Sankhyā : The Indian Journal of Statistics
2016, Volume 78-A, Part 1, pp. 133-153
© 2015, Indian Statistical Institute

Comparing Two Mixing Densities in Nonparametric Mixture Models

Denys Pommeret
Aix Marseille Université, Marseille Cedex 9, France

Abstract

In this paper we consider two nonparametric mixtures of quadratic natural exponential families with unknown mixing densities. We propose a statistic to test the equality of these mixing densities when the two natural exponential families are known. The test is based on moment characterizations of the distributions. The number of moments is retained automatically by a data driven technique. Some examples and simulations of implementation of the procedure are provided.

AMS (2000) *subject classification.* Primary 62G10; Secondary 62G07.
Keywords and phrases. Akaike's rule, Natural exponential families, Nonparametric mixture, Quadratic variance function, Schwarz's rule, Smooth test

1 Introduction

Let X_1 and X_2 be two random variables taking values in some real spaces $\mathcal{X}_1$ and $\mathcal{X}_2$. We assume that there exists two dominating probability measures, μ_1 on $\mathcal{X}_1$ and μ_2 on $\mathcal{X}_2$, such that the densities of X_1 and X_2 with respect to μ_1 and μ_2 are

$$f_1(x) = \int_{M_1} f_1(x|m)\pi_1(dm) \quad \text{and} \quad f_2(x) = \int_{M_2} f_2(x|m)\pi_2(dm), \tag{1.1}$$

where $f_1(.|m)$ and $f_2(.|m)$ are two known densities with respect to μ_1 and μ_2, respectively, π_1 and π_2 are two unknown distributions on some given sets M_1 and M_2. Models (1.1) belong to the class of nonparametric mixture models. This class can be illustrated by different situations as described in van and de Geer (2003), Groeneboom (1984), & Griffin (2010), among others. The most common situation being the convolution model, when $X = U + V$, with U and V two independent and unobservable real-valued random variables; V having a known density g, and the distribution π of U being unknown. Then the density f of X is the convolution of g with π; that is, $f(x) = \int f(x|m)\pi(dm)$, where $f(x|m) = g(x-m)$. In the one-sample case, the Maximum Likelihood Estimator (MLE) of π have been studied in van and de Geer (2003), when $f(.|m)$ is known. Recently, Chee and Wang (2013) (see also Wang, 2007) considered nonparametric estimation of mixing distributions by minimizing the quadratic distance between the empirical and the mixture distribution. They obtained the convergence of their estimator in L_2-norm. Asymptotic properties of MLE could be used for testing the mixing density in the one-sample case. More generally, classical powerful tests can be used for testing $\pi = \pi_0$ since under this equality the distribution of X is completely known. This fact is not still true in the two-sample case that we consider and we propose a method to avoid this problem of unobserved distributions.

We can illustrate the two-sample situation problem with a real data set, the Framingham Study on coronary heart disease described by Carroll et al. (2010). The data consist of measurements of two systolic blood pressures obtained at two different examinations. The two dates of examination coincide with two different populations and each measurement is considered as a convolution of the random systolic blood with an error due to the random measurement effect. The systolic blood distributions, denoted by π_1 and π_2 in (1.1), are unknown and the noise densities, $f_1(.|m)$ and $f_2(.|m)$, due to the measurement are assumed to be known and normally distributed (see Wang and Wang, 2011). This data set raises two different problems. First the reconstruction of the true systolic distributions, without error measurements. This is done in Wang and Wang (2011) with a deconvolution kernel method. Second, the comparison of these two systolic distributions which are not observed directly but contaminated through mixtures. This test of equality between π_1 and π_2 is the purpose of this paper.

Another model related to (1.1) is the well known compound model widely studied in the literature. It can be illustrated in actuarial science when two insurers observe the aggregate claims amounts of two non-life insurance portfolios over a given time period given by

$$X_1 = \sum_{j=1}^{N_1} U_{1j} \quad \text{and} \quad X_2 = \sum_{j=1}^{N_2} U_{2j}, \tag{1.2}$$

where the counting variables N_1, N_2 models the number of claims, and (U_{1j}, U_{2j}) is a sequence of i.i.d. couples of non-negative random variables that model severities incurred in the two portfolios (see for instance Asmussen and Albrecher 2010 for a complete review of such models). The distributions of the number of claims have been largely studied in the actuarial literature. Binomial negative, Poisson, generalized Poisson, or Poisson-Goncharov distributions (see for instance Denuit, 1997) canprovide probability model to describe the number of claims incurred by an insured motorist, for instance. Another interesting problem is to compare the number of claims on two different periods; that is, the two distributions π_1 and π_2 of N_1 and N_2, respectively. We will illustrate this problem on a real data set from a French insurance company. The data are anonymously extracted from observations of aggregate claims amounts during two periods 1999-2001 and 2002-2004.

REMARK 1.1. In the discrete case, the most famous infinite mixture is the Gamma-Poisson mixture yielding to the negative binomial distribution. In the continuous case, infinite Gaussian mixtures have been extensively used in Bayesian works (see for instance Griffin, 2010). More generally, nonparametric Bayesian approaches put their prior on infinitely complex models. For instance, infinite Gaussian mixtures can be used to model Protein sequences (see Dubey et al., 2004). Infinite von Mises-Fisher Mixture is also used to model treatment in radiation therapy (see Bangert et al., 2010).

In this paper we consider model (1.1) and the following two-sample problem

$$H_0 : \pi_1 = \pi_2 \quad \text{against} \quad H_1 : \pi_1 \neq \pi_2, \tag{1.3}$$

based on iid copies of X_1 and X_2.

To construct a test statistic we will restrict our attention to the case where the densities $f_i(.|m)$ $(i = 1, 2)$ belong to the class of Natural Exponential Families with Quadratic Variance Functions (NEF-QVF). This class of distributions contain six types of distributions: Poisson, normal, gamma, binomial, negative binomial and hyperbolic, and all their transformations by affinities and convolutions. It includes then some of the most popular distributions that can be used in (1.1), leading for instance to Poisson-gamma mixtures, and finite or infinite Gaussian mixtures. The complete description of NEF-QVF is given in Morris (1982) and we will give a short review in Section 2.

The way to construct the test statistic is to use some conditional properties of the densities $f_i(.|m)$ which are inherent to the NEF-QVF. Such properties were used by Lindsay (1989) in the case of discrete finite support point mixing distributions. The author obtained an estimator of the number of points of the support. In the framework of mixture of NEF-QVF, Pommeret (2005) obtained an expression of the distance between a mixture distribution and its parent distributions. In this paper we use polynomial representations of NEF-QVF to get simple expressions of the moments of π_1 and π_2 that we compare under the null hypothesis. Under the assumption that all moments characterize π_1 and π_2, we construct a smooth goodness of fit test as proposed in Ghattas et al. (2011) (see Rayner and Best, 2001 for more references). The number of moments to be compared is selected automatically by a penalized criterion. Schwarz (1978) criterion is considered as in Inglot et al. (1997) (see also Ledwina, 1994). Akaike criterion (see Akaike, 1974) will be also discussed as in Inglot and Ledwina (2006). Asymptotic properties of the test are obtained and its comportment is measured through simulations. The paper is organized a follows: In Section 2, we recall some basic results on NEF-QVF. In Section 3 we present the construction of the test. Convergence properties are given in Section 4 and an improvement of the statistic is proposed in Section 5. The paired case is considered in Section 6. Section 7 is devoted to the simulation study and Section 8 contains two real data illustrations.

2 Mixture of Natural Exponential Families

2.1. Natural Exponential Family with Quadratic Variance Function. In this Section we point out some definition concerning NEF (see Barndorff-Nielsen 1978 for more details). Let μ be a positive measure on $\mathbb{R}$, not a mass of Dirac and such that the interior of the domain of its Laplace transform, denoted by Θ_μ, is not empty. Write

$$k_\mu(\theta) \quad = \quad \log\left\{\int \exp(\theta x)\mu(dx)\right\},$$

for the *cumulant function* of μ and denote by ψ_μ the inverse function of the gradient k'_μ. The Natural Exponential Family (NEF) generated by μ is the set

$$F = \{P(m, F); m \in M = k'_\mu(\Theta_\mu)\},$$

where each $P(m, F)$ is a probability with mean m and with density w.r.t. μ given by

$$f_\mu(x|m) \quad = \quad \exp\{\psi_\mu(m)x - k_\mu(\psi_\mu(m))\}. \tag{2.1}$$

Note that the generating measure μ is not necessary a probability measure. Moreover a NEF F can be generated by all its elements, and if $\mu = P(m_0, F)$ we have $f_\mu(x|m_0) = 1$.

The variance of $P(m, F)$ is denoted by $V_F(m)$ and V_F is called the *variance function* of F. Note that the variance $V_F(m)$ coincide with the hessian

$k''_\mu(\psi_\mu(m))$. The family F is a NEF with Quadratic Variance Function (QVF) if its variance function is a second order polynomial, namely

$$V_F(m) = am^2 + bm + c. \tag{2.2}$$

This class of NEF-QVF is entirely described in Morris (1982) and it contains six types of distributions families: Poisson (the case $a = c = 0, b = 1$), normal (the case $a = b = 0$), Gamma (the case $b = c = 0$), binomial (the case $a < 0, b = 1, c = 0$), negative binomial (the case $a > 0, b = 1, c = 0$) and hyperbolic (the other cases). We have compiled some basic facts in Annex. In Table 4, the densities $f_\mu(x|m)$ are those of normal with mean m and variance σ^2, Poisson with mean m, Gamma with form parameter λ and with mean m, and negative binomial with convolution parameter λ and with mean m, respectively. Each density is considered w.r.t. a generating measure μ.

Associated orthogonal polynomials are constructed by deriving successively the density of $P(m, F)$. More precisely for all $(m, n) \in M \times \mathbb{N}$, we define

$$Q_n(x, m_0) = \frac{\partial^n}{\partial m^n}\{f_\mu(x|m)\}|_{m_0}, \tag{2.3}$$

which is a nth degree polynomial in x. If F is a NEF-QVF, the sequence $(Q_n(x, m_0))_{n\in\mathbb{N}}$ forms a $P(m_0, F)$ orthogonal basis (see Morris, 1982); that is,

$$\int Q_j(x, m_0)Q_k(x, m_0)P(m_0, F)(dx) = \delta_{jk}\|Q_j(., m_0)\|^2,$$

where $\delta_{jk} = 1$ if $j = k$ and 0 otherwise. Such classical orthogonal polynomials are described in Table 5 to make the paper self-contained (see Abramowitz and Stegun, 1965 for notation). Since we will use the mean parametrization for the negative binomial distribution, we will write $\mathcal{M}^{\lambda,(m)}$ instead of $\mathcal{M}^{\lambda,c}$ for the associated Meixner polynomials where $c = m/(\lambda+m)$. The calculation of the norms $\|Q_n(., m)\|^2 = \int Q_n^2(x, m)P(m, F)(dx)$, are given in Table 6.

2.2. Some Properties of NEF-QVF. Let μ_1 and μ_2 be two real probability measures with mean m_1 and m_2, respectively, generating two NEF-QVF denoted by F_1 and F_2. Let $f_1(.|m)$ and $f_2(.|m)$ be two density in F_1 and F_2 satisfying (1.1) and such that $M_1 \cap M_2$ is not empty. For $i = 1, 2$ we will denote by $\mathcal{Q}_i = \{Q_{i,j}(., m_i); j \in \mathbb{N}\}$ the set of μ_i-orthogonal polynomials defined as in (2.3) and we will write

$$P_{i,j}(x, m_i) = Q_{i,j}(x, m_i)/\|Q_{i,j}(., m_i)\|^2.$$

If X_i has density $f_i(.|m)$ $(i = 1, 2)$, conditioning by m we have the following property, (see (Morris, 1982) eq. (8.8))

$$\mathbb{E}(P_{i,j}(X_i, m_i)|m) \quad = \quad (m - m_i)^j / j!. \tag{2.4}$$

Finally, considering (1.1) as a mixture of NEF-QVF and writing Y_i, $i = 1, 2$, a random variable with distribution π_i, an integration of the previous equality (2.4)yields

$$\mathbb{E}((Y_i - m_i)^j) \quad = \quad j!\mathbb{E}(P_{i,j}(X_i, m_i)). \tag{2.5}$$

This equality (2.5) is the key of the construction of the test statistic since it permits to evaluate the non observed moments of π_1 and π_2 through polynomial transformations of the observations.

Remark 2.1. It is worth pointing out that such properties can be adapted to more general models than (1.1). For instance consider the mixture (1.2) when the U_{1j} are iid $\Gamma(a_1, b_1)$ distributed and the U_{2j} are iid $\Gamma(a_2, b_2)$ distributed. We obtain

$$f_1(x) = \int_{M_1} f_1(x|a_1 m)\pi_1(dm) \quad \text{and} \quad f_2(x) = \int_{M_2} f_2(x|a_2 m)\pi_2(dm).$$

In this case (2.4) and (2.5) become

$$\mathbb{E}(P_{i,j}(X_i, m_i)|m) \quad = \quad a_i^j (m - m_i)^j / j!, \quad i = 1, 2 \tag{2.6}$$

$$\mathbb{E}((Y_i - m_i)^j) \quad = \quad a_i^{-j} j!\mathbb{E}(P_{i,j}(X_i, m_i)), \quad i = 1, 2. \tag{2.7}$$

This situation is considered in the study of an insurance data set in Section 8.

3 Construction of the Test Statistic

3.1. Statistics Based on the Moments. Consider two iid samples, $(X_{11}, \cdots, X_{1n})$ and $(X_{21}, \cdots, X_{2s})$ from a mixture of NEF-QVF having the form (1.1). Note that X_1 and X_2 can be paired, with $n = s$, assuming that $\mathbb{V}(X_1 - X_2) > 0$. We first restrict our attention to the non-paired case where the two samples are independent. The paired case will be treated in Section 6.

We assume that all the moments characterize the probability measures π_1 and π_2. This assumption is clearly true if their supports are bounded. If their supports are unbounded, it can be assumed that the Carleman condition (see Carleman, 1926) is satisfied, implying the divergence of the series $\sum m_{2n}^{a_{2n}}$, where m_n is the nth moment and $a_n = -1/n$. Another condition should be that π_1 and π_2 belong to exponential families. Then, testing H_0 consists in testing the equalities of all moments of π_1 and π_2. From (2.5), when $m_1 = m_2 = m$, it is equivalent to compare both coefficients

$$\mathbb{E}(P_{1,j}(X_1, m)) \quad \text{and} \quad \mathbb{E}(P_{2,j}(X_2, m)), \quad \text{for} \quad j = 1, 2, 3, \cdots .$$

For this reason we will chose

$$m_1 = m_2 = m.$$

Our test statistic is then based on the estimators of their k first differences. Write

$$U_k \;=\; \left(\frac{1}{n} \sum_{i=1}^{n} P_{1,j}(X_{1i}, m) - \frac{1}{s} \sum_{i=1}^{s} P_{2,j}(X_{2i}, m) \right)_{j=1,\cdots,k}$$

If we assume that $n/(n+s) \to a \in]0;1[$, we have the following convergence to a centered normal distribution with covariance W:

$$\sqrt{\frac{ns}{n+s}}\, U_k \;\to\; N(0, W),$$

where

$$\begin{aligned} W &= (1-a)W_1 + aW_2, \\ diag(W_1) &= (\mathbb{V}(P_{1,1}(X_1, m)), \cdots, \mathbb{V}(P_{1,k}(X_1, m))), \\ diag(W_2) &= (\mathbb{V}(P_{2,1}(X_2, m)), \cdots, \mathbb{V}(P_{2,k}(X_2, m))). \end{aligned}$$

The test statistic is given by

$$T_k \;=\; \frac{ns}{n+s}\, U_k \widehat{W_k}^{-1} U_k, \tag{3.1}$$

where $\widehat{W_k} = diag(\widehat{v}_1, \cdots, \widehat{v}_k)$, and $\widehat{v}_i$ are convergent estimators of $v_i = \mathbb{V}(P_{1,i}(X_1, m_1)) + \mathbb{V}(P_{2,i}(X_2, m_2))$, with a trimming $e_{n,s} > 0$ such that $e_{n,s} \to 0$ when n, s tend to infinity. More precisely we chose $\widehat{v}_i = \widehat{w}_i + e_{n,s}$, where $\widehat{w}_i$ is the empirical estimator of v_i.

REMARK 3.1. The choice $m_1 = m_2 = m$ lead to a simplification in (2.5) which permits to write

$$\pi_1 = \pi_2 \;\Leftrightarrow\; \mathbb{E}(P_{1,j}(X_1, m)) = \mathbb{E}(P_{2,j}(X_2, m)), \quad \text{for} \quad j = 1, 2, 3, \cdots .$$

If $m_1 \neq m_2$ we must combine all moments given by (2.5) which would necessitate to combine several estimators leading to a loss of efficiency.

3.2. Data Driven Selection. The idea underlying smooth tests is to model the difference between π_1 and π_2 under H_1 as a series expansion along a k-dimensional family of functions. Testing H_0 thus reduces to test the nullity of the coefficients in the expansion. The number k of selected coefficient turns out to be a nuisance parameter to be optimized. A pioneer paper about smooth tests is Neyman (1937). Many references to this work can be found in Hart (1997). The performances of the test rely on the right choice of k (for discussions, see D'Agostino and Stephens, 1986; Rayner and Best, 1989). For a long time, many authors have restricted attention to study $k \in \{1, 2, 3, 4\}$. However, such choice can be misleading in some situations (see Inglot et al., 1990). More generally, since the right choice depends on the type of alternative, which is unknown, a deterministic choice of k is not satisfactory. A solution to this problem has been first proposed by Ledwina (1994) who studied a data-based method for choosing k between 1 and some arbitrary fixed K, inspired from Schwarz (1978)'s criterion selection rule. Afterward, some refinements have been proposed by several authors, among whose Kallenberg and Ledwina (1995a), Inglot et al. (1997). In these papers, K is allowed to tend to infinity at a given rate as the sample size tends to infinity. Once the optimal k is chosen, the corresponding test statistic is performed and used for testing H_0. Beyond Fan (1996) & Janic-Wróblewska and Ledwina (2000), smooth tests theory as well as data-driven methods for the selection of k have been adapted to the two-sample testing problem by Ghosh (2001) & Albers et al. (2001), in the independent case, and Ghattas et al. (2011) in the paired case.

To select the number k of components in the test statistic we follow Inglot et al. (1997) considering an increasing sequence of number of components $k(n, s)$ such that $\lim_{n,s\to\infty} k(n, s) = \infty$. Our selection rule is based on the Schwarz (Schwarz, 1978) criterion. Write

$$S_{n,s} \quad = \quad \min\left\{ \underset{1\le k\le k(n,s)}{\text{argmax}} \left(T_k - k\log(ns/(n+s))\right)\right\}.$$

The proposed data driven test statistic is given by

$$T(n, s) = T_{S_{n,s}} = \frac{ns}{n+s}\, U_{S_{n,s}} \widehat{W_{S_{n,s}}}^{-1} U_{S_{n,s}}.$$

The idea of this data driven criterion is that, asymptotically, $S_{n,s}$ will detect the most significant difference among all the moments of π_1 and π_2. But under H_0, since there is no difference, $S_{n,s}$ will converge to 1.

REMARK 3.2. In order to make clearer the application of Schwarz's criterion in our setting, let us notice that the testing problem (1.3) can be rewritten as $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$, where $\theta = \mathbb{E}(P_{1,i}(X_1, m) - P_{2,i}(X_2, m))_{i=1,\cdots,k}$. Suppose that the MLE $\widehat{\theta}$ of θ equals an empirical mean, that is $\widehat{\theta} = U_k$, as it is the case for instance when the distribution of $P_{1,i}(X_1, m) - P_{2,i}(X_2, m)$ belongs to an exponential family. Then, T_k is the score statistic.

4 Convergence of the Test Statistic

We will need the following two assumptions

(A) $k(n, s) = o(\sqrt{e_{n,s} \log(ns/(n + s))/(\alpha_n + \beta_s)})$,
where $\alpha_n = \max_{k\leq n}(\mathbb{V}(P_{1,k}(X_1, m)))$ and $\beta_s = \max_{k\leq s}(\mathbb{V}(P_{2,k}(X_2, m)))$.

(B) There exists $a \in]0; 1[$ such that $n/(n + s) \to a$, when n and s tend to infinity.

REMARK 4.1. Condition (A) implies that $e_{n,s} \log(ns/(n + s)) \to +\infty$. But for finite n and s we can choose a small trimming, much smaller than $1/\log(ns/(n + s))$ for instance, as the asymptotic result is up to a factor. Then in our simulation study, with n and $s \in \{30, 50, 100, 200\}$, we chose arbitrarily $e_{n,s} = 1/(n + s)$ in order to attempt to stabilize slightly the test statistics. In the same way, it is clear that the speed of convergence does not give the exact value of $k(n, s)$ because it is up to a constant factor. In practice we chose a large value: $k(n, s) = 10$. Our choice is voluntarily raised, in order to try to detect high coefficients having difference.

Finally, the trimming was added essentially to stabilize the test statistic for small n and s, when taking into account the estimator of the variance W_k in (3.1). But we can note that some authors as Munk (2011), or Doukhan et al. (2015), in another univariate context, showed that such data driven smooth test statistics can be used without the normalization by $\widehat{W_k}^{-1}$. However, in this case the empirical levels are unstable for small values of the sample size.

PROPOSITION 4.1. *Let Assumptions* (A) *and* (B) *hold. Then, under* H_0, $S_{n,s}$ *converges to 1 in probability.*

PROOF. By definition of $S_{n,s}$, using the positivity of T_1, we have

$$\mathbb{P}(S_{n,s} \geq 2) = \sum_{k=2}^{k(n,s)} \mathbb{P}(S_{n,s} = k) \leq \sum_{k=2}^{k(n,s)} \mathbb{P}\left(T_k^{1/2} \geq \sqrt{(k-1)\log ns/(n + s)}\right).$$

By Markov inequality, we obtain

$$\mathbb{P}\Big(T_k^{1/2} \geq \sqrt{(k-1)\log(ns/(n+s))}\Big) \leq \frac{\Big(ns/(n+s)\,\mathbb{E}\big(\|U_k\|^2\big)\Big)^{1/2}}{\sqrt{(k-1)e_{n,s}\log(ns/(n+s))}}.$$

Using the independence of (X_{1i}, X_{2i}) and (X_{1j}, X_{2j}) for $i \neq j$ we get

$$\begin{aligned}\mathbb{E}\Big(U_k\|^2\Big) &= \sum_{j=1}^{k}\mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n}P_{1,j}(X_{1i}) - \frac{1}{s}\sum_{i=1}^{s}P_{2,j}(X_{2i})\Big)^2 \\ &\leq \sum_{j=1}^{k}\Big(\frac{\alpha_n}{n} + \frac{\beta_s}{s}\Big).\end{aligned}$$

It follows that

$$\mathbb{P}(S_{n,s} \geq 2) \leq \sqrt{2}\frac{(\alpha_n + \beta_s)^{1/2}k(n,s)}{\sqrt{e_{n,s}\log(ns/(n+s))}},$$

which tends to zero.

For the particular case $k = 1$, it follows from the Central Limit Theorem that under H_0, T_1 tends to a Chi-square distribution with degree one. We then deduce the following result.

COROLLARY 4.1. *Let Assumptions (A) and (B) hold. Then, under H_0, $T(n,s)$ converges in distribution to a Chi-squared random variable with one degree of freedom.*

5 Improving the Test Statistic

The choice of the Schwartz criterion was influenced by the pioneer work of Ledwina (1994). But other penalized statistics with smaller penalties, such as Akaike (see (Akaike, 1974)) or modified forms of BIC, can be proposed.

Following Inglot and Ledwina (2006), the data driven test statistic should be either $T(n,s) = T_{S_{n,s}}$ or

$$T'(n,s) = T_{A_{n,s}}, \tag{5.1}$$

where

$$A_{n,s} = \min\Big\{\operatorname*{argmax}_{1\leq k\leq k(n,s)}(T_k - 2k)\Big\},$$

is based on the Akaike criterion. Hence, for relatively large n and s, the penalty $2k$ will be less than the penalty $k\log(ns/(n+s))$ and will permit to detect alternatives with large difference in higher order Fourier coefficients. But this smaller penalty may induce a bias in the test level. Then the idea is to use $T_{A_{n,s}}$ only when the alternative is very distant from H_0. Otherwise Inglot and Ledwina (2006) propose to use $T_{S_{n,s}}$. The decision is based on the threshold rule, for a fixed $c > 0$:

$$I_{n,s}(c) \quad = \quad \mathbb{I}_{\max_{1\le j\le k(n,s)}\left\{\left|\sqrt{ns/(n+s)}u_{kj}\right|\le\sqrt{c\log(ns/(n+s))}\right\}},$$

where $\mathbb{I}$ stands for the indicator function, and u_{kj} stands for the jth component of $U_k\widehat{W}_k^{-1/2}$ which is asymptotically a standard normal vector. The rule is based on a classical result for a sequence of independent and identically distributed $\mathcal{N}(0,1)$ variables $Z_1,\cdots,Z_n$ for which $\mathbb{P}(\max|Z_i|\ge\sqrt{2\log(n)})\to 0$, as $n\to\infty$. Writing $t(j,n,s)=j\log(ns/(n+s))\ I_{n,s}(c)+2j\ (1-I_{n,s}(c))$ (Inglot and Ledwina, 2006) considered the statistic

$$\tilde{S}_{n,s} = \min\{1\le k\le k(n,s) : T_k - t(k,n,s)\ge T_j - t(j,n,s); j=1,\cdots,k(n,s)\},$$

and its associated test statistic

$$T''(n,s) \quad = \quad T_{\tilde{S}_{n,s}}. \tag{5.2}$$

They suggested a value of $c=2.4$. We will compare empirically $T(n,s)$ and $T''(n,s)$ in our numerical studies, with this constant $c=2.4$.

6 Adaptation to the Paired Case

We consider the iid paired sample $((X_{11},X_{21}),\cdots,(X_{1n},X_{2n}))$, where X_{1i} and X_{2i} are dependent, and we write

$$\begin{aligned} Z_j(X_1,X_2) &= P_{1,j}(X_1,m)-P_{2,j}(X_2,m),\\ V_k(X_1,X_2) &= (Z_1(X_1,X_2),\cdots,Z_k(X_1,X_2)). \end{aligned}$$

Then H_0 coincides with $\mathbb{E}(Z_j(X_1,X_2))=0$, for all $j=1,2,\cdots$. The test statistic is given by

$$T_k = nU_k\widehat{W}_k^{-1}U_k^T,$$

with

$$U_k = \frac{1}{n}\sum_{i=1}^{n} V_k(X_{1,i},X_{2,i}) \quad \text{and} \quad \widehat{W}_k = diag(\widehat{v}_1,\cdots,\widehat{v}_k),$$

where $\widehat{v}_i$ are convergent estimators of $\mathbb{V}(Z_i(X_1, X_2))$, that we will simply denote by $\mathbb{V}(Z_i)$, with an additional trimming $e_n > 0$ such that $e_n \to 0$. Write

$$S_n = \min\left\{ \underset{1\le s\le k(n)}{\operatorname{argmax}} (T_s - s\log(n)) \right\}.$$

The proposed data driven test statistic is given by

$$T(n) = T_{S_n} = nU_{S_n}\widehat{W}_{S_n}^{-1}U_{S_n}^T.$$

We will need the following assumption

(C) $k(n) = o_{\mathbb{P}_0}(\sqrt{e_n\log(n)/M_n})$, where $M_n = \max_{i=1,\cdots,n}\mathbb{V}(Z_i)$.

Remark 6.1. Condition (C) can be satisfied as soon as all variances $\mathbb{V}(Z_k)$ exist and if we use a trimming such that $1/e_n = o(\log(n))$.

Proposition 4.1 can be adapted as follows.

Proposition 6.1. *Let Assumptions (C) holds. Then, under H_0, S_n converges to 1 in probability and T_{S_n} is asymptotically Chi-squared distributed with one degree of freedom.*

7 Simulations

7.1. Calibrating the Test Statistics. In our simulation study we will consider the case where $n = s$ and for simplicity of notation we write $T(n)$ instead of $T(n,n)$. First simulations showed that for small sample sizes the empirical levels are slightly higher than the theoretical one (we fixed $\alpha = 5$ %). Since $T(n) \ge T_1$, it is clear that under H_0, $\mathbb{P}(T(n) \le x)$ is overestimated by its asymptotic approximation and using the α-upper percentile of the limiting distribution leads to underestimate the actual level and power of the test. Then we will use a sharper approximation for $\mathbb{P}(T(n) \le x)$, as proposed in Kallenberg and Ledwina (1995). Denoting by G_1 and G_2 two standard independent Gaussian variables, we have under H_0

$$\mathbb{P}(T(n) \le x) \simeq \mathbb{P}(G_1^2 \le x) - \mathbb{P}(G_1^2 \le x, G_1^2 + G_2^2 \ge x, G_2^2 \ge \log n).$$

which can be straightforwardly reformulated as

$$\mathbb{P}(T(n) \le x) \simeq \begin{cases} \left(2\Phi(\sqrt{x})-1\right)\left(2\Phi(\sqrt{\log n}) - 1\right) & \text{if x} \le \log \text{n} \\ \left(2\Phi(\sqrt{x})-1\right)\left(2\Phi(\sqrt{\log n}) - 1\right) + 2\Phi(-\sqrt{\log n}) & \text{if x} \ge 2\log \text{n} \\ \text{linearize} & \text{else,} \end{cases}$$

where Φ is the cumulative distribution function of a standard Gaussian variable.

7.2. Models and Alternatives. We considered six models, denoted by **M1, ..., M6**, and seven alternatives, denoted **A1,A2 ..., A7**, for simulation. They are summarized in Table 1.

Notations $\mathcal{N}(m, \sigma^2)$ denotes the normal distribution with mean m and variance σ^2, χ^2_d denotes the Chi-squared distribution with d degrees of freedom, $Exp(\theta)$ stands for the exponential distribution with mean $1/\theta$, $\gamma(a, b)$ is the gamma distribution with shape parameter a and scale parameter b, $\mathcal{P}(m)$ is the Poisson distribution with mean m, $\mathcal{U}(a, b)$ is the continuous uniform distribution on the interval (a, b).

Model **M1** is a normal mixture of normal distributions. The family F_1 is the NEF of normal distributions with variance 2, and the family F_2 is the NEF of normal distributions with variance 1. Alternatives **A1** and **A2** coincide with location and scale deviations.

Model **M2** is an exponential mixture of normal distributions. Its alternative **A3** consists in a change of the exponential parameter.

Model **M3** is an chi-squared mixture of normal distributions. Its alternative **A4** consists in a change of the chi-squared freedom.

Model **M4** is a compound Poisson model. Alternative **A5** is a change of the Poisson parameter.

Model **M5** is composed of a uniform mixture of normal distributions and a uniform mixture of Poisson distributions. The family F_1 is the NEF of normal distributions with variance 2 and the family F_2 is the NEF of

Table 1: Models under the null hypothesis and alternatives

Model	μ_1	μ_2	π_1	π_2
M1	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$
A1	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,2)$
A2	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$\mathcal{N}(0,1)$	$\mathcal{N}(1,1)$
M2	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$Exp(1)$	$Exp(1)$
A3	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$Exp(1)$	$Exp(1/2)$
M3	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	χ^2_1	χ^2_1
A4	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	χ^2_1	χ^2_2
M4	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$\mathcal{P}(1)$	$\mathcal{P}(1)$
A5	$\mathcal{N}(0,2)$	$\mathcal{N}(0,1)$	$\mathcal{P}(1)$	$\mathcal{P}(2)$
M5	$\mathcal{N}(1,2)$	$\mathcal{P}(1)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$
A6	$\mathcal{N}(1,2)$	$\mathcal{P}(1)$	$\mathcal{U}(0,2)$	χ^2_1
M6	$\mathcal{P}(1)$	$\mathcal{P}(1)$	$\mathcal{U}(0,2)$	$\mathcal{U}(0,2)$
A7	$\mathcal{P}(1)$	$\mathcal{P}(1)$	$\mathcal{U}(0,2)$	χ^2_1

Poisson distributions. Alternative **A6** is a change of mixing distribution from uniform to chi-squared.

For model **M6**, under H_0, X_1 and $X2$ have the same distribution. Then we can compare our test with a two-sample test as the Neyman smooth test proposed in Ghattas et al. (2011) (see also Janic-Wróblewska and Ledwina, 2000; Wylupek, 2010).

7.3. Empirical Levels. For each $n = s \in \{30, 50, 100, 200\}$, we investigate the empirical levels of the tests based on $T(n)$ and $T''(n) = T''(n,n)$ defined by (5.2). We also show empirical levels obtained with $T'(n) = T'(n,n)$ given by (5.1) to better understand the difference between $T(n)$ and $T''(n)$. Empirical levels are obtained as the percentage of rejection of the null hypothesis over 1 000 replications of the test statistic under H_0. The orthogonal polynomials are the Hermite ones for model M1-M4, and the Charlier ones for model M5, and both Hermite and Charlier for model M6. These polynomials are obtained by standard transformations of those given in Table 6. For instance if $\mathcal{H}_n(x)$ are Hermite polynomials $\mathcal{N}(0,1)$-orthogonal, then $\mathcal{H}_n((x-m)/\sigma)$ are $\mathcal{N}(m,\sigma^2)$-orthogonal. The trimming was fixed at $e_{n,s} = 1/(n+s)$ and the maximal number of components was fixed equal to $k(n,s) = 10$, for all sample sizes. The empirical levels are displayed in Table 2. For $T'(n)$ it can be seen that empirical levels are generally greater that the nominal 5 %. The empirical levels of $T(n)$ seem to be the most close to 5 %. Naturally, it appears that $T''(n)$ offers a compromise between $T(n)$ and $T'(n)$, with empirical levels higher than 5 % for some cases, especially for **M1** and **M4**. It is worth noting that for $n > 200$, empirical levels associated to $T(n)$ and $T''(n)$ are both close to 5 %. For such sample sizes it seems to be interesting to use $T''(n)$ which will provide a better power. In Table 2 we indicated the empirical level for $n = 30$ without trimming, that is, when $e_{n,s} = 0$. Its increased relatively slightly the levels.

7.4. Empirical powers. For each $n \in \{30, 50, 100, 200\}$, we investigate the empirical levels of the test based on $T(n)$ and $T''(n)$. They are obtained as the percentage of rejection of the null hypothesis over 1 000 replications of the test statistic under alternatives. Figures 1–2 show the empirical powers obtained with $T(n)$ under all alternatives. Figure 1 shows the alternatives largely detected by the test. Figure 2 shows the alternatives with smaller empirical powers. The standard smooth two-sample test of Neyman (see for instance (Ghattas et al., 2011)) is also used for alternative A7. As expected, it gave similar results than the proposed test based on $T(n)$.

Figures 3–4 show the empirical powers obtained with $T''(n)$ under all alternatives. As expected all empirical powers are greater than those obtained with the initial statistic $T(n)$. The average gain is around 10 %, sometimes

Table 2: Empirical levels (in %) for Models **M1-M6** with sample sizes $n = s \in \{30, 50, 100, 200\}$, with trimming $e_{n,n} = 1/(2n)$ (except for the first column where $e_{n,n} = 0$) and $k(n,n) = 10$

Model	Statistic	$n = 30$ (with $e_{n,s} = 0$)	$n = 30$	$n = 50$	$n = 100$	$n = 200$
M1	$T(n)$	9.2	8.5	6.3	5.5	5.5
	$T'(n)$	17.1	13.1	15	6.2	15.1
	$T''(n)$	10.4	9.2	6.9	5.9	5.9
M2	$T(n)$	7.0	6.9	5.9	5	4.9
	$T'(n)$	9.4	7.2	5.4	10.1	10.2
	$T''(n)$	8.9	6.9	5.8	5.6	5.4
M3	$T(n)$	8.1	8.1	6.1	4.9	4.8
	$T'(n)$	13.1	9.1	13.2	9.1	9.0
	$T''(n)$	11.3	8.3	6.4	5.9	5.3
M4	$T(n)$	9.0	8.9	7.9	5.9	4.9
	$T'(n)$	17.9	17.2	15.4	11.1	16.0
	$T''(n)$	11.4	11.1	8.8	8.5	5.8
M5	$T(n)$	8.4	8.2	5.4	6.2	5.4
	$T'(n)$	5.3	5.3	5.2	4.8	9.1
	$T''(n)$	8.3	8.0	5	6.0	5.7
M6	$T(n)$	5.5	5.6	4.7	4.4	5.4
	$T'(n)$	6.8	6.3	5.4	4.7	9.0
	$T''(n)$	6.6	5.9	4.8	4.5	5.7

more. In conclusion we could suggest the use of $T(n)$ for moderate sample size ($n < 200$) and the use of $T''(n)$ when the empirical level should be stabilized, that is for $n \geq 200$.

8 Real Data Example

8.1. Framingham data. The Framingham data is taken from a study on coronary heart disease described by Carroll et al. (2010). The data consist of measurements of systolic blood pressure (SBP) obtained at two different dates. The study concerned $n = 1615$ males on an 8-year follow-up. At both dates the SBP was measured twice for each individual and we denote by $X_1(1), X_1(2)$ the two measurements at the first date, and by $X_2(1), X_2(2)$ the two measurements at the second date. It is known (see for instance (Wang and Wang, 2011) for a recent study) that these data can be considered as a convolution and we propose the model

$$X_i(j) \quad = \quad SBP_i + V_{ij}, \quad i, j = 1, 2$$

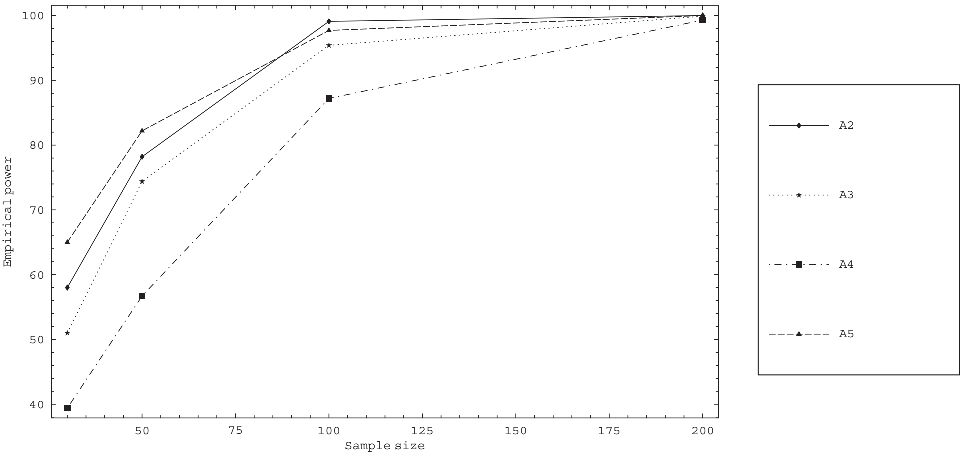

Figure 1: Empirical powers associated to the statistic $T(n)$ for alternatives A2 (♦), A3 (★), A4 (■) and A5 (▲) with sample sizes $n = 30, 50, 100, 200$

where V_{ij} are independent centered normal errors with variance σ_i^2, and SBP_i stands for the SBP of the individual at examination i. This is clearly a convolution model and we are interested in testing the equality of the systolic distributions π_i of SBP_i at time $i = 1, 2$, that is $H_0 : \pi_1 = \pi_2$.

We can use the difference between $X_i(1)$ and $X_i(2)$ to evaluate the variance of the errors σ_i^2. Indeed, $\mathbb{V}(X_i(1) - X_i(2)) = \mathbb{V}(V_{i1} - V_{i2}) = 2\sigma_i^2$.

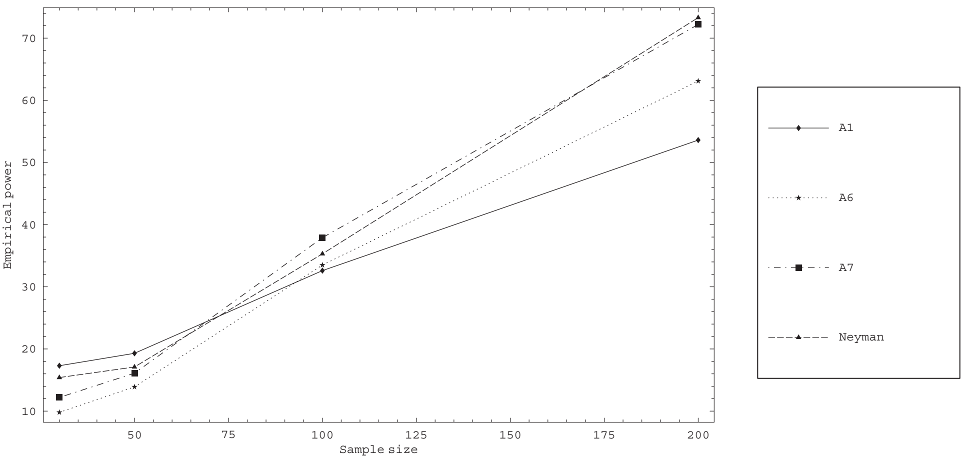

Figure 2: Empirical powers associated to the statistic $T(n)$ for alternatives A1 (♦), A6 (★), A7 (■) and A7 with Neyman test (▲) with sample sizes $n = 30, 50, 100, 200$

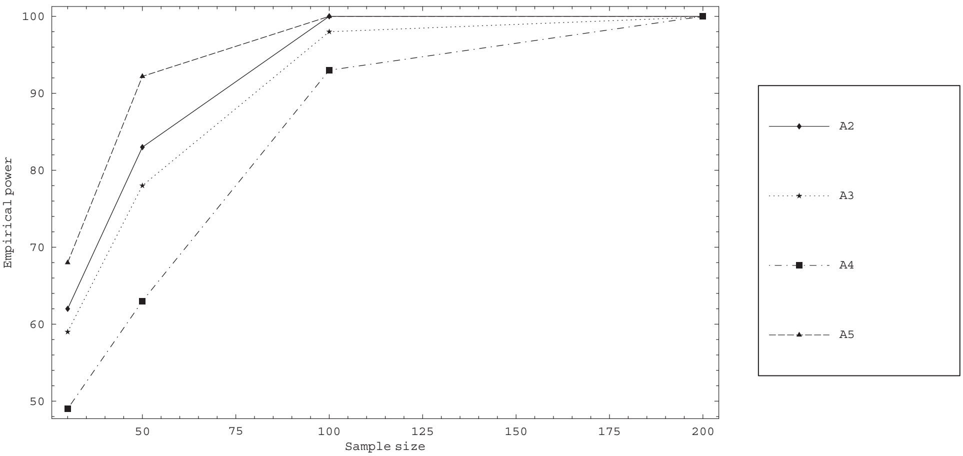

Figure 3: Empirical powers associated to the statistic $T''(n)$ for alternatives A2 (♦), A3 (★), A4 (■) and A5 (▲) with sample sizes $n = 30, 50, 100, 200$

Based on the 1615 observations we get $\sigma_1^2 = 55.0$ and $\sigma_2^2 = 58.5$. To construct our test statistic we consider the two samples: $X_{11}, \cdots, X_{1n}$ and $X_{21}, \cdots, X_{2n}$, from $X_i = (X_i(1) + X_i(2))/2$, $i = 1, 2$, with $n = 1615$. Clearly $X_i = SBP_i + \tilde{V}_i$ with $Var(\tilde{V}_i) = \sigma_i^2/2$. The densities of X_1 and X_2 have the form (1.1), with $f_1(.|m)$ and $f_2(.|m)$ the normal densities with variances $\sigma_1^2/2 = 27.5$ and $\sigma_2^2/2 = 29.25$, respectively. To apply our the test procedure we used Hermite polynomials, orthogonal w.r.t. μ the standard normal distribution, and we fixed $k(n) = 10$ and $e(n) = 1/(2n)$. We

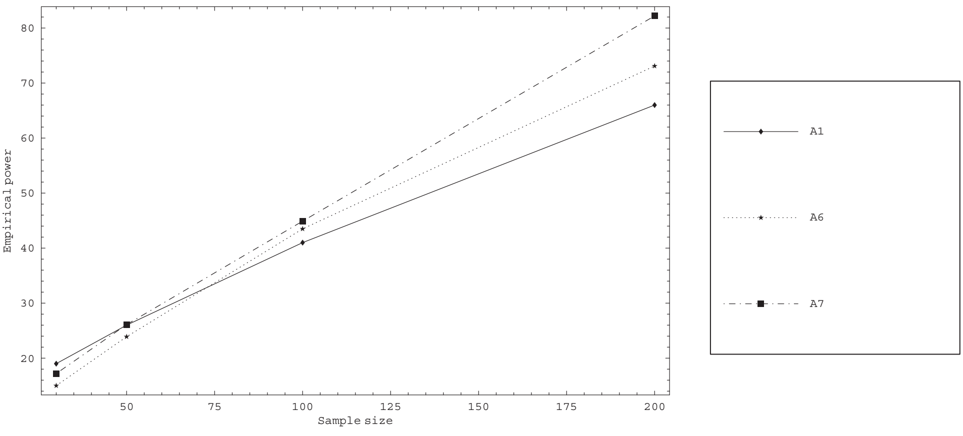

Figure 4: Empirical powers associated to the statistic $T''(n)$ for alternatives A1 (♦), A6 (★), A7 (■) with sample sizes $n = 30, 50, 100, 200$

obtained $T''(n) = T(n) = T_1$ with $S_n = 1$ and an approximated p-value equal to 0.034. Then the systolic blood pressures appear to have different distributions π_1 and π_2 at the two different dates. The data driven statistics S_n detects the degree of the most significant difference between the moments of π_1 and π_2. Here $S_n = 1$ indicating that the most significant difference between the systolic blood pressures is due to their first moments. Then the means of π_1 and π_2 differ significatively.

8.2. Claim amounts. Thirty six aggregated claim amounts were recorded over two periods: 1999-2001 and 2002-2004, by a branch of a French insurance company. The model assumed that claim amounts over 1999-2001 was gamma distributed with mean 29.91 (hundred euros) and variance 38.30, that is a $\Gamma(a_1 = 23.35, b_1 = 0.78)$. The claim amounts over 2002-2004 was gamma distributed with mean 31.11 (hundred euros) and variance 42.40, that is a $\Gamma(a_2 = 22.82, b_2 = 0.73)$. The observed aggregated claims over the two periods have compound distributions as in (1.2), with unknown distributions π_1 and π_2. Writing N_1 and N_2 the number of claims, we are interested in comparing their distributions; that is to test $H_0 : \pi_1 = \pi_2$. To apply our the test procedure we used two Laguerre polynomials families, orthogonal w.r.t. $\mu_1 \sim \Gamma(1/b_1, b_1)$ and $\mu_2 \sim \Gamma(1/b_2, b_2)$, respectively. We fixed $k(n) = 10$ and $e(n) = 1/(2n)$. Using our test procedure we get $T''(n) = T(n) = T_1$ and we obtained a p-value equal to 0.11. Hence there is no evidence that the two claims numbers distributions differ.

To illustrate the power of the test for such a model we simulated several Poisson compound models with the previous gamma margins $\Gamma(a_1, b_1)$ and $\Gamma(a_2, b_2)$. Although the distributions π_1 and π_2 are not necessarily Poisson, this gives an idea of the test behavior. We chose $\pi_1 \sim \mathcal{P}(m_1)$ and $\pi_2 \sim \mathcal{P}(m_2)$, and we fixed $m_1 = 500$, close to the observed mean in the data set. Table 3 gives the empirical level and empirical powers of the test with respect to the values of m_2 for $n = 36$, the sample size of the previous data set. For $m_1 = m_2 = 500$ it coincides with the empirical level of the test.

Table 3: Empirical level (in %) and empirical powers (in %) for a Poisson-gamma compound model associated with the gamma parameters observed in the previous data set in insurance, with $\pi_1 \sim \mathcal{P}(500)$ and $\pi_2 \sim \mathcal{P}(m_2)$

m_2	470	480	490	495	498	500
Empirical power or level	100.0	97.8	45.0	13.8	7.4	5.6

Annex

Table 4: Relations for cumulant functions: μ = generating measure, k_μ = cumulant, ψ_μ = inverse of k_μ, f_μ = density w.r.t. μ, $F(\mu)$ = NEF generated by μ

NEF $F(\mu)$	$\mu(dx)$	$k_\mu(\theta)$	$\psi_\mu(m)$	$f_\mu(x,m)$
Normal	$\frac{\exp(-x^2/(2\sigma^2))}{\sqrt{2\pi}\sigma}$	$\theta^2\sigma^2/2$	m/σ^2	$\exp\{(2xm - m^2)/(2\sigma^2)\}$
Poisson	$\sum \delta_n(dx)/n!$	$\exp(\theta)$	$\log(m)$	$\exp\{\log(m)x - m\}$
Gamma	$x^{\lambda-1}/\Gamma(\lambda)(dx)$	$-\lambda\log(-\theta)$	$-\lambda/m$	$\exp\{-\lambda x/m$ $+\lambda\log(\lambda/m)\}$
Negative binomial	$\sum \begin{pmatrix} \lambda \\ n \end{pmatrix} \delta_n(dx)$	$-\lambda\log(1$ $-\exp(\theta))$	$\log(\frac{m}{\lambda+m})$	$\exp\{\log(m/(\lambda+m))x$ $+\lambda\log(1 - m/(\lambda+m))\}$

Table 5: Some quadratic NEF, denoted by F, their variance functions, denotes by V_F, and their orthogonal polynomials ($\mathcal{H}_n$=Hermite, $\mathcal{L}_n^\lambda$=Laguerre, $\mathcal{C}_n^m$=Charlier, with the notation of Abramowitz and Stegun, 1972)

NEF F	$V_F(m)$	Polynomials $Q_n(x,m)$	$\|Q_n(.,m)\|^2$
Normal	σ^2	$(2\sigma^2)^{-n/2}\mathcal{H}_n(x/\sqrt{2\sigma^2})$	$n!\sigma^{-2n}$ (here $m=0$)
Poisson	m	$\mathcal{C}_n^m(x)$	$n!m^{-n}$
Gamma	m^2/λ	$\mathcal{L}_n^\lambda(\lambda x/m)$	$m^{-2n}n!\Gamma(n+\lambda)/\Gamma(\lambda)$
Negative binomial	$m^2/\lambda + m$	$\mathcal{M}_n^{\lambda,(m)}(x)$	$\lambda^{-n}(m^2/\lambda+m)^{-n}n!$ $\Gamma(n+\lambda)/\Gamma(\lambda)$

Table 6: Some classical orthogonal polynomials

Name	Notation	First terms	Recurrence relations
Hermite	$\mathcal{H}_n$	$\mathcal{H}_0 = 1$ $\mathcal{H}_1(x) = 2x$	$2x\mathcal{H}_n(x) = \mathcal{H}_{n+1}(x) + 2n\mathcal{H}_{n-1}(x)$
Charlier	$\mathcal{C}_n^\alpha$ $(\alpha > 0)$	$\mathcal{C}_0^\alpha = 1$ $\mathcal{C}_1^\alpha(x)$ $= (\alpha - x)/\alpha$	$x\mathcal{C}_n^\alpha(x) = -\alpha\mathcal{C}_{n+1}^\alpha(x) + (n+\alpha)\mathcal{C}_n^\alpha(x)$ $-n\mathcal{C}_{n-1}^\alpha(x)$
Laguerre	$\mathcal{L}_n^\alpha$ $(\alpha > -1)$	$\mathcal{L}_0^\alpha = 1$ $\mathcal{L}_1^\alpha(x)$ $= -x + \alpha + 1$	$-x\mathcal{L}_n^\alpha(x) = (n+1)\mathcal{L}_{n+1}^\alpha(x)$ $-(2n+\alpha+1)\mathcal{L}_n^\alpha(x)$ $+(n+\alpha)\mathcal{L}_{n-1}^\alpha(x)$
Meixner (first type)	$\mathcal{M}_n^{c,\beta}$ $(c \neq 1)$ $(\beta \in \mathbb{R})$	$\mathcal{M}_0^{c,\beta} = 1$ $\mathcal{M}_1^{c,\beta}(x)$ $= x - \beta c/(1-c)$	$x\mathcal{M}_n^{c,\beta}(x) = \mathcal{M}_{n+1}^{c,\beta}(x)$ $+\frac{(1+c)n+\beta c}{1-c}\mathcal{M}_n^{\beta,c}(x)$ $+\frac{cn(n+\beta-1)}{1-c}\mathcal{M}_{n-1}^{\beta,c}(x)$

References

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions.* Dover Publications.

AKAIKE, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 6, 716–723.

ALBERS, W., KALLENBERG, W. and MARTINI, F. (2001). Data driven rank tests for classes of tail alternatives. *J. Am. Stat. Assoc.* **96**, 685–696.

ASMUSSEN, S. and ALBRECHER, H. (2010) Ruin Probabilities. Advanced Series on Statistical Science Applied Probability. World Scientific.

BANGERT, M., HENNIG, P. and OELFKE, U. (2010) Using an infinite von mises-fisher mixture model to cluster treatment beam directions in external radiation therapy. In: *International Conference on Machine Learning and Applications.*

BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families in Statistical Theory.* Wiley, Chichester.

CARLEMAN, T. (1926). *Les fonctions quasi analytiques.* Collection de Monographies sur la Théorie des Fonctions, Gauthier-Villars, Paris.

CARROLL, R., RUPPERT, D., STEFANSKI, L. and CRAINICEANU, C. (2010). *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition.* Taylor & Francis.

CHEE, C.-S. and WANG, Y. (2013). Minimum quadratic distance density estimation using nonparametric mixtures. *Comput. Stat. Data Anal.* **57**, 1, 1–16.

D'AGOSTINO, R. B. and STEPHENS, M. A. (1986). *Goodness-Of-Fit Techniques.* Marcel Dekker, New York.

DENUIT, M. (1997). A new distribution of poisson-type for the number of claims. *ASTIN Bulletin* **27**, 229–242.

DOUKHAN, P., POMMERET, D. and REBOUL, L. (2015). Data driven smooth test of comparison for dependent sequences. *J. Multivar. Anal.* **139**, 147–165.

DUBEY, A., HWANG, S., RANGEL, C., RASMUSSEN, C., GHAHRAMANI, Z. and WILD, D. L. (2004). Clustering protein sequence and structure space with infinite gaussian mixture models. *World Scientific*, 399–410.

FAN, J. (1996). Test of significance based on wavelet thresholding and Neyman's truncation. *J. Am. Stat. Assoc.* **91**, 674–688.

GHATTAS, B., POMMERET, D., REBOUL, L. and YAO, A. (2011). Data driven smooth test for paired populations. *Journal of Statistical Planning and Inference* **141**, 1, 262–275.

GHOSH, A. (2001). *Testing equality of two densities using Neyman's smooth test.* Tech. rep., Department of Economics, University of Illinois.

GRIFFIN, J. E. (2010). Default priors for density estimation with mixture models. *Bayesian Analysis* **5**, 1, 45–64.

GROENEBOOM, P. (1984) *Estimating a Monotone Density.* Report MS / Centrum voor Wiskunde en Informatica, Department of Mathematical Statistics. Stichting Math. Centrum.

HART, J. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer, New York.

INGLOT, T., JURLEWICZ, T. and LEDWINA, T. (1990). On neyman-type smooth tests of fit. *Statistics* **21**, 549–568.

INGLOT, T., KALLENBERG, W. and LEDWINA, T. (1997). Data-driven smooth tests for composite hypotheses. *Ann. Stat.* **25**, 1222–1250.

INGLOT, T. and LEDWINA, T. (2006). Towards data driven selection of a penalty function for data driven neyman test. *Linear Algebra Appl.* **417**, 124–133.

JANIC-WRÓBLEWSKA, J. and LEDWINA, T. (2000). Data driven rank test for two-sample problem. *Scand. J. Stat.* **27**, 281–297.

KALLENBERG, W. and LEDWINA, T. (1995). On data-driven Neyman's tests. *Probab. Math. Stat.* **15**, 409–426.

KALLENBERG, W. and LEDWINA, T. (1995a). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *Ann. Stat.* **23**, 1594–1608.

LEDWINA, T. (1994). Data-driven version of Neyman's smooth test of fit. *J. Am. Stat. Assoc.* **89**, 1000–1005.

LINDSAY, B. G. (1989). Moment matrices: Applications in mixtures. *Ann. Stat.* **17**, 2, 722–740.

MORRIS, C. (1982). Natural exponential families with quadratic variance functions. *Ann. Stat.* **10**, 65–80.

MUNK, AXEL S.J.-P.V.J.G.G. (2011). Neyman smooth goodness-of-fit tests for the marginal distribution of dependent data. *Ann. Inst. Statist. Math.* **63**, 5, 939–959.

NEYMAN, J. (1937). Smooth test for goodness of fit. *Skandinavisk Aktuarietidskrift* **20**, 149–199.

POMMERET, D. (2005). Distance of a mixture from its parent distribution. *Sankhyā* **67**, 4, 699–714.

RAYNER, J. and BEST, D. (1989). *Smooth Tests of Goodness of Fit.* Oxford University Press, New York.

RAYNER, J. and BEST, D. (2001). *A Contingency Table Approach to Nonparametric Testing.* Chapman and Hall/CRC.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Ann. Stat.* **6.2**, 461–464.

VAN and DE GEER, S. (2003). Asymptotic theory for maximum likelihood in nonparametric mixture models. *Comput. Stat. Data Anal.* **41**, 453–464.

WANG, J. (2007). A linearization procedure and a vdm/ecm algorithm for penalized and constrained nonparametric maximum likelihood estimation for mixture models. *Comput. Stat. Data Anal.* **51**, 2946–2957.

WANG, X.-F. and WANG, B. (2011). Deconvolution estimation in measurement error models: The r package decon. *J. Stat. Softw.* **39**, 10, 1–24.

WYLUPEK, G. (2010). Data driven k sample test. *Technometrics* **52**, 107–123.

DENYS POMMERET
INSTITUT DE MATHÉMATIQUES DE
MARSEILLE
AIX MARSEILLE UNIVERSITÉ
13288 MARSEILLE CEDEX 9, FRANCE
E-mail: denys.pommeret@univ-amu.fr

Paper received: 16 February 2014.