Curves comparison using wavelet

Jean-François Angers*

CRM-2833

February 2002

*Département de mathématiques et de statistique; Université de Montréal; C.P. 6128, Succ. "Centre-ville"; Montréal, Québec H3C 3J7. jean-francois.angers@umontreal.ca.

Abstract

This paper discusses the problem of detecting the difference between two functions. The first step is to estimate these functions based on two independent samples. No restriction is made on the design, *i.e.* the sample size and the covariate values can be different. However, the covariate should be the same for both functions. Using a empirical Bayesian nonparametric regression set up, the functions are represented using wavelet decomposition and a prior density is put on the wavelet coefficients. Three tests are proposed and compared with those discussed in the literature. A simulation study investigate the behavior of these tests for moderate sample sizes and a method to compute the critical values is proposed.

Résumé

Dans cet article, nous cherchons à déterminer si deux fonctions sont égales ou non. Dans un premier temps, nous devons estimer ces fonctions en se basant sur deux échantillons indépendants. Il n'y a aucune restriction sur le design, la taille des échantillons et les fonctions peuvent être observées à des temps différents. Les fonctions sont représentées par une série d'ondelettes et le problème est réécrit sous la forme d'un modèle linéaire. En utilisant une approche bayésienne empirique, nous estimons les coefficients de la série d'ondelettes pour chacune des fonctions. Par la suite, nous proposons trois tests pour déterminer si les deux fonctions sont égales ou non. Une étude de simulation a été faite afin de les comparer à d'autres tests proposés dans la littérature.

1 Introduction

A common problem in experimental work is the comparison of two or more regression functions. For exemple, an experimenter may want to compare the measured responses of a treatment group of subjects with those of a control group. If in addition, a covariate is measured, the experimenter may choose to regress the response for each group and compare the two regression curves. The classical methods for comparison of regression curves for two or more populations use parametric models for the regression functions and compare the resulting model parameters. A disadvantage of this approach is that it requires the specification of a parametric model, which is often difficult.

Nonparametric methods have been proposed. However, for most of these methods, the critical values have to be computed using simulation (*cf.* Hall and Hart, 1990; Kulasekera, 1995) or some asymptotic calculation (*cf.* Koul and Schick, 1997). Another approach is to transform the problem using Fourier or wavelet transforms (*cf.* Fan and Lin, 1998). Often in order to apply this alternative approach, the data have to be sampled at equispaced design points (or some binning or interpolation methods has to be used to preprocess the data). Finally, Fan and Lin (1998) proposed a smoothed principal component analysis which involves choosing smoothing parameters and studying the effect of the projected data on the estimated principal axes.

In this paper, several procedures to test the equality of two regression functions are proposed. Only smoothness conditions on the regression functions are needed. Consequently, the proposed tests can be used for a wide class of functions.

The considered setting is the following. We observe data in the form $\{(Y_{1i}, x_{1i}), i = 1, 2, \ldots, n_1\}$ and $\{(Y_{2i}, x_{2i}), i = 1, 2, \ldots, n_2\}$ with

$$Y_{1i} = f_1(x_{1i}) + \varepsilon_{1i}, \qquad \begin{array}{l} i = 1, 2, \ldots, n_1, \\ x_{1i} \in \mathcal{T}, \end{array} \tag{1}$$

and

$$Y_{2i} = f_2(x_{2i}) + \varepsilon_{2i}, \qquad \begin{array}{l} i = 1, 2, \ldots, n_2, \\ x_{2i} \in \mathcal{T}, \end{array} \tag{2}$$

where $\mathcal{T}$ is a bounded subset of $\mathbb{R}$, $\epsilon_{\ell,i} \sim N(0, \sigma_\ell^2)$, $i = 1, 2, \ldots, n_\ell$, $\ell = 1, 2$ are independent random errors for both groups. We are interested in testing the hypothesis

$$H_0 : f_1 = f_2 \text{ versus } H_1 : f_1 \neq f_2 \tag{3}$$

over the domain $\mathcal{T}$.

The problem of hypothesis testing for curves has been considered in time series literature (*cf.* Shumway, 1988, and references therein). Brillinger (1973, 1980) developed various techniques for analyzing functional data. The literature on nonparametric curve testing includes Hall and Hart (1990); Kulasekera (1995); Koul and Schick (1997). More references can be found in Fan and Lin (1998).

This paper is divided along the following sections. In Section 2, the nonparametric regression using wavelets, as introduced in Angers and Delampady (2001), is presented. However, in order to speed up and simplify the calculation in the simulation, an empirical Bayesian version of this nonparametric regression estimator is proposed instead of the full Bayesian approach. In Section 3, three tests are proposed. The first one is a straightforward Bayesian hypothesis test using Bayes factor while the two others can be viewed as alternative procedures. The second test is based on the posterior distribution of the $\mathcal{L}_2(\mathbb{R})$ distance between both regression functions. This posterior

distribution is approximated using Edgeworth expansion under both the null and the alternative hypothesis. The third test is based on the bootstrap of the $\mathcal{L}_2(\mathbb{R})$ distance of the nonparametric regression functions under the null hypothesis.

2 Bayesian wavelet estimator of a function

In this section, a brief review of the nonparametric regression technique proposed in Angers and Delampady (2001) is presented. Here, the subscript $\ell = 1, 2$ referring to the set of observations is omitted in order to ease the notation.

2.1 Description of the model

First, we decompose the regression functions given in equations (1) and (2) into a linear combination of a wavelet basis. We use a compactly supported wavelet function $\psi \in \mathcal{C}^s$, the set of real-valued functions with continous derivatives up to order s. Note that any function f in $\mathcal{L}_2(\mathbb{R})$ has the wavelet decomposition

$$f(x) = \sum_{|k|\leq K_0} \alpha_k \phi_k(x) + \sum_{j\geq 0} \sum_{|k|\leq K_j} \beta_{jk} \psi_{j,k}(x), \tag{4}$$

with

$$\begin{aligned} \phi_k(x) &= \phi(x-k), \text{ and} \\ \psi_{j,k}(x) &= 2^{j/2}\psi(2^j x - k), \end{aligned}$$

where ϕ is the scaling function ('father wavelet') corresponding to the 'mother wavelet' ψ. Since we suppose that ϕ and ψ have compact support, there exist a K_j such that $\phi_k(x)$ and $\psi_{j,k}(x)$ vanish $\forall x \in \mathcal{T}$ whenever $|k| > K_j$, For any specified resolution level J, equation (4) can be written as

$$\begin{aligned} f(x) &= \sum_{|k|\leq K_0} \alpha_k \phi_k(x) + \sum_{j=0}^{J} \sum_{|k|\leq K_j} \beta_{jk} \psi_{j,k}(x) + \sum_{j=J+1}^{\infty} \sum_{|k|\leq K_j} \beta_{jk} \psi_{j,k}(x) \\ &= g_J(x) + R_J(x), \end{aligned} \tag{5}$$

where

$$\begin{aligned} g_J(x) &= \sum_{|k|\leq K_0} \alpha_k \phi_k(x) + \sum_{j\geq 0}^{J} \sum_{|k|\leq K_j} \beta_{jk} \psi_{j,k}(x), \text{ and} \\ R_J(x) &= \sum_{j=J+1}^{\infty} \sum_{|k|\leq K_j} \beta_{jk} \psi_{j,k}(x). \end{aligned} \tag{6}$$

Delyon and Juditsky (1995) consider a similar model but adopt an approach for analysis which is very different from ours. Some other good references for general wavelet related materials and applications are Daubechies (1992); Härdle et al. (1998); Mallat (1998); Ogden (1997); Vidakovic (1999).

2.2 Prior model

The above wavelet decomposition (*cf.* equation (5)) is exploited to give the prior model for f_ℓ in (1) and (2). The strength of wavelet basis over others is that wavelets can identify local features much more effectively and efficiently. Note in the representation (5) that there are ϕ_k functions to detect the global features, and once this is done, there are the $\psi_{j,k}$ functions to check for local details. Such a representation is especially useful when we do not assume that the regression function has any global smoothness features, and hence one may not find "spline type models" very appropriate. Details on related Bayesian approaches to wavelet based function estimation can be found in Vidakovic (1998); Müller and Vidakovic (1999).

Note that at the resolution level J, equations (1) and (2) can both be rewritten as

$$y_i \;=\; g_J(x_i) + \eta_i + \varepsilon_i,$$

where $\eta_i = R_J(x_i)$. The resolution level should satisfy the following equation

$$l_X 2^{J+1} + J(l_\psi + 1) + (l_\phi + l_\psi + 2) \leq \min\{n_1, n_2\}$$

where l_X denotes the length of $\mathcal{T}$, l_ϕ and l_ψ, the length of the support of ϕ and ψ respectively (detailed can be found in Angers and Delampady (2001)). Failure to do so can make the design matrix (see next subsection) singular and hence the estimated values of α_k and β_{jk} may be strongly dependent on the choice of the prior.

To provide a joint prior distribution on α_k and β_{jk} we assume that they are all independent normal random variables with mean 0. The prior variance of α_k is assumed to be τ^2, whereas to accommodate the decreasing effect of the "detail" coefficients β_{jk}, we assume that their variance is $\tau^2/2^{2js}$, (*cf.* Abromovich and Sapatinas (1999) for further justification for the choice of the first stage prior variances). Since σ^2 and τ^2 are unknown and we use an empirical Bayes approach, they will have to be estimated from the marginal density of the observations.

2.3 Posterior densities

Let $\gamma = (\alpha', \beta')'$ where $\alpha = (\alpha_k)_{|k|\leq K_0}$, and $\beta = (\beta_{jk})_{0\leq j\leq J, |k|\leq K_j}$. The prior specified above indicates that $\gamma|\tau^2 \sim N(0, \tau^2\Gamma)$ where

$$\Gamma = \begin{pmatrix} I_{2K_0+1} & 0 \\ 0 & \Delta_{M_\beta} \end{pmatrix},$$

where $M_\beta = \sum_{j=0}^{J}(2K_j + 1)$ and with Δ being the variance-covariance matrix of β (which is also diagonal, with the diagonal entries as specified in the previous section). Also $(\eta_1, \ldots \eta_n)'|\tau^2 \sim N(0, \tau^2 Q_n)$, where Q_n is formed from the variance structure of $\{\beta_{jk}\}_{J+1\leq j<\infty, |k|\leq K_j}$. Since the β_{jk} are assumed to be independent, the (i, l) entry of Q_n is given by

$$\begin{aligned}
(Q_n)_{i,l} &= \tau^{-2}\mathbb{C}ov(\eta_i, \eta_l) \\
&= \tau^{-2}\mathbb{C}ov\left(\sum_{j\geq J+1}\sum_{|k|\leq K_j}\beta_{jk}\psi_{jk}(x_i), \sum_{p\geq J+1}\sum_{|q|\leq K_j}\beta_{pq}\psi_{pq}(x_l)\right) \\
&= \tau^{-2}\sum_{j\geq J+1}\sum_{p\geq J+1}\sum_{|k|\leq K_j}\sum_{|q|\leq K_j}\mathbb{C}ov(\beta_{jk}, \beta_{pq})\psi_{jk}(x_i)\psi_{pq}(x_l) \\
&= \sum_{j\geq J+1}\sum_{|k|\leq K_j} 2^{-2js}\psi_{jk}(x_i)\psi_{jk}(x_l). \qquad (7)
\end{aligned}$$

Since the ψ function is bounded, it can be shown easily that the covariance matrix of $(\eta_1, \ldots, \eta_n)'$ is well defined (*cf.* Angers and Delampady, 2001). In Angers and Delampady (2001) a sensitivity study on the choice of Q_n is done and it is shown that this choice of Q_n does not unduly influence the wavelet smoother.

Let $X = (\Phi', S')$ with the ith row of Φ' being $\{\phi_k(x_i)\}'_{|k|\leq K_0}$ and the ith row of S' being $\{\psi_{jk}(x_i)\}'_{0\leq j\leq J, |k|\leq K_j}$. Given γ, σ^2 and τ^2, we have the following linear model for the observations vector $Y = (y_1, \ldots, y_n)'$:

$$Y = X\gamma + u,$$

where $u = \eta + \varepsilon \sim N(0, \Sigma)$ with $\Sigma = \sigma^2 I_n + \tau^2 Q_n$. This follows from the fact that

$$\begin{aligned} Y|\gamma, \eta, \sigma^2, \tau^2 &\sim N(X\gamma + \eta, \sigma^2 I_n), \\ \eta|\tau^2 &\sim N(0, \tau^2 Q_n). \end{aligned} \tag{8}$$

From (8) and using standard hierarchical Bayes techniques (*cf.* Lindley and Smith, 1972) and matrix identities (*cf.* Searle, 1982), it follows that

$$\begin{aligned} Y|\sigma^2, \tau^2 &\sim N(0, \sigma^2 I_n + \tau^2 [X\Gamma X' + Q_n]), \\ \gamma|Y, \sigma^2, \tau^2 &\sim N(AY, B), \end{aligned}$$

where

$$A = \tau^2 \Gamma X' \left(\sigma^2 I_n + \tau^2 [X\Gamma X' + Q_n]\right)^{-1}, \tag{9}$$

$$B = \tau^2 \Gamma - \tau^4 \Gamma X' \left(\sigma^2 I_n + \tau^2 [X\Gamma X' + Q_n]\right)^{-1} X\Gamma. \tag{10}$$

In order to proceed to the estimation of σ^2 and τ^2, some algebraic simplifications are needed (*cf.* Angers and Delampday, 1992). Spectral decomposition yields $X\Gamma X' + Q_n = HDH'$, where $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ is the matrix of eigenvalues and H is the orthogonal matrix of eigenvectors. Thus,

$$\begin{aligned} \sigma^2 I_n + \tau^2 [X\Gamma X' + Q_n] &= H\left(\sigma^2 I_n + \tau^2 D\right) H' \\ &= \tau^2 H \left(v I_n + D\right) H', \end{aligned}$$

where $v = \sigma^2/\tau^2$. Using this spectral decomposition, the marginal density of Y given τ^2 and v can be written as

$$\begin{aligned} m(Y \mid \tau^2, v) &= \frac{1}{(2\pi\tau^2)^{n/2}} \frac{1}{\det(vI_n + D)^{1/2}} \\ &\quad \times \exp\left\{-\frac{1}{2\tau^2} Y'H(vI_n + D)^{-1}H'Y\right\} \\ &= \frac{1}{(2\pi\tau^2)^{n/2}} \frac{1}{\prod_{i=1}^n (v + d_i)^{1/2}} \exp\left\{-\frac{1}{2\tau^2} \sum_{i=1}^n \frac{s_i^2}{v + d_i}\right\}, \end{aligned} \tag{11}$$

where $s = (s_1, \ldots, s_n)' = H'Y$.

To estimate σ^2 and τ^2 from equation (11) we proceed as follows:

1. Let $v = \sigma^2/\tau^2$.

2. Solve, in terms of v,

$$\sum_{i=1}^{n} \frac{s_i^2}{(v+d_i)^2} \left[\sum_{j=1}^{n} \left(\frac{d_j - d_i}{v+d_j} \right) \right] = 0.$$

Denote this solution by $\widehat{v}$.

3. Compute

$$\widehat{\tau}^2 = n^{-1} \sum_{i=1}^{n} (s_i^2)/(\widehat{v}+d_i),$$
$$\widehat{\sigma^2} = \widehat{v}\widehat{\tau}^2.$$

Hence, if we replace τ^2 and σ^2 by $\widehat{\tau}^2$ and $\widehat{\sigma}^2$ into the prior model, the marginal and posterior densities are

$$Y \sim N\left(0, \widehat{\tau}^2 H\left(\widehat{v} I_n + D\right) H'\right) \tag{12}$$

$$\gamma|Y \sim N\left(\Gamma X' H \left(\widehat{v} I_n + D\right)^{-1} H' Y, \right.$$
$$\left. \widehat{\tau}^2 \left(\Gamma - \Gamma X' H \left(\widehat{v} I_n + D\right)^{-1} H' X \Gamma\right)\right). \tag{13}$$

Under the squared-error loss, the Bayesian nonparametric estimate of $f(\cdot)$ is given by

$$\begin{aligned} \widehat{f}(x) &= \sum_{|k| \le K_0} \widehat{\alpha}_k \phi_k(x) + \sum_{j=0}^{J} \sum_{|k| \le K_j} \widehat{\beta}_{jk} \psi_{jk}(x) \\ &= D(x) \widehat{\gamma}, \end{aligned}$$

where

$$\begin{aligned} D(x) &= \left(\{\phi_k(x)\}'_{|k| \le K_0}, \{\psi_{jk}(x)\}'_{0 \le j \le J, |k| \le K_0} \right), \\ \widehat{\gamma} &= \Gamma X' H \left(\widehat{v} I_n + D\right)^{-1} H' Y. \end{aligned}$$

2.4 Observed distance between two functions

Suppose that the functions f_1 and f_2 given in equations (1) and (2) are estimated by

$$\begin{aligned} \widehat{f}_\ell(x) &= \sum_{|k| \le K_0} \widehat{\alpha}_{(\ell)k} \phi_k(x) + \sum_{j=0}^{J} \sum_{|k| \le K_j} \widehat{\beta}_{(\ell)jk} \psi_{jk}(x) \\ &= D(x) \widehat{\gamma}_\ell, \end{aligned} \tag{14}$$

where

$$\widehat{\gamma}_\ell = \Gamma X'_\ell H_\ell \left(\widehat{v}_\ell I_{n_\ell} + D_\ell\right)^{-1} H'_\ell Y_\ell \tag{15}$$

for $\ell = 1, 2$. (The symbol $X_\ell, H_\ell, v_\ell, D_\ell$ and Y_ℓ refer to the same quantities defined in this previous section. However they are computed with the observations from the ℓ^{th} model.) Hence the $\mathcal{L}_2$-distance between $\widehat{f}_1$ and $\widehat{f}_2$ is given by :

$$d(\widehat{\gamma}_1, \widehat{\gamma}_2) = \int_{\mathcal{T}} \|\widehat{f}_1(x) - \widehat{f}_2(x)\|^2 dx$$

$$
\begin{aligned}
&= \int_{\mathcal{T}} \| \sum_{|k|\leq K_0} \widehat{\alpha}_{(1)k}\phi_k(x) + \sum_{j=0}^{J}\sum_{|k|\leq K_j} \widehat{\beta}_{(1jk}\psi_{jk}(x) \\
&\qquad - \sum_{|k|\leq K_0} \widehat{\alpha}_{(2)k}\phi_k(x) - \sum_{j=0}^{J}\sum_{|k|\leq K_j} \widehat{\beta}_{(2jk}\psi_{jk}(x)\|^2 dx \\
&= \int_{\mathcal{T}} \| \sum_{|k|\leq K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})\phi_k(x) \\
&\qquad + \sum_{j=0}^{J}\sum_{|k|\leq K_j} (\widehat{\beta}_{(1)jk} - \widehat{\beta}_{(2)jk})\psi_{jk}(x)\|^2 dx \\
&= \sum_{|k|\leq K_0} (\widehat{\alpha}_{(1)k} - \widehat{\alpha}_{(2)k})^2 + \sum_{j=0}^{J}\sum_{|k|\leq K_j} (\widehat{\beta}_{(1)jk} - \widehat{\beta}_{(2)jk})^2 \qquad (16) \\
&= (\widehat{\gamma}_1 - \widehat{\gamma}_2)'(\widehat{\gamma}_1 - \widehat{\gamma}_2). \qquad (17)
\end{aligned}
$$

since $\{\phi_k, \psi_{jk}\}_{j\geq 0, k\in\mathbb{Z}}$ is an orthogonal basis of $\mathcal{L}_2(\mathbb{R})$.

Based on this distance, three tests are derived and presented in the next section.

3 Three possible tests

In this section, three tests to choose between the hypotheses given in (3) are presented. The first one is based on Bayes factor while to two others are using the observed distance given by equation (17).

3.1 Test based on the Bayes factor

The first test compares the posterior ratio to the prior ratio under both hypothesis given in (3). It can be shown that the Bayes factor is given by

$$
B(y_1, y_2) = \frac{m_0(y_1, y_2)}{m_1(y_1, y_2)},
$$

where $m_l(y_1, y_2)$ denotes the marginal of the observations under the hypothesis H_l, $l = 0, 1$. Using equation (12) for both sets of observations, the marginal of (Y_1, Y_2), under $H_1 : f_1 \neq f_2$, is given by

$$
\begin{aligned}
m_1(y_1, y_2) = &\frac{1}{(2\pi)^{(n_1+n_2)/2}\tau_1^{n_1}\tau_2^{n_2}} |\widehat{v}_1 I_{n_1} + D_1|^{-1/2} |\widehat{v}_2 I_{n_2} + D_2|^{-1/2} \\
&\times \exp\left\{ -\frac{1}{2}\left(y_1' H_1 (\widehat{v}_1 I_{n_1} + D_1)^{-1} H_1' y_1 + y_2' H_2 (\widehat{v}_2 I_{n_2} + D_2)^{-1} H_2' y_2 \right) \right\}.
\end{aligned}
$$

In order to write the marginal under the null hypothesis, let $Y_0 = (Y_1', Y_2')'$. The design matrix for this problem is $X_0 = (X_1', X_2')'$ and let $Q_{n_1+n_2}$ be defined by equation (7) where $x_i, x_\ell \in \{x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}, x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}\}$. Let $H_0 D_0 H_0' = X_0 \Gamma X_0' + Q_{n_1+n_2}$. If $\widehat{v}_0$ denotes the solution of

$$
\sum_{i=1}^{n_1+n_2} \frac{s_{0i}^2}{(v + d_{0i})^2} \left[\sum_{j=1}^{n_1+n_2} \left(\frac{d_{0j} - d_{0i}}{v + d_{0j}} \right) \right] = 0,
$$

where $s_0 = (s_{01}, s_{02}, \ldots, s_{0(n_1+n_2)})' = H_0' Y_0$. Hence, the marginal under the null hypothesis is given by:

$$m_0(y_1, y_2) = \frac{1}{(2\pi)^{(n_1+n_2)/2} \tau_0^{n_1+n_2}} |\widehat{v}_0 I_{n_1+n_2} + D_0|^{-1/2} \times \exp\left\{ -\frac{1}{2} \left((y_1', y_2') H_0 (\widehat{v}_0 I_{n_1+n_2} + D_0)^{-1} H_0' (y_1', y_2')' \right) \right\}, \quad (18)$$

where

$$\widehat{\tau}_0^2 = \frac{1}{n_1 + n_2} \sum_{i=1}^{n_1+n_2} \frac{s_{0i}^2}{\widehat{v}_0 + d_{0i}}.$$

and we accept $H_0 : f_1 = f_2$ if $B(y_1, y_2)$ is large enough.

3.2 Distance between two functions

In section 2.4, the observed distance between the two curves is given by equations (16) or (17). Using equation (5) and since $\{\phi_k, \psi_{jk}\}_{j\geq 0, k\in\mathbb{Z}}$ is an orthogonal basis of $\mathcal{L}_2(\mathbb{R})$, we can defined the $\mathcal{L}_2$-distance between f_1 and f_2 (*cf.* Zhao, 2000) as:

$$\begin{aligned}
\mathrm{d}(f_1, f_2) &= \int_{\mathcal{T}} \|g_{J1}(x) - g_{J2}(x)\|^2 dx \\
&= \int_{\mathcal{T}} \| \sum_{|k|\leq K_0} \alpha_{(1)k}\phi_k(x) + \sum_{j\geq 0} \sum_{|k|\leq K_j} \beta_{(1jk}\psi_{jk}(x) \\
&\quad - \sum_{|k|\leq K_0} \alpha_{(2)k}\phi_k(x) - \sum_{j\geq 0} \sum_{|k|\leq K_j} \beta_{(2jk}\psi_{jk}(x) \|^2 dx \\
&= \int_{\mathcal{T}} \| \sum_{|k|\leq K_0} (\alpha_{(1)k} - \alpha_{(2)k})\phi_k(x) \\
&\quad + \sum_{j\geq 0} \sum_{|k|\leq K_j} (\beta_{(1)jk} - \beta_{(2)jk})\psi_{jk}(x) \|^2 dx \\
&= \sum_{|k|\leq K_0} (\alpha_{(1)k} - \alpha_{(2)k})^2 + \sum_{j\geq 0} \sum_{|k|\leq K_j} (\beta_{(1)jk} - \beta_{(2)jk})^2 \qquad (19) \\
&= d(\gamma_1, \gamma_2) + \sum_{j\geq (J+1)} \sum_{|k|\leq K_j} (\beta_{(1)jk} - \beta_{(2)jk})^2 \\
&= d(\gamma_1, \gamma_2) + O(2^{-2(J+1)s}).
\end{aligned}$$

Hence $d(\gamma_1, \gamma_2)$ can be considered as a good approximation of equation (19).

Let $\xi = d(\gamma_1, \gamma_2)$. Even though the exact density of ξ can be written as an explicit series (*cf.* Johnson and Kotz, 1970, Chapter 29), it is difficult to compute accurately. However, its posterior cumulents can be easily computed. Using (Searle, 1971, Section 2.5), the r^{th} posterior cumulent of ξ is given by

$$\begin{aligned}
C_r(y_1, y_2) &= 2^{r-1}(r-1)! \left[\mathrm{tr}(V^r(y_1, y_2)) \right. \\
&\quad \left. + r\theta(y_1, y_2)' V^{r-1}(y_1, y_2)\theta(y_1, y_2) \right], \qquad (20)
\end{aligned}$$

where $V(y_1, y_2) = B_1 + B_2$ and $\theta(y_1, y_2) = A_1 y_1 + A_2 y_2$ (*cf.* equations (9) and (10). Using the Edgeworth expansion (*cf.* Lange, 1998, Section 17.5), the density of ξ, can be approximated by:

$$\pi_\Xi(\xi \mid y_1, y_2) = \varphi\left(\frac{\xi-\mu}{\sigma}\right)\left[1 + \frac{\rho_3}{6}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\rho_4}{24}H_4\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\rho_3^2}{72}H_6\left(\frac{\xi-\mu}{\sigma}\right)\right],$$

and its c.d.f. by

$$\Pi_\Xi(\xi \mid y_1, y_2) = \Phi\left(\frac{\xi-\mu}{\sigma}\right) - \varphi\left(\frac{\xi-\mu}{\sigma}\right)\left[\frac{\rho_3}{6}H_2\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\rho_4}{24}H_3\left(\frac{\xi-\mu}{\sigma}\right) + \frac{\rho_3^2}{72}H_5\left(\frac{\xi-\mu}{\sigma}\right)\right],$$

where $\varphi(\cdot)$ denotes the standard normal density, $\Phi(\cdot)$, its c.d.f. $H_k(\cdot)$, the k^{th} Hermite polynomial (*cf.* Lange, 1998, Section 15.4) and

$$\begin{aligned} \mu &= C_1(y_1, y_2), \\ \sigma &= \sqrt{C_2(y_1, y_2)}, \\ \rho_k &= \frac{C_k(y_1, y_2)}{\sigma^k}. \end{aligned}$$

3.3 Test based of the c.d.f. of the distance

Although the test given in section 3.1 is the proper Bayesian thing to do, it is very computationally demanding since it requires the spectral decomposition of an $(n_1 + n_2)$ squared matrix. In this section and the following one, alternative tests are proposed.

The second test proposed is based on the cdf of ξ and it is defined as:

$$D(y_1, y_2) = \sup_{x \in \mathcal{T}} \left[\Pi_{0,\Xi}(x) - \Pi_{1,\Xi}(x)\right], \tag{21}$$

where $\Pi_{i,\Xi}(x)$ denotes the c.d.f. of the ξ under the hypothesis H_i $i = 0, 1$. The rational behind this test is that the distance between f_1 and f_2 should be larger if H_1 is true than if H_0 is true. Consequently, $\Pi_{0,\Xi}$ should be stochastically larger than $\Pi_{1,\Xi}$ and $D(y_1, y_2)$ should be large if H_1 is true (*cf.* Lehmann, 1959). Hence, we will reject $H_0 : f_1 = f_2$ is $D(y_1, y_2)$ is large enough. The power of this test will be discussed in section 4.2. The critical level of this test is obtained using a bootstrap method similar to the one discussed in the next section.

3.4 Bootstrap test

The last test is based on the bootstrap algorithm (*cf.* Davison and Hinkley, 1997, Section 6.3). This test is obtained using the following algorithm.

Step 1 Compute the estimated value of γ_1 and γ_2 using equation (15).

Step 2 Estimate the distance under H_1 using equation (17) and denote this distance by d^*.

Step 3 Compute the estimated value of $f_\ell(x_{\ell i})$ for $i = 1, 2, \ldots, n_1$, $\ell = 1, 2$ using equation (14).

Step 4 Compute the residual values $e_{\ell i} = y_{\ell i} - \widehat{f}_\ell(x_{\ell i})$ for $i = 1, 2, \ldots, n_\ell$ and $\ell = 1, 2$.

Step 5 Combining both samples, estimate the common function under the null hypothesis, denoted by $\widehat{f}_0$.

Step 6 Set $m = 1$.

Step 7 Select at random with replacement n_ℓ values among $\{e_{\ell 1}, e_{\ell 2}, \ldots, e_{\ell n_\ell}\}$, denoted by $e_{\ell i}(m)$ and define $y_{\ell i}(m) = \widehat{f}_0(x_{\ell i}) + e_{\ell i}(m)$ for $i = 1, 2, \ldots, n_\ell$ and $\ell = 1, 2$.

Step 8 Using the bootstrap observations $y_{\ell i}(m)$, estimate the values of γ_1 and γ_2 and compute the distance using equation (17), denoted by $d(\widehat{\gamma}_1(m), \widehat{\gamma}_2(m))$.

Step 9 Set $m = m + 1$. If $m \leq M$, go back to step **7**, otherwise, go to step **10**.

Step 10 Compute the bootstrap *p-value* defined as

$$p\text{-}value = \frac{1}{M+1}\left[1 + \sum_{m=1}^{M} \mathbb{I}\{d(\widehat{\gamma}_1(m), \widehat{\gamma}_2(m)) \geq \mathrm{d}^*\}\right],$$

where $\mathbb{I}\{d(\widehat{\gamma}_1(m), \widehat{\gamma}_2(m)) \geq \mathrm{d}^*\}$ equals 1 if $d(\widehat{\gamma}_1(m), \widehat{\gamma}_2(m)) \geq \mathrm{d}^*$, and 0 otherwise.

4 Simulation and example

In this section, a simulation study is made in order to compare the power of these tests with those proposed in Kulasekera (1995). An example using a real data set is also considered. A simulation to study the power of the proposed tests is also done based on the real data set example.

4.1 Set up

This simulation study of level and power for the three proposed tests is similar to the one proposed in Kulasekera (1995). This enable us to compare the performance of the proposed tests with those proposed in Kulasekera (1995). The functions used in this simulation study are given in Table 1. The sample sizes are chosen to be equal ($n_1 = n_2 = 20, 30$ and 50). The design points were generated using a uniform distribution on the unit interval. The errors are distributed according to a normal with 0 mean and standard deviation equaled to 0.1.

4.2 Simulation results

In Table 1, it can be seen that the critical region for the c.d.f. (Section 3.3) test does not vary much across functions for small sample size ($n = 20$) (from 0.0625 to 0.09). However for $n = 30$ and $n = 50$, the critical region depend on the considered functions (from 0.0425 to 0.19 for $n = 30$ and 0.0325 to 0.1325 for $n = 50$). For the bootstrap test (Section 3.4), the critical regions does not vary much for all sample sizes considered. If the test based of the Bayes factor (Section 3.1) is to reject H_0 if $BF > 1$, then this test is too conservative.

In Tables 2 and 3, we reported the empirical power of the three proposed test along with the "best" test proposed in Kulasekera (1995). (The best test is defined as the one among those presented in Kulasekera (1995) with the largest power for each case.) The tests discussed in Kulasekera (1995) are based on quasi-residuals obtained after the functions are estimated using a kernel based methods. For all tests considered, their power is almost constants. Beside the test based on the Bayes factor (whose level is uncontroled) the bootstrap test is the best, followed by the one based on the c.d.f.

Table 1: Empirical level (times 1000)

$f_1(x) = f_2(x)$	n	Bayes factor	Distance	Cut off	Bootstrap	Cut off
$\sqrt{x}$	20	4	53	0.0625	50	0.04
	30	0	49	0.0825	49	0.0225
	50	0	48	0.1	52	0.0275
x^2	20	11	52	0.09	52	0.025
	30	1	51	0.12	50	0.0275
	50	2	50	0.1325	55	0.02
$\cos(\pi x)$	20	0	51	0.09	52	0.0625
	30	0	46	0.0425	48	0.0275
	50	0	50	0.0725	50	0.0275
$\cos(2\pi x)$	20	7	51	0.09	49	0.0425
	30	0	51	0.0425	48	0.0225
	50	0	53	0.0525	52	0.0225
$\cos(4\pi x)$	20	44	50	0.0725	51	0.045
	30	0	50	0.1025	49	0.0275
	50	0	49	0.0325	48	0.025
$\cos^2(2\pi x)$	20	32	51	0.075	52	0.05
	30	2	50	0.19	45	0.0025
	50	0	49	0.045	48	0.025

distance. The test from Kulasekera (1995) is always the test with the lowest power. However, the power of the bootstrap test and the one based on the c.d.f. are closed to each other and the difference between them and the power of Kulasekera (1995) test decreased as n and/or the "true" distance increased.

Even though the bootstrap test is slightly more powerful, the use of the test based on equation (21) is recommended because:

i) $\mathrm{d}(f_1, f_2)$ depend on the scale of the observations while $D(y_1, y_2) \in [-1, 1]$ for any functions f_1 and f_2;

ii) $D(y_1, y_2)$ can be easily visualized by plotting the posterior c.d.f. of ξ under both hypotheses (*cf.* Figure 2). Hence, the c.d.f. test is easier to interpret.

4.3 Ethanol example

In this section, we apply the proposed test to a real data set. This data set pertains to an exhaust study using ethanol as fuel in an experimental engine (*cf.* Cleveland, 1993) (available as a S-Plus data set). The response variable Y is the concentration of nitric oxide plus the concentration of nitrogen dioxide in the exhaust of an experimental engine when it is set at different equivalence ratios (x) and compression ratios (c). The goal of this experiment is to determine the dependency of the concentration of oxides of nitrogen on various engine settings. The compression ratio takes only few values, which is categorized as "Low" ($c < 10$) and "High" ($c \geq 10$). Hence, Y is modelled as a function of x for the two types of compression ratios and we want to compare the two functions. The "low" group has 39 observations and the "high" group has 49. These observations along with the estimated functions for both groups are given in Figure 1. The estimated functions are very similar to the one obtained using kernel estimators given in Kulasekera (1995). Looking at Figure 1, it is

Table 2: Empirical Power (×1000) of the different tests at 5% level

Functions	Size	BF(1)	CDF	Bootstrap	W_{best}
$f(x)=\sqrt{x}; g(x)=x^2$	20	1000	981	985	418
true distance=0.129	30	1000	998	998	572
	50	1000	992	1000	728
$f(x)=\sqrt{x}; g(x)=f(x)+x$	20	1000	976	996	593
true distance=0.333	30	1000	990	996	664
	50	1000	1000	1000	781
$f(x)=x^2; g(x)=f(x)+x$	20	1000	984	995	586
true distance=0.333	30	1000	996	998	654
	50	1000	1000	1000	772
$f(x)=\cos(\pi x); g(x)=f(x)+x$	20	1000	976	991	616
true distance=0.333	30	1000	995	999	677
	50	1000	992	1000	827
$f(x)=\cos(2\pi x); g(x)=f(x)+x$	20	1000	743	894	683
true distance=0.333	30	1000	975	990	740
	50	1000	1000	1000	861

Table 3: Empirical Power (×1000) of the different tests at 5% level (ctnd)

Functions	Size	BF(1)	CDF	Bootstrap	W_{best}
$f(x)=\cos^2(2\pi x); g(x)=\sin^2(2\pi x)$	20	997	924	953	445
true distance=0.5	30	993	994	880	474
	50	1000	1000	1000	523
$f(x)=\sqrt{x}; g(x)=f(x)+1$	20	1000	991	995	593
true distance=1	30	1000	1000	1000	968
	50	1000	1000	1000	990
$f(x)=\cos(\pi x); g(x)=f(x)+1$	20	1000	987	994	905
true distance=1	30	1000	1000	1000	967
	50	1000	1000	1000	997
$f(x)=\cos(2\pi x); g(x)=f(x)+1$	20	1000	743	894	683
true distance=1	30	1000	975	990	740
	50	1000	992	1000	861
$f(x)=\cos(4\pi x); g(x)=\sin(4\pi x)$	20	1000	831	873	629
true distance=1	30	1000	899	982	736
	50	1000	999	1000	863

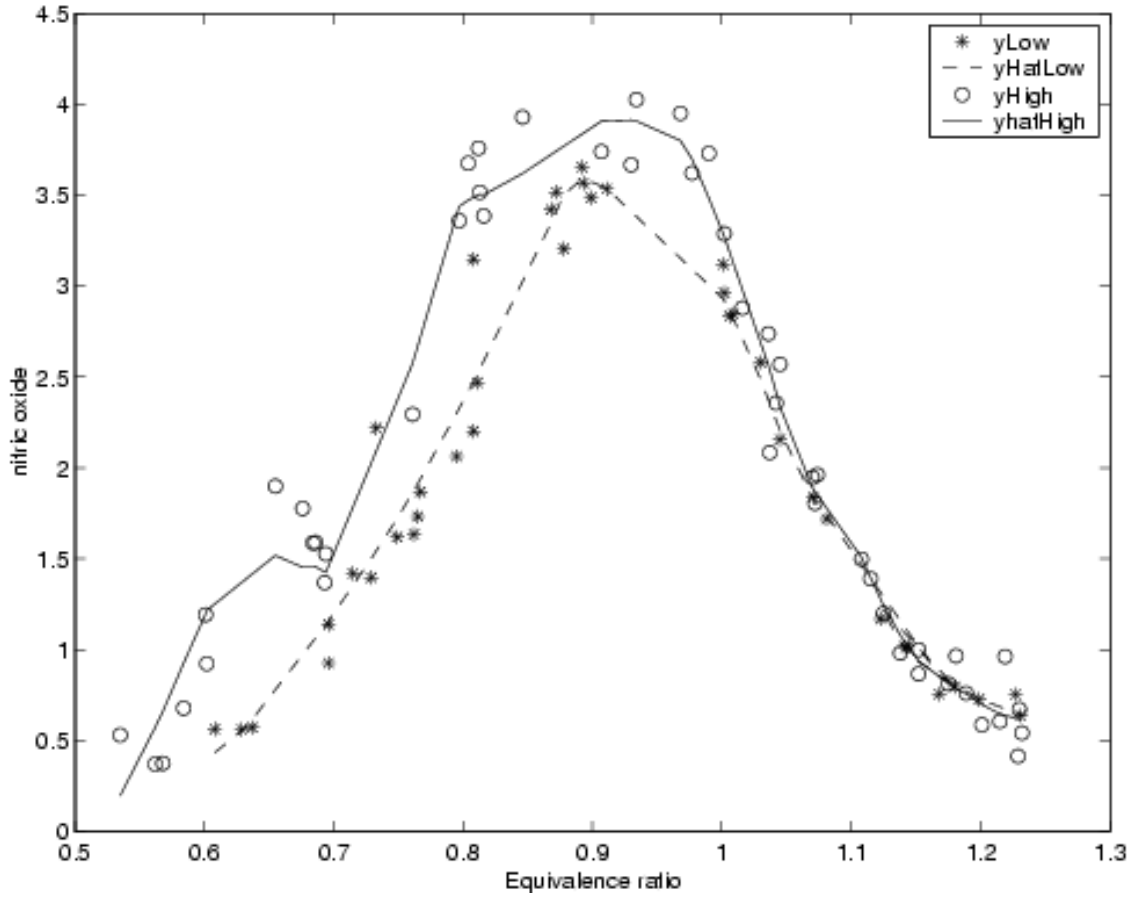

Figure 1: Data points and the estimated functions

Table 4: Empirical *p-values* for the Ethanol data

Test	*p-values*		
	$\forall x$	$x < 1$	$x \geq 1$
Bayes	0.000	0.000	0.000
CDF	0.008	0.000	0.132
Bootstrap	0.042	0.355	0.598

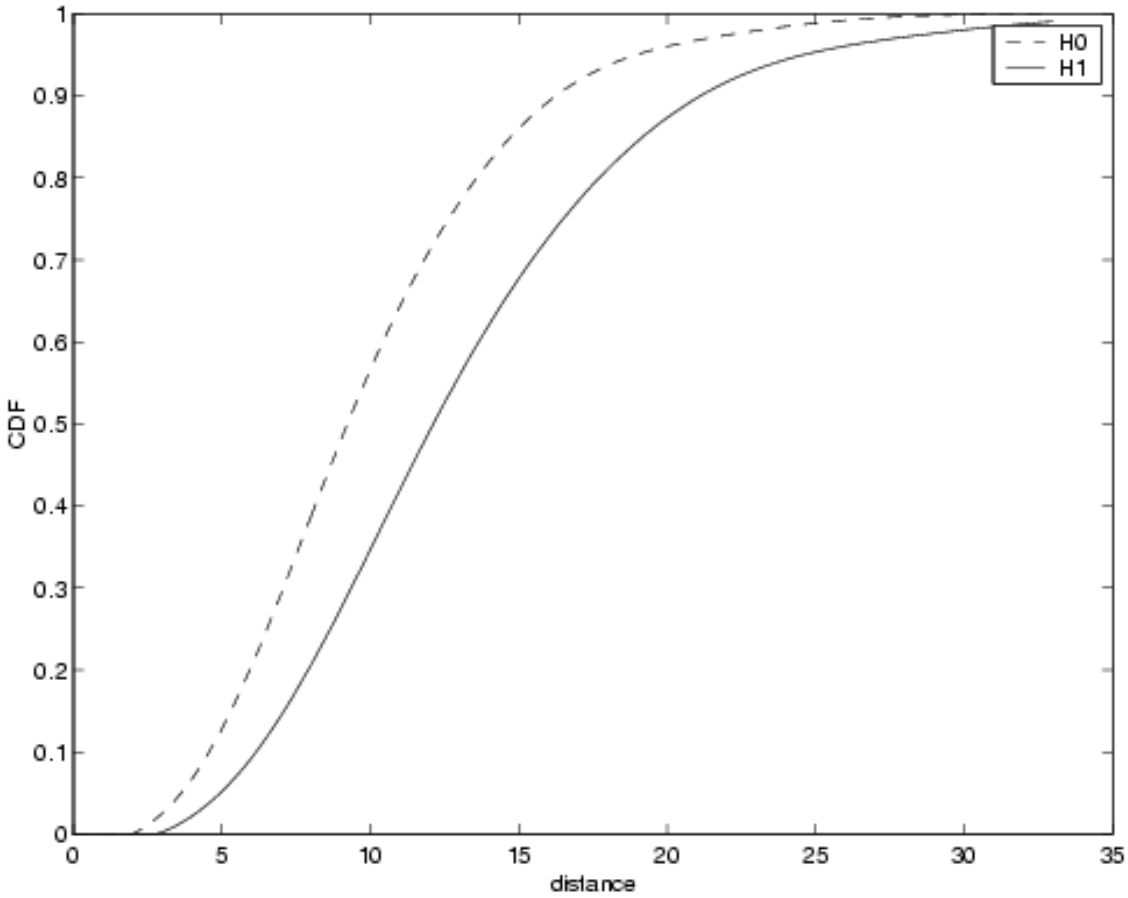

Figure 2: c.d.f. of the distance under H_0 and H_1 hypotheses

obvious that these two functions are different, specially when the equivalence ratio is less than 1, conclusion shared by Kulasekera (1995). For this data set, $D(y_1, y_2) = 0.272$ and $d(\widehat{\gamma}_1, \widehat{\gamma}_2) = 0.224$. Based on a simulation of 250 iterations, the empirical *p-values* are given in Table 4. In this table, we also reported the *p-values* for the same tests but for the cases where the equivalence ratio is less and, larger or equal to 1. From Table 4, it can be seen that, using all the observations, all tests agree that both functions are different. However the *p-value* for bootstrap test is much larger than the other ones. It can also be seen that the Bayes test is too sensitive and its *p-value* is always equal to 0 even when the equivalence ratio is larger or equal to 1. When the data set is divide into two, there is not enough observations for the bootstrap test to pick up the difference between the two functions for $x < 1$. Hence, the test based on the c.d.f. of the distance does somehow better.

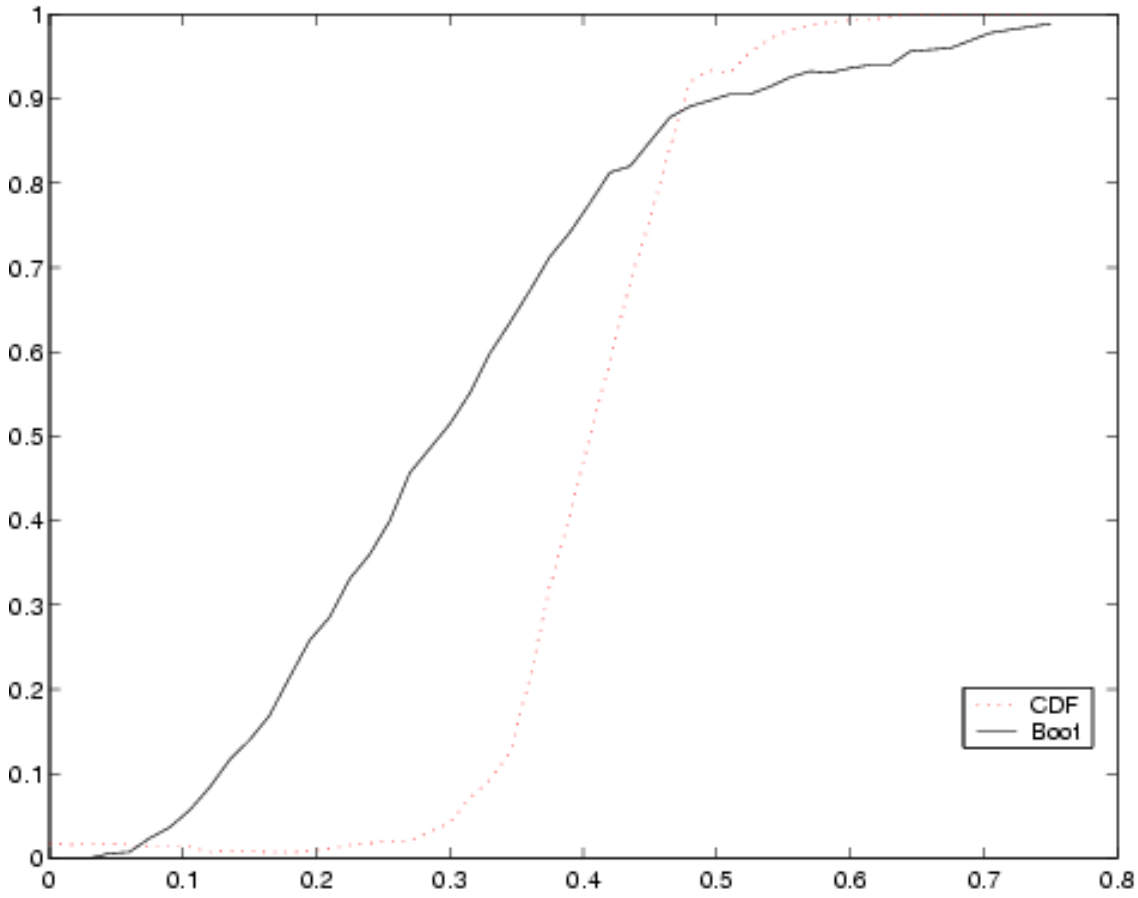

Figure 3: c.d.f. of the distance under H_0 and H_1 hypotheses

4.4 Power simulation

In this section, we study the power of the bootstrap and c.d.f. tests in a more realistic set up. (We did not consider the Bayes test because its level is very difficult to control.) Let $\widehat{f}_1$ and $\widehat{f}_2$ denote the estimated function from the low and high compression ratios of the previous section and define

$$\widetilde{f}_2(x) = D(x)\left[\widehat{\gamma}_1 + \theta\left(\widehat{\gamma}_2 - \widehat{\gamma}_1\right)\right], \tag{22}$$

for $0 \leq \theta \leq 0.8$. (This is a similar set up as in Fan and Lin (1998).) The pseudo-observations for the second group of data are defined at

$$\widetilde{Y}_{2,j} = \widetilde{f}_2(x_{2,j}) + r_{2,j},$$

where $r_{2,j} = Y_{2,j} - \widehat{f}_2(x_{2,j}) = Y_{2,j} - D(x_{2,j})\widehat{\gamma}_2$ for $j = 1, 2, \ldots, 49$.

Figure 3 depicts the power function of the bootstrap and c.d.f. tests at 5% level. Each point is obtained using a bootstrap procedure with 500 iterations and θ (*cf.* equation (22)) goes from 0 to 0.8 by step of 0.025. From this figure, we can see that the bootstrap test is better for intermediate values of θ (between 0.1 and 0.45) while the test based on the c.d.f. is better for large values of θ (larger than 0.45).

5 Concluding Remarks

In this paper, we want to test if two functions are equal or not. In a Bayesian setting, the test based on the Bayes factor should be preferred. However it can be difficult to evaluate it because it requires the spectral decomposition of an $(n_1 + n_2)$ squared matrix. Two alternative tests are proposed which are faster to compute and required less computing power.

Based on the simulation done for this paper, we find that both alternatives (the test based on the c.d.f. of the $\mathcal{L}_2(\mathbb{R})$-distance and the bootstrap test) are more or less equivalent. However, the c.d.f. test is preferred because it does not depend on the scale of the observations making it easier to interpret.

References

Abromovich, F. and Sapatinas, T. (1999), "Bayesian approach to wavelet decomposition and shrinkage," in *Bayesian Inference in wavelet-based models*, eds. Müller, P. and Vidakovic, B., New York: Springer, vol. 141 of *Lecture Notes in Statistics*, pp. 33–50.

Angers, J.-F. and Delampday, M. (1992), "Hierarchical Bayesia estimation and curve fitting," *The Canadian Journal of Statistics*, 20, 35–49.

— (2001), "Bayesian nonparametric regression using wavelets," *To appear in Sankhyä, Series B*.

Brillinger, D. R. (1973), "The analysis of time series collected in an experiment design," in *Multivariate Analysis III*, ed. Krishnaiah, P., New York: Academic Press, pp. 241–256.

— (1980), "Some aspects of the analsis of evoked response experiments," in *Statistics and related topics*, eds. Csörgö, M., Dawson, D., Rao, J., and Saleh, A., Amsterdam: North-Holland, pp. 15–168.

Cleveland, W. (1993), *Visualizing data*, Summit, N.J.: Hobart Press.

Daubechies, I. (1992), *Ten lectures on wavelets*, Philadelphia: SIAM.

Davison, A. and Hinkley, D. (1997), *Bootstrap methods and their application*, New York: Cambridge University Press.

Delyon, B. and Juditsky, A. (1995), "Estimating wavelet coefficients," in *Wavelets and Statistics, Lecture Notes in Statistics*, eds. Antoniadis, A. and Oppenheim, G., New York: Springer, vol. 103, pp. 151–167.

Fan, J. and Lin, S.-K. (1998), "Test of significance when data are curves," *Journal of the American Statistical Association*, 93, 1007 – 1021.

Hall, P. and Hart, J. D. (1990), "Bootstrap test for difference between means in nonparametric regression," *Journal of the American Statistical Association*, 85, 1039–1049.

Härdle, W., Kerkyachaian, G., Picard, D., and Tsybakov, A. (1998), *Wavelets, appoximation, and statistical applications*, vol. 129 of *Lecture Notes in Statistics*, New York: Springer.

Johnson, N. L. and Kotz, S. (1970), *Continuous univariate distribution – 2*, Boston: Houghton Mifflin Company.

Koul, H. L. and Schick, A. (1997), "Testing for the equality of two nonparametric regression curves," *Journal of Statistical Planning and Inference*, 65, 293–314.

Kulasekera, K. B. (1995), "Comparison of regression curve using quasi-residuals," *Journal of the American Statistical Association*, 90, 1085–1093.

Lange, K. (1998), *Numerical analysis for statisticians*, New York: Springer.

Lehmann, E. (1959), *Testing statistical hypothesis*, New York: John Wiley and Sons.

Lindley, D. V. and Smith, A. F. (1972), "Bayes estimates for the linear model," *Journal of the Royal Statistical Society, Series B*, 34, 1–41.

Mallat, S. (1998), *A wavelet tour of signal processing*, San Diego: Academic Press.

Müller, P. and Vidakovic, B. (1999), "An introduction to wavelets," in *Bayesian inference in wavelet-based models, Lecture Notes in Statistics*, eds. Müller, P. and Vidakovic, B., New York: Springer, vol. 141, pp. 1–18.

Ogden, R. (1997), *Essential wavelets for statistical applications and data analysis*, Boston: Birkhäuser.

Searle, S. (1971), *Linear models*, New York: John Wiley and Sons.

— (1982), *Matrix algebra useful for statistics*, New York: Wiley.

Shumway, R. (1988), *Applied statistical time series analysis*, Englewood Cliffs, NJ: Prentice-Hall.

Vidakovic, B. (1998), "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *Journal of the American Statistical Association*, 93, 173–179.

— (1999), *Statistical modeling by wavelets*, New York: John Wiley.

Zhao, L. (2000), "Bayesian aspects of some nonparametric Problems," *The Annals of Statistics*, 28, 532–552.