
GAMMA KERNEL ESTIMATION OF THE DENSITY DERIVATIVE ON THE POSITIVE SEMI-AXIS BY DEPENDENT DATA

Author: L.A. MARKOVICH
– Institute of Control Sciences, Russian Academy of Sciences,
Moscow, Russia, and
Institute for Information Transmission Problems, Moscow,
Bolshoy Karetny per. 19, build. 1, Moscow 127051, Russia
kimo1@mail.ru

Received: December 2014 Revised: January 2015 Accepted: February 2015

Abstract:

- We estimate the derivative of a probability density function defined on $[0, \infty)$. For this purpose, we choose the class of kernel estimators with asymmetric gamma kernel functions. The use of gamma kernels is fruitful due to the fact that they are non-negative, change their shape depending on the position on the semi-axis and possess good boundary properties for a wide class of densities. We find an optimal bandwidth of the kernel as a minimum of the mean integrated squared error by dependent data with strong mixing. This bandwidth differs from that proposed for the gamma kernel density estimation. To this end, we derive the covariance of derivatives of the density and deduce its upper bound. Finally, the obtained results are applied to the case of a first-order autoregressive process with strong mixing. The accuracy of the estimates is checked by a simulation study. The comparison of the proposed estimates based on independent and dependent data is provided.

Key-Words:

- *density derivative; dependent data; gamma kernel; nonparametric estimation.*

AMS Subject Classification:

- 60G35, 60A05.

1. INTRODUCTION

Kernel density estimation is a non-parametric method to estimate a probability density function (pdf) $f(x)$. It was originally studied in [20], [22] for symmetric kernels and univariate independent identically distributed (i.i.d.) data. When the support of the underlying pdf is unbounded, this approach performs well. If the pdf has a support on $[0, \infty)$, the use of classical estimation methods with symmetric kernels yield a large bias on the zero boundary and leads to a bad quality of the estimates [30]. This is due to the fact that symmetric kernel estimators assign nonzero weight at the interval $(-\infty, 0]$. There are several methods to reduce the boundary bias effect, for example, the data reflection [25], boundary kernels [19], the hybrid method [14], the local linear estimator [18], [17] among others. Another approach is to use asymmetric kernels. In case of univariate nonnegative i.i.d. random variables (r.v.s), the pdf estimators with gamma kernels were proposed in [8]. In [5] the gamma-kernel estimator was developed for univariate dependent data. The gamma kernel is nonnegative and it changes its shape depending on the position on the semi-axis. Estimators constructed with gamma kernels have no boundary bias if $f''(0) = 0$ holds, i.e. when the underlying density $f(x)$ has a shoulder at $x = 0$ (see formula (4.3) in [31]). This shoulder property is fulfilled particularly for a wide exponential class of pdfs which satisfy important integral condition

$$(1.1) \quad \int_0^{\infty} x^{-1/2} f(x) dx < \infty$$

assumed in [8]. In [31] the half normal and standard exponential pdfs are considered as examples such that the boundary kernel $K_c(t)$ (p. 553 in [31]) gives the better estimate than the gamma-kernel estimator considered in [8]. At the same time, the exponential distribution does not satisfy both the shoulder condition and the condition (1.1). The half normal density satisfies the shoulder condition, but it does not satisfy (1.1). Since (1.1) is not valid for the latter pdfs, such comparison is not appropriate.

Alternative asymmetrical kernel estimators like inverse Gaussian and reciprocal inverse Gaussian estimators were studied in [24]. The comparison of these asymmetric kernels with the gamma kernel is given in [6].

Along with the density estimation it is often necessary to estimate the derivative of a pdf. Derivative estimation is important in the exploration of structures in curves, comparison of regression curves, analysis of human growth data, mean shift clustering or hypothesis testing. The estimation of the density derivative is required to estimate the logarithmic derivative of the density function. The latter has a practical importance in finance, actuary mathematics, climatology and signal processing. However, the problem of the density derivative estimation has received less attention. It is due to a significant increasing complexity of

calculations, especially for the multivariate case. The boundary bias problem for the multivariate pdf becomes more solid [4]. The pioneering papers devoted to univariate symmetrical kernel density derivative estimation are [7], [26].

The paper does not focus on the boundary performance but on finding of the optimal bandwidth that is appropriate for the pdf derivative estimation in case of dependent data satisfying a strong mixing condition. In [30] an optimal mean integrated squared error (MISE) of the kernel estimate of the first derivative of order $n^{-\frac{4}{7}}$ was indicated. This corresponds to the optimal bandwidth of order $n^{-\frac{1}{7}}$ for symmetrical kernels. The estimation of the univariate density derivative using a gamma kernel estimator by independent data was proposed in [11], [12]. This allows us to achieve the optimal MISE of the same order $n^{-4/7}$ with a bandwidth of order $n^{-\frac{2}{7}}$.

1.1. Contributions of this paper

It is shown that in the case of dependent data, assuming strong mixing, we can estimate the derivative of the pdf using the same technique that has been applied for independent data in [11]. Lemma 2.1, Section 2.1 contains the upper bound of the covariance. The mathematical technic applied for the derivative estimation is similar to one applied for the pdf. However, formulas became much more complicated, particularly because one has to deal with the special Digamma function that includes the bandwidth b . Thus, one has to pick out the order by b from complicated expressions containing logarithms and the special function. In Section 2.2 we find the optimal bandwidth $b \sim n^{-2/7}$ which is different from the optimal bandwidth $b_2^* \sim n^{-2/5}$ proposed for the pdf estimation (see [8], p.476).

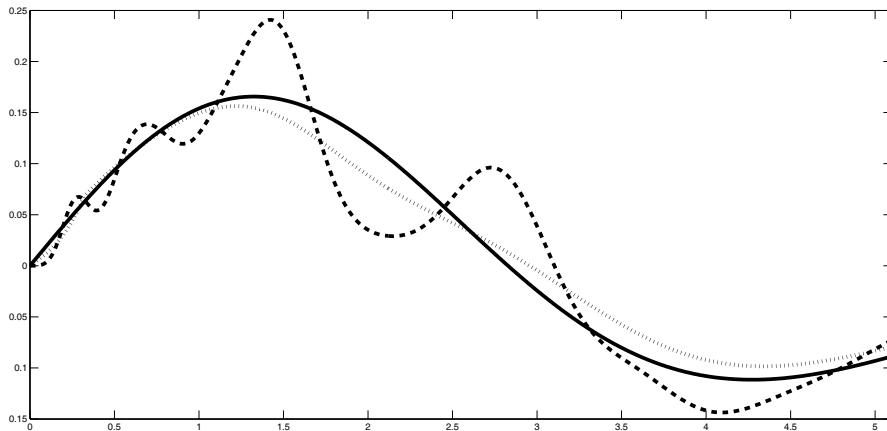


Figure 1: Nonparametric gamma-kernel estimation of Maxwell density derivative function for sample size $n=2000$. The pdf derivative (solid line), the estimate with b (dotted gray line), the estimate with b_2^* (dashed line).

In Fig. 1 it is shown that the use of b_2^* to estimate the pdf derivative leads to a bad quality (for simplicity the i.i.d. data were taken). We prove that the optimal *MISE* of the pdf derivative has the same rate of convergence to the true pdf derivative as for the independent case, namely $O(n^{-4/7})$. We show in Section 2.3 that for the strong mixing autoregressive process of the first order (AR(1)) all results are valid without additional conditions. In Section 3 a simulation study for i.i.d. and dependent samples is performed. The flexibility of the gamma kernel allows us to fit accurately the multi-modal pdf derivatives.

1.2. Practical motivation

In practice it is often necessary to deal with sequences of observations that are derived from stationary processes satisfying the strong mixing condition. As an example of such processes one can take autoregressive processes like in Section 2.3. Along with the evaluation of the density function and its derivative by dependent samples, the estimation of the logarithmic derivative of the density is an actual problem. The logarithmic pdf derivative is the ratio of the derivative of the pdf to the pdf itself. The pdf derivative estimation is necessary for an optimal filtering in the signal processing and control of nonlinear processes where only the exponential pdf class is used, [10]. Moreover, the pdf derivative gives information about the slope of the pdf curve, its local extremes, significant features in data and it is useful in regression analysis [9]. The pdf derivative also plays a key role in clustering via mode seeking [23].

1.3. Theoretical background

Let $\{X_i; i = 1, 2, \dots\}$ be a strongly stationary sequence with an unknown probability density function $f(x)$, which is defined on $x \in [0, \infty)$. We assume that the sequence $\{X_i\}$ is α -mixing with coefficient

$$\alpha(i) = \sup_k \sup_{\substack{A \in \mathcal{F}_k^k(X) \\ B \in \mathcal{F}_{k+i}^\infty(X)}} |P(A \cap B) - P(A)P(B)|.$$

Here, $\mathcal{F}_i^k(X)$ is the σ -field of events generated by $\{X_j, i \leq j \leq k\}$ and $\alpha(i) \rightarrow 0$ as $i \rightarrow \infty$. For these sequences we will use a notation $\{X_j\}_{j \geq 1} \in \mathcal{S}(\alpha)$. Let $f_i(x, y)$ be a joint density of X_1 and X_{1+i} , $i = 1, 2, \dots$

Our objective is to estimate the derivative $f'(x)$ by a known sequence of observations $\{X_i\}$. We use the non-symmetric gamma kernel estimator that was

defined in [8] by the formula

$$(1.2) \quad \hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^n K_{\rho_b(x),b}(X_i).$$

Here

$$(1.3) \quad K_{\rho_b(x),b}(t) = \frac{t^{\rho_b(x)-1} \exp(-t/b)}{b^{\rho_b(x)} \Gamma(\rho_b(x))}$$

is the kernel function, b is a smoothing parameter (bandwidth) such that $b \rightarrow 0$ as $n \rightarrow \infty$, $\Gamma(\cdot)$ is a standard gamma function and

$$(1.4) \quad \rho_b(x) = \begin{cases} \rho_1(x) = x/b, & \text{if } x \geq 2b, \\ \rho_2(x) = (x/(2b))^2 + 1, & \text{if } x \in [0, 2b]. \end{cases}$$

The use of gamma kernels is due to the fact that they are nonnegative, change their shape depending on the position on the semi-axis and possess better boundary bias than symmetrical kernels. The boundary bias becomes larger for multivariate densities. Hence, to overcome this problem the gamma kernels were applied in [4]. Earlier the gamma kernels were only used for the density estimation of identically distributed sequences in [4], [8] and for stationary sequences in [5].

To our best knowledge, the gamma kernels have been applied to the density derivative estimation at first time in [11]. In this paper the derivative $f'(x)$ was estimated under the assumption that $\{X_1, X_2, \dots, X_n\}$ are i.i.d. random variables as derivative of (1.2). This implies that

$$(1.5) \quad \hat{f}'_n(x) = \frac{1}{n} \sum_{i=1}^n K'_{\rho_b(x),b}(X_i)$$

holds, where

$$(1.6) \quad K'_{\rho_b(x),b}(t) = \begin{cases} K'_{\rho_1(x),b}(t) = \frac{1}{b} K_{\rho_1(x),b}(t) L_1(t), & \text{if } x \geq 2b, \\ K'_{\rho_2(x),b}(t) = \frac{x}{2b^2} K_{\rho_2(x),b}(t) L_2(t), & \text{if } x \in [0, 2b], \end{cases}$$

is the derivative of $K_{\rho(x),b}(t)$,

$$(1.7) \quad \begin{aligned} L_1(t) &= L_1(t, x) = \ln t - \ln b - \Psi(\rho_1(x)), \\ L_2(t) &= L_2(t, x) = \ln t - \ln b - \Psi(\rho_2(x)), \end{aligned}$$

Here $\Psi(x)$ denotes the Digamma function (the logarithmic derivative of the gamma function). The unknown smoothing parameter b was obtained as the minimum of the mean integrated squared error (*MISE*) which, as known, is equal to

$$MISE(\hat{f}'_n(x)) = \mathbf{E} \int_0^{\infty} (f'(x) - \hat{f}'_n(x))^2 dx.$$

Remark 1.1. The latter integral can be splitted into two integrals \int_0^{2b} and \int_{2b}^∞ . In the case when $x \geq 2b$ the integral \int_0^{2b} tends to zero when $b \rightarrow 0$. Hence, we omit the consideration of this integral in contrast to [31]. The first integral has the same order by b as the second one, thus it cannot affect on the selection of the optimal bandwidth.

The following theorem has been proved.

Theorem 1.1 ([11]). *If $b \rightarrow 0$ and $nb^{3/2} \rightarrow \infty$ as $n \rightarrow \infty$, the integrals*

$$\int_0^\infty P(x)dx, \quad \int_0^\infty x^{-3/2}f(x)dx$$

are finite and $\int_0^\infty P(x)dx \neq 0$, then the leading term of a MISE expansion of the density derivative estimate $\hat{f}'(x)$ is equal to

$$(1.8) \quad \begin{aligned} MISE(\hat{f}'_n(x)) &= \frac{b^2}{16} \int_0^\infty P(x)dx \\ &+ \int_0^\infty \frac{n^{-1}b^{-3/2}x^{-3/2}}{4\sqrt{\pi}} \left(f(x) + b \left(\frac{f(x)}{2x} - \frac{f'(x)}{2} \right) \right) dx \\ &+ o(b^2 + n^{-1}(b^{-3/2})), \end{aligned}$$

where

$$P(x) = \left(\frac{f(x)}{3x^2} + f''(x) \right)^2.$$

Taking the derivative of (1.8) in b leads to equation

$$(1.9) \quad \begin{aligned} \frac{b}{8} \int_0^\infty \left(\frac{f(x)}{3x^2} + f''(x) \right)^2 dx - \frac{3n^{-1}b^{-\frac{5}{2}}}{8\sqrt{\pi}} \int_0^\infty x^{-\frac{3}{2}}f(x)dx \\ + \frac{n^{-1}b^{-\frac{3}{2}}}{16\sqrt{\pi}} \int_0^\infty x^{-\frac{3}{2}} \left(\frac{f(x)}{x} - f'(x) \right) dx = 0. \end{aligned}$$

Neglecting the term with $b^{-3/2}$ as compared to the term $b^{-5/2}$, the equation becomes simpler and its solution is equal to the optimal global bandwidth

$$(1.10) \quad b_0 = \left(\frac{3 \int_0^\infty x^{-3/2}f(x)dx}{\sqrt{\pi} \int_0^\infty \left(\frac{f(x)}{3x^2} + f''(x) \right)^2 dx} \right)^{2/7} n^{-2/7}.$$

The substitution of b_0 into (1.8) yields an optimal MISE with the rate of convergence $O(n^{-\frac{4}{7}})$. The unknown density and its second derivative in (1.10) were estimated by the rule of thumb method [12].

In [30], p.49, it was indicated an optimal *MISE* of the first derivative kernel estimate $n^{-\frac{4}{7}}$ with the bandwidth of order $n^{-\frac{1}{7}}$ for symmetrical kernels. Nevertheless, our procedure achieves the same order $n^{-4/7}$ with a bandwidth of order $n^{-\frac{2}{7}}$. Moreover, our advantage concerns the reduction of the bias of the density derivative at the zero boundary by means of asymmetric kernels. Gamma kernels allow us to avoid boundary transformations which is especially important for multivariate cases.

Further results presented in Section 2.2 will be based on Theorem 1.1.

2. MAIN RESULTS

2.1. Estimation of the density derivative by dependent data

Here, we estimate the density derivative by means of the kernel estimator (1.5) by dependent data. Thus, its mean squared error is determined as

$$(2.1) \quad MSE(\hat{f}'_n(x)) = (Bias(\hat{f}'_n(x)))^2 + var(\hat{f}'_n(x)),$$

where, due to the stationarity of the process X_i , the variance is given by

$$\begin{aligned} var(\hat{f}'_n(x)) &= var\left(\frac{1}{n} \sum_{i=1}^n K'_b(X_i)\right) = \frac{1}{n^2} var\left(\sum_{i=1}^n K'_b(X_i)\right) \\ &= \frac{1}{n^2} \left(\sum_{i=1}^n var(K'_b(X_i)) + 2 \sum_{1 \leq i < j \leq n} cov(K'_b(X_i), K'_b(X_j)) \right) \\ &= \frac{1}{n} var(K'_b(X_i)) + \frac{2}{n^2} \sum_{1 \leq i < j \leq n} cov(K'_b(X_i), K'_b(X_j)) \\ &= \frac{1}{n} var(K'_b(X_i)) + \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) cov(K'_b(X_1), K'_b(X_{1+i})) \\ &= V(x) + C(x). \end{aligned}$$

For simplicity we use here and further the notation $K'_{\rho_b(x),b}(t) = K'_b(t)$ in (1.5).

Thus, (2.1) can be written as

$$(2.2) \quad MSE(\hat{f}'_n(x)) = B(x)^2 + V(x) + C(x),$$

where

$$B(x) = Bias(\hat{f}'_n(x)).$$

The bias of the estimate does not change, but the variance contains a covariance. The next lemma is devoted to its finding.

Lemma 2.1. *Let*

1. $\{X_j\}_{j \geq 1} \in \mathcal{S}(\alpha)$ and $\int_1^\infty \alpha(\tau)^v d\tau < \infty$, $0 < v < 1$ hold,
2. $f(x)$ be a twice continuously differentiable function,
3. $b \rightarrow 0$ and $nb^{-(v+1)/2} \rightarrow \infty$ as $n \rightarrow \infty$.

Then the covariance $C(x)$ is bounded by

$$\begin{aligned}
 |C(x)| &= \left| \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \text{cov}(K'_{\rho_b(x),b}(X_1), K'_{\rho_b(x),b}(X_{1+i})) \right| \\
 (2.3) \quad &\leq \left(2^{-\frac{v+3}{2}} \pi^{\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+1}{2}}}{n} \left(b^2 C_2(v, x) + b C_1(v, x) + C_3(v, x) \right)^{1-v} \right. \\
 &\quad \left. + o(b^2) \right) \int_1^\infty \alpha(\tau)^v d\tau,
 \end{aligned}$$

where $K'_{\rho_b(x)}$ is defined by (1.6) and $C_1(v, x)$, $C_2(v, x)$ and $C_3(v, x)$ are given by (4.8).

A similar lemma was proved in [10] for symmetrical kernels and not strictly positive x .

2.2. Mean integrated squared error of $\hat{f}'_n(x)$

Using the upper bound (2.3) we can obtain the upper bound of the MISE and find the expression of the optimal bandwidth b as the minimum of the latter.

Theorem 2.1. *If the conditions of Theorem 1.1 and Lemma 2.1 hold, then the MISE expansion for the estimate $\hat{f}'_n(x)$ of the density derivative is equal to*

$$\begin{aligned}
 (2.4) \quad \text{MISE}(f'(x)) &\leq \int_0^\infty \frac{n^{-1} b^{-\frac{3}{2}} x^{-\frac{3}{2}}}{4\sqrt{\pi}} \left(f(x) + \frac{b}{2} \left(\frac{f(x)}{x} - f'(x) \right) \right) dx \\
 &\quad + \int_0^\infty \left(2^{-\frac{v+3}{2}} \pi^{\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+1}{2}}}{n} C_3(v, x)^{1-v} \right) \int_1^\infty \alpha(\tau)^v d\tau dx \\
 &\quad + \frac{b^2}{16} \int_0^\infty P(x) dx + o(b^2 + n^{-1}(b^{-\frac{3}{2}})).
 \end{aligned}$$

and the optimal bandwidth is $b_{opt} = o(n^{-2/7})$ and the $\text{MISE}_{opt} = O(n^{-4/7})$.

Remark 2.1. It is evident from the formula (2.4) that the term responsible for the covariance has the order $\frac{b^{-\frac{v+1}{2}}}{n}$, $0 < v < 1$. Thus, it does not influence the order of MISE irrespective of the mixing coefficient $\alpha(\tau)$.

The proof is given in Appendix 4.

2.3. Example of a strong mixing process

We use the first-order autoregressive process as an example of a process that satisfies Theorem 1.1. X_i determines a first-order autoregressive (AR(1)) process with the innovation r.v. ϵ_0 and the autoregressive parameter $\rho \in (-1, 1)$ if

$$(2.5) \quad X_i = \rho X_{i-1} + \epsilon_i, \quad i = \dots - 1, 0, 1, \dots,$$

holds and ϵ_i is a sequence of i.i.d. r.v.s. Let AR(1) process (2.5) be strong mixing with mixing numbers $\alpha(\tau)$, $\tau = 1, 2, \dots$

$$(2.6) \quad \alpha(\tau) \leq \tilde{\alpha}(\tau) \equiv \begin{cases} 2(C+1)E|X_i|^\nu |\rho|^\tau, & \text{if } \tau \geq \tau_0, \\ 1, & \text{if } 1 \leq \tau < \tau_0, \end{cases}$$

where $\nu = \min\{p, q, 1\}$ and $p > 0, q > 0, C > 0, \tau_0 > 0$ hold. In [2] it was proved that with some conditions AR(1) is a strongly mixing process.

In Appendix 4 we prove the following lemma.

Lemma 2.2. *Under the conditions (2.6) the AR(1) process (2.5) satisfies Lemma 2.1 and Theorem 2.1.*

3. SIMULATION RESULTS

To investigate the performance of the gamma-kernel estimator we select the following positive defined pdfs: the Maxwell ($\sigma = 2$), the Weibull ($a = 1, b = 4$) and the Gamma ($\alpha = 2.43, \beta = 1$) pdf,

$$f_M(x) = \frac{\sqrt{2}x^2 \exp(-x^2/2\sigma^2)}{\sigma^3\sqrt{\pi}},$$

$$f_W(x) = sx^{s-1} \exp(-x^s),$$

$$f_G(x) = \frac{x^{\alpha-1} \exp(-x/\beta)}{\beta^\alpha \Gamma(\alpha)}.$$

Their derivatives

$$(3.1) \quad \begin{aligned} f'_M(x) &= -\frac{\sqrt{2}x \exp(-x^2/2\sigma^2)(x^2 - 2\sigma^2)}{\sigma^5\sqrt{\pi}}, \\ f'_W(x) &= -sx^{s-2} \exp(-x^s)(sx^s - s + 1), \\ f'_G(x) &= \frac{x^{\alpha-2} \exp(-x/\beta)(\beta + x - \alpha\beta)}{\beta^{\alpha+1}\Gamma(\alpha)} \end{aligned}$$

are to be estimated. The Weibull and the Gamma pdfs are frequently used in a wide range of applications in engineering, signal processing, medical research, quality control, actuarial science and climatology among others. For example, most total insurance claim distributions are shaped like gamma pdfs [13]. The gamma distribution is also used to model rainfalls [1]. Gamma class pdfs, like Erlang and χ^2 pdfs are widely used in modeling insurance portfolios [15].

We generate Maxwell, Weibull and Gamma i.i.d. samples with sample sizes $n \in \{100, 500, 1000, 2000\}$ using standard Matlab generators. To get the dependent data we generate Markov chains with the same stationary distributions using the Metropolis–Hastings algorithm [16]. Due to the existence of the probability of rejecting a move from the previous point to the next one, the variance of such Markov sequence $\{X_t\}$ is corrupted by the function of the latter rejecting probability (see [27], Theorem 3.1). The Metropolis–Hastings Markov chains [16] are geometrically ergodic for the underlying light-tailed distributions. Hence, they satisfy the strong mixing condition [21].

The gamma kernel estimates (1.2) with the optimal bandwidth (1.10) for the derivatives (3.1) can be seen in Figures 2–4. The optimal bandwidth (1.10) is counted for every replication of the simulation using the rule of thumb method, where as a reference density we take the gamma pdf.

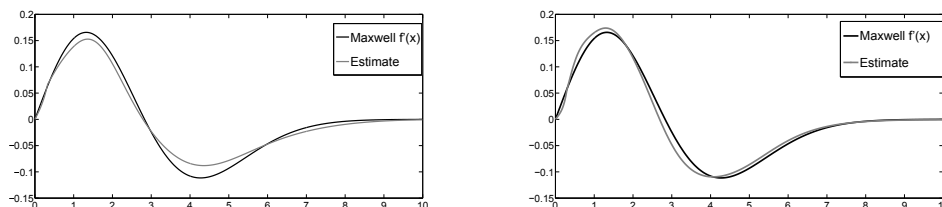


Figure 2: Estimates of the Maxwell pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_M(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

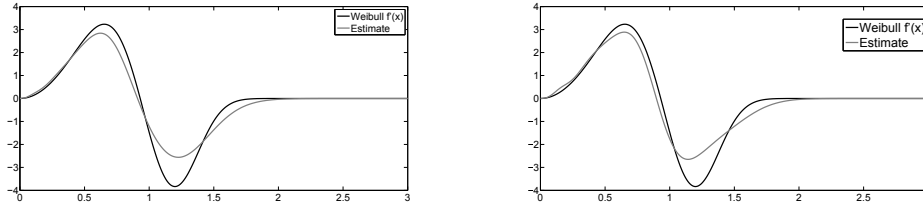


Figure 3: Estimates of the Weibull pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_W(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

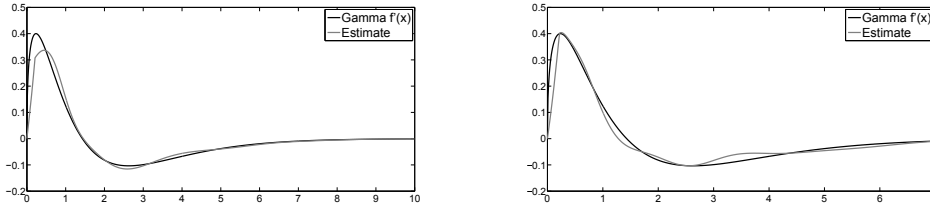


Figure 4: Estimates of the Gamma pdf derivative by i.i.d. data (left) and by dependent data (right): the $f'_G(x)$ (black line), gamma kernel estimate from the rule of thumb (grey line) for the sample size $n = 2000$.

The estimation error of the pdf derivative is calculated by the following formula

$$m = \int_0^{\infty} (f'(x) - \hat{f}'(x))^2 dx,$$

where $f'(x)$ is a true derivative and $\hat{f}'(x)$ is its estimate. Values of m 's averaged over 500 simulated samples and the standard deviations for the underlying distributions are given in Table 1 for i.i.d. r.v.s and in Table 2 for dependent data.

Table 1: Mean errors m and standard deviations for i.i.d. r.v.s.

Distribution	n			
	100	500	1000	2000
Gamma	0.032792 (0.011967)	0.015208 (0.0044094)	0.010675 (0.0027815)	0.0074668 (0.0016452)
Weibull	2.0056 (0.52931)	1.1987 (0.25172)	0.9157 (0.18333)	0.69155 (0.12178)
Maxwell	0.0077597 (0.0033915)	0.0035692 (0.0015351)	0.0028675 (0.00099263)	0.0020923 (0.00068739)

Table 2: Mean errors m and standard deviations for strong mixed r.v.s.

Distribution	n			
	100	500	1000	2000
Gamma	0.039226 (0.015824)	0.018124 (0.006055)	0.01252 (0.0038485)	0.0086675 (0.0023361)
Weibull	2.2052 (1.1585)	1.3009 (0.5957)	0.97509 (0.41041)	0.75382 (0.28755)
Maxwell	0.0077694 (0.006793)	0.0039277 (0.0028336)	0.002878 (0.0020021)	0.0027313 (0.0016573)

As expected, the mean error and the standard deviation decrease when the sample size rises, and this holds both for i.i.d. and the dependent case. The performance of the gamma kernel changes when dependence is introduced, but the results in both tables are close. The mean errors are very close due to the fact the bandwidth parameter is selected to minimize this error. However, the standard deviations for the dependent data are higher than for the i.i.d. r.v.s. For example, for the sample size of 500 the mean errors and the standard deviations for the Maxwell pdf for the i.i.d. r.v.s are 0.0035692 (0.0015351) and for dependent r.v.s 0.0039277 (0.0028336). They differ due to the contribution of the Metropolis–Hastings rejecting probability. This difference is less pronounced for larger sample sizes.

The Metropolis–Hastings algorithm gives opportunity to generate AR processes with known pdfs. As a consequence we know their derivatives and can find mean errors and standard deviations of the gamma-kernel density derivatives estimates for the dependent data. In the case when we consider the noise distribution $\{\epsilon\}$ of the AR model (2.5) and the autoregressive parameter ρ that influences on the dependence rate (2.6), we cannot indicate in general the true pdf of the process. Hence, we consider the histogram based on 200000 observations as a true pdf. As the noise distribution $\{\epsilon\}$ let us take the Gamma distribution ($\alpha = 1.5, \beta = 1$) and the Maxwell distribution ($\sigma = 1$). In [5] it was proved that, as in the i.i.d. case, the gamma-kernel estimator of the pdf achieves the same optimal rate of convergence in terms of the mean integrated squared error as for strongly mixed r.v.s. For the various parameters $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ the gamma estimates for the densities of the AR models are given in Figures 5–6.

Since the gamma-kernel estimators perform good for the various dependence rates it is also true for the gamma-kernel pdf derivative estimators, but the bandwidth parameter must be selected differently.

Hence, this findings confirms the fact that the covariance term (2.3) of the pdf derivative is negligible in comparison with its variance and implies that one

can use the same optimal bandwidth (1.10), both for independent and strongly mixed dependent data.

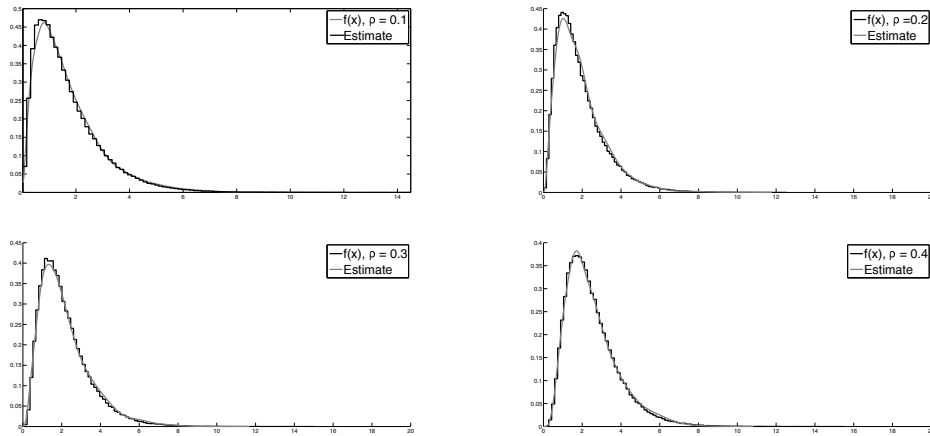


Figure 5: Gamma-kernel estimates of the pdf of the AR model with the Gamma noise and $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ for the sample size $n = 2000$.

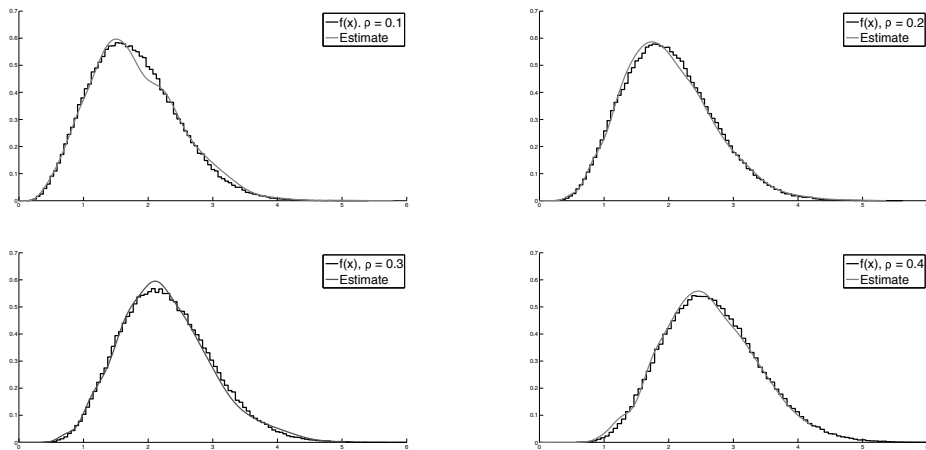


Figure 6: Gamma-kernel estimates of the pdf of the AR model with the Maxwell noise and $\rho \in \{0.1, 0.2, 0.3, 0.4\}$ for the sample size $n = 2000$.

4. APPENDIX

Proof of Lemma 2.1: Taking an integral from (2.2) we get

$$(4.1) \quad MISE(\hat{f}'(x)) = \int_0^\infty (B(x)^2 + V(x) + C(x))dx,$$

where

$$(4.2) \quad C(x) = \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \text{cov}(K'_b(X_1), K'_b(X_{1+i})).$$

To evaluate the covariance we shall apply Davydov's inequality

$$(4.3) \quad |\text{cov}(K'_b(X_1), K'_b(X_{1+i}))| \leq 2\pi\alpha(i)^{1/r} \|K'_b(X_1)\|_q \|K'_b(X_{1+i})\|_p,$$

where $p^{-1} + q^{-1} + r^{-1} = 1$, $1 \leq p, q, r \leq \infty$, [3].

The latter norm for the case $x \geq 2b$ is determined by

$$(4.4) \quad \begin{aligned} \|K'_b(X_1)\|_q &= \left(\int \left(\frac{1}{b} K(y) L_1(y) \right)^q f(y) dy \right)^{1/q} \\ &= \frac{1}{b} \left(\mathbf{E} \left(K(\xi_1)^{q-1} L_1(\xi_1)^q f(\xi_1) \right) \right)^{1/q}, \end{aligned}$$

where $L_1(t)$ is introduced in (1.7). The kernel $K(\xi_1)$ was used in (4.4) as a density function and ξ_1 is a $Gamma(\rho_1(x), b)$ random variable.

In the case $x \in [0, 2b)$, similarly we have

$$(4.5) \quad \begin{aligned} \|K'_b(X_1)\|_q &= \left(\int \left(\frac{x}{2b^2} K(y) L_2(y) \right)^q f(y) dy \right)^{1/q} \\ &= \frac{x}{2b^2} \left(\mathbf{E} \left(K(\xi_2)^{q-1} L_2(\xi_2)^q f(\xi_2) \right) \right)^{1/q}, \end{aligned}$$

where $L_2(t)$ is determined by (1.7), and ξ_2 is a $Gamma(\rho_2(x), b)$ random variable. Expressions (4.4) and (4.5) are constructed similarly, thus to a certain point, we will not make differences between them.

By the standard theory of the gamma distribution it is known that $\mu = \mathbf{E}(\xi) = \rho_b(x)b$ and the variance is given by $\text{var}(\xi) = \rho_b(x)b^2$. For simplicity, we further use the notation ρ instead of $\rho_b(x)$ defined in (1.4).

The Taylor expansion of both mathematical expectations in (4.4), (4.5) in the neighborhood of μ is represented by

$$\begin{aligned} \mathbf{E} \left(K(\xi)^{q-1} L(\xi)^q f(\xi) \right) &= K(\mu)^{q-1} L(\mu)^q f(\mu) + (K(\xi)^{q-1} L(\xi)^q f(\xi))' |_{\xi=\mu} \mathbf{E}(\xi - \mu) \\ &\quad + (K(\xi)^{q-1} L(\xi)^q f(\xi))'' |_{\xi=\mu} \frac{\mathbf{E}(\xi - \mu)^2}{2} + o(\mathbf{E}(\xi - \mu)^2). \end{aligned}$$

In the case when $x \geq 2b$, $\mu = \rho b = x$, $var(\xi) = \rho b^2 = xb$, we get

$$\begin{aligned} & \mathbb{E} (K(\xi)^{q-1} L(\xi)^q f(\xi)) = \\ &= \frac{K(x)^{q-1}}{b} \left(qL(x)^{q+1} f'(x) - L(x)^q f(x) L'(x) \right. \\ & \quad - L(x)^{q+1} f'(x) + bL(x)^q f''(x) + q^2 L(x)^q f(x) L'(x) + bq^2 L(x)^{q-2} (L'(x))^2 f(x) \\ & \quad \left. + 2bqL(x)^{q-1} L'(x) f'(x) + bqL(x)^{q-1} f(x) L''(x) - bqL(x)^{q-2} (L'(x))^2 \right) \\ & \quad + \frac{K(x)^{q-1} L(x)(q-1)}{b^2} \left((q-1)f(x)L(x)^{q+1} + bL(x)^q f'(x) \right. \\ & \quad \left. + bqL(x)^{q-1} f(x)L'(x) \right) + o(b^2). \end{aligned}$$

Using Stirling’s formula

$$\Gamma(z) = \sqrt{\frac{2\pi}{z}} \left(\frac{z}{e}\right)^z \left(1 + O\left(\frac{1}{z}\right)\right),$$

we can rewrite the kernel function as

$$K(t) = \frac{t^{\rho-1} \exp(-t/b)}{b^\rho \Gamma(\rho)} = \frac{t^{\rho-1} \exp(-t/b) \exp(\rho)}{b^\rho \sqrt{2\pi} \rho^{\rho-\frac{1}{2}} (1 + O(1/\rho))}.$$

Taking $\rho = \rho_1(x)$ according to (1.4), $t = x$, it holds

$$K(\rho_1(x)b) = \frac{1}{\sqrt{2\pi}} \frac{x^{x/b-1} \exp((x-x)/b)}{b^{\frac{x}{b}} \frac{x}{b}^{-\frac{1}{2}} (1 + O(b/x))} = \frac{x^{-\frac{1}{2}} b^{-\frac{1}{2}}}{\sqrt{2\pi} (1 + O(b/x))}.$$

Hence, its upper bound is given by

$$(4.6) \quad K(x) \leq \frac{1}{\sqrt{2\pi x b}}.$$

Next, using the property of the Digamma function $\Psi(x) = \ln(x) - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} + O(1/x^6)$, the first equation in (1.7) can be rewritten as

$$(4.7) \quad L_1(\rho_1 b) = \ln(\rho_1 b) - \ln(b) - \Psi(\rho_1) = \frac{b}{2x} + \frac{b^2}{12x^2} + o(b^2).$$

Then substituting (4.6) in (4.4) and using the expressions (4.6) and (4.7), we deduce

$$\| K'_b(X_1) \|_q \leq \pi^{\frac{1-q}{2q}} (2x)^{\frac{1-q}{2q}-1} b^{\frac{1-q}{2q}} \left(b^2 C_2(q, x) + b C_1(q, x) + C_3(q, x) \right)^{1/q} + o(b^2),$$

where we used the notations

$$\begin{aligned}
 (4.8) \quad C_1(q, x) &= -f(x) \frac{2q^3 - 9q^2 + 4q - 33}{24x} - f'(x) \frac{q+1}{2} + f''(x) \frac{x}{2}, \\
 C_2(q, x) &= f(x) \frac{2q + 54x - q^2x + 21q^3x + q^4x + 93qx}{144x^3} \\
 &\quad - f'(x) \frac{(q+1)^2}{12x} + f''(x) \frac{q+1}{12}, \\
 C_3(q, x) &= -f(x) \frac{(q+1)(q-2)}{2}.
 \end{aligned}$$

The same steps can be done for $\|K'_b(X_{1+i})\|_p$ from (4.3). Then, if $p = q$ holds, one can represent Davydov's inequality (4.3) as

$$\begin{aligned}
 (4.9) \quad & |cov(K'_b(X_1), K'_b(X_{1+i}))| \leq \\
 & \leq 2\pi\alpha(i)^{\frac{1}{r}} \pi^{\frac{1-q}{q}} (2x)^{\frac{1-q}{q}-2} b^{\frac{1-q}{q}} \left(b^2 C_2(q, x) + b C_1(q, x) + C_3(q, x) \right)^{2/q} + o(b^2).
 \end{aligned}$$

Using (4.9) and taking $p = q = 2 + \delta$, $r = \frac{2+\delta}{\delta}$ it can be deduced that the covariance (4.2) is given by

$$\begin{aligned}
 |C(x)| &= \left| \frac{2}{n} \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) cov(K'_b(X_1), K'_b(X_{1+i})) \right| \\
 &\leq \left(2^{-\frac{2\delta+3}{\delta+2}} \pi^{\frac{1}{\delta+2}} x^{-\frac{3\delta+5}{\delta+2}} \frac{b^{-\frac{\delta+1}{\delta+2}}}{n} \left(b^2 C_2(\delta, x) + b C_1(\delta, x) + C_3(\delta, x) \right)^{\frac{2}{2+\delta}} \right) \\
 &\quad \cdot \sum_{i=1}^{n-1} \left(1 - \frac{i}{n}\right) \alpha(i)^{\frac{\delta}{2+\delta}} \Big| + o(b^2).
 \end{aligned}$$

Then we can estimate the covariance by the previous expressions

$$\begin{aligned}
 |C(x)| &\leq S(b, x, \delta, n) \sum_{\tau=2}^n \left(1 - \frac{\tau-1}{n}\right) \alpha(\tau-1)^{\frac{\delta}{2+\delta}} + o(b^2) \\
 &\leq S(b, x, \delta, n) \sum_{\tau=2}^{\infty} \alpha(\tau-1)^{\frac{\delta}{2+\delta}} + o(b^2) \leq S(b, x, \delta, n) \int_1^{\infty} \alpha(\tau)^{\frac{\delta}{2+\delta}} d\tau + o(b^2),
 \end{aligned}$$

where we used the following notation

$$S(b, x, \delta, n) = 2^{-\frac{2\delta+3}{\delta+2}} \pi^{\frac{1}{\delta+2}} x^{-\frac{3\delta+5}{\delta+2}} \frac{b^{-\frac{\delta+1}{\delta+2}}}{n} \left(b^2 C_2(\delta, x) + b C_1(\delta, x) + C_3(\delta, x) \right)^{\frac{2}{2+\delta}}.$$

Let us denote $\frac{\delta}{2+\delta} = v$, $0 < v < 1$. Then, in this notations, we get the estimate of the covariance

$$|C(x)| \leq \left(2^{-\frac{v+3}{2}} \pi^{\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+1}{2}}}{n} \left(b C_1(v, x) + C_3(v, x) \right)^{1-v} + o(b^2) \right) \int_1^{\infty} \alpha(\tau)^v d\tau.$$

By $0 < v < 1$ then it follows

$$|C(x)| \sim \frac{1}{n} b^{-\frac{v+1}{2}}. \quad \square$$

Remark 4.1. The main contribution to MISE (4.1) is provided by the part corresponding to $x \geq 2b$, so we will not do similar calculations here and further for $x \in [0, 2b)$ as $b \rightarrow 0$.

Proof of Theorem 2.1: Regarding the dependent case it is known that the MISE contains the bias, the variance and the covariance. By (1.8) it follows that the integrated sum of the squared bias and variance is the following expression

$$(4.10) \quad \begin{aligned} \int_0^{\infty} (B(x)^2 + V(x)) dx &= \frac{b^2}{16} \int_0^{\infty} P(x) dx \\ &+ \int_0^{\infty} \frac{n^{-1} b^{-\frac{3}{2}} x^{-\frac{3}{2}}}{4\sqrt{\pi}} \left(f(x) + \frac{b}{2} \left(\frac{f(x)}{x} - f'(x) \right) \right) dx \\ &+ o(b^2 + n^{-1} b^{-\frac{3}{2}}). \end{aligned}$$

This corresponds to the independent case.

By integration of (2.3) we get the upper bound of the integrated covariance

$$(4.11) \quad \int_0^{\infty} C(x) dx \leq \int_0^{\infty} \left(2^{-\frac{v+3}{2}} \pi^{-\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+1}{2}}}{n} C_3(v, x)^{1-v} + o(b^2) \right) \int_1^{\infty} \alpha(\tau)^v d\tau dx.$$

Combining (4.10) and (4.11), one can write

$$\begin{aligned} MISE(f'(x)) &\leq \int_0^{\infty} \frac{n^{-1} b^{-3/2} x^{-3/2}}{4\sqrt{\pi}} \left(f(x) + \frac{b}{2} \left(\frac{f(x)}{x} - f'(x) \right) \right) dx \\ &+ \int_0^{\infty} 2^{-\frac{v+3}{2}} \pi^{-\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+1}{2}}}{n} C_3(v, x)^{1-v} dx \int_1^{\infty} \alpha(\tau)^v d\tau \\ &+ \frac{b^2}{16} \int_0^{\infty} P(x) dx + o(b^2 + n^{-1} b^{-\frac{5}{2}}). \end{aligned}$$

The derivative of this expression in b leads to

$$(4.12) \quad \begin{aligned} &\frac{b}{8} \int_0^{\infty} P(x) dx - \frac{3n^{-1} b^{-\frac{5}{2}}}{8\sqrt{\pi}} \int_0^{\infty} x^{-\frac{3}{2}} f(x) dx + \frac{n^{-1} b^{-\frac{3}{2}}}{16\sqrt{\pi}} \int_0^{\infty} x^{-\frac{3}{2}} \left(\frac{f(x)}{x} - f'(x) \right) dx \\ &- \int_0^{\infty} \frac{v+1}{2} 2^{-\frac{v+3}{2}} \pi^{-\frac{1-v}{2}} x^{-\frac{v+5}{2}} \frac{b^{-\frac{v+3}{2}}}{n} C_3(v, x)^{1-v} dx \int_1^{\infty} \alpha(\tau)^v d\tau = 0. \end{aligned}$$

Since $0 < \nu < 1$ holds as in Lemma 2.1, the third term in (4.12) by b has the worst rate

$$c_1 b^{-\frac{\nu+3}{2}} = O\left(b^{-\frac{3}{2}}\right),$$

where c_1 is a constant.

Neglecting terms with $b^{-3/2}$ and $b^{-\frac{\nu+3}{2}}$ in comparison to the term containing $b^{-5/2}$, we simplify the equation

$$\frac{b^{7/2}}{8} \int_0^\infty P(x) dx - \frac{3n^{-1}}{8\sqrt{\pi}} \int_0^\infty x^{-3/2} f(x) dx + o(b^{7/2}) = 0.$$

The optimal $b = o(n^{-2/7})$ is the same as in (1.10). Let us insert such b in (2.4)

$$\begin{aligned} (4.13) \quad MISE_{opt}(\hat{f}'(x)) &= \int_0^\infty \frac{P(x)n^{-4/7}}{16} T^{4/7} dx + \int_0^\infty \frac{n^{-4/7} T^{-3/7} x^{-3/2}}{4\sqrt{\pi}} f(x) dx \\ &+ \int_0^\infty \frac{n^{-6/7} T^{-1/7} x^{-3/2}}{8\sqrt{\pi}} \left(\frac{f(x)}{x} - f'(x) \right) dx \\ &+ \int_0^\infty \left(2^{-\frac{\nu+3}{2}} \pi^{-\frac{1-\nu}{2}} x^{-\frac{\nu+5}{2}} \frac{T^{-\frac{\nu+1}{7}}}{n^{\frac{6-\nu}{7}}} C_3(\nu, x)^{1-\nu} dx \int_1^\infty \alpha(\tau)^\nu d\tau, \right. \end{aligned}$$

where

$$T = \frac{3 \int_0^\infty x^{-3/2} f(x) dx}{\sqrt{\pi} \int_0^\infty \left(\frac{f(x)}{3x^2} + f''(x) \right)^2 dx}.$$

The last term in (4.13) has the rate $o(n^{\frac{\nu-6}{7}})$. By $0 < \nu < 1$ we get that the optimal rate of convergence of MISE is given by $MISE_{opt}(\hat{f}'(x)) = O(n^{-4/7})$. \square

Proof of Lemma 2.2: We have to prove that $\alpha(\tau)$ defined by (2.6) satisfies the conditions of Lemma 2.1. Conditions 2 and 3 of Lemma 2.1 only refer to the density distribution. Thus, we remain to check only the first condition of Lemma 2.1.

To this end, using (2.6) we get

$$\begin{aligned} (4.14) \quad \int_1^\infty \alpha(\tau)^\nu d\tau &\leq \int_1^{\tau_0} d\tau + \int_{\tau_0}^\infty (2(C+1)E|X_i|^\nu |\rho^\nu|^\tau)^\nu d\tau \\ &= \tau_0 - 1 + (2(C+1)E|X_i|^\nu)^\nu \int_{\tau_0}^\infty (|\rho^\nu|^\tau)^\nu d\tau. \end{aligned}$$

The integral in (4.14) can be taken in general as

$$\int_{\tau_0}^{\infty} (|\rho^\nu|^\tau)^v d\tau = \frac{|\rho^\nu|^{\tau v}}{v \ln(|\rho^\nu|)} \Big|_{\tau_0}^{\infty}$$

Thus, to satisfy the first condition of Lemma 2.1, it must be

$$(4.15) \quad |\rho^\nu|^{\tau v} \Big|_{\tau=\infty} < \infty.$$

Since $\rho \in (-1, 1)$ holds, it follows $|\rho| \in [0, 1)$. For $\rho = 0$ (4.15) is satisfied. For $|\rho| \in (0, 1)$ one can rewrite (4.15) as

$$\left(\frac{1}{\xi}\right)^{\nu\tau v} \Big|_{\tau=\infty} < \infty, \quad \xi > 1,$$

which is valid as $\nu v > 0$. The latter is true since $0 < v < 1$ and $\nu = \min\{p, q, 1\} > 0$. Thus, the strong mixing AR(1) process (2.5) satisfies Lemma 2.1. Hence, it satisfies the conditions of Theorem 2.1. \square

ACKNOWLEDGMENTS

I am grateful to my supervisor DrSci Alexander Dobrovidov for an interesting topic. The work was partly supported by the Russian Foundation for Basic Research grant 16-08-01285 A and the Russian Science Foundation grant 14-50-00150.

REFERENCES

- [1] AKSOY, H. (2000). Use of Gamma distribution in hydrological analysis, *Turk. J. Engin. Environ. Sci.*, **24**, 419–428.
- [2] ANDREWS, D.W.K. (1983). *First order autoregressive processes and strong mixing*, Yale University, New Haven, Connecticut.
- [3] BOSQ, S. (1996). *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, Springer, New York.
- [4] BOUEZMARNIA, T. and ROMBOUTS, J.V.K. (2007). Nonparametric density estimation for multivariate bounded data, *Journal of Statistical Planning and Inference*, **140**(1), 139–152.

- [5] BOUEZMARNIA, T. and ROMBOUITS, J.V.K. (2010). Nonparametric density estimation for positive times series, *Computational Statistics and Data Analysis*, **54**(2), 245–261.
- [6] BOUEZMARNIA, T. and SCAILLET, O. (2003). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory*, **21**, 390–412.
- [7] BHATTACHARYA, P.K. (1967). Estimation of a probability density function and its derivatives, *The Indian Journal of Statistics*, **A 29**, 373–382.
- [8] SONG XI CHEN (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics*, **54**, 471–480.
- [9] DE BRABANTER, K. and DE BRABANTER, J. and DE MOOR, B. (2011). *Nonparametric derivative estimation*, “Proc. of the 23rd Benelux Conference on Artificial Intelligence (BNAIC)”, Gent, Belgium, 75–81.
- [10] DOBROVIDOV, A.V. and KOSHKIN, G.M. and VASILIEV, V.A. (2012). *Non-Parametric State Space Models*, Kendrick press, USA.
- [11] DOBROVIDOV, A.V. and MARKOVICH, L.A. (2013). *Nonparametric gamma kernel estimators of density derivatives on positive semi-axis*, “Proc. of IFAC MIM 2013”: Petersburg, Russia, June 19–21, 944–949.
- [12] DOBROVIDOV, A.V. and MARKOVICH, L.A. (2013). *Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis*, “Proc. of IFAC International Workshop on Adaptation and Learning in Control and Signal Processing Caen”, France, 500–505.
- [13] FURMAN, E. (2008). On a multivariate gamma distribution, *Statist. Probab. Lett.*, **78**, 2353–2360.
- [14] HALL, P. and WEHRLY, T.E. (1991). A geometrical method for removing edge effects from kernel-type nonparametric regression estimators, *J. Amer. Statist. Assoc.*, **86**, 665–672.
- [15] HÜRLIMANN, W. (2001). Analytical evaluation of economic risk capital for portfolios of gamma risks, *ASTIN Bulletin*, **31**, 107–122.
- [16] HASTINGS, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**(1), 97–109.
- [17] JONES, M.C. (1993). Simple boundary correction for density estimation kernel, *Statistics and Computing*, **3**, 135–146.
- [18] LEJEUNE, M. and SARDA, P. (1992). Smooth estimators of distribution and density functions, *Computational Statistics and Data Analysis*, **14**, 457–471.
- [19] MÜLLER, H.G. (1991). Smooth optimum kernel estimators near endpoints, *Biometrika*, **78**(3), 521–530.
- [20] PARZEN, E. (1962). On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, **33**(3), 1065.
- [21] ROBERTS, G.O.; ROSENTHAL, J.S.; SEGERS, J. and SOUSA, B. (2007). Extremal indices, geometric ergodicity of Markov chains, and MCMC, *Extremes*, **9**(3-4), 213–229.
- [22] ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*, **27**(3), 832.

- [23] SASAKI, H.; HYVÄRINEN, A. and SUGIYAMA, M. (2014). *Clustering via mode seeking by direct estimation of the gradient of a log-density*. In “Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2014)”, to appear.
- [24] SCAILLET, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics*, **16**, 217–226.
- [25] SCHUSTER, E.F. (1985) Incorporating support constraints into nonparametric estimators of densities, *Commun. Statist. Theory Methods*, **14**, 1123–1136.
- [26] SCHUSTER, E.F. (1969) Estimation of a probability function and its derivatives, *Ann. Math. Statist.*, **40**, 1187–1195.
- [27] SKÖLD, M. and ROBERTS, G.O. (2003). Density estimates from the Metropolis–Hastings Algorithm, *Scand. J. Stat.*, **30**, 699–718.
- [28] TSYPKIN, YA.Z. (1985). Optimality in adaptive control systems. Uncertainty and Control, *Springer, Lecture Notes in Control and Information Sciences Berlin, Heidelberg*, **70**, 153–214.
- [29] TURLACH, B.A. (1993). Bandwidth Selection in Kernel Density Estimation: A Review, *CORE and Institut de Statistique*.
- [30] WAND, M.P. and JONES, M.C. (1995). Kernel Smoothing, *Chapman and Hall*, London.
- [31] ZHANG, S. (2010). A note on the performance of the gamma kernel estimators at the boundary, *Statis. Probab. Lett.*, **80**, 548–557.