Tests d’ajustement fondés sur la méthode Monte Carlo randomisée pour des distributions exponentielles

Jean-Marie Dufour, Abdeljelil Farhat

Tests d'ajustement fondés sur la méthode Monte Carlo randomisée pour des distributions exponentielles

Jean-Marie Dufour
Université MCGill

Abdeljelil Farhat
Université de Monastir

Première version: Juin 2007
Cette version: Janvier 2009

Résumé

Les distributions Exponentielles sont largement utilisées pour la modélisation des données dans la durée des événements statistiques, l'économétrie et de la finance. Ainsi, les tests d'exponentialité sont un important problème lorsque l'on s'intéresse à l'étude des données. Cependant, la plupart des tests proposés sont limités à la distribution exponentielle avec un seul paramètre et les tableaux des valeurs critiques sont disponibles uniquement pour un nombre limité des tailles des échantillons. Ceci peut être la source d'importants problèmes de niveaux et de puissances des tests.

Dans cette étude, nous proposons d'abord l'utilisation de la technique des tests de Monte Carlo randomisé, en vue de contrôler la la taille des différents tests d'exponentialité. Nous montrons aussi sur le plan théorique et par simulation que les tests obtenus de cette façon sont exacts pour toute taille d'échantillon. Nous proposons aussi des modifications de la procédure fondée sur l'estimateur des moments et nous montrons que ce dernier puisse devenir plus puissant que d'autres tests proposés dans la littérature antérieure.

Deuxièmement, nous étendons l'étude des tests proposés pour la distribution exponentielle avec un paramètre, tels que les tests basés sur les coefficients de Gini (Gail et Gastwirth, 1978, JASA) et sur les distances de Kullback-Leibler (Ebrahimi, Habibullah et Soofi, 1992, JRSS B), à des tests pour les distributions Exponentielles à deux paramètres. Dans une troisième étape, nous proposons l'utilisation des tests d'exponentialité basés sur l'estimation de la densité par la méthode du noyau. Les résultats obtenus montrent que ces derniers ont des meilleures puissances par rapport aux tests basés sur la fonction de répartition empirique ou celui de Shapiro et Wilk (1972, Technometrics).

Cette étude est achevée par la proposition de nouveaux tests basés sur les procédures des combinaisons de plusieurs statistiques de tests. Ces tests ont montré de très bonnes propriétés de puissance.

Mots clées: Test Monte Carlo Randomisé; test d'ajustement; méthode non parametrique; Tests d'exponentialité; procédure de combinaison de tests.

ABSTRACT

Exponential distributions are widely used for modeling duration data in statistics, econometrics and finance. Thus testing for exponentiality is an important problem when studying such data. But, most of the tests proposed are limited to exponential distributions with only one parameter, and the available tables only consider a limited number of sample sizes and levels, which can be the source of size and power problems. In this paper, we first propose to use the technique of Monte Carlo tests in order to control the size of various exponentiality tests. We show both theoretically and by simulation that the tests obtained in this way are exact for any given sample size. Second, we propose modified procedures based on simple moment estimators and show that the latter yield power improvements over alternative tests proposed in the earlier literature. Thirdly, we extend some tests proposed for exponential distributions with one parameter, such as tests based on Gini coefficients (Gail and Gastwirth, 1978, JASA) and Kullback-Leibler distances (Ebrahimi, Habibullah and Soofi, 1992, JRSS B), to exponential distributions with two parameters. Fourth, we propose and show how to obtain exact exponentiality tests based on kernel-type density estimators, which appear to have better power than tests based on empirical distribution functions and the Shapiro and Wilk (1972, Technometrics) test. Fifth, test procedures which combine several tests are proposed and shown to very good power properties. Finally, the procedures studied are compared in a Monte Carlo experiment.

Key words: Monte Carlo test; goodness-of-fit test; nonparametric methods; Exponentiality Tests, combined test procedure.

Introduction et sommaire

Dans cette étude, nous proposons d'abord l'utilisation de la technique des tests de Monte Carlo Randomisée (MCR) (voir Dufour (2006)), en vue de contrôler le niveau des différents tests d'exponentialité. Nous montrons aussi sur le plan théorique et par simulation que les tests obtenus de cette façon sont exacts pour toute taille d'échantillon. Nous proposons aussi des modifications de la procédure fondée sur l'estimateur des moments et nous montrons que ce dernier puisse devenir plus puissant que d'autres tests proposés dans la littérature antérieure.

En résumé, les principaux objectifs de ce travail sont les suivants:
(i) étendre l'utilisation des tests de MCR pour effectuer des tests d'exponentialité;
(ii) montrer que les tests de MCR ont toujours un niveau exact, tandis que les tests originaux ont souvent un niveau biaisé;
(iii) proposer des procédures de combinaison fondées sur les propriétés des tests de MCR pour obtenir des tests plus performants dans le cas où la taille d'échantillon est réduite;
(iv) effectuer la comparaison des performances du point de vue de la puissance pour différents tests afin de sélectionner les meilleurs selon plusieurs contre-hypothèses courantes.

Après une introduction, nous discutons la procédure des tests de MCR. Dans la section 3, nous décrivons les statistiques de test considérées ainsi que la procédure de combinaison proposée. La section 4 est consacrée à la discussion des résultats obtenus dans nos expériences de simulation. Dans la section 5, nous énonçons quelques conclusions.

Pour confronter ce sujet, considérons un échantillon aléatoire X_1, ..., X_n associé à une variable aléatoire X dont la fonction de répartition (f.r.) est $F(x; \alpha, \beta)$, où $\alpha \in R$ et $\beta \in R^+$ sont, respectivement, deux paramètres inconnus de position et d'échelle. Le problème, dans cette étude, est de tester

$$H_0 : F(x; \alpha, \beta) = F_0\left(\frac{x-\alpha}{\beta}\right) = 1 - exp\left\{-\frac{(x-\alpha)}{\beta}\right\},\ x > \alpha. \tag{0.1}$$

contre l'alternative

$$H_1 : F(x; \alpha, \beta) \neq F_0\left(\frac{x-\alpha}{\beta}\right),\ \text{pour au moins une valeur de } x, \tag{0.2}$$

où F_0 est la f.r. de la distribution exponentielle (E).

Si α et β ont été connus, il serait équivalent à vérifier si $Y_i = (X_i - \alpha)/\beta,\ 1 \leq i \leq n,$, sont régi par la f.r. $F_0(x)$. Mais, étant donné qu'ils ne sont pas connus, α et β seront remplacée par des estimations $\hat{\alpha}_X = \hat{\alpha}(X_1, \ldots, X_n)$ et $\hat{\beta}_X = \hat{\beta}(X_1, \ldots, X_n)$. Les statistiques que nous examinerons seront des fonctions des variables observables $W_i = (X_i - \hat{\alpha}_X)/\hat{\beta}_X,\ 1 \leq i \leq n$. Elles ne sont pas mutuellement indépendants et ne suivent pas la distribution de F_0, mais leur distribution est libre des paramètres inconnus α and β. Cela découle du fait que, si on note $\hat{\alpha}_Y = \hat{\alpha}(Y_1, \ldots, Y_n)$ et $\hat{\beta}_Y = \hat{\beta}(Y_1, \ldots, Y_n)$, nous avons

$$\frac{Y_i - \hat{\alpha}_Y}{\hat{\beta}_Y} = \frac{(X_i - \alpha)/\beta - (\hat{\alpha}_X - \alpha)/\beta}{\hat{\beta}_X/\beta} = \frac{X_i - \hat{\alpha}_X}{\hat{\beta}_X},\ 1 \leq i \leq n. \tag{0.3}$$

Par conséquent, toute statistique construite à partir des W_i est pivotale.

Nous savons aussi que, pour la majorité des tests d'ajustement (goodness-of-fit tests (GOF)), la distribution de la statistique de test, sous l'hypothèse nulle, est très compliquée ou même parfois inconnue. Il s'ensuit que les méthodes habituelles sont limitées, soit à des procédures asymptotiques dont la performance n'est pas très fiable, ou d'exiger la construction de tables spécialisées. Pour la majorité, ces tables ont été obtenues par simulation et leur utilisation génère des problèmes selon la nature des données observées. En outre, les tableaux ne peuvent pas couvrir toutes les tailles d'échantillon (voir, par exemple, Lilliefors (1969) ou Maurer and Margolin (1976)).

Pour résoudre ce problème, nous optons pour la mise en place des procédures basées sur les méthodes MCR initialement proposées par Dwass (1957) et Barnard (1963) et utilisées par Birnbaum (1974), Dufour, Farhat, Gardiol, and Khalaf (1998), Dufour and Farhat (2002) et Dufour (2006)].

Pour l'estimation des paramètres inconnus α et β, Stephens (1974) a, souvent, fait recours aux estimateurs du maximum de vraisemblance (Maximum Likelihood Estimates (MLE)). En dépit des qualités reconnues de ces estimateurs, nous notons en cette recherche, où les tailles d'échantillons sont réduites, que l'utilisation des estimateurs des moments améliore clairement, dans plusieurs situations, la puissance des tests étudiés. Cela nous a permis aussi d'introduire des tests de MCR combinés qui ont prouvé, pour la majorité des cas, d'être nettement plus puissants que le tests initiaux.

1. Présentation théorique:

Pour effectuer des tests de l'hypothèse nulle H_0, nous considérons les statistiques de test suivantes:

1) Tests basées sur la fonction de répartition empirique

Les tests basées sur la fonction de répartition empirique (f.r.e.) mesurent des distances entre la f.r.e. et la f.r. de la distribution sous l'hypothèse nulle H_0. Les plus connues sont celui de Kolmogorov-Smirnov (KS) [Kolmogorov (1933), Smirnov (1939)], le test de Cramér-von Mises (CM) [Cramér (1928)], le test d'Anderson-Darling (AD) [Anderson and Darling (1954)] et de Watson (W) [Watson (1961)]. Pour un échantillon fini, les distributions des statistiques de AD et CM sont assez compliquées, mais une théorie asymptotique est disponible. Pour la statistique de KS, les distribution exacte et limite sont non standard et même la distribution asymptotique, les points critiques doivent être estimés.

Nous avons examiné aussi les statistiques obtenues par les modifications des trois dernières statistiques ci-dessus proposées par Stephens(1974).

Pour les quatre premières statistiques, nous allons utiliser les valeurs critiques figurant dans Tableau 4.15 de D'Agostino and Stephens (1986) alors que les valeurs critiques correspondant à la modification des statistiques sont présentées dans le tableau 4.14 de D'Agostino and Stephens (1986).

2) Tests basés sur la Corrélation

Nous allons aussi examiner les statistiques fondées sur le coefficient de corrélation entre le vecteur $W_{(.)} = (W_{(1)}, .., W_{(n)})$ et un vecteur de scores $c = (c_1, .., c_n)$, où $c_1 \leq ... \leq c_n$, liée à F_0. Le coefficient de corrélation est défini comme

$$R(W_{(.)}, c) = \frac{\sum_{i=1}^{n}(W_{(i)} - \bar{W})(c_i - \bar{c})}{\left[\sum_{i=1}^{n}(W_{(i)} - \bar{W})^2 \sum_{i=1}^{n}(c_i - \bar{c})^2\right]^{1/2}} \tag{1.4}$$

et, en ce qui concerne les scores, nous allons nous concentrer sur le choix entre deux options, à savoir,

$$m_i = E\left(F_0^{-1}(U_{(i)})\right), i = 1, ..., n, \tag{1.5}$$

où $U_{(1)} < ... < U_{(n)}$ sont des statistiques d'ordre associées à un échantillon aléatoire de taille n issu d'une distribution $U(0,1)$, et

$$h_i = F_0^{-1}\left(E(U_{(i)})\right) = F_0^{-1}(\frac{i}{n+1}), i = 1, ..., n, \tag{1.6}$$

Et, comme les $W_i, 1 \leq i \leq n$, ne dépendent ni de α ni de β, il est clair que toute statistique définie à partir du coefficient de corrélation (1.4) est pivotale.

Nous considérons les deux statistiques basées sur le coefficient de corrélation:

$$Z(W_{(.)}, m) = n\{1 - R^2(W_{(.)}, m)\}$$

où $m_i = E\left\{-log(1 - U_{(i)})\right\} = \sum_{j=1}^{i}(n-j+1)^{-1}, 1 \leq i \leq n$, [voir Savage (1956)], sont les scores exacts, et

$$Z(W_{(.)}, h) = n\{1 - R^2(W_{(.)}, h)\}$$

où $h_i = -log\left\{1 - E(U_{(i)})\right\} = -log\left\{1 - i/(n+1)\right\}, 1 \leq i \leq n$, sont les scores approximatifs. Ces deux tests ont été considérés par Stephens qui a évalué les valeurs critiques [D'Agostino and Stephens (1986, Tables 5.6 and 5.7)].

3) Test de Shapiro-Wilk

Nous prenons la statistique de Shapiro-Wilk (SW)], telle que définie par [Shapiro and Wilk (1965)] et étudiée par [Metz, Haccou, and Meelis (1994)]. Cette statistique est:

$$SW = \frac{n(\bar{X} - X_{(1)})}{(n-1)S^2} \tag{1.7}$$

où $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2$. Pour points les critiques, nous allons utiliser les valeurs figurant dans le tableau 1 de [Shapiro and Wilk (1965)]

4) Test de Gini

La statistique du test de Gini telle que définie par [Gail and Gastwirth (1978)] est:

$$GI = \frac{\sum_{i=1}^{n-1} i(n-i)(X_{(i+1)} - X_{(i)})}{(n-1)\sum_{i=1}^{n} X_i} \tag{1.8}$$

GIM est la diminution de la transformation monotone GINI statistique, où
GIM = 1/GI

5) Test de Kullback-Leibler

Nous utilisons la statistique de test normalisé

$$KL_{mn} = exp(-I_{mn}) = exp(H_{mn} - Log(\bar{W}) - 1) \tag{1.9}$$

basée sur l'estimateur de Vasicek H_{mn} pour estimer la fonction de Kullback-Leibler de discrimination d'information entre deux distributions de données par

$$I(F : F_0) = \int_0^\infty f(x)Log\{\frac{f(x)}{f_0(x)}\}dx. \tag{1.10}$$

L'estimateur de Vasicek est donné par

$$H_{mn} = n^{-1}\sum_{i=1}^{n} Log\{\frac{n}{2m}(W_{(i+m)} - W_{(i-m)})\}, \tag{1.11}$$

où la taille de la fenêtre m est un nombre entier positif plus petit que $n/2$, $W(j) = W_{(1)}$ si $j < 1, W(j) = W_{(n)}$ si $j > n$. Ce test a été examiné par Ebrahimi, Habibullah et Soofi qui a évalué les valeurs critiques [Ebrahimi, Habibullah, and Soofi (1992, Tables 1 and 2)].

5) Tests de Monte Carlo standardisés

Une fois l'étude de simulation sur la base des statistiques ci-dessus a été réalisée, nous avons remarqué qu'un groupe de tests MCR ont donné lieu à la puissance importante pour un premier sous-ensemble d'alternatives, mais très faible pour un deuxième sous-ensemble. Pour exploiter ce fait, nous allons examiner ici les tests basés sur le maximum de plusieurs des statistiques normalisées. La normalisation vise à assurer comparabilité entre les statistiques et consiste simplement à soustrayant la moyenne empirique de chaque statistique et en divisant le résultat par l'erreur type correspondant empirique, où la moyenne empirique et l'erreur type sont calculées à partir des valeurs observées et simulées de la test statistique.

2. Etude de simulation

Pour évaluer les niveaux empiriques des tests étudiés, nous avons généré des échantillons de tailles relativement faibles ($<$50)de $E(0,1)$ et afin d'évaluer leurs compétences empiriques, nous avons généré des échantillons provenant des distributions: valeur absolue d'une normale standard $|N|$, $B(1,1)$, $\Gamma(2)$, lognormal ($\Lambda(4,\ 1.5^2)$), $|t|$ (student(5) en valeur absolue) et U (uniforme[0,1]).

References

Anderson, T. W., and D. A. Darling (1954): "A Test of Goodness-of-Fit," *Journal of the American Statistical Association*, 49, 765–769.

Barnard, G. A. (1963): "Comment on 'The Spectral Analysis of Point Processes' by M. S. Bartlett," *Journal of the Royal Statistical Society, Series B*, 25, 294.

Birnbaum, Z. W. (1974): "Computers and Unconventional Test-Statistics," in *Reliability and Biometry*, ed. by F. Proschan, and R. J. Serfling, pp. 441–458. SIAM, Philadelphia, PA.

Cramér, H. (1928): "On the Composition of Elementary Errors," *Skandinavisk Aktuarietidskrift*, 11, 141–180.

D'Agostino, R. B., and M. A. Stephens (eds.) (1986): *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

Dufour, J.-M. (2006): "Monte Carlo Tests with Nuisance Parameters: A General Approach to Finite-Sample Inference and Nonstandard Asymptotics in Econometrics," *Journal of Econometrics*, 133, 443–477.

Dufour, J.-M., and A. Farhat (2002): "Exact Nonparametric Two-Sample Homogeneity Tests," in *Proceedings of the 2000 International Workshop on "Goodness-of-fit Tests and Validity of Models"*, ed. by C. Huber-Carol, N. Balakhrishnan, M. Nikulin, and M. Mesbah, chap. 33, pp. 435–448. Birkhaüser, Boston, Massachusetts.

Dufour, J.-M., A. Farhat, L. Gardiol, and L. Khalaf (1998): "Simulation-Based Finite Sample Normality Tests in Linear Regressions," *The Econometrics Journal*, 1, 154–173.

Dwass, M. (1957): "Modified Randomization Tests for Nonparametric Hypotheses," *Annals of Mathematical Statistics*, 28, 181–187.

Ebrahimi, N., M. Habibullah, and E. S. Soofi (1992): "Testing Exponentiality Based on Kullback-Leibler Information," *Journal of the Royal Statistical Society, Series B*, 54(3), 739–748.

Gail, M. H., and J. L. Gastwirth (1978): "A Scale-Free Goddness-of-Fit Test for the Exponential Distribution Based on the Lorenz Curve (in Theory and Methods)," *Journal of the American Statistical Association*, 73, 787–793.

KOLMOGOROV, A. N. (1933): "Sulla determinazione empirica di una legge di distribizione," *Giornale dell' Instituto Italiano degli Atturai*, 4, 83–91.

LILLIEFORS, W. H. (1969): "On the Kolmogorov-Smirnov Test for the Exponential Distribution with Mean Unknown," *Journal of the American Statistical Association*, 64, 387–389.

MAURER, W., AND B. H. MARGOLIN (1976): "Tests of Kolmogorov-Smirnov Type for Exponential Data with Unknown Scale, and Related Problems," *Biometrika*, 63, 149–160.

METZ, J. A. J., P. HACCOU, AND E. MEELIS (1994): "On the Shapiro-Wilk Test and Darling's Test for Exponentiality," *Biometrics*, 50, 527–530.

SAVAGE, I. R. (1956): "Contributions to the Theory of Rank Order Statistics: Two-Sample Case," *Annals of Mathematical Statistics*, 27, 590–615.

SHAPIRO, S. S., AND M. B. WILK (1965): "An Analysis of Variance Test for the Exponential Distribution (Complete Samples)," *Technometrics*, 14, 355–370.

SMIRNOV, N. V. (1939): "Sur les écarts de la courbe de distribution empirique (Russian/French Summary)," *Matematičeskiĭ Sbornik N.S.*, 6, 3–26.

STEPHENS, M. A. (1974): "EDF Statistics for Goodness-of-Fit and some Comparisons," *Journal of the American Statistical Association*, 69, 730–737.

WATSON, G. S. (1961): "Goodness-of-Fit Tests on a Circle," *Biometrika*, 48, 109–114.