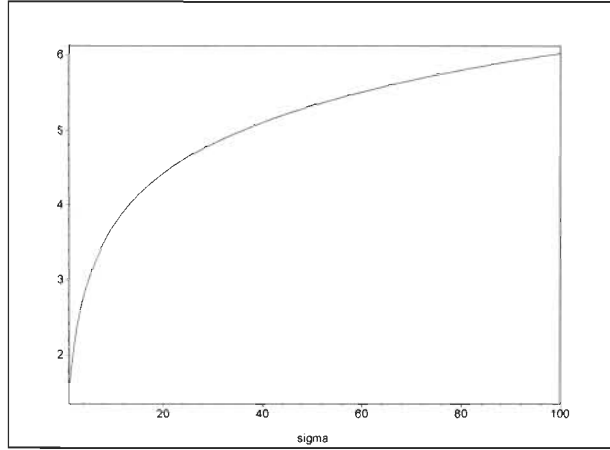


Figure 4.1: Entropie de la densité normale en fonction de l'écart-type σ

4.2.2 Test de Vasicek

Une idée de test de normalité consiste à comparer une version empirique de l'entropie à la valeur théorique sous \mathcal{H}_0 . L'hypothèse nulle est alors rejetée pour de grandes différences observées. On estime $\mathbb{E}(f)$ par

$$\mathbb{E}_{mn} = \frac{1}{n} \sum_{i=1}^n \log \left\{ \frac{n}{2m} (X_{(i+m)} - X_{(i-m)}) \right\},$$

où $m < n/2$ est un entier positif et $X_{(i)}$ est la i -ème statistique d'ordre. Par convention, on pose $X_{(i)} = X_{(1)}$ si $i < 1$ et $X_{(i)} = X_{(n)}$ si $i > n$.

Pour élaborer une statistique de test pour la normalité, on note que

$$\exp \{ \mathbb{E}(\phi_{\mu,\sigma}) \} = \sqrt{2\pi}e\sigma \quad \text{et} \quad \exp(\mathbb{E}_{mn}) = \frac{n}{2m} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}.$$

En prenant le ratio, on a

$$\frac{\exp(\mathbb{E}_{mn})}{\exp \{ \mathbb{E}(\phi_{\mu,\sigma}) \}} = \frac{n}{2m\sqrt{2\pi}e\sigma} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}$$

En négligeant la constante $\sqrt{2\pi e}$ et en estimant σ par son estimateur naturel $\hat{\sigma}$, on retrouve la statistique de test

$$V_{mn} = \frac{n}{2mS} \left\{ \prod_{i=1}^n (X_{(i+m)} - X_{(i-m)}) \right\}^{1/n}.$$

Les valeurs critiques du test de normalité basé sur V_{mn} pour différentes valeurs de n et m se retrouvent dans l'article de Vasicek (1976). Une étude de puissance conduite par cet auteur montre que la puissance de V_{mn} est souvent comparable aux tests de Shapiro–Wilk et de Anderson–Darling.

CHAPITRE 5

ÉVALUATION DE LA PUISSANCE POUR DES TAILLES FINIES

5.1 Alternatives locales à la loi normale

Dans ce chapitre, des contre-hypothèses à la distribution normale seront décrites. Les densités associées à ces alternatives sont de la forme ℓ_δ , où $\lim_{\delta \rightarrow 0} \ell_\delta(z) = \phi(z)$. Ainsi, il sera possible d'étudier l'efficacité de tests de normalité sous des hypothèses très semblables à la normale.

5.1.1 Mélanges de lois normales

Une façon de créer des alternatives à une loi consiste à considérer des mélanges de lois. Dans le cas de la loi normale, une possibilité consiste à considérer des densités de la forme

$$\ell_\delta(z) = (1 - \delta)\phi(z) + \frac{\delta}{\gamma}\phi\left(\frac{z - \nu}{\gamma}\right). \quad (5.1)$$

Un tel mélange de lois fut utilisé par Thadewald & Büning (2007) pour étudier la puissance de quelques tests de normalité sous de petits échantillons. La figure 5.1 présente des mélanges de lois normales pour $\gamma = 1$ pour différentes valeurs de δ .

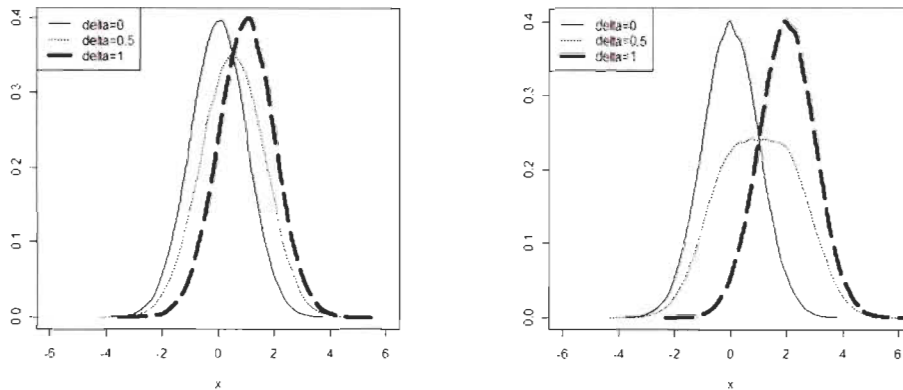


Figure 5.1: Mélanges de lois $\mathcal{N}(0, 1)$ et $\mathcal{N}(\nu, 1)$ pour différentes valeurs de δ . À gauche, $\nu = 1$; à droite, $\nu = 2$

On voit sur la figure 5.1 que plus la moyenne ν augmente, plus la distribution s'éloigne de la normalité. Ainsi, la capacité des tests de normalité à rejeter \mathcal{H}_0 devrait croître en fonction de ν . Pour ce qui est du paramètre δ , on note que les valeurs $\delta = 0$ et $\delta = 1$ amène des modèles sous \mathcal{H}_0 . Ainsi, la puissance des tests devrait être faible près de ces valeurs. La différence la plus marquée par rapport à la normalité se produit lorsque $\delta = 1/2$.

La figure 5.2 présente des mélanges de lois normales lorsque $\nu = 0$ et $\gamma = 2$, et ce pour différentes valeurs de δ . Cette fois, le paramètre δ n'agit pas sur l'asymétrie de la distribution, mais plutôt sur son coefficient d'aplatissement.

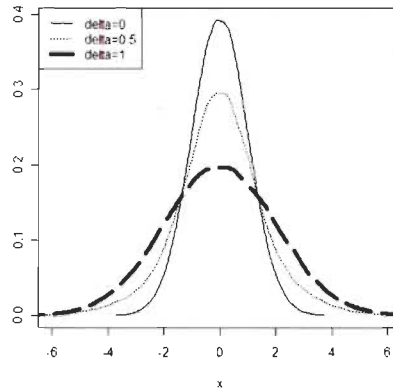


Figure 5.2: Mélanges de lois $\mathcal{N}(0, 1)$ et $\mathcal{N}(0, 4)$ pour différentes valeurs de δ

5.1.2 Alternative à queues lourdes

Soient des mélanges de la forme

$$l_\delta(z) = \begin{cases} \phi(z), & z < 0; \\ (1 - \delta)\phi(z) + \delta g(z), & z \geq 0. \end{cases} \quad (5.2)$$

où g est une certaine densité symétrique par rapport à 0. La figure 5.3 présente les densités obtenues lorsque g est la densité de Student avec $\nu = 5$ et $\nu = 10$ degrés de liberté. On remarque d'emblée une grande similarité entre ces distributions et la distribution normale, même en faisant varier les valeurs de δ . Les différences se retrouvent principalement au niveau des ailes des distributions, mais celles-ci s'amointrissent à mesure que ν augmente. Cette famille de distributions alternatives permettra d'évaluer la puissance des tests de normalité sous des alternatives ayant des ailes plus lourdes que celles de la loi normale.

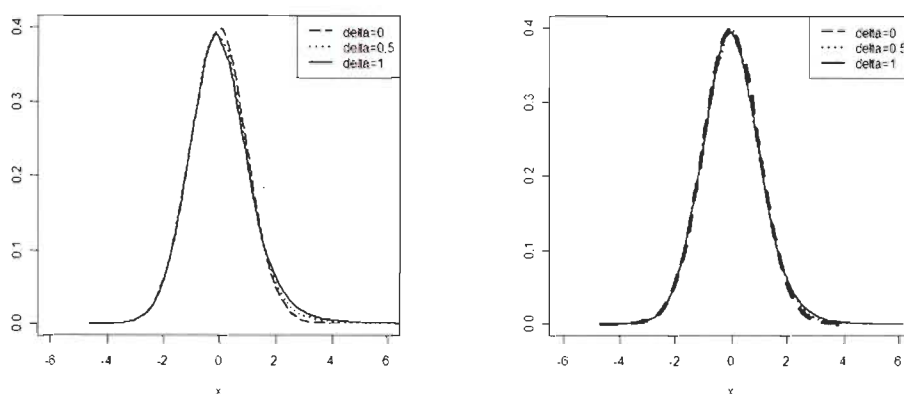


Figure 5.3: Alternative à queues lourdes lorsque g est la densité de Student. À gauche, $\nu = 5$; à droite, $\nu = 10$

5.1.3 Une alternative asymétrique

Azzalini (1985) a introduit la famille de densités

$$\ell_{\delta}(z) = 2\phi(z)\Phi(\delta z). \quad (5.3)$$

Comme cas particulier, on retrouve la densité normale pour $\delta = 0$ car

$$\lim_{\delta \rightarrow 0} \ell_{\delta}(z) = 2\phi(z)\Phi(0) = \phi(z),$$

puisque $\Phi(0) = P(Z \leq 0) = 1/2$, où $Z \sim \mathcal{N}(0, 1)$. La figure 5.4 présente cette densité pour quelques valeurs de δ . À mesure que δ augmente, on voit que la densité devient de plus en plus asymétrique. Ces lois permettront de juger de l'efficacité des tests à rejeter des hypothèses présentant de l'asymétrie.

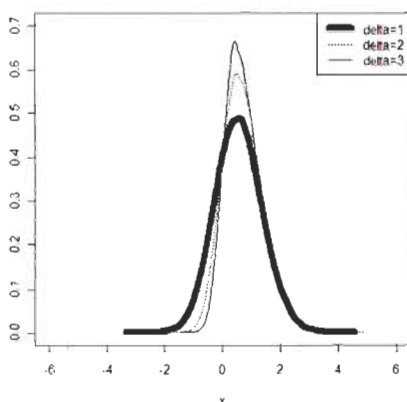


Figure 5.4: Densités asymétriques avec diverses valeurs de δ

5.2 Résultats des simulations

Étudier la puissance des tests est une étape cruciale pour qu'un statisticien puisse choisir une méthode appropriée parmi la panoplie de tests disponibles. Plusieurs études de ce genre ont été effectuées, notamment par Shapiro et al. (1968), Stephens (1974) et Pearson et al. (1977). Dans toutes ces investigations, les distributions alternatives considérées sont passablement éloignées de la distribution sous \mathcal{H}_0 , à savoir la loi normale. Il en résulte, tel qu'attendu, des puissances élevées pour la plupart des procédures. On serait beaucoup plus intéressé à connaître le comportement des tests dans des cas où la loi est proche de la normalité.

Dans cette section, une étude de puissance pour quelques-uns des tests décrits aux chapitres 2, 3 et 4 sera élaborée. Les alternatives considérées seront de

type *local*. Précisément, on considérera des contre-hypothèses de la forme (5.1), (5.2) et (5.3). Les procédures considérées seront basées sur les statistiques S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n . Pour effectuer les tests, les valeurs critiques obtenues d'une simulation de type Monte-Carlo seront utilisées. Les résultats se retrouvent au tableau 6.1 du chapitre 6.

5.2.1 Puissance sous des mélanges de lois normales

Soient les densités de la forme

$$\ell_\delta(z) = (1 - \delta)\phi(z) + \delta\phi(z - \nu).$$

Il s'agit d'un cas particulier du modèle (5.1) lorsque $\gamma = 1$. Si $X \sim \ell_\delta$, alors

$$E(X^j) = (1 - \delta)E(Z^j) + \delta E\{(Z + \nu)^j\},$$

où $Z \sim \mathcal{N}(0, 1)$. Ceci permet de calculer $E(X) = \delta\nu$, $E(X^2) = 1 + \delta\nu^2$, $E(X^3) = \delta(3\nu + \nu^3)$ et $E(X^4) = 3 + \delta\nu^2(\nu^2 + 6)$. Ainsi, $\text{var}(X) = 1 + \delta(1 - \delta)\nu^2$ et les coefficients d'asymétrie et d'aplatissement sont respectivement

$$\beta_1 = \frac{\delta(1 - \delta)(1 - 2\delta)\nu^3}{\{1 + \delta(1 - \delta)\nu^2\}^{3/2}},$$

$$\beta_2 = \frac{3 + \delta(1 - \delta)\nu^2 \{6 + (3\delta^2 - 3\delta + 1)\nu^2\}}{\{1 + \delta(1 - \delta)\nu^2\}^2}.$$

La figure 5.5 montre le comportement de β_1 et β_2 en fonction du paramètre δ lorsque $\nu = 1$. On retrouve les valeurs typiques du modèle normal, à savoir $\beta_1 = 0$ et $\beta_2 = 3$, quand $\delta = 0$ et $\delta = 1$, ce qui s'explique par le fait qu'on se retrouve sous \mathcal{H}_0 . Cependant, ces valeurs sont également visitées sous \mathcal{H}_1 . En effet, $\beta_1 = 0$ pour $\delta = 1/2$, alors que $\beta_2 = 3$ pour $\delta = 1/2 \pm \sqrt{3}/6$.

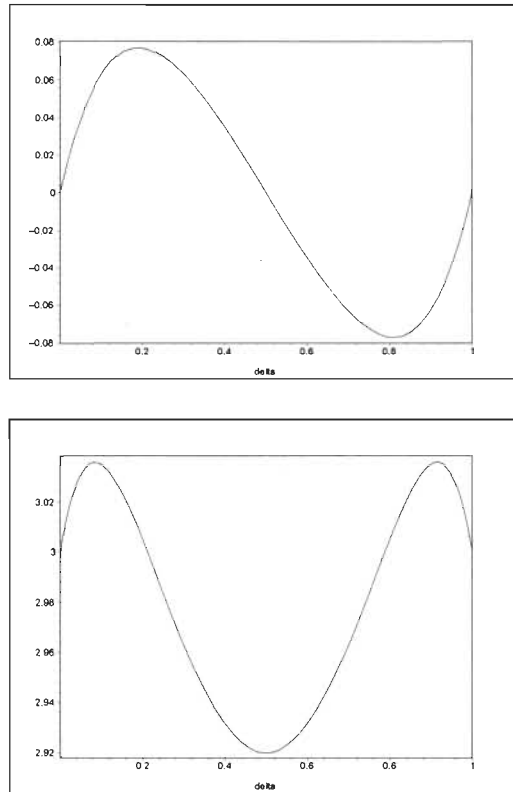


Figure 5.5: Coefficients d'asymétrie (en haut) et d'aplatissement (en bas) en fonction de δ dans le modèle (5.1) quand $\nu = 1$ et $\gamma = 1$

Le tableau 5.1 présente les puissances estimées sous le modèle (5.1) avec $\nu = \gamma = 1$. En examinant les résultats pour $\delta = 0$ et $\delta = 1$, on constate que toutes les statistiques conservent leur seuil sous \mathcal{H}_0 , et ce pour toutes les tailles d'échantillons. Sauf pour S_n et K_n , les puissances des tests atteignent leur apogée lorsque $\delta = 0.5$, c'est-à-dire lorsque l'alternative (5.1) est la plus éloignée de l'hypothèse nulle. Les faibles puissances de S_n et K_n s'expliquent par les faibles variations des coefficients d'asymétrie et d'aplatissement, tel qu'on peut le constater à la figure 5.5. Globalement, R_n est la meilleure.

Tableau 5.1: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.1) avec $\nu = 1$ et $\gamma = 1$

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	0	0.0487	0.0487	0.0502	0.0530	0.0491	0.0486	0.0474
	0.25	0.0485	0.0468	0.3743	0.3654	0.1995	0.2328	0.1498
	0.5	0.0433	0.0425	0.2657	0.3393	0.1810	0.2025	0.1430
	0.75	0.0481	0.0464	0.1249	0.1789	0.0955	0.1005	0.0861
	1	0.0504	0.0486	0.0463	0.0480	0.0504	0.0526	0.0492
100	0	0.0488	0.0508	0.0512	0.0559	0.0569	0.0582	0.0568
	0.25	0.0558	0.0450	0.5650	0.5774	0.3210	0.3799	0.2360
	0.5	0.0395	0.0433	0.4098	0.5514	0.3075	0.3410	0.2247
	0.75	0.0536	0.0473	0.1712	0.2857	0.1433	0.1517	0.1156
	1	0.0507	0.0481	0.0503	0.0562	0.0563	0.0577	0.0545
500	0	0.0522	0.0575	0.0489	0.0551	0.0476	0.0474	0.0508
	0.25	0.0968	0.0558	0.9836	0.9821	0.8886	0.9383	0.7399
	0.5	0.0423	0.0581	0.9354	0.9885	0.9098	0.9377	0.7824
	0.75	0.0987	0.0540	0.4504	0.7676	0.4906	0.5105	0.3593
	1	0.0504	0.0608	0.0473	0.0540	0.0530	0.0529	0.0562

5.2.2 Puissance sous l'alternative à queues lourdes

Les tableaux 5.2, 5.3, 5.4 et 5.5 présentent les puissances estimées des statistiques sous l'alternative 5.2. Spécifiquement, le modèle à queues lourdes fournit par la densité de Student à $\nu = 5, 10, 20$ et 30 degrés de liberté est utilisé pour construire les contre-hypothèses.

On remarque d'abord que les puissances, globalement, augmentent avec les tailles des échantillons, tel qu'attendu. De plus, les puissances observées décroissent en fonction du paramètre ν . Cela s'explique par le fait que la densité de Student s'approche de celle de la loi normale standard, donc de \mathcal{H}_0 , à mesure que ν augmente. Aussi, contrairement au modèle 5.1, les puissances sont croissantes en fonction de δ . En fait, plus $\delta \in [0, 1]$ prend une valeur élevée, plus le poids accordé à la densité de Student est important, d'où la plus grande capacité à rejeter \mathcal{H}_0 .

Contrairement aux résultats sous des mélanges de lois normales, toutes les statistiques montrent des puissances appréciables. Néanmoins, les procédures basées sur JB_n et R_n sont les meilleures, et ce pour tous les scénarios considérés. La statistique de Kolmogorov–Smirnov KS_n est la moins puissante parmi les sept statistiques étudiées. On remarque également que malgré leur grande simplicité, les statistiques S_n et K_n font aussi bien, sinon mieux, que les fonctionnelles du processus empirique \mathbb{F}_n , c'est-à-dire CVM_n , AD_n et KS_n .

Tableau 5.2: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.2) avec $\nu = 5$

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	0.25	0.1228	0.1250	0.1971	0.1683	0.1041	0.1192	0.0857
	0.5	0.2040	0.1832	0.3070	0.2807	0.1612	0.1896	0.1342
	0.75	0.2648	0.2312	0.3804	0.3684	0.2220	0.2521	0.1748
	1	0.3229	0.2710	0.4208	0.4410	0.2761	0.3072	0.2147
100	0.25	0.1705	0.1795	0.3096	0.2643	0.1438	0.1620	0.1197
	0.5	0.2991	0.2895	0.4868	0.4454	0.2535	0.2982	0.1970
	0.75	0.4100	0.3665	0.5775	0.5652	0.3492	0.4063	0.2679
	1	0.5113	0.4351	0.6467	0.6799	0.4485	0.4955	0.3491
500	0.25	0.4517	0.4850	0.7278	0.6273	0.3231	0.3825	0.2292
	0.5	0.7707	0.7498	0.9322	0.9045	0.6733	0.7417	0.5243
	0.75	0.9204	0.8724	0.9823	0.9844	0.8891	0.9008	0.7694
	1	0.9750	0.9350	0.9955	0.9966	0.9662	0.9535	0.9053

Tableau 5.3: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.2) avec $\nu = 10$

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	0.25	0.0701	0.0701	0.0979	0.0876	0.0595	0.0663	0.0580
	0.5	0.0960	0.0895	0.1435	0.1252	0.0795	0.0872	0.0717
	0.75	0.1169	0.1027	0.1747	0.1615	0.0901	0.1042	0.0769
	1	0.1488	0.1246	0.2011	0.1938	0.1075	0.1229	0.0890
100	0.25	0.0864	0.0895	0.1342	0.1175	0.0705	0.0761	0.0646
	0.5	0.1266	0.1247	0.2033	0.1798	0.0957	0.1070	0.0814
	0.75	0.1679	0.1590	0.2599	0.2417	0.1207	0.1404	0.0990
	1	0.2040	0.1796	0.3107	0.3072	0.1498	0.1706	0.1157
500	0.25	0.1523	0.1758	0.2845	0.2228	0.0892	0.0980	0.0721
	0.5	0.2966	0.3000	0.4943	0.4314	0.1734	0.2037	0.1254
	0.75	0.4615	0.4145	0.6454	0.6171	0.2903	0.3406	0.1992
	1	0.6009	0.5063	0.7478	0.7552	0.4234	0.4881	0.2892

Tableau 5.4: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.2) avec $\nu = 20$

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	0.25	0.0520	0.0525	0.0636	0.0626	0.0531	0.0540	0.0504
	0.5	0.0670	0.0661	0.0862	0.0796	0.0586	0.0599	0.0554
	0.75	0.0698	0.0656	0.0966	0.0926	0.0617	0.0653	0.0583
	1	0.0824	0.0736	0.1137	0.1060	0.0658	0.0693	0.0611
100	0.25	0.0619	0.0597	0.0818	0.0767	0.0608	0.0631	0.0590
	0.5	0.0716	0.0709	0.1095	0.0979	0.0666	0.0677	0.0613
	0.75	0.0833	0.0853	0.1287	0.1224	0.0739	0.0772	0.0711
	1	0.1001	0.0941	0.1513	0.1512	0.0806	0.0888	0.0732
500	0.25	0.0706	0.0872	0.0825	0.0776	0.0522	0.0521	0.0534
	0.5	0.1160	0.1178	0.1120	0.1203	0.0603	0.0632	0.0574
	0.75	0.1662	0.1625	0.1565	0.1527	0.0672	0.0720	0.0612
	1	0.2184	0.1989	0.1936	0.1980	0.0812	0.0880	0.0704

Tableau 5.5: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.2) avec $\nu = 30$

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	0.25	0.0496	0.0483	0.0609	0.0567	0.0470	0.0485	0.0511
	0.5	0.0575	0.0557	0.0712	0.0670	0.0557	0.0564	0.0543
	0.75	0.0639	0.0584	0.0837	0.0783	0.0532	0.0537	0.0531
	1	0.0667	0.0632	0.0885	0.0873	0.0600	0.0621	0.0584
100	0.25	0.0572	0.0577	0.0683	0.0684	0.0551	0.0572	0.0539
	0.5	0.0628	0.0620	0.0861	0.0813	0.0601	0.0652	0.0572
	0.75	0.0647	0.0711	0.0936	0.0951	0.0634	0.0677	0.0589
	1	0.0746	0.0739	0.1090	0.1115	0.0664	0.0694	0.0642
500	0.25	0.0637	0.0696	0.0829	0.0825	0.0575	0.0559	0.0553
	0.5	0.0803	0.0897	0.1179	0.1221	0.0662	0.0668	0.0614
	0.75	0.0982	0.1030	0.1557	0.1530	0.0681	0.0710	0.0619
	1	0.1326	0.1277	0.1923	0.1998	0.0811	0.0880	0.0706

5.2.3 Puissance sous une alternative asymétrique

Soient les densités introduites à l'équation (5.3). Dans ce cas, on a

$$E(X^k) = \int_{\mathbb{R}} x^k \ell_{\delta}(x) dx,$$

où $\ell_{\delta}(x) = 2\phi(x)\Phi(\delta x)$. En effectuant le changement de variable $u = -x$ et en notant que $\ell_{\delta}(-u) = 2\phi(u)\{1 - \Phi(\delta u)\}$, on trouve

$$\begin{aligned} E(X^k) &= \int_{\mathbb{R}} 2(-1)^k u^k \phi(u) \{1 - \Phi(\delta u)\} du \\ &= 2(-1)^k \int_{\mathbb{R}} u^k \phi(u) du - (-1)^k \int_{\mathbb{R}} u^k \ell_{\delta}(u) du \\ &= 2(-1)^k E(Z^k) - (-1)^k E(X^k), \end{aligned}$$

où $Z \sim \mathcal{N}(0, 1)$. Ainsi, si k est pair, on a $E(X^k) = E(Z^k)$. En particulier, $E(X^2) = 1$ et $E(X^4) = 3$. En utilisant MAPLE, on trouve

$$\mu = E(X) = \sqrt{\frac{2}{\pi}} \frac{\delta}{\sqrt{\delta^2 + 1}} \quad \text{et} \quad E(X^3) = \sqrt{\frac{2}{\pi}} \frac{\delta(2\delta^2 + 3)}{(\delta^2 + 1)^{3/2}}.$$

Un peu d'algèbre permet de déduire que

$$\sigma^2 = \frac{(\pi - 2)\delta^2 + \pi}{\pi(\delta^2 + 1)}.$$

Les expressions pour β_1 et β_2 en fonction de δ ont une forme relativement compliquée, bien qu'elles puissent facilement s'obtenir avec MAPLE. La figure 5.6 montre d'ailleurs l'influence du paramètre δ sur l'asymétrie et l'aplatissement de la densité (5.3). Comme on peut le constater, l'asymétrie à droite est fortement influencée par l'augmentation de δ . Pour le coefficient d'aplatissement, celui-ci s'écarte moins de la valeur sous \mathcal{H}_0 , soit $\beta_2 = 3$.

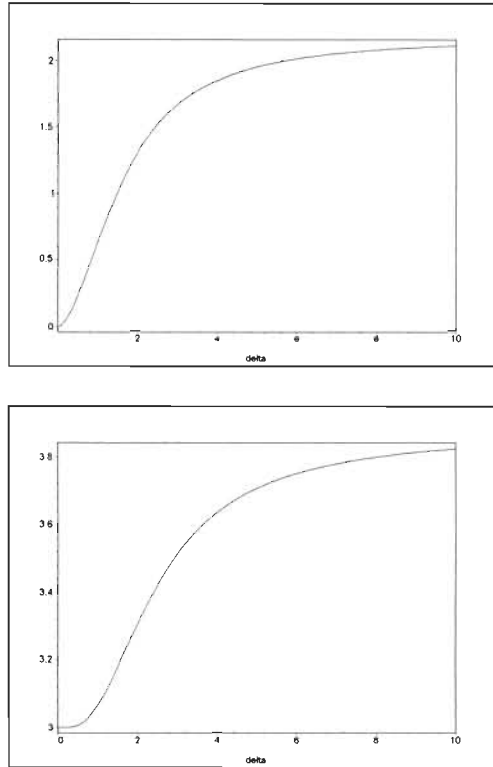


Figure 5.6: Coefficients d'asymétrie (en haut) et d'aplatissement (en bas) en fonction de δ dans le modèle (5.3)

Au tableau 5.6, les puissances des sept statistiques sont estimées sous le modèle (5.3). Globalement, le test basé sur S_n est le plus puissant. Il est suivi par les procédures basées sur les statistiques de Jarque–Bera et de Anderson–Darling, respectivement JB_n et AD_n . Cette excellente performance de S_n s'explique par l'influence importante de δ sur le coefficient d'asymétrie, tel qu'on peut le constater à la figure 5.6. La statistique de Gel et al. (2007), à savoir R_n , est la moins performante. La statistique K_n , basée sur le coefficient d'aplatissement, a également une faible puissance. Cela s'explique par la relative insensibilité de β_2 en fonction de δ , tel que vu à la figure 5.6.

Tableau 5.6: Estimation, basée sur 10 000 répétitions, de la puissance de S_n , K_n , JB_n , R_n , CVM_n , AD_n et KS_n sous l'alternative (5.3)

n	δ	S_n	K_n	JB_n	R_n	CVM_n	AD_n	KS_n
50	1	0.0663	0.0570	0.0688	0.0596	0.0614	0.0634	0.0588
	2	0.2279	0.1021	0.1880	0.0961	0.1534	0.1687	0.1381
	3	0.4113	0.1450	0.3408	0.1249	0.3148	0.3516	0.2533
100	1	0.0835	0.0619	0.0862	0.0674	0.0722	0.0757	0.0734
	2	0.4091	0.1298	0.3537	0.1298	0.2878	0.3196	0.2407
	3	0.7197	0.2018	0.6404	0.1728	0.5887	0.6438	0.4797
500	1	0.2416	0.0870	0.1934	0.0826	0.1333	0.1469	0.1159
	2	0.9812	0.3128	0.9597	0.2299	0.8942	0.9273	0.7967
	3	0.9999	0.5028	1.0000	0.3465	0.9993	0.9997	0.9931

CHAPITRE 6

ASYMPTOTIC POWER OF TESTS OF NORMALITY UNDER LOCAL ALTERNATIVES

6.1 Introduction

Due mainly to its mathematical tractability, the univariate normal distribution is the base to many statistical models. Then the question rises to know wether a random sample can be considered to be drawn from a population whose density is of the form

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right\}, \quad x \in \mathbb{R},$$

for some $(\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$. If f denotes the density of a given population, then the hypotheses of normality are $\mathcal{H}_0 : f(x) = \phi_{\mu,\sigma}(x)$ and $\mathcal{H}_1 : f(x) \neq \phi_{\mu,\sigma}(x)$ for at least one $x \in \mathbb{R}$, where μ and σ are left unspecified.

Many procedures exist to decide on the acceptance or rejection of \mathcal{H}_0 . The first contribution seems to date back to Geary (1947), where the ratio of

the absolute mean and standard deviation is considered as a criteria to assess normality. This test is formally described by Gel et al. (2007). Other methods are based on the empirical versions of skewness and kurtosis. While each of these two statistics can be used for inferential purposes, the idea of D'Agostino & Pearson (1973) was to combine them. The resulting procedure is often referred to the Jarque–Bera test, which has been later popularized by Jarque & Bera (1980, 1987). Finally, an important family of tests is composed of the functionals of the empirical process with estimated parameters, considered in a general setting by Durbin (1973). The Cramér–von Mises, Anderson–Darling and Kolmogorov–Smirnov procedures are members of this class.

In order for a statistician to choose a suitable method among the available tests, it is crucial to study their power under alternative hypotheses. This has been done by several authors, including Shapiro et al. (1968), Stephens (1974) and Pearson et al. (1977). However, many of these studies consider hypotheses that are so far away from normality that even a simple look at an histogram would lead to the rejection of \mathcal{H}_0 . Furthermore, as mentioned by Durio & Nikitin (2003), many power investigations for goodness-of-fit tests are based on location-scale alternatives, which does not cover the wide range of possible departures from the null hypothesis.

If one admits that a desirable property of a statistical procedure is its ability to reject the null hypothesis under small departures from \mathcal{H}_0 , so-called *local alternatives* must be considered. These alternatives enable to study both the small sample and asymptotic efficiency of tests for normality. Some attempts

have been made to compute some kinds of asymptotic relative efficiencies, but the latter are generally based on Bahadur's approach. Although elegant formulas can be discovered from this method, the results are difficult to interpret and hardly relate to daily practice since one considers the power of the tests as the size $\alpha \rightarrow 0$.

Using the asymptotic theory of van der Vaart & Wellner (1996) and repeated applications of Lecam's third lemma, explicit representations of seven statistics for normality will be obtained under local sequences of distributions. This will enable the comparison of the tests via their local power functions and in the light of an asymptotic relative efficiency measure recently proposed by Berg & Quesy (2009). As will be seen, the conclusions and recommendations that one can extract from these asymptotic investigations can be linked directly to real-world situations, i.e. in cases of small and moderate sample sizes.

The paper is organized as follows. In section 2, the seven tests under investigation are introduced. In section 3, their asymptotic distributions under a general class of local alternatives are obtained. These results are used, in section 4, to characterize the large sample behavior of the tests under three kinds of local alternatives. In section 5, formulas to compute local power curves and asymptotic relative efficiencies are given and comparisons among the seven tests based on these tools are made in section 6. Relationships with power results obtained from a Monte-Carlo simulation for small and moderate sample sizes are described in section 7.

6.2 Test statistics for non-normality

In this section, the test statistics whose efficiencies will be analyzed later are presented. These procedures are suitable for the composite null hypothesis $\mathcal{H}_0 : f = \phi_{\mu, \sigma}$, where (μ, σ) are unspecified. From a random sample X_1, \dots, X_n , the latter are estimated by the sample mean and variance, namely

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2.$$

Note that all the tests that are described next are invariant with respect to μ and σ since they are based on the normalized pseudo-observations

$$\hat{X}_{i,n} = \frac{X_i - \hat{\mu}}{\hat{\sigma}}.$$

6.2.1 Tests based on sample skewness and kurtosis

For $k \geq 2$, let μ_k denote the population's k -th central moment, i.e. $\mu_k = E\{(X - EX)^k\}$. The skewness and kurtosis measures are defined respectively by $\beta_1 = \mu_3/\sigma^3$ and $\beta_2 = \mu_4/\sigma^4$. Their natural estimators are

$$\hat{\beta}_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right)^3 = \frac{1}{n} \sum_{i=1}^n \hat{X}_{i,n}^3$$

and

$$\hat{\beta}_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right)^4 = \frac{1}{n} \sum_{i=1}^n \hat{X}_{i,n}^4.$$

When the observations are drawn from a population whose density belongs to the normal family, it is well-established that $\sqrt{n}\hat{\beta}_1 \rightsquigarrow \mathcal{N}(0, 6)$, $\sqrt{n}(\hat{\beta}_2 - 3) \rightsquigarrow \mathcal{N}(0, 24)$ and the latter are independent. Here and in the sequel, \rightsquigarrow means convergence in distribution. Simple tests of normality are then based on the normalized versions

$$S_n = \frac{\sqrt{n}\hat{\beta}_1}{\sqrt{6}} \quad \text{and} \quad K_n = \frac{\sqrt{n}(\hat{\beta}_2 - 3)}{\sqrt{24}}.$$

Specifically, procedures based on the asymptotic distributions would reject \mathcal{H}_0 whenever

$$S_n > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad \text{or} \quad K_n > \Phi^{-1}\left(1 - \frac{\alpha}{2}\right),$$

where Φ is the cdf of the standard normal distribution.

A potentially more efficient test could emerge if one considers a combination of both S_n and K_n . This is the idea behind the Jarque–Bera test statistic, popularized by Jarque & Bera (1980, 1987). The test statistic, which has been originally proposed by D’Agostino & Pearson (1973), is simply

$$JB_n = S_n^2 + K_n^2.$$

Bowman & Shenton (1975) used the fact that JB_n is the sum of two asymptotically independent standardized normals to deduce that $JB_n \rightsquigarrow \chi_2^2$ under the null hypothesis of normality. Thus, a test based on this asymptotic result leads to the rejection of \mathcal{H}_0 if $JB_n > \chi_{2,\alpha}^2$, where $\chi_{2,\alpha}^2$ is the α -th percentile of the χ_2^2 distribution. As noted by many authors, however, the rate of convergence of JB_n toward its asymptotic distribution is very slow, which prevent from using $\chi_{2,\alpha}^2$ as a critical value when the sample size is small or moderate. It is better to use estimated critical values in that case.

6.2.2 The test of Gel et al. (2007)

This test is based on the ratio of two estimators of the standard deviation in the normal model. Specifically, the test statistic is

$$R_n = \sqrt{n} \left(\frac{\hat{\sigma}}{\tilde{\sigma}} - 1 \right),$$

where $\tilde{\sigma} = \sqrt{\pi/2} n^{-1} \sum_{i=1}^n |X_i - \hat{\mu}|$. An alternate definition is

$$R_n = \sqrt{n} \left(\frac{1}{\tilde{\sigma}^*} - 1 \right), \quad \text{where } \sigma^* = \sqrt{\pi/2} \frac{1}{n} \sum_{i=1}^n |\hat{X}_{i,n}|,$$

which emphasizes on the fact that R_n is based on the normalized observations.

Under the null hypothesis of normality, Gel et al. (2007) showed that $R_n \rightsquigarrow \mathcal{N}(0, \sigma_R^2)$, where $\sigma_R^2 = (\pi - 3)/2$. Thus, \mathcal{H}_0 is rejected whenever

$$\frac{|R_n|}{\sigma_R} > \Phi^{-1} \left(1 - \frac{\alpha}{2} \right).$$

6.2.3 Functionals of an empirical goodness-of-fit process

First define the pseudo-observations

$$\hat{T}_{i,n} = \Phi \left(\frac{X_i - \hat{\mu}}{\hat{\sigma}} \right) = \Phi \left(\hat{X}_{i,n} \right), \quad 1 \leq i \leq n.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $T = \Phi\{(X - \mu)/\sigma\}$ follows a standard uniform distribution. It is thus natural to base a test on the empirical goodness-of-fit process

$$\mathbb{F}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{I} \left(\hat{T}_{i,n} \leq t \right) - t \right\}.$$

This process was first considered by Durbin (1973), where its asymptotic behavior is established under a quite general setting. The result is recalled here under lemma 6.1 in the special case of normality. An alternate proof of this result using the theory of empirical processes for pseudo-observations due to Ghoudi & Rémillard (1998) is given.

Lemme 6.1. *Under \mathcal{H}_0 ,*

$$\mathbb{F}_n(t) \rightsquigarrow \mathbb{F}_0(t) = \beta(t) + \phi \left\{ \Phi^{-1}(t) \right\} \left\{ W_1 + \Phi^{-1}(t) \frac{W_2}{\sqrt{2}} \right\},$$

where β is a brownian bridge, i.e. a continuous and centered Gaussian process with $E\{\beta(s)\beta(t)\} = \min(s, t) - st$ and W_1, W_2 are independent $\mathcal{N}(0, 1)$ random variables.

Proof. In order to follow the notation of Ghoudi & Rémillard (1998), let

$$H_n(x) = \Phi \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right) \quad \text{and} \quad H(x) = \Phi \left(\frac{x - \mu}{\sigma} \right),$$

so that

$$\mathbb{F}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{I}\{H_n(X_i) \leq t\} - K(t)],$$

where $K(t) = P\{H(X) \leq t\} = t$. From classical arguments in the theory of empirical processes,

$$\beta_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{I}\{H(X_i) \leq t\} - K(t)]$$

converges to a standard brownian bridge β . Also, it is easy to show that $\mathbb{H}_n(x) = \sqrt{n}\{H_n(x) - H(x)\}$ converges to

$$\mathbb{H}(x) = -\phi \left(\frac{x - \mu}{\sigma} \right) \left\{ W_1 + \left(\frac{x - \mu}{\sigma} \right) \frac{W_2}{\sqrt{2}} \right\},$$

where W_1, W_2 are independent $\mathcal{N}(0, 1)$ random variables. Thus

$$\mu_t(\mathbb{H}) = \text{E} \{ \mathbb{H}(X) | H(X) = t \} = -\phi \{ \Phi^{-1}(t) \} \left\{ W_1 + \Phi^{-1}(t) \frac{W_2}{\sqrt{2}} \right\}.$$

Theorem 2.1 of Ghoudi & Rémillard (1998) implies that $\mathbb{F}_n(t) \rightsquigarrow \beta(t) - \mu_t(\mathbb{H})$, which can be seen to be of the announced form. \square

From this result, one obtains that the limiting covariance function of \mathbb{F}_n is

$$\text{E} \{ \mathbb{F}_0(s) \mathbb{F}_0(t) \} = \min(s, t) - st - \phi \{ \Phi^{-1}(s) \} \phi \{ \Phi^{-1}(t) \} \left\{ 1 + \frac{\Phi^{-1}(s) \Phi^{-1}(t)}{2} \right\},$$

which is exactly the expression that one can deduce from Durbin et al. (1975).

One can see that the limit is free of μ and σ . As noted by Durbin (1973), this is the case under any location/scale families.

Three popular test statistics that can be written as functionals of \mathbb{F}_n will be considered in the sequel. They are based on the Cramér–von Mises, Anderson–Darling and Kolmogorov–Smirnov distances, leading to

$$\text{CVM}_n = \int_0^1 \{ \mathbb{F}_n(t) \}^2 dt, \quad \text{AD}_n = \int_0^1 \frac{ \{ \mathbb{F}_n(t) \}^2 }{ t(1-t) } dt \quad \text{and} \quad \text{KS}_n = \sup_{t \in [0,1]} | \mathbb{F}_n(t) |.$$

The fact that the limiting distributions of these statistics do not depend on the unknown mean and variance under \mathcal{H}_0 enables to obtain the critical values of the tests by simple Monte Carlo simulations. These values are presented in table 6.1. For large sample sizes, one can use the asymptotic critical values of CVM_n et de AD_n obtained by Stephens (1976), namely .126 and .751 respectively. Another possibility is to compute p -values from a parametric bootstrap, whose validity in univariate goodness-of-fit testing is established by Jogesh Babu & Rao (2004).

Table 6.1: Estimation, based on 100 000 replicates, of the 95-th percentile of the distribution of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under the normality assumption

Test statistics	$n = 50$	$n = 100$	$n = 250$	$n = \infty$
$ S_n $	1.818	1.895	1.926	1.960
$ K_n $	1.531	1.686	1.818	1.960
JB_n	4.516	4.925	5.568	5.992
$ R_n $	0.531	0.528	0.522	0.522
CVM_n	0.128	0.123	0.126	0.126
AD_n	0.753	0.735	0.750	0.751
KS_n	0.883	0.889	0.898	0.898

Remark 6.1. *Note that many goodness-of-fit test statistics for normality that have been proposed in the literature are based on the empirical process*

$$\tilde{\mathbb{F}}_n(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \mathbb{I}(X_i \leq x) - \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) \right\}.$$

The latter is linked to \mathbb{F}_n via

$$\tilde{\mathbb{F}}_n(x) = \mathbb{F}_n \left\{ \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) \right\} \rightsquigarrow \mathbb{F}_0 \left\{ \Phi\left(\frac{x - \mu}{\sigma}\right) \right\}.$$

After the simple change of variable $t = \Phi\{(x - \mu)/\sigma\}$, one sees that the three statistics based on $\tilde{\mathbb{F}}_n$ would converge to the same limit.

In the next section, the asymptotic behavior of the seven tests described in this section is investigated under local sequences of alternatives.

6.3 Asymptotic behavior of the test statistics under local sequences

Since the test statistics encountered in this paper are invariant with respect to the mean and standard deviation, it can be assumed, without loss of generality, that $\mu = 0$ and $\sigma = 1$. Then local alternatives to \mathcal{H}_0 are characterized by densities ℓ_δ that satisfy

$$\lim_{\delta \downarrow 0} \ell_\delta(z) = \phi(z),$$

where $\phi = \phi_{0,1}$. It is also assumed that ℓ_δ is not a member of the class of parametric normal densities, otherwise the model would be uninteresting for power investigations. For a fixed $\delta > 0$, the power of any consistent test for normality will tend to 1 as the sample size increases. Comparing the asymptotic power of tests for normality would then be irrelevant. A way to circumvent this problem is to consider a sequence of alternatives that is build in such a way that the asymptotic distributions are non-degenerate. Such sequences of densities are those of the form $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$ that satisfy

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \left\{ \sqrt{n} \left(\sqrt{\ell_{\delta/\sqrt{n}}(z)} - \sqrt{\phi(z)} \right) - \frac{\delta \dot{\ell}_0(z)}{2\sqrt{\phi(z)}} \right\}^2 dz = 0, \quad (6.1)$$

where $\dot{\ell}_0(z) = \lim_{\delta \rightarrow 0} \partial \ell_\delta(z) / \partial \delta$. This criterion, which can be found in van der Vaart & Wellner (1996), ensures that the law induced by $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$ is contiguous with respect to the law induced by the standard normal distribution. A consequence is that the asymptotic behavior of the test statistics under such sequences can be deduced from that under \mathcal{H}_0 . This is the aim of this

section to provide the asymptotic distributions of the seven test statistics under study.

The first result concerns the statistics based on skewness and kurtosis, namely S_n , K_n and JB_n . To this end, define the k -th local moment by

$$\mu'_k = \int_{\mathbb{R}} z^k \dot{\ell}_0(z) dz.$$

Proposition 6.1. *Under a sequence of densities $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$ that satisfies (6.1),*

$$S_n \rightsquigarrow \mathcal{N}\left(\frac{\delta\mu'_3 - 3\delta\mu'_1}{\sqrt{6}}, 1\right), K_n \rightsquigarrow \mathcal{N}\left(\frac{\delta\mu'_4 - 6\delta\mu'_2}{\sqrt{24}}, 1\right) \text{ and } JB_n \rightsquigarrow \chi_2^2(\delta^2\gamma),$$

where

$$\gamma = \frac{(\mu'_3 - 3\mu'_1)^2}{6} + \frac{(\mu'_4 - 6\mu'_2)^2}{24}.$$

Proof. If one assumes without loss of generality that $\mu = 0$ and $\sigma = 1$, one obtains that S_n and K_n have the asymptotic representations

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i^3 - 3X_i}{\sqrt{6}} \right) + o_{\mathbb{P}}(1) \text{ and } K_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i^4 - 6X_i^2 + 3}{\sqrt{24}} \right) + o_{\mathbb{P}}(1).$$

under the null hypothesis of normality. From the bivariate central limit theorem, (S_n, K_n) converges in law, under \mathcal{H}_0 , to a standard bivariate normal distribution. On the other side, condition (6.1) implies

$$L_n = \log \frac{\ell_{\delta/\sqrt{n}}(\mathbf{X})}{\phi(\mathbf{X})} = \frac{\delta}{\sqrt{n}} \sum_{i=1}^n \frac{\dot{\ell}_0(X_i)}{\phi(X_i)} + \frac{\delta^2}{2n} \sum_{i=1}^n \left\{ \frac{\ddot{\ell}_0(X_i)}{\phi(X_i)} - \left(\frac{\dot{\ell}_0(X_i)}{\phi(X_i)} \right)^2 \right\} + o_{\mathbb{P}}(1), \quad (6.2)$$

where $\mathbf{X} = (X_1, \dots, X_n)$. Hence, the multivariate central limit theorem enables to establish that (S_n, K_n, L_n) converges in law to a normal distribution

with mean $(0, 0, 0)'$ and covariance matrix

$$\Sigma_{S,K,L} = \begin{pmatrix} 1 & 0 & \frac{\delta(\mu'_3 - 3\mu'_1)}{\sqrt{6}} \\ 0 & 1 & \frac{\delta(\mu'_4 - 6\mu'_2)}{\sqrt{24}} \\ \frac{\delta(\mu'_3 - 3\mu'_1)}{\sqrt{6}} & \frac{\delta(\mu'_4 - 6\mu'_2)}{\sqrt{24}} & \sigma_L^2 \end{pmatrix}.$$

An application of Lecam's third lemma yields that under the sequence of alternatives $(\ell_{\delta/\sqrt{n}})$, the pair (S_n, K_n) converges in law to a bivariate normal distribution with mean vector $\delta((\mu'_3 - 3\mu'_1)/\sqrt{6}, (\mu'_4 - 6\mu'_2)/\sqrt{24})'$ and covariance matrix $\text{diag}(1, 1)$. The consequences are immediate on S_n , K_n and $\text{JB}_n = S_n^2 + K_n^2$. \square

Now the result of Gel et al. (2007) about the asymptotic behavior of R_n is generalized to the case of contiguous alternatives.

Proposition 6.2. *Under a sequence of densities $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$ that satisfies (6.1),*

$$R_n \rightsquigarrow \mathcal{N}\left(\frac{\delta(\mu'_2 - \tilde{\mu}'_2)}{2}, \sigma_R^2\right),$$

where

$$\tilde{\mu}'_2 = \sqrt{\frac{2}{\pi}} \int_{\mathbb{R}} |z| \dot{\ell}_0(z) dz.$$

Proof. Assuming, without loss of generality, that $\mu = 0$ and $\sigma = 1$, one has

$$\sqrt{n}\tilde{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\sqrt{\frac{\pi}{2}} |X_i| - 1 \right) + o_{\mathbb{P}}(1)$$

and

$$\sqrt{n}\hat{\sigma} = \frac{1}{2\sqrt{n}} \sum_{i=1}^n (X_i^2 - 1) + o_{\mathbb{P}}(1).$$

Since $\tilde{\sigma}$ converges in probability to 1, Slutsky's lemma implies that

$$\mathbb{R}_n = \sqrt{n}(\hat{\sigma} - \tilde{\sigma}) + o_{\mathbb{P}}(1) = \frac{1}{2\sqrt{n}} \sum_{i=1}^n \left(1 + X_i^2 - \sqrt{2\pi}|X_i|\right) + o_{\mathbb{P}}(1).$$

Lecam's third lemma and equation (6.2) entails that \mathbb{R}_n converges in law to a normal distribution with variance $\sigma_{\mathbb{R}}^2$ and mean

$$\begin{aligned} \text{cov} \left\{ \frac{1 + X^2 - \sqrt{2\pi}|X|}{2}, \frac{\delta \dot{\ell}_0(X)}{\phi(X)} \right\} &= \frac{\delta}{2} \int_{\mathbb{R}} \left(1 + x^2 - \sqrt{2\pi}|x|\right) \dot{\ell}_0(x) dx \\ &= \frac{\delta}{2} \left\{ \int_{\mathbb{R}} x^2 \dot{\ell}_0(x) dx - \sqrt{2\pi} \int_{\mathbb{R}} |x| \dot{\ell}_0(x) dx \right\} \\ &= \frac{\delta}{2} (\mu'_2 - \tilde{\mu}'_2), \end{aligned}$$

as claimed. \square

The asymptotic behavior of the empirical goodness-of-fit process \mathbb{F}_n under sequences of local distributions is described next. Durbin (1973) obtained the result under another set of conditions and local alternatives. A similar result is to be found in Durbin et al. (1975).

Proposition 6.3. *Under a sequence of densities $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$ that satisfies (6.1),*

$$\mathbb{F}_n(t) \rightsquigarrow \mathbb{F}_0(t) + \delta \xi(t),$$

where

$$\xi(t) = \dot{\mathcal{L}}_0 \{ \Phi^{-1}(t) \} + \phi \{ \Phi^{-1}(t) \} \left(\mu'_1 + \frac{\mu'_2 \Phi^{-1}(t)}{2} \right)$$

and

$$\dot{\mathcal{L}}_0(z) = \int_{(-\infty, z]} \dot{\ell}_0(s) ds.$$

As a consequence,

$$\text{CVM}_n \rightsquigarrow \int_0^1 \{\mathbb{F}_0(t) + \delta\xi(t)\}^2 dt,$$

$$\text{AD}_n \rightsquigarrow \int_0^1 \frac{\{\mathbb{F}_0(t) + \delta\xi(t)\}^2}{t(1-t)} dt,$$

$$\text{KS}_n \rightsquigarrow \sup_{t \in [0,1]} |\mathbb{F}_0(t) + \delta\xi(t)|.$$

Proof. Recall that without loss of generality, one can assume that $\mu = 0$ and $\sigma = 1$. Note that $\mathbb{F}_n(t) = \tilde{\mathbb{F}}_n\{\Phi^{-1}(\hat{\sigma}t + \hat{\mu})\}$, where $\tilde{\mathbb{F}}_n(x) = \tilde{\mathbb{F}}_{n1}(x) - \tilde{\mathbb{F}}_{n2}(x)$, with

$$\tilde{\mathbb{F}}_{n1}(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \{\mathbb{I}(X_i \leq x) - \Phi(x)\}, \quad \tilde{\mathbb{F}}_{n2}(x) = \sqrt{n} \left\{ \Phi\left(\frac{x - \hat{\mu}}{\hat{\sigma}}\right) - \Phi(x) \right\}.$$

From van der Vaart & Wellner (1996), $\tilde{\mathbb{F}}_{n1}(x) \rightsquigarrow \tilde{\mathbb{F}}_1(x) = \beta\{\Phi(x)\} + \delta\dot{\mathcal{L}}_0(x)$, where $\dot{\mathcal{L}}_0(x) = \int_{-\infty}^x \dot{\ell}_0(s) ds$. Next, note that $\tilde{\mathbb{F}}_{n2}$ is asymptotically equivalent to $-\phi(x)(W_{n1} + xW_{n2}/\sqrt{2})$, where

$$\begin{aligned} W_{n1} &= \sqrt{n} \frac{\hat{\mu}}{\hat{\sigma}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i + o_{\mathbb{P}}(1), \\ W_{n2} &= \sqrt{2n}(\hat{\sigma} - 1) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{X_i^2 - 1}{\sqrt{2}} \right) + o_{\mathbb{P}}(1). \end{aligned}$$

Equation (6.2) combined with Lecam's third lemma yield $W_{n1} \rightsquigarrow W_1 + \delta\mu'_1$ and $W_{n2} \rightsquigarrow W_2 + \delta\mu'_2/\sqrt{2}$, so that

$$\tilde{\mathbb{F}}_{n2}(x) \rightsquigarrow \tilde{\mathbb{F}}_2(x) = -\phi(x) \left(W_1 + \frac{xW_2}{\sqrt{2}} \right) - \delta\phi(x) \left(\mu'_1 + \frac{x\mu'_2}{2} \right).$$

Since $\sup_{0 \leq t \leq 1} |\Phi^{-1}(\hat{\sigma}t + \hat{\mu}) - \Phi^{-1}(t)|$ converges in probability to 0, the result follows from $\mathbb{F}_n(t) = \tilde{\mathbb{F}}_n\{\Phi^{-1}(\hat{\sigma}t + \hat{\mu})\} \rightsquigarrow \tilde{\mathbb{F}}_1\{\Phi^{-1}(t)\} - \tilde{\mathbb{F}}_2\{\Phi^{-1}(t)\}$. \square

Durbin et al. (1975) obtained the asymptotic representations of CVM_n and AD_n in terms of a weighted sum of non-central chi-square random variables. They were generalizing an earlier result of Durbin & Knott (1972) in the case when μ and σ are specified. The types of alternatives used by Durbin et al. (1975) are however of a complicated form and are difficult to interpret.

Propositions 6.1, 6.2 and 6.3 give the general representations of the seven statistics under sequences of densities that satisfy assumption (6.1). In the next section, three such models are considered for which explicit formulas are given for their asymptotic representations.

6.4 Three kinds of local alternatives

In this section, three families of alternatives to the normal distribution will be introduced. In all cases, the associated densities are mixtures of the form $l_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$, where h is a density. It is then clear that $\lim_{\delta \rightarrow 0} l_\delta(z) = \phi(z)$. It is easy to compute that $\dot{l}_0(z) = h(z) - \phi(z)$. Hence, $\dot{\mathcal{L}}_0(z) = H(z) - \Phi(z)$, where H is the cdf associated to h . Moreover, $\mu'_k = E(Y^k) - E(Z^k)$ and $\tilde{\mu}'_2 = \sqrt{2\pi}E|Y| - 2$, where $Y \sim h$ and $Z \sim \phi$. The three models that will be considered in the sequel are

\mathcal{A}_1 . bimodal alternatives : $h(z) = \phi(x - \mu)$, $\mu > 0$;

\mathcal{A}_2 . kurtosis alternatives : $h(z) = (1/\sigma)\phi(x/\sigma)$, $\sigma > 1$;

\mathcal{A}_3 . Heavy-tailed alternatives : $h(z) = \phi(z)\mathbb{I}(z < 0) + g_\nu(z)\mathbb{I}(z \geq 0)$, where g_ν is Student's density with ν degrees of freedom.

The alternatives (1) and (2) have been used by Thadewald & Büning (2007) to study the power of some tests of normality in small samples. The local alternatives (3) have tails that are heavier than that of the normal distribution.

A corollary to propositions 6.1, 6.2 and 6.3 is now given for the class \mathcal{A}_1 of models. Obviously, the drift terms, i.e. those who appears in the asymptotic distributions as a consequence of contiguity, vanish whenever $\mu = 0$.

Corollaire 6.1. *Under the sequence of alternatives \mathcal{A}_1 ,*

(i) $S_n \rightsquigarrow \mathcal{N}\left(\frac{\delta\mu^3}{\sqrt{6}}, 1\right)$, $K_n \rightsquigarrow \mathcal{N}\left(\frac{\delta\mu^4}{\sqrt{24}}, 1\right)$ and $JB_n \rightsquigarrow \chi_2^2(\delta^2\gamma)$, where $\gamma = \mu^6(\mu^2 + 4)/24$;

(ii) $R_n \rightsquigarrow \mathcal{N}(\delta a(\mu), \sigma_R^2)$, where $a(\mu) = \frac{\mu^2}{2} + \sqrt{\frac{\pi}{2}}\mu + 1 - \sqrt{2\pi}\{\mu\Phi(\mu) + \phi(\mu)\}$;

(iii) $\mathbb{F}_n(t) \rightsquigarrow \mathbb{F}_0(t) + \delta \left[\Phi\{\Phi^{-1}(t) - \mu\} - t + \phi\{\Phi^{-1}(t)\} \left\{ \mu + \frac{\mu^2\Phi^{-1}(t)}{2} \right\} \right]$.

Proof. Let $Y \sim \phi(z - \mu)$, so that Y is equal in distribution to $Z + \mu$, where $Z \sim \phi$. Hence, $\mu'_k = \mathbb{E}\{(Z + \mu)^k - Z^k\}$. Straightforward computations yield

$\mu'_1 = \mu$, $\mu'_2 = \mu^2$, $\mu'_3 = \mu(\mu^2 + 3)$ and $\mu'_4 = \mu^2(\mu^2 + 6)$. Moreover,

$$\begin{aligned} \mathbb{E}|Z + \mu| &= \int_{-\infty}^{\infty} |z + \mu| \phi(z) dz \\ &= - \int_{-\infty}^{-\mu} (z + \mu) \phi(z) dz + \int_{-\mu}^{\infty} (z + \mu) \phi(z) dz \\ &= - \int_{-\infty}^{-\mu} \phi'(z) dz - \mu \int_{-\infty}^{-\mu} \phi(z) dz - \int_{-\mu}^{\infty} \phi'(z) dz + \mu \int_{-\mu}^{\infty} \phi(z) dz \\ &= 2\phi(\mu) + \mu \{2\Phi(\mu) - 1\}, \end{aligned}$$

so that $\tilde{\mu}'_2 = 2\sqrt{2\pi}\phi(\mu) + \sqrt{2\pi}\mu\{2\Phi(\mu) - 1\} - 2$. \diamond

The next corollary is an analog to corollary 6.1 for the class \mathcal{A}_2 .

Corollaire 6.2. *Under the sequence of alternatives \mathcal{A}_2 ,*

(i) $S_n \rightsquigarrow \mathcal{N}(0, 1)$, $K_n \rightsquigarrow \mathcal{N}\left(\frac{\delta(\sigma^2 - 1)^2}{\sqrt{8/3}}, 1\right)$ and $JB_n \rightsquigarrow \chi^2_2(\delta^2\gamma)$, where $\gamma = 3(\sigma^2 - 1)^4/8$;

(ii) $R_n \rightsquigarrow \mathcal{N}\left(\frac{\delta(\sigma - 1)^2}{2}, \sigma_R^2\right)$;

(iii) $\mathbb{F}_n(t) \rightsquigarrow \mathbb{F}_0(t) + \delta \left[\Phi \left\{ \frac{\Phi^{-1}(t)}{\sigma} \right\} - t + \left(\frac{\sigma^2 - 1}{2} \right) \phi \left\{ \Phi^{-1}(t) \right\} \Phi^{-1}(t) \right]$.

Proof. First of all, $\mu'_k = \mathbb{E}\{(\sigma Z)^k - Z^k\} = (\sigma^k - 1)\mathbb{E}(Z^k)$. Hence, $\mu'_1 = \mu'_3 = 0$, $\mu'_2 = \sigma^2 - 1$ and $\mu'_4 = 3(\sigma^4 - 1)$. Also, $\tilde{\mu}'_2 = \sqrt{2\pi}\mathbb{E}|\sigma Z| - 2 = 2(\sigma - 1)$.

\diamond

The fact that S_n , under local sequences of the type \mathcal{A}_2 , has the same asymptotic distribution as under \mathcal{H}_0 is noteworthy. This is easily explained by the fact that the alternatives are symmetric.

Corollaire 6.3. *Under the sequence of alternatives \mathcal{A}_3 ,*

(i) $S_n \rightsquigarrow \mathcal{N}(\delta a_1(\nu), 1)$, $K_n \rightsquigarrow \mathcal{N}(a_2(\nu), \delta, 1)$ and $JB_n \rightsquigarrow \chi_2^2(\delta^2 \gamma)$, where

$$a_1(\nu) = \frac{\sqrt{\nu}(9-\nu)\Gamma\left(\frac{\nu-3}{2}\right)}{4\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} + \frac{1}{\sqrt{2\pi}}, \quad a_2(\nu) = \frac{3(\nu+4)}{(\nu-2)(\nu-4)}$$

and $\gamma = a_1^2(\nu) + a_2^2(\nu)$;

(ii) $R_n \rightsquigarrow \mathcal{N}(\delta a_3(\nu), \sigma_R^2)$, where $a_3(\nu) = \frac{1}{2} \left\{ \frac{1}{\nu-2} - \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} + \frac{1}{\sqrt{2\pi}} \right\}$;

(iii) $F_n(t) \rightsquigarrow F_0(t) + \delta [G_\nu \{ \Phi^{-1}(t) \} - t] \mathbb{I}(t \geq 1/2) - \delta \phi \{ \Phi^{-1}(t) \} \left\{ \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} - \frac{1}{\sqrt{2\pi}} + \frac{1}{2(\nu-2)} \Phi^{-1}(t) \right\}$.

Proof. First note that $\ell_0(z) = \{g_\nu(z) - \phi(z)\} \mathbb{I}(z \geq 0)$. Then, from well-known results about the moments of the Student distribution that one can find, for example, in Casella & Berger (1990), one computes

$$\mu'_k = \begin{cases} \frac{\nu^{1/2}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} - \frac{1}{\sqrt{2\pi}}, & k = 1; \\ \frac{1}{\nu-2}, & k = 2 \\ \frac{\nu^{3/2}\Gamma\left(\frac{\nu-3}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} - \sqrt{\frac{2}{\pi}}, & k = 3; \\ \frac{3(3\nu-4)}{(\nu-2)(\nu-4)}, & k = 4. \end{cases}$$

After straightforward algebra, one finds

$$\begin{aligned} a_1(\nu) &= \frac{\mu'_3 - 3\mu'_1}{\sqrt{6}} = \frac{\sqrt{\nu}(9 - \nu)\Gamma\left(\frac{\nu-3}{2}\right)}{4\sqrt{6\pi}\Gamma\left(\frac{\nu}{2}\right)} + \frac{1}{\sqrt{12\pi}}, \\ a_2(\nu) &= \frac{\mu'_4 - 6\mu'_2}{\sqrt{24}} = \frac{3(\nu + 4)}{\sqrt{24}(\nu - 2)(\nu - 4)}. \end{aligned}$$

Note also that

$$\tilde{\mu}'_2 = \sqrt{2\pi} \int_0^\infty z \{g_\nu(z) - \phi(z)\} dz = \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} - \frac{1}{\sqrt{2\pi}}.$$

Hence,

$$a_3(\nu) = \frac{\mu'_2 - \tilde{\mu}'_2}{2} = \frac{1}{2} \left\{ \frac{1}{\nu - 2} - \frac{\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{2\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} + \frac{1}{\sqrt{2\pi}} \right\}.$$

◇

In section 6, corollaries 6.1, 6.2 and 6.3 will be used in order to compute the local power functions of the tests. It will then be possible to make comparisons in view of their local efficiencies. Before that, the notions of local power curve and local efficiency, as well as strategies to compute them, will be given. This is the subject of the next section.

6.5 Asymptotic power and efficiency

Many attempts have been made to generalize the notion of Pitman's asymptotic relative efficiency outside the asymptotic normality. Rothe (1981), for example, extended the idea to asymptotically chi-squared distributions. It is however unapplicable when dealing with goodness-of-fit procedures based

on functionals of empirical processes. Another possibility is to consider the limiting Bahadur efficiency, which was shown by Wieand (1976) to correspond to the limiting Pitman efficiency under certain conditions. This was the approach privileged by Koziol (1986), where the limiting Bahadur efficiency of CVM_n and AD_n is computed under the local alternatives of Durbin et al. (1975). As mentioned in the introduction, such results are difficult to interpret since one considers the power of the tests as $\alpha \rightarrow 0$.

The approach that will be employed here is based on local power curves. Let W_{n1} and W_{n2} be two test statistics for a given null hypothesis \mathcal{H}_0 . Their local power curves are defined by

$$\beta_i(\delta) = \lim_{n \rightarrow \infty} P(\text{the test based upon } W_{ni} \text{ rejects } \mathcal{H}_0 \mid \ell_{\delta/\sqrt{n}}), \quad i = 1, 2.$$

The asymptotic relative efficiency measure proposed by Berg & Quessy (2009) is

$$\text{ARE}(W_{n1}, W_{n2}) = \left(\lim_{M \rightarrow \infty} \frac{\int_0^M \{1 - \beta_2(\delta)\} d\delta}{\int_0^M \{1 - \beta_1(\delta)\} d\delta} \right)^2. \quad (6.3)$$

The first motivation behind this definition is the fact that it generalizes Pitman's notion of ARE when W_{n1} and W_{n2} are asymptotically normal. To see this, suppose $W_i \rightsquigarrow \mathcal{N}(\delta\mu_i, \sigma_i^2)$, $i = 1, 2$, under a given local alternative. In that case, the local power curves are

$$\beta_i(\delta) = 1 + \Phi\left(-z_{\alpha/2} - \delta \left| \frac{\mu_i}{\sigma_i} \right|\right) - \Phi\left(z_{\alpha/2} - \delta \left| \frac{\mu_i}{\sigma_i} \right|\right), \quad (6.4)$$

where $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$. The proof of the next lemma can be found in Berg & Quessy (2009).

Lemme 6.2. *Let $W_n \rightsquigarrow \mathcal{N}(\delta\mu, \sigma^2)$. Then*

$$\int_0^\infty \{1 - \beta_W(\delta)\} d\delta = \frac{z_{\alpha/2}}{|\mu/\sigma|}.$$

A consequence of lemma 6.2 is that

$$\text{ARE}(W_{n1}, W_{n2}) = \left(\frac{\mu_2/\sigma_2}{\mu_1/\sigma_1} \right)^2$$

when W_{n1} and W_{n2} are asymptotically normal. This is exactly the definition of Pitman's relative efficiency measure.

It is also possible to derive an explicit expression for statistics that are asymptotically non-central chi-squared. The result will be useful to investigate the local power of JB_n , though it is of an independent interest.

Lemme 6.3. *Let $W_n \rightsquigarrow \chi_k^2(\delta^2\gamma)$. Denote by $G(x, \ell)$ the cdf of a central chi-square distribution with ℓ degrees of freedom and $\chi_{\ell, \alpha}^2$ its corresponding percentile of order $1 - \alpha$. Then*

$$\beta_W(\delta) = 1 - \sum_{j=0}^{\infty} \frac{e^{-\delta^2\gamma/2} (\delta^2\gamma/2)^j}{j!} G(\chi_{k, \alpha}^2; 2j + k) \quad (6.5)$$

and

$$\int_0^\infty \{1 - \beta_W(\delta)\} d\delta = \frac{1}{\sqrt{2\gamma}} \int_0^{\chi_{k, \alpha}^2/2} e^{-u} \left\{ \sum_{j=0}^{\infty} \frac{\Gamma(j + 1/2) u^{j+k/2-1}}{j! \Gamma(j + k/2)} \right\} du.$$

In particular,

$$\int_0^\infty \{1 - \beta_W(\delta)\} d\delta = \frac{z_{\alpha/2}}{\sqrt{\gamma}}$$

when $k = 1$ and

$$\int_0^\infty \{1 - \beta_W(\delta)\} d\delta = \frac{1}{\sqrt{2\gamma}} \int_0^{\chi_{2, \alpha}^2/2} e^{-u} \left\{ \sum_{j=0}^{\infty} \frac{\Gamma(j + 1/2) u^j}{j! j!} \right\} du$$

when $k = 2$.

Proof. The expression for β_W is standard. Next, since

$$\int_0^\infty e^{-\delta^2\gamma/2} (\delta^2\gamma/2)^j d\delta = \frac{1}{\sqrt{2\gamma}} \int_0^\infty y^{j-1/2} e^{-y} dy = \frac{1}{\sqrt{2\gamma}} \Gamma\left(j + \frac{1}{2}\right),$$

one computes

$$\int_0^\infty \{1 - \beta(\delta)\} d\delta = \frac{1}{\sqrt{2\gamma}} \sum_{j=0}^\infty \frac{\Gamma(j + \frac{1}{2})}{j!} G(\chi_{k;\alpha}^2; 2j + k).$$

Using the fact that

$$G(x; 2j + k) = \int_0^x \frac{y^{j+k/2-1} e^{-y/2}}{\Gamma(j + k/2) 2^{j+k/2}} dy = \int_0^{x/2} \frac{u^{j+k/2-1} e^{-u}}{\Gamma(j + k/2)} du,$$

one obtains the result after straightforward computations. \square

Note that the preceding result is in accordance with lemma 6.2 when $k = 1$ since in that case, $W_n \rightsquigarrow \chi_1^2(\delta^2\gamma) \equiv X^2$, where $X \sim \mathcal{N}(\delta\sqrt{\gamma}, 1)$. Moreover, lemma 6.3 enables to obtain that $\text{ARE}(W_{n1}, W_{n2}) = \gamma_2/\gamma_1$ when $W_{ni} \rightsquigarrow \chi_k^2(\delta^2\gamma_i)$, $i = 1, 2$. This result can then be used to compare the asymptotic relative efficiency of testing procedures that converge to the chi-squared distribution with the same number of degrees of freedom.

It does not seem possible to obtain an explicit expression for $\int_0^\infty \{1 - \beta(\delta)\} d\delta$ when β is the local power curve of either CVM_n , AD_n or KS_n . An approximate numerical method is then needed. Since $\mathbb{F}_n(t) \rightsquigarrow \mathbb{F}_0(t) + \delta\xi(t)$ under a local sequence that satisfies (6.1), one can estimate the limit of \mathbb{F}_n by $\hat{\mathbb{F}}_0(t) + \delta\xi(t)$, where for a sufficiently large N ,

$$\hat{\mathbb{F}}_0(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \left\{ \mathbb{I}(\hat{T}_{i,N} \leq t) - t \right\},$$

with $\hat{T}_{i,N} = \Phi\{(X_i - \bar{X}_N)/S_N\}$, X_1, \dots, X_N are i.i.d. $\mathcal{N}(0, 1)$ and \bar{X}_N, S_N are the empirical mean and standard deviation. Replacing \mathbb{F}_n by $\hat{\mathbb{F}}_0(t) +$

$\delta\xi(t)$ in the expressions for the Cramér–von Mises and Anderson–Darling statistics that one can find in equations (3.4) and (3.6), approximations under contiguity are given by

$$\widehat{\text{CVM}}(\delta) = \frac{1}{M} \sum_{m=1}^M \left\{ \hat{\mathbb{F}}_0 \left(\frac{m}{M} \right) + \delta\xi \left(\frac{m}{M} \right) \right\}^2$$

and

$$\widehat{\text{AD}}(\delta) = M \sum_{m=1}^M \left\{ \hat{\mathbb{F}}_0 \left(\frac{m}{M} \right) + \delta\xi \left(\frac{m}{M} \right) \right\}^2 / m(M-m),$$

where M has to be chosen arbitrarily large to ensure sufficient accuracy. For the Kolmogorov–Smirnov statistic, a possible approximation is

$$\widehat{\text{KS}}(\delta) = \max_{1 \leq m \leq M} \left| \hat{\mathbb{F}}_0 \left(\frac{m}{M} \right) + \delta\xi \left(\frac{m}{M} \right) \right|.$$

In order to ease the computations, note that

$$\hat{\mathbb{F}}_0 \left(\frac{m}{M} \right) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbb{I} \left(\hat{T}_{i,N} \leq \frac{m}{M} \right) - \frac{\sqrt{Nm}}{M}.$$

Now suppose that for an arbitrary statistic W_n and a fixed value of $\delta > 0$, one has a sample $\hat{W}^{(1)}(\delta), \dots, \hat{W}^{(K)}(\delta)$ of observations that are distributed, asymptotically, as the limiting distribution of W_n under $(\ell_{\delta/\sqrt{n}})_{n \geq 1}$. An approximation for the local power curve of W_n is then given by

$$\hat{\beta}_W(\delta) = \frac{1}{K} \sum_{k=1}^K \mathbb{I} \left\{ \hat{W}^{(k)}(\delta) > q_\alpha \right\}. \quad (6.6)$$

If $\hat{\beta}_W(i/T)$, $i \in \{1, \dots, D\}$, are available, where D is chosen such that $\hat{\beta}_W(D/T) = 1$. An approximation of the asymptotic efficiency at level α is

then given by

$$\widehat{\text{AE}}(W) = \frac{D}{T} + \frac{1 + \alpha}{2T} - \frac{1}{T} \sum_{i=0}^D \hat{\beta}_W \left(\frac{i}{T} \right). \quad (6.7)$$

In the next section, these formulas are used to compute the local power curves and the local efficiencies of the tests.

6.6 Investigation of the asymptotic efficiency of the seven tests

In this section, comparisons of the asymptotic efficiency of the tests are made using the results of sections 4–5.

6.6.1 Local power curves

Here, the local power curves of the seven test statistics will be compared under models \mathcal{A}_1 , \mathcal{A}_2 and \mathcal{A}_3 . For the asymptotically normally distributed statistics S_n , K_n and R_n , these functions are of the form (6.4), while that of JB_n derives from equation (6.5). For CVM_n , AD_n and KS_n , the curves are obtained from (6.6) with $N = 2\,500$, $M = 500$ and 10 000 replicates.

Model \mathcal{A}_1

In figure 6.1, the local power curves of the seven tests are presented when $\mu = 1$ and $\mu = 2$. For $\mu = 1$, JB_n and S_n have very similar local power curves and are the best statistics. They are followed by the three functionals of \mathbb{F}_n and by K_n , with a slight advantage to the Anderson–Darling statistics AD_n . The test based on R_n has a weak power. For $\mu = 2$, the Jarque–Bera test emerges as the best procedure, while S_n is still a good alternative too. Note the equivalence between K_n and S_n in that case. The functionals of \mathbb{F}_n and the test based on R_n have quite similar powers, AD_n being, again, the best among the group.

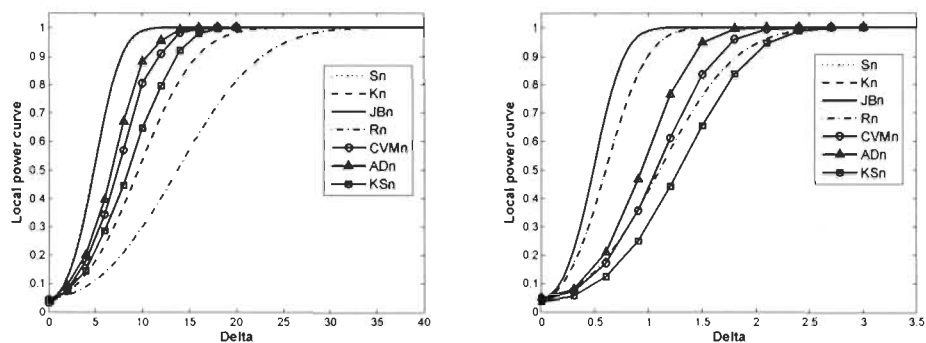


Figure 6.1: Local power curves of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under model \mathcal{A}_1 . Left panel : $\mu = 1$; right panel : $\mu = 2$

Model \mathcal{A}_2

In figure 6.2, the local power curves of the seven tests are presented when $\sigma = 1.5$ and $\sigma = 2$. In both situations, the tests based on K_n and JB_n are the best choices, with a mild advantage to K_n . The test statistic R_n performs rather well, too, especially when σ is small. The test statistics based on functionals of \mathbb{F}_n are less powerful. Similarly as in model \mathcal{A}_1 , AD_n is the best and KS_n is the worst among the three. Note that S_n has no power at all since its behavior under these local alternatives is the same as under \mathcal{H}_0 , as stated in corollary 6.2.

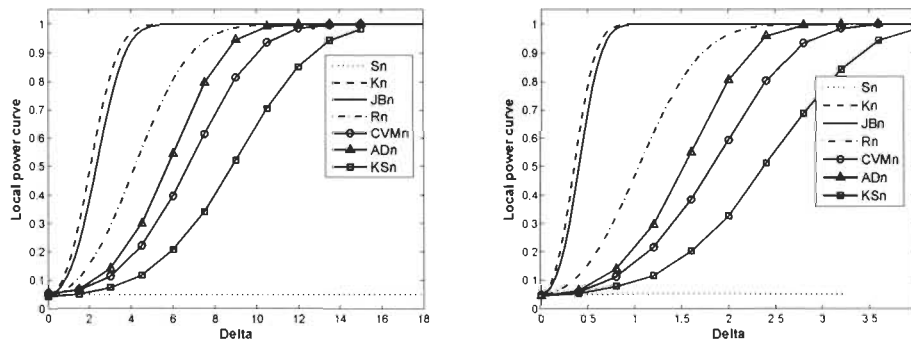


Figure 6.2: Local power curves of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under model \mathcal{A}_2 . Left panel : $\sigma = 1.5$; right panel : $\sigma = 2$

Model \mathcal{A}_3

In figure 6.3, the local power curves of the seven tests are presented when $\nu = 5$ and $\nu = 10$. When $\nu = 5$, the best procedures are those based on

K_n and JB_n . The least powerful test statistics are CVM_n and KS_n . It is more ambiguous in the case of AD_n , S_n and R_n . Indeed, while S_n has a slight advantage over R_n , the local power curve of the Anderson–Darling statistic outperforms the other two for large values of δ . When $\nu = 10$, note the surprisingly good performance of AD_n , especially for large values of δ . However, the differences among the seven tests are less notable than for $\nu = 5$. The asymptotic relative efficiency computations presented in the next subsection will enable to decide on the more powerful statistic in situations when the local power curves intersect.

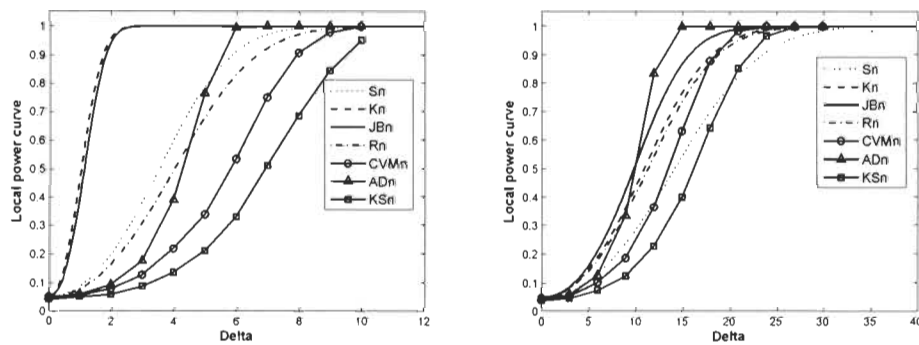


Figure 6.3: Local power curves of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under model \mathcal{A}_3 . Left panel : $\nu = 5$; right panel : $\nu = 10$

6.6.2 Relative efficiency

For a given statistic W_n with an associated local power curve β_W , let $AE_W = 1 / \int_0^\infty \{1 - \beta_W(\delta)\} d\delta$ denote its asymptotic efficiency. The latter intervenes in the definition of asymptotic relative efficiency that one can find in equa-

tion (6.3). The higher AE_W is, the more powerful a statistic is locally. For S_n , K_n , JB_n and R_n , the asymptotic efficiencies can be readily computed. They can be written as functions of the parameters that appear in models \mathcal{A}_1 , \mathcal{A}_2 and \mathcal{A}_3 . These relationships are presented in figures 6.4–6.6.

In figure 6.4, the asymptotic efficiencies as a function of μ of the four statistics in model \mathcal{A}_1 are presented. The cases $\mu \in [0, 1]$ and $\mu \in [1, 3]$ are offered separately since the conclusions differ according to whether μ is small or large. For small values of μ (left panel), S_n and JB_n are the most powerful. In the right panel of figure 6.4, one sees that as μ increases, S_n loses some of its efficiency to the advantage of K_n , which performs almost as good as the Jarque–Bera test statistic. Finally, R_n is the least powerful for all values of μ . These observations are in agreement with the conclusions drawn from figure 6.1.

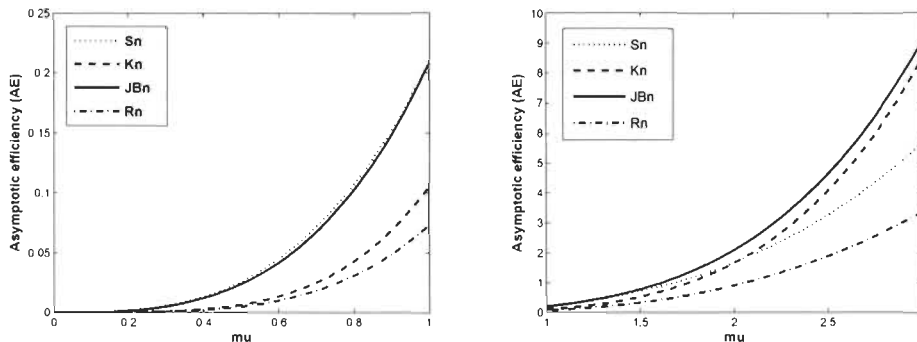


Figure 6.4: Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of μ in model \mathcal{A}_1 . Left panel : $\mu \in [0, 1]$; right panel : $\mu \in [1, 3]$

In figure 6.5, one offers the asymptotic efficiencies in model \mathcal{A}_2 as a function

of σ . As was noted previously, the statistic S_n based on skewness has no power, which explains its absence. The test statistic K_n , based on kurtosis, has the highest asymptotic efficiency, with a slight advantage over JB_n . This is explained by the fact that JB_n , apart from being based on K_n , is also a function of S_n , which does not contribute to the power. The performance of R_n is very weak, especially as the alternative moves away from \mathcal{H}_0 , i.e. as σ increases. Again, these conclusions are in perfect agreement with the local power curves presented in figure 6.2.

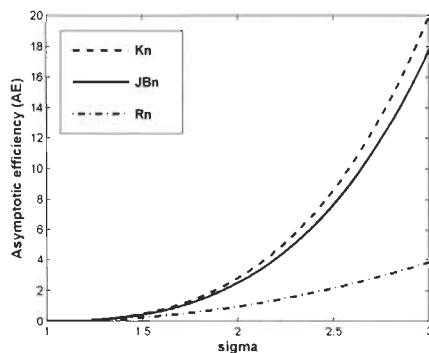


Figure 6.5: Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of $\sigma \in [1, 3]$ in model \mathcal{A}_2

The asymptotic efficiencies that one can find in figure 6.6 indicate the superiority of JB_n and K_n over S_n and R_n in model \mathcal{A}_3 , especially for small values of ν . Obviously, all efficiencies are decreasing functions of ν , since model \mathcal{A}_3 converges to the standard normal as $\nu \rightarrow \infty$. The decrease in efficiency of R_n is however slower than for its competitors, so that R_n exceeds K_n around $\nu = 10$. This is, too, in agreement with figure 6.3.

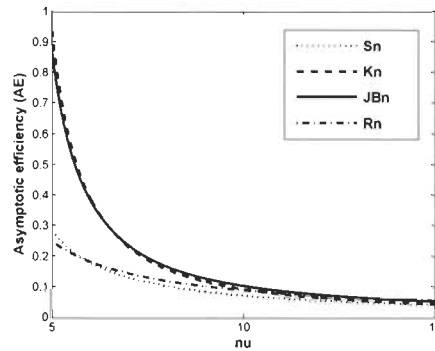


Figure 6.6: Asymptotic efficiency of S_n , K_n , JB_n and R_n as a function of $\nu \in [5, 15]$ in model \mathcal{A}_3

To close this section, the asymptotic relative efficiency measure of Berg & Quesy (2009) presented in (6.3) is computed for the seven test statistics. This will complement the analyses of the local power curves presented in figures 6.1–6.3. For S_n , K_n , JB_n and R_n , the computations are based on lemma 6.2 and lemma 6.3. For the functionals of the empirical process \mathbb{F}_n , the asymptotic efficiencies are approximated with formula (6.7). The results are in table 6.2.

Aside from some cases, the Jarque–Bera test statistics is the best procedure. A notable exception is under model \mathcal{A}_2 , where K_n is about 26% more efficient than JB_n . One can also note the deceptively low power of CVM_n , AD_n and KS_n . These results give other arguments in favor of the Jarque–Bera test for normality. It also confirms the poor behavior of the functionals of \mathbb{F}_n .

Table 6.2: Asymptotic relative efficiency of S_n , K_n , R_n , CVM_n , AD_n and KS_n with respect to the Jarque–Bera test statistic JB_n

Model	parameter	S_n	K_n	R_n	CVM_n	AD_n	KS_n
\mathcal{A}_1	$\mu = 1$	1.0109	0.2527	0.1227	0.4404	0.5324	0.3281
\mathcal{A}_1	$\mu = 2$	0.6317	0.6317	0.1864	0.2084	0.2781	0.1433
\mathcal{A}_2	$\sigma = 1.5$	0.0000	1.2637	0.3046	0.1263	0.1738	0.0731
\mathcal{A}_2	$\sigma = 1$	0.0000	1.2639	0.1469	0.0518	0.0741	0.0304
\mathcal{A}_3	$\nu = 5$	0.1041	1.1596	0.0806	0.0439	0.0826	0.0302
\mathcal{A}_3	$\nu = 10$	0.4709	0.7928	0.7363	0.5806	1.0883	0.4003

6.7 Discussion and link with empirical results

In order to complement the asymptotic results presented in the preceding section, a standard Monte–Carlo simulation study has been designed. The results are to be found in tables 6.3–6.5, where the ability of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n to reject the null hypothesis of normality is approximated under mixture distributions of the form \mathcal{A}_1 , \mathcal{A}_2 and \mathcal{A}_3 . Three sample sizes were considered, namely $n = 50, 100, 250$. The approximate critical values of table 6.1 were used. This was particularly necessary for the Jarque–Bera statistics since its convergence toward its asymptotic distribution is very slow.

First of all, it can be said that the empirical powers match the asymptotic results very well. This is particularly true for small values of δ , because the asymptotic study considered the mixture parameter δ going to infinity at the rate $1/\sqrt{n}$. Hence, the conclusions which were stated in section 6 also apply,

to a large extent, to small and moderate sample sizes. Thus, the Jarque–Bera test is the most recommendable, while the functionals of \mathbb{F}_n , although leading to consistent procedures, are to be avoided.

Some results in tables 6.3–6.5 can be explained from the characteristic of the models considered. For example, note that for the densities in the class \mathcal{A}_1 of models, the central moment of order k is

$$E(X^k) = (1 - \delta)E(Z^k) + \delta E\{(Z + \mu)^k\}.$$

This enables to compute $E(X) = \delta\mu$, $E(X^2) = 1 + \delta\mu^2$, $E(X^3) = \delta(3\mu + \mu^3)$ and $E(X^4) = 3 + \delta\mu^2(\mu^2 + 6)$. Hence, $\text{var}(X) = 1 + \delta(1 - \delta)\mu^2$,

$$\beta_1 = \frac{\delta(1 - \delta)(1 - 2\delta)\mu^3}{\{1 + \delta(1 - \delta)\mu^2\}^{3/2}} \quad \text{and} \quad \beta_2 = \frac{3 + \delta(1 - \delta)\mu^2 \{6 + (3\delta^2 - 3\delta + 1)\mu^2\}}{\{1 + \delta(1 - \delta)\mu^2\}^2}.$$

The fact that β_1 vanishes at $\delta = 1/2$ explains the biases of S_n , which is based on the sample skewness, that one can observe in tables 6.3–6.5. This has also an influence on the Jarque–Bera statistics, especially when $n = 50, 100$. Under model \mathcal{A}_2 , one has $E(X^j) = (1 - \delta + \delta\sigma^j)E(Z^j)$, where $Z \sim \mathcal{N}(0, 1)$. As a consequence, $E(X) = E(X^3) = 0$, $E(X^2) = 1 - \delta + \delta\sigma^2$ and $E(X^4) = 3(1 - \delta + \delta\sigma^4)$. Then, in that case, $\text{var}(X) = 1 - \delta + \delta\sigma^2$, $\beta_1 = 0$ and

$$\beta_2 = \frac{3 + 3\delta(\sigma^4 - 1)}{\{1 + \delta(\sigma^2 - 1)\}^2}.$$

The weak power of S_n is then explained by the fact that these alternatives have a null coefficient of skewness.

Table 6.3: Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 50$

Statistics	$\delta = .1$			$\delta = .3$			$\delta = .5$		
	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$
S_n	5.9	17.2	58.6	4.8	8.4	20.2	4.3	1.4	0.7
K_n	5.3	9.1	26.2	4.7	5.2	13.8	4.6	9.9	50.7
JB_n	5.8	15.3	52.9	4.3	5.5	15.3	3.8	1.0	2.4
R_n	5.5	10.0	30.1	4.5	3.2	7.3	4.0	4.0	36.6
CVM_n	5.1	11.5	43.8	4.7	12.1	56.3	4.2	6.3	42.2
AD_n	5.3	12.8	49.1	4.8	12.2	57.5	4.2	6.8	43.3
KS_n	5.3	10.3	36.5	5.0	11.0	46.5	4.5	6.4	31.4
Statistics	$\sigma = 1.5$			$\sigma = 2$			$\sigma = 2.5$		
	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$
S_n	8.9	21.4	35.7	11.4	24.2	36.6	10.5	17.5	23.7
K_n	9.7	24.8	45.4	12.7	32.8	56.8	10.6	23.5	37.0
JB_n	10.3	26.6	46.8	13.9	34.7	57.1	12.2	25.0	37.9
R_n	8.8	21.5	40.2	12.3	34.8	61.8	12.0	31.0	54.2
CVM_n	5.9	12.0	24.4	7.1	19.5	41.7	7.0	16.8	35.0
AD_n	6.3	14.5	29.9	7.8	22.9	47.1	7.8	18.4	37.5
KS_n	5.6	9.7	19.2	6.6	15.2	32.0	6.6	13.9	26.9
Statistics	$\nu = 5$			$\nu = 10$			$\nu = 15$		
	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$
S_n	8.3	5.9	5.5	13.8	7.8	6.4	20.6	9.9	7.7
K_n	8.2	5.8	5.2	13.5	7.4	6.2	18.7	8.8	6.9
JB_n	8.5	5.7	5.3	15.0	8.3	6.5	21.4	10.4	7.6
R_n	7.0	5.9	5.4	12.0	7.1	6.1	16.1	8.5	6.4
CVM_n	5.7	5.2	4.7	8.2	5.4	4.9	11.2	6.2	5.1
AD_n	6.1	5.3	4.8	9.4	5.8	5.0	13.1	7.1	5.4
KS_n	5.6	5.3	4.8	7.3	5.6	5.2	10.2	6.2	5.1

Table 6.4: Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 100$

Statistics	$\delta = .1$			$\delta = .3$			$\delta = .5$		
	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$
S_n	6.9	32.9	89.1	5.9	17.3	51.4	4.6	1.5	0.6
K_n	8.3	15.2	49.2	7.2	10.7	31.8	7.1	27.8	90.9
JB_n	7.7	30.5	87.3	6.0	14.6	70.7	4.5	4.9	5.6
R_n	5.7	12.6	47.8	4.6	4.2	18.0	3.9	12.7	85.4
CVM_n	4.8	18.4	74.0	5.1	22.7	90.7	4.4	12.2	80.0
AD_n	5.1	21.5	80.5	5.4	24.2	91.4	4.7	13.1	81.3
KS_n	5.7	16.6	65.1	5.7	20.5	82.6	5.1	10.9	64.9
Statistics	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$
	S_n	11.3	28.8	46.3	14.1	30.4	41.5	13.0	21.9
K_n	16.4	43.8	71.1	21.3	59.9	86.5	19.3	46.0	68.3
JB_n	15.9	44.0	70.7	21.6	58.5	84.4	19.7	44.3	64.1
R_n	10.2	30.7	57.6	15.5	54.1	85.6	16.2	48.5	80.6
CVM_n	5.8	15.6	36.7	8.1	31.8	68.7	8.4	28.1	61.6
AD_n	6.7	19.9	45.0	9.4	38.0	74.7	9.7	32.3	65.2
KS_n	6.3	12.5	28.8	7.8	24.1	56.1	8.1	22.6	49.2
Statistics	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$
	S_n	11.7	7.1	6.9	22.0	10.1	7.9	32.6	13.7
K_n	13.5	9.1	8.2	23.4	11.7	9.5	31.7	14.6	10.9
JB_n	13.3	8.2	7.5	24.9	11.6	8.8	35.5	15.5	10.6
R_n	8.7	5.7	5.4	16.0	7.7	6.1	22.1	8.9	6.9
CVM_n	6.6	4.5	4.8	11.0	5.7	5.1	16.6	6.4	5.3
AD_n	7.3	4.8	5.2	13.1	6.3	5.4	19.2	7.4	5.7
KS_n	6.8	5.4	5.4	10.3	5.9	6.1	15.2	6.9	6.0

Table 6.5: Estimation, based on 10 000 replicates, of the power of S_n , K_n , JB_n , R_n , CVM_n , AD_n and KS_n under alternatives of the form $\ell_\delta(z) = (1 - \delta)\phi(z) + \delta h(z)$; $n = 250$

Statistics	$\delta = .1$			$\delta = .3$			$\delta = .5$		
	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 1$	$\mu = 2$	$\mu = 3$
S_n	7.6	63.1	99.7	6.0	39.5	93.2	3.8	1.2	0.3
K_n	5.9	21.8	79.8	4.2	10.7	50.3	4.3	45.2	99.7
JB_n	7.2	56.2	99.7	4.6	36.9	99.3	3.4	18.6	98.3
R_n	5.7	20.4	83.1	4.4	10.1	49.0	4.3	44.7	100.0
CVM_n	6.2	42.8	99.1	6.0	55.7	100.0	5.2	31.7	99.9
AD_n	6.4	48.7	99.5	6.3	58.3	100.0	5.2	34.4	99.9
KS_n	6.0	34.5	96.6	5.8	44.3	99.8	5.3	22.3	98.4
Statistics	$\sigma = 1.5$			$\sigma = 2$			$\sigma = 2.5$		
	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$	$\sigma = 1.5$	$\sigma = 2$	$\sigma = 2.5$
S_n	12.6	32.9	52.9	14.6	31.5	42.8	12.0	21.0	27.4
K_n	21.7	70.1	94.7	32.5	88.8	99.5	28.2	75.1	95.4
JB_n	22.2	68.9	94.0	31.1	86.4	99.2	26.6	70.4	92.6
R_n	13.5	51.8	86.6	25.0	85.1	99.4	25.7	82.1	99.1
CVM_n	7.7	27.6	67.4	12.4	63.4	97.4	12.8	61.6	95.7
AD_n	8.5	35.4	76.7	14.7	71.9	98.4	15.2	66.5	96.8
KS_n	6.5	18.7	51.9	9.3	46.6	91.6	10.0	45.6	88.3
Statistics	$\nu = 5$			$\nu = 10$			$\nu = 15$		
	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$	$\nu = 5$	$\nu = 10$	$\nu = 15$
S_n	13.8	7.1	5.7	34.8	12.9	9.0	54.0	19.4	12.2
K_n	16.3	7.8	6.2	36.3	14.3	9.1	52.3	19.4	11.9
JB_n	17.1	8.3	6.2	39.9	15.6	10.3	59.0	23.2	14.2
R_n	10.5	6.2	5.5	24.4	9.5	6.5	37.2	12.4	8.3
CVM_n	6.9	5.7	5.0	16.3	6.8	5.6	29.3	8.9	6.3
AD_n	7.5	5.7	4.8	18.5	7.4	5.9	33.3	10.3	7.0
KS_n	6.5	5.4	4.9	13.3	6.2	5.4	23.3	7.5	6.1

CONCLUSION

Dans ce mémoire, on s'est intéressé à une panoplie de tests de normalité. Plus particulièrement, on a cherché à calculer la puissance de plusieurs des tests les plus populaires. Deux situations ont été considérées, à savoir les puissances à tailles finies et les puissances asymptotiques.

La première contribution de ce travail se situe au chapitre 5. On y retrouve les résultats d'une vaste étude de simulation pour évaluer la puissance de sept tests de normalité. L'originalité tient aux types d'alternatives qui ont été considérées, à savoir des contre-hypothèses locales. Celles-ci permettent d'évaluer la l'efficacité des procédures à rejeter l'hypothèse de normalité sous de faibles écarts à cette loi.

La deuxième contribution fait l'objet du chapitre 6, où le cas asymptotique est considéré. Plus précisément, le comportement asymptotique des statistiques étudiées au chapitre 5 est obtenu sous plusieurs classes d'alternatives dites locales. Ceci permet de calculer des courbes de puissance locales, et ainsi de comparer théoriquement les procédures de test.

Les conclusions des chapitres 5 et 6 concordent remarquablement bien. Autrement dit, les puissances asymptotiques, donc théoriques, sont pratiquement identiques à celles obtenues dans le cas à tailles finies. Ainsi, les calculs

d'efficacité asymptotique effectuées au chapitre 6 peuvent guider un praticien dans le choix d'un test de normalité approprié. À ce titre, au vu des contre-hypothèses utilisées dans ce mémoire, c'est la statistique de Jarque–Bera qui montre, en général, les meilleures puissances.

Une voie possible de recherches futures serait de considérer une version pondérée de la statistique de Jarque–Bera, à savoir

$$JB_n^\omega = 2\omega S_n^2 + 2(1 - \omega)K_n^2,$$

où $\omega \in [0, 1]$. Cette statistique est une généralisation de la statistique de Jarque–Bera puisque $JB_n = JB_n^{1/2}$. Elle permettrait de donner plus ou moins d'importance aux facteurs d'asymétrie et d'aplatissement. Ainsi, le paramètre ω pourrait être ajusté par un utilisateur selon le type d'écart à la normalité qui est attendu, ou souhaitable à détecter.

ANNEXE A

PROGRAMMES EN MATLAB

A.1 Calcul des statistiques de test

A.1.1 Statistique S_n

```
1 function S = Sn(X)
2
3 a = size(X); n = a(1);
4
5 % Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 % Calcul du coefficient d'asymétrie empirique
10 Betal = 0;
11 for i=1:n
12     Betal = Betal + Z(i)^3;
13 end
14 Betal = Betal/n;
15
16 % Calcul de la statistique de test
17 S = sqrt(n)*abs(Betal)/sqrt(6);
18
19 clear Z;
```

A.1.2 Statistique K_n

```
1 function K = Kn(X)
2
3 a = size(X); n = a(1);
4
5 %Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 %Calcul du coefficient d'aplatissement empirique
10 Beta2 = 0;
11 for i=1:n
12     Beta2 = Beta2 + Z(i)^4;
13 end
14 Beta2 = Beta2/n;
15
16 %Calcul de la statistique de test
17 K = sqrt(n)*abs(Beta2-3)/sqrt(24);
18
19 clear Z;
```

A.1.3 Statistique JB_n

```
1 function JB = JBn(X)
2
3 a = size(X); n = a(1);
4
5 % Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 % Calcul des coefficients d'asymétrie
10 % et d'aplatissement empiriques
11 Beta1 = 0; Beta2 = 0;
12 for i=1:n
13     Beta1 = Beta1 + Z(i)^3;
14     Beta2 = Beta2 + Z(i)^4;
15 end
16 Beta1 = Beta1/n; Beta2 = Beta2/n;
17
18 % Calcul de la statistique de test
19 JB = n*Beta1^2/6 + n*(Beta2 - 3)^2/24;
20
21 clear Z;
```


A.1.4 Statistique R_n

```
1 function R = Rn(X)
2
3 a = size(X); n = a(1);
4
5 % calcul de l'écart-type empirique
6 Sn = std(X);
7
8 % Calcul de l'erreur absolue moyenne (EAM)
9 Xbar = mean(X);
10 EAM = 0;
11 for i=1:n
12     EAM = EAM + abs(X(i)-Xbar);
13 end
14 EAM = sqrt(pi/2)*Jn/n;
15
16 % calcul de la statistique de test
17 R = sqrt(n)*abs((Sn/EAM)-1);
```

A.1.5 Statistique CVM_n

```
1 function CVM = CVMn(X)
2
3 a = size(X); n = a(1);
4
5 % Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 % Calcul des pseudo-observations
10 T = normcdf(Z);
11
12 % Calcul de la statistique de test
13 sum1 = 0; sum2 = 0;
14 for i=1:n
15     sum1 = sum1 + T(i)^2;
16     for j=1:n
17         sum2 = sum2 + max(T(i),T(j));
18     end
19 end
20 CVM = (n/3) + sum1 - (sum2/n);
21
22 clear Z T;
```

A.1.6 Statistique AD_n

```
1 function AD = ADn(X)
2
3 a = size(X); n = a(1);
4
5 % Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 % Calcul des pseudo-observations
10 T = normcdf(Z);
11
12 % Calcul de la statistique de test
13 SUM1 = 0; SUM2 = 0;
14 for i=1:n
15     SUM1 = SUM1 + log(1-T(i));
16     for j=1:n
17         temp = max(T(i),T(j));
18         SUM2 = SUM2 + log((1/temp)-1);
19     end
20 end
21 AD = (SUM2/n) - 2*SUM1 - n;
22
23 clear Z T;
```

A.1.7 Statistique KS_n

```
1 function KS = KSn(X)
2
3 a = size(X); n = a(1);
4
5 %Standardisation des observations
6 Xbar = mean(X); Sn = std(X);
7 Z = (X-Xbar)/Sn;
8
9 %Calcul des pseudo-observations
10 T = normcdf(Z);
11
12 %Calcul de la statistique de test
13 T = sort(T); L1 = zeros(n,1);
14 for i=1:n
15     L1(i) = max(abs(T(i)-(i/n)), abs(T(i)-((i-1)/n)));
16 end
17 KS = sqrt(n)*max(L1);
18
19 clear Z T L1
```

A.2 Programmes de simulation selon des alternatives locales

A.2.1 Mélanges de lois normales

```
1 function X = sim_mixture_norm(n, delta, mu, sigma)
2
3 %Création de l'espace mémoire pour les données
4 X = zeros(n,1);
5
6 %Algorithme de simulation
7 for i=1:n
8     U = rand();
9     if ( U < delta )
10        X(i) = normrnd(mu, sigma, 1, 1);
11    elseif ( U >= delta )
12        X(i) = normrnd(0, 1, 1, 1);
13    end
14
15    clear U;
16 end
```

A.2.2 Mélanges de lois normale et Student

```
1 function X = sim_mixture_student(n, delta , nu)
2
3 %Création de l'espace mémoire pour les données
4 X = zeros(n,1);
5
6 %Algorithme de simulation
7 for i=1:n
8     U = rand();
9     if (U < 1/2)
10        Y = normrnd(0,1,1,1);
11        X(i) = -abs(Y);
12    elseif (U >= 1/2)
13        V = rand();
14        if ( V < 1-delta )
15            Y = normrnd(0,1,1,1);
16            X(i) = abs(Y);
17        elseif ( V >= 1-delta )
18            Y = trnd(nu,1,1);
19            X(i) = abs(Y);
20        end
21    end
22 end
23
24 clear Y U V;
```

A.2.3 Alternative asymétrique

```
1 function X = sim_skew(n, delta)
2
3 % Création de l'espace mémoire pour les données
4 X = zeros(n, 1);
5
6 % Algorithme de simulation (méthode du rejet)
7 for i=1:n
8     U=1; temp=0;
9     while (U > temp)
10        Y = normrnd(0, 1);
11        temp = normcdf(delta*Y);
12        U = rand();
13    end
14    X(i) = Y;
15 end
```

A.3 Estimation des valeurs critiques

```
1 function [qCVM1, qCVM2, qKS1, qKS2, qJB, qR] = quantiles(n,IT)
2
3 CVM1 = zeros(IT,1); CVM2 = zeros(IT,1); KS1 = zeros(IT,1);
4 KS2 = zeros(IT,1); JB = zeros(IT,1); R = zeros(IT,1);
5
6 for i=1:IT
7     X = normrnd(0,1,n,1);
8     [CVM1(i),CVM2(i),KS1(i),KS2(i),JB(i),R(i)] = tests_normality(X);
9     clear X;
10 end
11
12 CVM1 = sort(CVM1);
13 CVM2 = sort(CVM2);
14 KS1 = sort(KS1);
15 KS2 = sort(KS2);
16 JB = sort(JB);
17 R = sort(R);
18
19 q = ceil(0.95*IT);
20
21 qCVM1 = CVM1(q);
22 qCVM2 = CVM2(q);
23 qKS1 = KS1(q);
24 qKS2 = KS2(q);
25 qJB = JB(q);
26 qR = R(q);
27
28 clear CVM1; clear CVM2; clear KS1;
29 clear KS2; clear JB; clear R;
```


A.4 Calculs de puissance

A.4.1 Puissance à tailles finies sous des mélanges de normales

```

1 function [pS,pK,pJB,pR,pCVM,pAD,pKS] = power(n,IT)
2
3 delta = [.1, .3, .5];
4 mu = [1,2,3]; sigma = 1;
5
6 %VC = {1.818, 1.531, 4.516, .531, .128, .753, .883}; VC = {1.926, 1.818, 5.568, .522, .126, .750,
7 .898};
8 pS = zeros(3,3);
9 pK = zeros(3,3);
10 pJB = zeros(3,3);
11 pR = zeros(3,3);
12 pCVM= zeros(3,3);
13 pAD = zeros(3,3);
14 pKS = zeros(3,3);
15
16 for i=1:3
17     for j=1:3
18         for k=1:IT
19             X = sim_mixture_norm(n, delta(i),mu(j),sigma);
20             if ( Sn(X) > VC(1) )
21                 pS(i,j) = pS(i,j) + 1/IT;
22             end
23             if ( Kn(X) > VC(2) )
24                 pK(i,j) = pK(i,j) + 1/IT;
25             end
26             if ( JBn(X) > VC(3) )
27                 pJB(i,j) = pJB(i,j) + 1/IT;
28             end
29             if ( Rn(X) > VC(4) )
30                 pR(i,j) = pR(i,j) + 1/IT;
31             end
32             if ( CVMn(X) > VC(5) )
33                 pCVM(i,j) = pCVM(i,j) + 1/IT;
34             end
35             if ( ADn(X) > VC(6) )
36                 pAD(i,j) = pAD(i,j) + 1/IT;
37             end
38             if ( KSn(X) > VC(7) )
39                 pKS(i,j) = pKS(i,j) + 1/IT;
40             end
41         end
42     end

```

A.4.2 Puissance à tailles finies sous des mélanges de normale et Student

```

1 function [pS,pK,pJB,pR,pCVM,pAD,pKS] = power(n,IT)
2
3 delta = [.1, .3, .5];
4 nu = [5,10,15];
5
6 %VC = (1.818, 1.531, 4.516, .531, .128, .753, .883); VC = (1.926, 1.818, 5.568, .522, .126, .750,
7 .898);
8 pS = zeros(3,3);
9 pK = zeros(3,3);
10 pJB = zeros(3,3);
11 pR = zeros(3,3);
12 pCVM= zeros(3,3);
13 pAD = zeros(3,3);
14 pKS = zeros(3,3);
15
16 for i=1:3
17     for j=1:3
18         for k=1:IT
19             X = sim_mixture_student(n, delta(i), nu(j));
20             if ( Sn(X) > VC(1) )
21                 pS(i,j) = pS(i,j) + 1/IT;
22             end
23             if ( Kn(X) > VC(2) )
24                 pK(i,j) = pK(i,j) + 1/IT;
25             end
26             if ( JBn(X) > VC(3) )
27                 pJB(i,j) = pJB(i,j) + 1/IT;
28             end
29             if ( Rn(X) > VC(4) )
30                 pR(i,j) = pR(i,j) + 1/IT;
31             end
32             if ( CVMn(X) > VC(5) )
33                 pCVM(i,j) = pCVM(i,j) + 1/IT;
34             end
35             if ( ADn(X) > VC(6) )
36                 pAD(i,j) = pAD(i,j) + 1/IT;
37             end
38             if ( KSn(X) > VC(7) )
39                 pKS(i,j) = pKS(i,j) + 1/IT;
40             end
41         end
42     end

```

A.4.3 Puissance à tailles finies sous l'alternative asymétrique

```

1 function [pS,pK,pJB,pR,pCVM,pAD,pKS] = power(n,IT)
2
3 delta = [1,2,3];
4
5 %VC = (1.818, 1.531, 4.516, .531, .128, .753, .883); VC = (1.926, 1.818, 5.568, .522, .126, .750,
6 .898);
7 pS = zeros(3,1);
8 pK = zeros(3,1);
9 pJB = zeros(3,1);
10 pR = zeros(3,1);
11 pCVM= zeros(3,1);
12 pAD = zeros(3,1);
13 pKS = zeros(3,1);
14
15 for i=1:3
16     for k=1:IT
17         X = sim_skew(n, delta(i));
18         if ( Sn(X) > VC(1) )
19             pS(i,j) = pS(i,j) + 1/IT;
20         end
21         if ( Kn(X) > VC(2) )
22             pK(i,j) = pK(i,j) + 1/IT;
23         end
24         if ( JBn(X) > VC(3) )
25             pJB(i,j) = pJB(i,j) + 1/IT;
26         end
27         if ( Rn(X) > VC(4) )
28             pR(i,j) = pR(i,j) + 1/IT;
29         end
30         if ( CVMn(X) > VC(5) )
31             pCVM(i,j) = pCVM(i,j) + 1/IT;
32         end
33         if ( ADn(X) > VC(6) )
34             pAD(i,j) = pAD(i,j) + 1/IT;
35         end
36         if ( KSn(X) > VC(7) )
37             pKS(i,j) = pKS(i,j) + 1/IT;
38         end
39     end
40 end

```

A.4.4 Courbes de puissance locale

```

1 function [beta_CVM,beta_AD,beta_KS] = LP_curve(model,a,N,M,IT,delta)
2
3 beta_CVM=0;
4 beta_AD=0;
5 beta_KS=0;
6
7 for i=1:IT
8     X = normrnd(0,1,N,1);
9     Z = (X-mean(X))/std(X);
10    T = normcdf(Z);
11
12    Fo=zeros(M,1);
13    for m=1:M
14        ml = m-.5;
15        F = sum(T<=(ml/M));
16        Fo(m) = (F/sqrt(N)) - sqrt(N)*ml/M;
17    end
18
19    Q1 = zeros(M,1); Q2 = zeros(M,1);
20    for m=1:M
21        ml = m-.5;
22        temp = Fo(m) + delta*xi(model,(m-.5)/M,a);
23        Q1(m) = temp;
24        Q2(m) = temp/sqrt(ml*(M-ml));
25    end
26
27    CVM_hat = Q1.'*Q1/M;
28    AD_hat = M*Q2.'*Q2;
29    KS_hat = max(abs(Q1));
30
31    clear Q1 Q2 Fo X Z T;
32
33    if CVM_hat > 0.126
34        beta_CVM = beta_CVM + 1/IT;
35    end
36    if AD_hat > 0.751
37        beta_AD = beta_AD + 1/IT;
38    end
39    if KS_hat > 0.898
40        beta_KS = beta_KS + 1/IT;
41    end
42 end

```

Références

- AZZALINI, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171–178.
- BERG, D. & QUESSY, J.-F. (2009). Local power analyses of goodness-of-fit tests for copulas. *Scand. J. Statist.* .
- BOWMAN, K. O. & SHENTON, L. R. (1975). Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2 . *Biometrika* **62**, 243–250.
- CASELLA, G. & BERGER, R. L. (1990). *Statistical inference*. The Wadsworth & Brooks/Cole Statistics/Probability Series. Pacific Grove, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
- CRAMÉR, H. (1928). On the composition of elementary errors. *Skand. Aktuarietidskr* **11**, 13–74.
- D'AGOSTINO, R. & PEARSON, E. S. (1973). Tests for departure from normality. Empirical results for the distributions of b_2 and $\sqrt{b_1}$. *Biometrika* **60**, 613–622.

- D'AGOSTINO, R. & STEPHENS, M. (1986). *Goodness-of-fit techniques*. CRC.
- DURBIN, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Ann. Statist.* **1**, 279–290.
- DURBIN, J. & KNOTT, M. (1972). Components of Cramér-von Mises statistics. I. *J. Roy. Statist. Soc. Ser. B* **34**, 290–307.
- DURBIN, J., KNOTT, M. & TAYLOR, C. C. (1975). Components of Cramér-von Mises statistics. II. *J. Roy. Statist. Soc. Ser. B* **37**, 216–237.
- DURIO, A. & NIKITIN, Y. Y. (2003). Local Bahadur efficiency of some goodness-of-fit tests under skew alternatives. *J. Statist. Plann. Inference* **115**, 171–179.
- GEARY, R. C. (1947). Testing for normality. *Biometrika* **34**, 209–242.
- GEL, Y. R. & GASTWIRTH, J. L. (2008). A robust modification of the Jarque-Bera test of normality. *Econom. Lett.* **99**, 30–32.
- GEL, Y. R., MIAO, W. & GASTWIRTH, J. L. (2007). Robust directed tests of normality against heavy-tailed alternatives. *Comput. Statist. Data Anal.* **51**, 2734–2746.
- GHOUDI, K. & RÉMILLARD, B. (1998). Empirical processes based on pseudo-observations. In *Asymptotic methods in probability and statistics (Ottawa, ON, 1997)*. Amsterdam: North-Holland, pp. 171–197.

- JARQUE, C. M. & BERA, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econom. Lett.* **6**, 255–259.
- JARQUE, C. M. & BERA, A. K. (1987). A test for normality of observations and regression residuals. *Internat. Statist. Rev.* **55**, 163–172.
- JOGESH BABU, G. & RAO, C. R. (2004). Goodness-of-fit tests when parameters are estimated. *Sankhyā* **66**, 63–74.
- KOLMOGOROV, A. (1933). On the empirical determination of a distribution. *Breakthroughs in statistics* **2**, 106–113.
- KOZIOL, J. A. (1986). Relative efficiencies of goodness of fit procedures for assessing univariate normality. *Ann. Inst. Statist. Math.* **38**, 485–493.
- PEARSON, E. S., D'AGOSTINO, R. B. & BOWMAN, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika* **64**, 231–246.
- ROTHER, G. (1981). Some properties of the asymptotic relative Pitman efficiency. *Ann. Statist.* **9**, 663–669.
- SHAPIRO, S. S., WILK, M. B. & CHEN, H. J. M. (1968). A comparative study of various tests for normality. *J. Amer. Statist. Assoc.* **63**, 1343–1372.
- SMIRNOV, E. & POLEJAEFF, W. (1934). Density of Population and Sterility of the Females in the Coccid *Lepidosaphes ulmi* L. *The Journal of Animal Ecology*, 29–40.

- STEPHENS, M. A. (1974). Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* **69**, 730–737.
- STEPHENS, M. A. (1976). Asymptotic results for goodness-of-fit statistics with unknown parameters. *Ann. Statist.* **4**, 357–369.
- THADEWALD, T. & BÜNING, H. (2007). Jarque-Bera test and its competitors for testing normality—a power comparison. *J. Appl. Stat.* **34**, 87–105.
- THODE JR., H. (2002). Testing for Normality, number 164 in ‘Statistics: Textbooks and Monographs’.
- VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. New York: Springer-Verlag. With applications to statistics.
- VASICEK, O. (1976). A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B* **38**, 54–59.
- VON MISES, R. (1931). Wahrscheinlichkeitsrechnung 316–335. *Deuticke, Leipzig*.
- WIEAND, H. S. (1976). A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.* **4**, 1003–1011.