Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Choosing the optimal bandwidth: The distribution function's kernel estimator method

Gregorio S. Atuncar, Lupércio F. Bessegato, Luiz H. Duczmal

Relatório Técnico
RTP-06/2003

Relatório Técnico
Série Pesquisa

Choosing the Optimal Bandwidth: The Distribution Function's Kernel Estimator Method

By GREGORIO S. ATUNCAR

Departament of Statistics, Federal University of Minas Gerais

gregorio@est.ufmg.br

LUPÉRCIO F. BESSEGATO

Departament of Statistics, Federal University of Minas Gerais

lupercio@est.ufmg.br

LUIZ DUCZMAL

Departament of Statistics, Federal University of Minas Gerais

duczmal@est.ufmg.br

Summary

We propose a plug-in method to estimate the optimal bandwidth to be used in the definition of the kernel estimator of a distribution function. The empirical characteristic function is used to define the estimator. Our method is based on the ideas given by Chiu (1991). We compared the results of our method with other ones such as Bowman et al (1998), Sarda (1993) and Altman & Leger (1995). Numerical experiments suggest that this method exhibits good convergence properties with less variance.

1. Introduction

There are several approaches to estimate a distribution function. The kernel estimator has been widely used for this purpose in contraposition to the parametric approach.

If $X_1, X_2, \ldots, X_n$ is a random sample from a random variable X whose distribuition function is F, the estimator of F, evaluated at the point x is defined as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} W\left(\frac{x - X_i}{h_n}\right) \tag{1}$$

where W is a distribution function called the kernel and h_n is called the bandwidth.

We assume that the density function w of W is bounded, symmetric, continuosly differentiable, compactly supported and $0 < \int t^2 w(t) dt = k_2 < \infty$. We will also assume that $h_n \to 0$ and $nh_n \to \infty$, when $n \to \infty$. The convergence rate and the smoothing of the kernel estimator depend on the choosing of h_n. In that sense, it is extremely relevant to study the estimator of the optimal bandwidth. From here, just to simplify we will write h instead of h_n and the integrals will be assumed over the real line unless otherwise is indicated.

This paper will be concentrated on the estimation of the optimal bandwidth, h_{opt}. Before starting the study of the problem, we cite some properties of $\hat{F}_n$. More details can be seen in

Nadaraya (1964) and Bessegato (2002). The expectance and the variance of $\hat{F}_n$ are given by $E\left[\hat{F}_n\right] = F(x) + h^2 C_2 + o\left(h^2\right)$, and $Var\left(\hat{F}_n\right) = \frac{1}{n}F(x)[1 - F(x)] - \frac{h}{n}C_1 + o\left(h/n\right)$ respectively where $C_1 = 2\,F'(x) \int zW(z)\,w(z)dz$ and $C_2 = \frac{k_2}{2}F''(x)$

We note that the estimator is biased and the bias does not depend directly on the sample size n, but depend on h.

The variance of the empirical distribution function , $\tilde{F}_n(x) = \frac{1}{n}\#\{i : X_i \leq x\}$ is given by $\frac{F(x)\{1-F(x)\}}{n}$, where n is the sample size. From this we see that $\hat{F}_n$ has smaller variance and the dominant term in the reduction is $C_1 h/n$, with $C_1 > 0$. In fact, as $W(z) > 1/2$, for $z > 0$ and $W(z) < 1/2$, for $z < 0$, we can prove that $\int z\,W(z)\,w(z)\,dz > 0$. This proves that $Var\left[\hat{F}_n\right] \leq Var\left[\tilde{F}_n\right]$.

Looking at $E(\hat{F}_n)$ and $Var(\hat{F}_n)$ we can see that the larger the bandwidth, the smaller will be the variance of the kernel estimator of F, and the larger will be the bias, and vice versa. So we need to choose h such that we can make a trade off between variance and bias. This is the paramount problem in kernel estimation.

In this work we study a plug - in method to estimate the optimal bandwidth. This method is based on the empirical characteristic function and we consider only the i.i.d. case but we believe that the method can be extended to other cases. We are working now in the markovian case to estimate the invariant distribution function.

Several authors have studied the properties of the estimator $\hat{F}_n(x)$, defined by using the integral of the kernel estimator of the density function. Nadaraya (1964), Winter (1973) and Yamato (1973) studied the convergence of $\hat{F}_n(x)$ to $F(x)$, with probability 1.

Jones (1990) studied the performance of several kernels to estimate the distribution function and verified that the choosing of the kernel does not represent a large impact on the performance of the estimator $\hat{F}_n(x)$. Falk (1983) also investigated whether a given kernel type estimator of a distribution function at a single point has asymptotically better performance than the empirical estimator.

In the following section we define the Plug - in Method analysed in this work. In section 3 we present the theoretical properties of the proposed estimator. In section 4 we show the results obtained through simulations. In section 5 we present some applications.

2. Bandwidth Selection

The choice of the kernel is not such an important thing as the choosing of the bandwidth. The last problem has been studied widely in the literature. As a general rule, h is chosen trying to minimize the Mean Integrated Squared Error (MISE) of $\hat{F}_n(x)$. The MISE is defined as

$$MISE(h) = E \int \left\{ \hat{F}_n(x) - F(x) \right\}^2 dx \,. \tag{2}$$

It is difficult to compute the MISE for finite samples. Jones

(1990) has derived an expression for an approximation of its value. This expression is given by

$$MISE(h) = n^{-1} \int F(x)\,[1 - F(x)]\,dx - C_3\,h\,n^{-1} + C_4\,h^4 + o\left(h\,n^{-1} + h^4\right) \,, \tag{3}$$

where:

$$C_3 = \int W(z)\,[1 - W(z)]\,dz \quad \text{and} \quad C_4 = \frac{k_2^2}{4} \int \left[F''(x)^2\right] dx \,.$$

From this, the optimal bandwidth is obtained as

$$h_{opt} = \left\{ \frac{\int W(z)[1 - W(z)]\,dz}{\left[\int z^2 dW(z)\right]^2 \int \left[F''(x)\right]^2\,dx} \right\}^{1/3} n^{-1/3} \tag{4}$$

Unfortunately the optimal bandwidth depends on the unknown distribution. In this work we construct an estimator for $H = \int \left[(F''(x)\right]^2 (x)dx$. Then we make a "plug-in" in the expression for h_{opt} to get an estimator of the optimal bandwidth. Before going on we observe that the optimal bandwidth to estimate the distribution function is of order $n^{-\frac{1}{3}}$, different than the optimal bandwidth to estimate the density function that is of order $n^{-\frac{1}{5}}$.

There are several methods to estimate the optimal bandwidth. Although in practice it is possible to choose this parameter in a subjective way, we believe that is better to choose the estimator as a function of the data.

The most studied method in the density function case is the Cross - validation method, proposed by Rudemo (1982) and Bowman (1984). In a way similar to the proposed in the density function method, Bowman et al (1998) presented a cross - validation method to estimate the optimal bandwidth in the distribution function case. In that method, the choosing of the optimal bandwidth is based on an unbiased estimator of the MISE.

It is well recognized that, in the density function case, the cross-validation method is subject to large sample variation. In simulation studies it was observed that this method chooses bandwidths smaller than those indicated by the asymptotic results.

Another possible approach to estimate the optimal bandwidth is the "Plug - in Method". This is the method we are going to study and is based on the estimation of the only unknown quantity in the expression defining the optimal bandwidth, (The integral $\int [f'']^2$,in the density function case and the integral $\int [F'']^2$, in the distribution function case). According to Chiu (1991), apparently the plug-in method has the advantage that does not need an optimization program.

In the density function case, Chiu (1991) proposed a plug -in method based on the empirical characteristic function. The empirical characteristic function is truncated in a value Λ, because for higher frequencies, the empirical characteristic function does not add significant information about the distribution and the value is dominated by noise. The goal of this work is proposing, for the distribution function case, a similar estimator for the optimal bandwidth. We shall prove asymptotic results for the proposed estimator. Besides the usual assumptions, we assume the additional conditions 3.1 to 3.3 in the following section.

Based on the mentioned assumptions, we estimate H. Previously, we find an alternative expression for H. By using the inversion formula (Theorem 6.2.3, in Chung (1974), pag.155), we have $f(x) = \frac{1}{2\pi} \int e^{-i\lambda x} \varphi(\lambda) d\lambda$. Hence:

$$
\begin{aligned}
f'(x) = F''(x) &= \frac{1}{2\pi} \int (-i\lambda) e^{-i\lambda x} \varphi(\lambda) d\lambda \\
&= \frac{1}{2\pi} \int e^{-i\lambda x} \left[-i\lambda \varphi(\lambda) \right] d\lambda \,.
\end{aligned}
$$

Then, by using the Parseval identity (Theorem 9.13, in Rudin (1987), pag. 187), we obtain:

$$
\begin{aligned}
H &= \int \left(F''(x) \right)^2 dx \\
&= \frac{1}{2\pi} \int \lambda^2 \left| \varphi(\lambda) \right|^2 d\lambda \,. \qquad (5)
\end{aligned}
$$

The estimator of H naturally will be expressed in terms of the empirical characteristic function $\hat{\varphi}(\lambda) = \frac{1}{n} \sum_{j=1}^{n} e^{i \ \lambda \ X_j}$.

In Brillinger (1981) it is established that $\left| \hat{\varphi}(\lambda) - \varphi(\lambda) \right|^2$ is approximately distributed according to an exponential with mean $\frac{1}{n} \left(1 - \left| \varphi(\lambda) \right|^2 \right)$ and the following result is also true:
$E \left| \left| \hat{\varphi}(\lambda) \right|^2 \right| \cong \frac{1}{n} \left[(n-1) \left| \varphi(\lambda) \right|^2 + 1 \right]$. By using that, we can approximate H by

$$
\hat{H} = \frac{1}{\pi} \int_0^{\Lambda} \lambda^2 \left[\left| \hat{\varphi}(\lambda) \right|^2 - \frac{1}{n} \right] d\lambda, \qquad (6)
$$

where $\Lambda = min\left\{\lambda : |\hat{\varphi}(\lambda)|^2 \leq \frac{c}{n}\right\}$, for some $c > 1$. The term $1/n$ is used to reduce the variation of $|\hat{\varphi}(\lambda)|^2$ for $\lambda > \Lambda$, identifying the term not containing information about F or φ.

Finally we substitute $\hat{H}$ instead of H in the expression (4) to obtain an estimator $\hat{h}_{opt}$, for h_{opt}.

3. Strong Consistency of $\hat{\mathbf{H}}$

In this section we establish the strong consistency of the proposed estimator. The prove will be given in the Appendix. According to Chiu (1991), the conditions about $f(x)$ and $w(x)$ are given below. Since the procedure to estimate the optimal bandwidth is based on the empirical characteristic function, the conditions are also given in terms of $\varphi(\lambda)$ and $\varphi_W(\lambda)$:

3.1 *There are positive constants M_1, M_2, K_1 e K_2, such that $M_1|\lambda|^{-K_1} \geq |\varphi(\lambda)|^2 \geq M_2|\lambda|^{-K_2}$ when $|\lambda| \to \infty$. We also assume that $|\varphi(\lambda)|^2 > 0$, for all λ.*

3.2 *The density $f(x)$ has a uniformly bounded derivative and satisfies the relation*
$\int_{|x|>M} f(x)\, dx \leq O\left(M^{-1}\right)$, *when $M \to \infty$.*

3.3 *The kernel $w(x)$ is a symmetric probability density function and satisfies $\int |x|^3 w(x)\, dx < \infty$. The characteristic function of $w(x)$, given by $\varphi_w(\lambda) = \int e^{i\lambda x} w(x)\, dx$, satisfies the following conditions: $\varphi_w(\lambda) = O(\lambda^{-3})$ and $\varphi'_w(\lambda) = O(\lambda^{-3})$, when $\lambda \to \infty$.*

Considering the assumptions cited above, we now cite two results from Chiu (1991):

Proposition 3.1 *Assume the condition 3.1 above, then for any* $\delta > 0$,
$P\left(\Lambda \leq n^{1/K_1+\delta}\right) \to 1$, *when* $n \to \infty$.
This is Lemma 2, from Chiu (1991), page 1900.

Proposition 3.2 *Assume the conditions 3.1 and 3.3 above, then for any* $\delta > 0$, $P\left(\Lambda \geq n^{1/K_2-\delta}\right) \to 1$, *when* $n \to \infty$.
This result implies that $\max_{0 \leq \lambda \leq n} |\hat{\varphi}_d(\lambda)| \leq Mn^{-1/2+\delta}$.
This is Lema 3, from Chiu (1991), page 1901.

Our main results then follow:

Proposition 3.3 Strong Consistency of $\hat{\mathbf{H}}$
Under conditions 3.1 and 3.2, with $K_1 > 6$, $\hat{H}$ *converges almost surely to* H, *when* $n \to \infty$.

Since $\hat{h}$ is a continuous function of $\hat{H}$, an immediate consequence is the following:

Corollary 3.4 Strong Convergency of $\hat{\mathbf{h}}_{\mathbf{opt}}$
Under hypotheses 1 and 2, $\hat{h}_{opt}$ *converges almost surely to* h_{opt}, *when* $n \to \infty$.

4. Computational Results

We made computer simulations in C language with the aim of checking the behavior of the method explained in the previous sections, for some well-known distributions. In order to make consistent comparisons with the literature, we opted to use the same parameters sets that appear in Bowman et al (1998).

For each simulation in the set we conducted the following steps: (i) Sample generation; (ii) Determination of the limit of integration Λ; (iii) $\hat{H}$ computation; (iv) $\hat{h}_{opt}$ computation; (v) $\hat{F}_n$ computation and graphic display; (vi) $ISE(h)$ computation; (vii) Computation of the distribution estimation from the density function estimation.

The value ($c = 3$) suggested in Chiu (1991) was adopted as the choice for the limit of integration Λ. We verified experimentally that the results were not significantly affected when c remained in the proposed interval $(-\log(0.15), -\log(0.05))$. We used the Simpson approximation for the numeric integration. See Mathews (1992). Special care must be taken when dealing with the upper limits of the truncated improper integral, that must be large enough to guarantee the correctness of our results. In the special case of the ISE computations, we simply used rectangular Riemman approximations for the integrals (with 0.01 sized sub-intervals). That procedure, which involved the integration of squares of differences, proved itself to be stable and reliable in our case.

Each simulation in the set consists of 1000 samples of sizes 25, 50, 100 and 200, for the standard $normal(0, 1)$ and $gamma(2, 1)$ distributions. We verified the behavior of the H estimator, the optimal bandwidth, and the $ISE(h)$ for each sample. Besides, we compared our estimates with the estimates obtained via the empirical characteristic function and the integral of the density function. This last one was obtained with the method proposed in Chiu (1991) and Damasceno (2000). The gaussian kernel was always used wherever necessary in the smoothing kernel techniques, except where indicated.

The experiments from Bowman et al (1998) were repeated again here, also with 100 samples for each sample size, in order to maintain the consistency in our comparisons.

The tables 1 and 2, respectively for the $normal(0,1)$ and $gamma(2,1)$, summarize our results regarding the 1000 samples simulations.

The values $\hat{H}$ and $\hat{h}_{opt}$ for the normal distribution are very close to the theoretical value, and their variability diminishes as the sample size increases. From the estimated MISE values, we verified the good performance of our distribution function's kernel estimator method, compared with the empirical function method and the kernel estimator of the density function integral method. In the case of the gamma distribution, we note that the discrepancy between the H estimate and its theoretical value induces a bias in the functional estimate. This happens because we have an obvious boundary problem, with the second derivative of the function $gamma(2,1)$ not defined in zero, jeopardizing the application

Table 1: Simulation Results - Normal Distribution

		$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
H	Theoretical	0.1411	0.1411	0.1411	0.1411	0.1411
	$\hat{E}(\hat{H})$	0.1436	0.1395	0,1426	0.1412	0.1413
	$\hat{\sigma}(\hat{H})$	0.1814	0.0710	0.0484	0.0313	0.0194
	$\hat{E}(\hat{H} - H)^2$	0.03289	0.00504	0.00234	0.00098	0.00038
h_{opt}	Theoretical	0.5436	0.4315	0.3425	0.2718	0.2003
	$\hat{E}(\hat{h}_{opt})$	0.5889	0.4500	0.3479	0.2740	0.2007
	$\hat{\sigma}(\hat{h}_{opt})$	0.1052	0.0593	0.0339	0.0187	0.0089
	$\hat{E}(\hat{h}_{opt} - h_{opt})^2$	0.01311	0.00386	0.00117	0.00035	0.00008
$\widehat{MISE}$	$\hat{F}_n$	0.018543	0.009711	0.004724	0.002454	0.000966
	$\int \hat{f}_n$	0.019269	0.010383	0.005189	0.002772	0.001146
	$\tilde{F}$	0.023131	0.011854	0.005698	0.002861	0.001099

Each entry in this table is the mean of 1000 experiments, for each sample size.
$\hat{F}_n$: Distribution kernel estimator; $\int \hat{f}_n$: Estimator of F by integration of f_n;$\tilde{F}$: Empirical distribution function

of the method. The variability of the estimator diminishes slowly as the sample size increases, the same thing happening with the differences between the estimates and the theoretical values.

Regarding the original Bowman experiment, we present in Figures (1) and (2) the box-plot graphics for the sample sizes relative to the log(ISE). We used the Epanechnikov's kernel, in the same way as the paper mentioned above, in order to maintain the original experiment characteristics. Under these conditions, the numerical experiments suggest a better performance using our proposed estimator.

Table 2: Simulation Results - Gamma Distribution

		$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
H	Theoretical	0.2500	0.2500	0.2500	0.2500	0.2500
	$\hat{E}(\hat{H})$	0.1202	0.1277	0.14324	0.1570	0.1731
	$\hat{\sigma}(\hat{H})$	0.1697	0.0872	0.0693	0.0549	0.0410
	$\hat{E}(\hat{H} - H)^2$	0.04561	0.02255	0.0.01621	0.01166	0.00759
h_{opt}	Theoretical	0.4486	0.35604	0.2826	0.2243	0.1653
	$\hat{E}(\hat{h}_{opt})$	0.6589	0.47807	0.3541	0.2680	0.1889
	$\hat{\sigma}(\hat{h}_{opt})$	0.1549	0.0851	0.0474	0.0276	0.0.0137
	$\hat{E}(\hat{h}_{opt} - h_{opt})^2$	0.06823	0.02211	0.00735	0.00267	0.00075
$\widehat{MISE}$	$\hat{F}_n$	0.025181	0.013025	0.006593	0.003309	0.001386
	$\int \hat{f}_n$	0.093018	0.044311	0.020821	0.010013	0.003834
	$\tilde{F}$	0.029735	0.015122	0.007471	0.003698	0.001510

Each entry in this table is the mean of 1000 experiments, for each sample size

$\hat{F}_n$: Distribution kernel estimator; $\int \hat{f}_n$: Estimator so F by integration of f_n; $\tilde{F}$: Empirical distribution function

5. Aplication

Tajima (1989) proposed a method to test the neutral mutation hypothesis problem (see details in the reference). That method is based on the comparison of two estimators of the parameter $\theta = 4N\mu$ where N is the effective population, μ is the mutation rate per generation. A statistic W is defined as a function of those two estimators. The Tajima test is based on the assumption that, under the null hypothesis, such statistic has a beta distribution.

Atuncar & Silva (2002) observed that for some values of sam-

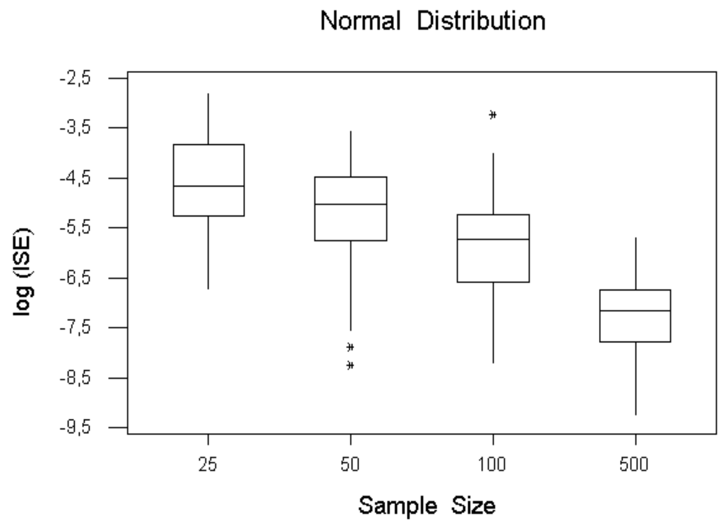

Figure 1: Normal sample Box-plot

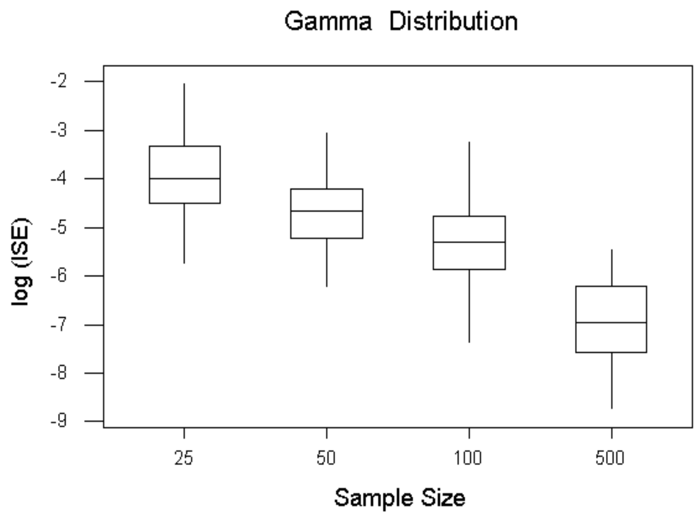

Figure 2: Gamma sample Box-plot

ple sizes,n and θ, the beta distribution is not appropriate to model the statistic W. For each pair of values (n, θ) with $n = 10, 15, ..., 50$ and $\theta = 1, 2, 3, ..., 50$, Silva (2001) simulated 7,000 genealogical trees under the neutral hypothesis and for each tree, he got the corresponding value of W. From this data, the density function of W was estimated by using the kernel estimator and a new critical region was defined by integration of the kernel estimator of the density function. The new critical region improved significatively the test.

Table 3: Estimated critical points, under neutrality hypothesis

		From Silva		Dist. kernel estimator	
N	θ	$\hat{D}_{0,025}$	$\hat{D}_{0,975}$	$\hat{D}_{0,025}$	$\hat{D}_{0,975}$
10	1	-1.614	1.772	-1.623	1.804
	5	-1.795	1.751	-1.751	1.693
20	1	-1.695	1.884	-1.642	1.832
	5	-1.803	1.822	-1.776	1.793
40	1	-1.650	1.975	-1.630	2.069
	5	-1.730	1.873	-1.750	1.832
50	1	-1.659	2.003	-1.618	1.958
	5	-1.740	1.906	-1.703	1.808

In our work, we estimated directly the distribution function of W by using the samples simulated by Silva. Considering the significance level of 5%, the critical points $W(0.025)$ and $W(0.975)$ were estimated,where:

$$W_{0,025} = min\{w : \ F(w) \geq 0,025\} \text{ and}$$

$$W_{0,975} = min\{w : \ F(w) \geq 0,975\}$$

The results are displayed on Table [3]. On the same table we present, for comparison, the results gotten by Silva. We observe that our values are close to the corresponding Silva's values and they are better than those obtained by Tajima.

Acknowledgements

The first author acknowledges the partial support from CAPES Proj. PROCAD 0094/01 - 3. We also acknowledge prof. F.B.Cruz for his help in computational issues related to LaTeX.

Appendix

Proof of Proposition 3.3

From (5) and (6), we have that:

$$\begin{aligned} \pi\hat{H} - \pi H &= \int_0^\Lambda \lambda^2 \left[|\hat{\varphi}(\lambda)|^2 - \frac{1}{n} \right] d\lambda - \int_0^\infty \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda \\ &= \int_0^\Lambda \lambda^2 \left[|\hat{\varphi}(\lambda)|^2 - \frac{1}{n} \right] d\lambda - \int_0^\Lambda \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda + B_1 \, , \end{aligned} \tag{7}$$

where

$$B_1 = - \int_\Lambda^\infty \lambda^2 \, |\varphi(\lambda)|^2 \; d\lambda \tag{8}$$

Defining $\hat{\varphi}_d(\lambda) = \hat{\varphi}(\lambda) - \varphi(\lambda)$, we have:

$$|\hat{\varphi}(\lambda)|^2 = |\hat{\varphi}_d(\lambda) + \varphi(\lambda)|^2 = |\varphi(\lambda)|^2 + |\hat{\varphi}_d(\lambda)|^2 + 2 \, Re \, [\varphi(\lambda) \, \hat{\varphi}_d(- \lambda)]$$

So, the expression(7) becomes:

$$\pi\hat{H} - \pi H = \int_0^{\Lambda} \lambda^2 \left[|\hat{\varphi}_d(\lambda)|^2 - \frac{1}{n} \right] d\lambda + 2\,Re \int_0^{\Lambda} \lambda^2 \varphi(\lambda)\hat{\varphi}_d(-\lambda)\, d\lambda + B_1 \,. \tag{9}$$

Define now

$$B_2 = 2\,Re \int_0^{\Lambda} \lambda^2 \varphi(\lambda)\hat{\varphi}_d(-\lambda)\, d\lambda \,, \tag{10}$$

$$B_3 = \int_0^{\Lambda} \lambda^2 \left[|\hat{\varphi}_d(\lambda)|^2 - \frac{1}{n} \right] d\lambda \,. \tag{11}$$

Convergence of B_1

By proposition 3.1 we have that $P\left(\Lambda \geq n^{1/K_2-\delta}\right) \to 1$, when $n \to \infty$.

Since $\int_{\Lambda}^{\infty} \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda = g_1(\Lambda)$ is a continuous and decreasing function in Λ, we have that $P\left[g_1(\Lambda) \leq g_1(n^{1/K_2-\delta})\right] \to 1$ almost surely. This implies:

$$\int_{\Lambda}^{\infty} \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda \leq \int_{n^{1/K_2-\delta}}^{\infty} \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda \,. \tag{12}$$

By using the condition 3.1, we obtain

$$\begin{aligned} |B_1| \leq \int_{n^{1/K_2-\delta}}^{\infty} \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda \; &\leq \; \int_{n^{1/K_2-\delta}}^{\infty} M_1 \lambda^{-K_1+2} \, d\lambda = A_n \\ &\leq \; \frac{M_1}{-3+K_1} n^{(1/K_2-\delta)(-K_1+3)} \,. \end{aligned}$$

With $K_1 > 3$, we choose δ such that $(1/K_2 - \delta)(-K_1 + 3) < 0$.

Define $A_n = \left| \int_{n^{1/K_2-\delta}}^{\infty} \lambda^2 \, |\varphi(\lambda)|^2 \, d\lambda \right|$. We will prove that $A_n \to 0$ almost surely.

Let ϵ be a positive number. Since $A_n \leq \frac{M_1}{k_1-3} n^{\alpha}$, for some $\alpha < 0$, there exists n_0 such that $A_n < \epsilon \,, \forall\, n > n_0$. Then

$$\sum_{n=1}^{\infty} P(A_n > \epsilon) = \sum_{n=1}^{n_0} P(A_n > \epsilon) < \infty \,.$$

Hence, by the Borel-Cantelli lemma , $A_n \to 0$ almost surely and as a consequence, $B_1 \to 0$ almost surely.

Convergency of B_2

Since $|\varphi(\lambda)| \leq 1$, from proposition 3.2, we have that $|\hat{\varphi}_d(-\lambda)| \leq \max_{0 \leq \lambda \leq n} |\hat{\varphi}_d(\lambda)| \leq M n^{-1/2+\delta}$. So,

$$2\, Re \int_0^{\Lambda} \lambda^2 \varphi(\lambda) \hat{\varphi}_d(-\lambda)\, d\lambda \leq 2 \int_0^{\Lambda} \lambda^2 M n^{-1/2+\delta}\, d\lambda \,.$$

This proves that $B_2 \to 0$ almost surely.

In fact, define $B_n = \int_0^{n^{1/K_1+\delta}} \lambda^2 M\, n^{-1/2+\delta}\, d\lambda$. We take $K_1 > 6$ and choose δ such that $(3/K_1 - 1/2 + 4\delta) < 0$. Then $B_n < n^{\alpha}$, with $\alpha < 0$. So, applying the Borel-Cantelli lemma as before, we prove that $B_2 \to 0$ almost surely.

Convergence of B_3

$$\begin{aligned} |B_3| &= \left| \int_0^{\Lambda} \lambda^2 \left[|\hat{\varphi}_d(\lambda)|^2 - \frac{1}{n} \right] d\lambda \right| \leq \int_0^{\Lambda} \lambda^2 \left| \left[|\hat{\varphi}_d(\lambda)|^2 - \frac{1}{n} \right] \right| d\lambda \leq \\ & \int_0^{\Lambda} \lambda^2 \left| \left[|\hat{\varphi}_d(\lambda)|^2 + \frac{1}{n} \right] \right| d\lambda \leq \int_0^{\Lambda} \lambda^2 \, |\hat{\varphi}_d(\lambda)|^2 \, d\lambda + \int_0^{\Lambda} \lambda^2 \frac{1}{n}\, d\lambda \end{aligned}$$

By Proposition 3.1, we have that for $\delta > 0$, $P\left(\Lambda \leq n^{1/K_1+\delta}\right) \to 1$, when $n \to \infty$. This proves that $B_3 \to 0$ almost surely.

We have already verified that B_1, B_2 and B_3 converge almost surely to zero. Then, $\hat{H}$ converges almost surely to H.

References

Altman, N. & Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *J. Statist. Plan. Inf.* **46,** 195–214.

Atuncar, G. S.; Silva, R. C. (2002). Evaluation and Improvement of Tajima's Test. *To be submitted.*

Bessegato, L. F. (2002). *Escolha do parâmetro de suavidade na estimativa da função de distribuição.* MsC Dissertation. Departament of Statistics UFMG.

Bowman, A. W. (1984). An alternative method of cross validation for the smoothing of density estimates. *Biometrika* **71,** 353–360.

Bowman, A. W.; Hall, P.; Prvan, T. (1998). Bandwidth selection for the smoothing of distribution functions. *Biometrika* **85,** 799–808.

Brillinger, D. R. (1981). *Time series data analysis and theory.* New York: Holt, Rinehart and Winston.

Chiu, S. T. (1991) Bandwidth selection for kernel density estimation. *The Annals of Statistics.* **33,** 1883–1905.

Chung, K. L. (1974). *A course in probability theory, second edition.* San Diego: Academic Press.

Damasceno, E. C. (2000). *Escolha do parâmetro de suavidade em estimação funcional.* MsC Dissertation. Departament of Statistics UFMG.

Falk, M. (1983). Relative efficiency and deficiency of kernel type estimators of smooth distribution functions. *Statist. Neer.* **37,** 73–83.

Jones, M. C. (1990). The performance of kernel density functions in kernel distribution function estimation. *Statistics and Probability Letters* **9,** 129–132.

Mathews, J. H. (1992). *Numerical methods for Mathematics, Science and Engineering.* Englewood Cliffs: Prentice Hall.

Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory Prob. Applic.* **9,** 497–500.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics.* **9,** 65–78.

Rudin, W. (1987) *Real and complex analysis, third edition.* New York: McGraw-Hill.

Sarda, P. (1993). Smoothing parameter selection for smooth distribution functions. *J. Statist. Plan. Inf.* **35,** 65–75.

Silva R. C. (2001). *Avaliação e melhoramento do Teste de Tajima para neutralidade de mutações.* MsC Dissertation. Departament of Statistics UFMG.

TAJIMA, F. (1989). Statistical method for testing the neutral mutation hyphotesis by DNA polymorphism. *Genetics.* **123,** 585–595.

WINTER, B. B. (1973). Strong uniform consistency of integrals of density estimators. *Canad. J. Statist.* **1,** 247–253.

YAMATO, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.* **15,** 69-78.