Data-driven bandwidth choice for gamma kernel estimates of density derivatives on the positive semi-axis

Alexander V. Dobrovidov, * Liubov A. Markovich **

* Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia (e-mail: dobrovid@ip.ru).
** Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia (e-mail: kimo1@mail.ru).

Abstract: In some applications it is necessary to estimate derivatives of probability densities defined on the positive semi-axis. The quality of nonparametric estimates of the probability densities and their derivatives are strongly influenced by smoothing parameters (bandwidths). In this paper an expression for the optimal smoothing parameter of the gamma kernel estimate of the density derivative is obtained. For this parameter data-driven estimates based on methods called "rule of thumb" and "cross-validation" are constructed. The quality of the estimates is verified and demonstrated on examples of density derivatives generated by Maxwell and Weibull distributions.

Keywords: Nonparametric estimation, density derivative, gamma kernel, data-driven bandwidth.

1. INTRODUCTION

In the paper of [4] a nonparametric estimator of the probability density function (pdf) for non-negative random variables was obtained. The evaluation was built using gamma kernels which are asymmetric. In some applications like financial mathematics it is necessary to estimate derivatives of the density. In this paper the estimate of the density derivative is constructed as a derivative of the gamma kernel estimate. In a previous paper of the authors [3] the statistical properties of these kernel estimates defined on the positive semi-axis were investigated. The asymptotic bias, variance and the integrated mean squared error (MISE) as \( n \to \infty \) were found. The smoothing parameter that minimizes MISE, is called the optimal smoothing parameter. It depends on functionals of the unknown true probability density and its derivatives. Therefore, it is impossible to calculate the exact value of this optimal smoothing parameter. However, the consistent estimates of this parameter can be constructed using realizations of the underlying random variable (r.v.) \( X \). In the paper these estimates for the density derivative are obtained by known methods called "the rule of thumb" (see [5]) and "cross-validation" (CV) (see [6], [7]).

2. BASIC DEFINITIONS AND PRELIMINARY RESULTS

Let a positive r.v. \( X \) be described by the pdf \( f(x) \), \( x > 0 \). If \( f(x) \) is unknown, then the standard statistical problem is to evaluate this density by an independent sample \( X^n = \{X_1, \ldots, X_n\} \). For the pdf defined on the positive semi-axis in the paper of [4] it is proposed a gamma kernel estimator

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{\rho(x),b}(X_i),
\]

where

\[
K_{\rho(x),b}(t) = \frac{t^{\rho(x)-1} \exp(-t/b)}{b^{\rho(x)} \Gamma(\rho(x))}
\]

is a gamma kernel. Here \( b \) is a smoothing parameter such that \( b \to 0 \) as \( n \to \infty \), \( \Gamma(\cdot) \) is the standard gamma function and

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

The support of the gamma kernel matches the support of the pdf to be estimated. For convenience let us introduce two kernel functions on the adjacent domains

\[
K_{\rho_1(x),b}(t) = \frac{t^{z/b-1} \exp(-t/b)}{b^{z/b} \Gamma((z/b)^2 + 1)}, \quad \text{if } x \geq 2b,
\]

\[
K_{\rho_2(x),b}(t) = \frac{t^{z/b+1} \exp(-t/b)}{b^{z/b+1} \Gamma((z/b)^2 + 1)}, \quad \text{if } x \in [0, 2b).
\]

The estimate of \( f'(x) \) is constructed as the derivative of \( f(x) \). Thus, we can write it as follows

\[
\hat{f}'(x) = \frac{1}{n} \sum_{i=1}^{n} K'_{\rho(x),b}(X_i),
\]

where

\[
K'_{\rho(x),b}(t) =
\begin{cases}
K'_{\rho_1(x),b}(t) = \frac{1}{b} K_{\rho_1(x),b}(t)L_1, & \text{if } x \geq 2b, \\
K'_{\rho_2(x),b}(t) = \frac{2}{b^2} K_{\rho_2(x),b}(t)L_2, & \text{if } x \in [0, 2b),
\end{cases}
\]
with 
\[
L_1 = L_1(t; x) = \ln t - \ln b - \Psi(p_1(x)), \\
L_2 = L_2(t; x) = \ln t - \ln b - \Psi(p_2(x)).
\]

Here \( \Psi(x) \) denotes the Digamma function (logarithm of the derivative of the Gamma function). The asymptotic properties of the density derivative estimate are determined by the following two lemmas.

**Lemma 1.** (Expectation) If \( b \to 0 \) then the leading term of mathematical expectation expansion for the density derivative estimate \( 2 \) equals
\[
E(\hat{f}'(x)) = \begin{cases} 
EK_{p_1(x), b}(X_1), & x \geq 2b, \\
EK_{p_2(x), b}(X_1), & x \in [0, 2b),
\end{cases}
\]
where
\[
EK_{p_1(x), b}(X_1) = (1/b)EK_{p_1(x), b}(X_1)L_1(X_1; x) \\
= f'(x) + b \left( \frac{1}{12x^2} f(x) + \frac{1}{4} f''(x) \right) + o(b), \\
EK_{p_2(x), b}(X_1) = (x/2b^2)EK_{p_2(x), b}(X_1)L_2(X_1; x) \\
= f'(x) \left( \frac{x}{2b} - \frac{b}{6x} \right) + f''(x) \left( \frac{7x}{48} + \frac{x^2}{2b} \right) + o(b).
\]

Note that under fixed \( b \) the estimate \( \hat{f}'(x) \) in the small area near zero has a bias, which grows at \( x \to 0 \). However in the asymptotic case when \( b \to 0 \) right border of this area \( x = 2b \) decreases to zero also. Therefore, it is interesting to know the bias limit when \( x \) and \( b \) converge to zero simultaneously, i.e. when the ratio \( x/b \) tends to some constant \( \kappa \) when \( b \to 0 \).

Let us proceed to calculate the variance of the derivative estimate.

**Lemma 2.** (Variance) If \( b \to 0 \) and \( nb^{5/2} \to \infty \), then the leading term of variance expansion for the density derivative estimate \( 2 \) equals
\[
Var(\hat{f}'(x)) = \frac{n^{-1}b^{-5/2}x^{-1/2}}{\sqrt{\pi}} \left[ f(2x) \left( \ln^2 x / 2 + \frac{1}{2} - \frac{1}{2x} \ln 2 \right) + f'(2x)(2 \ln 2 - \ln^2 x) \right] + o(b^2).
\]

It should be noted that in function \( f(2x) \) the double argument comes from the operation of squaring and averaging with a gamma density. The next problem is to calculate the mean squared error \( MSE(x) \) in accordance to the well known formula. Then
\[
MSE(\hat{f}'(x)) = \frac{b^2}{16} \left( \frac{f''(x)}{3x^2} + f''(x) \right)^2 + \frac{n^{-1}b^{-5/2}x^{-1/2}}{\sqrt{\pi}} \left[ f(2x) \left( \ln^2 x / 2 + \frac{1}{2} - \frac{1}{2x} \ln 2 \right) + f'(2x)(2 \ln 2 - \ln^2 x) \right] + o(b^2)
\]
where
\[
P(x) = \frac{f(x)}{3x^2} + f''(x), \\
M(x) = f(2x) \left( \frac{1}{2} - \frac{1}{2x} \ln 2 \right) + f'(2x)(2 \ln 2 - \ln^2 x) \ln 2, \\
N(x) = f(2x)(\ln 2)^2.
\]

Now we proceed to the global performance \( 3 \) and we will receive the integrated optimal bandwidth which does not depend on \( x \).

**Theorem (MISE).** If \( b \to 0 \) and \( nb^{5/2} \to \infty \), and the integrals
\[
\int_0^\infty \left( \frac{f(x)}{3x^2} + f''(x) \right)^2 dx, \int_0^\infty x^{-3/2} \ln x f(2x) dx, \int_0^\infty f''(x)^2 dx
\]
are finite and \( \int_0^\infty \left( \frac{f(x)}{3x^2} + f''(x) \right)^2 dx \neq 0 \), then the leading term of a MISE expansion for the density derivative estimate \( \hat{f}'(x) \) equals
\[
IBias^2(\hat{f}'(x)) = \frac{b^2}{16} \int_0^\infty \left( \frac{f(x)}{3x^2} + f''(x) \right)^2 dx + o(b^2).
\]
\[ MISE(f'(x)) = \frac{b^2}{16} \int_0^\infty \left( \frac{f(x)}{3x^2} + f''(x) \right)^2 \, dx + \left( \frac{n^{-1} x^{-1/2}}{\sqrt{\pi}} \right) \left[ b^{-3/2} M(x) + b^{-5/2} N(x) \right] \, dx + o(b^2 + n^{-1} (b^{-3/2} + b^{-5/2})) \]  

Minimization of (4) in \( b \) leads to a global optimal bandwidth

\[ b_0 = \left( \frac{20 \int_0^\infty x^{-1/2} N(x) \, dx}{\sqrt{\pi} \int_0^\infty \left( \frac{f(x)}{3x^2} + f''(x) \right)^2 \, dx} \right)^{2/9} n^{-2/9}. \]  

Its substitution into (4) yields the optimal MISE

\[ \text{MISE}_{opt}(f'(x)) = \frac{P(x) n^{-1/4}}{16} \left( \frac{20 \int_0^\infty x^{-1/2} N(x) \, dx}{\sqrt{\pi} \int_0^\infty P(x) \, dx} \right)^{2/9} + \frac{x^{-1/2}}{\sqrt{\pi}} \left( n^{-4/9} M(x) \frac{20 \int_0^\infty x^{-1/2} N(x) \, dx}{\sqrt{\pi} \int_0^\infty P(x) \, dx} \right)^{-1/9} + n^{-4/9} N(x) \frac{20 \int_0^\infty x^{-1/2} N(x) \, dx}{\sqrt{\pi} \int_0^\infty P(x) \, dx} \right)^{-1/9} \, dx. \]

The integral restrictions of the Theorem are fulfilled, for example, for a family of \( \chi^2 \)-distributions with the number of degrees of freedom \( m \geq 3 \) and also for a Maxwell and Weibull distributions, which will be investigated as true distributions in a simulation below. Proofs of the assertions are presented in [?].

From the expression for \( \text{MISE}_{opt} \) it follows that the nonparametric estimate (2) converges in mean square to the true density derivative with the rate \( O(n^{-4/9}) \). This rate is certainly less than the rate of convergence for the density \( O(n^{-4/5}) \), but the class of derivatives of the density is more complex than the class of densities. A similar decrease in the rate of convergence for the derivatives compared with the densities was observed in the use of Gaussian kernel functions on the whole line.

3. DATA-DRIVEN BANDWIDTH CHOICE

The optimal smoothing parameter for the density derivative estimate, defined by the formula (5), depends on the unknown true density \( f(x) \) and its derivatives. Therefore, it is impossible to calculate the true value of this parameter. However, one can construct a non-parametric estimate of this parameter. Quite a lot of methods for bandwidth estimation from the sample \( X_i \) are known. The simplest and most convenient (in authors’ opinion) are methods called the rule of thumb (see [?]) and cross-validation (see [?]). The first one is a parametric method. It gives a rough estimate by using in (5) instead of the unknown true density a so-called reference function, i.e. a density in the form of the kernel function. Parameters of the latter can be set by any classical method of parameter estimation. In this paper we use the method of moments.

The second approach for bandwidth estimation is to represent the integrals in (5) in a form of expectation of some function \( \varphi(\cdot) \), i.e. as \( E \varphi(X) \). The function \( \varphi(\cdot) \) may depend also on the unknown true density \( f(x) \) and its derivatives. We shall mark it as \( \varphi(x, f^{(\alpha)}(x)) \), \( \alpha = 0, 1, ... \). The expectation of \( \varphi(\cdot) \) equals to

\[ E \varphi \left( X, f^{(\alpha)}(X) \right) = \int_0^\infty \varphi \left( x, f^{(\alpha)}(x) \right) \, f(x) \, dx \]

by definition and it can be approximated by the arithmetic mean

\[ E \varphi \left( X, f^{(\alpha)}(X) \right) \approx \frac{1}{n} \sum_{i=1}^n \varphi \left( X_i, f^{(\alpha)}(X_i) \right), \]

converging to it as \( n \to \infty \). The unknown density and its derivatives in (6) are replaced by their gamma kernel estimates, but in the form of Cross-validation

\[ j^{(\alpha)}(X_i) = 1 \frac{1}{n-1} \sum_{j=1 \atop j \neq i}^n K_{\rho(X_i)}(X_j), \quad \alpha = 0, 1, \ldots \]  

These estimates will also depend on the unknown smoothing parameters which are substituted by rough estimates from the rule of thumb. Experience has shown that roughness of bandwidth estimation on the second level does not affect too much on the accuracy of estimation of densities and their derivatives, [?].

Let us find the value of the smoothing parameter following the rule of thumb. For this suppose we choose the gamma density

\[ f(x) = \frac{x^{\alpha-1} \exp(-\frac{x}{\rho})}{\rho^\alpha \Gamma(\rho)} \]

as a reference function. The first moment and the variance of it are \( pb \) and \( pb^2 \), respectively. According to the method of moments, we have to equate them to the first sample moment \( \bar{m} = n^{-1} \sum_{i=1}^n X_i \) and the sample variance \( \bar{D} = n^{-1} \sum_{i=1}^n (X_i - \bar{m})^2 \), correspondingly. Then we obtain for the parameters of (8) following simple expressions

\[ b_m = \bar{D} / \bar{m}, \quad \rho_m = (\bar{m})^2 / \bar{D}. \]  

Fig. 1 gives an idea of the relative closeness of the true (Maxwell) and references (gamma) densities, when the first two moments of the corresponding distributions are almost identical.

Now one can replace the unknown density in (5) by a gamma density (8) with known parameters (9). Then the numerator of \( b_0 \) is
The denominator of (5) contains three integrals. Let us consider them separately, for the first one the function is \( \varphi(x) = \frac{f(x)}{9x^4} \), then

\[
I_{d_1} = \int_0^\infty \left( \frac{f(x)}{3x^2} \right)^2 dx = \mathbb{E} \left( \frac{f(x)}{9x^4} \right) \approx \frac{1}{n} \sum_{i=1}^{n} \frac{f(X_i)}{9X_i^4}. \tag{14}
\]

The unknown density in (14) is replaced by its cross-validation estimate (7):

\[
I_{d_2} = \int_0^\infty \left( \frac{f''(x)f(x)}{3x^2} \right) dx = \mathbb{E} \left( \frac{f''(x)}{f(x)} \right) \approx \frac{1}{n} \sum_{i=1}^{n} \frac{f''(X_i)}{f_i(X_i)} \approx \frac{1}{n(n-1)}. \tag{15}
\]

For the second one it holds

\[
I_{d_3} = \int_0^\infty \left( \frac{f''(x)f(x)}{3x^2} \right) dx = \mathbb{E} \left( \frac{f''(x)}{f(x)} \right)
\]

with its derivative

\[
\hat{f}_i(X_i) = \frac{1}{n-1} \sum_{j \neq i} \frac{X_j^{-\frac{m-1}{2}} \exp\left(-\frac{X_j}{b}\right)}{\Gamma\left(\frac{X_j}{b}\right)b^{X_j} \Gamma\left(\frac{X_j}{b}\right)}. \tag{16}
\]

More accurate result can be received by using the cross-validation method. According to (6) the function \( \varphi(x) \) in the numerator of (5) equals

\[
\varphi(x) = \frac{(\ln 2)^2}{\sqrt{2} \pi^{1/2}}.
\]

Then the integral in the numerator yields

\[
I_{n1} = \int_0^\infty \frac{(\ln 2)^2}{\sqrt{2} \pi^{1/2}} f(x) dx \approx \frac{(\ln 2)^2}{\sqrt{2}} \sum_{i=1}^{n} X_i^{-1/2}. \tag{13}
\]
Density derivative estimates corresponding to calculated bandwidths in both methods used are represented in Fig. 3 and Fig. 4. Here solid line corresponds to the true density derivative to be estimated, dashed line corresponds to rule of thumb estimates and dash-dotted line corresponds to CV estimates.

### Table 1. Bandwidth $b$ for the Maxwell density derivative

| $m$   | rule of thumb | CV  |
|-------|---------------|-----|
| 100   | 0.335         | 0.209|
| 1000  | 0.203         | 0.246|
| 2000  | 0.175         | 0.047|

### Table 2. Bandwidth $b$ for the Weibull density derivative

| $m$   | rule of thumb | CV  |
|-------|---------------|-----|
| 100   | 0.167         | 0.090|
| 1000  | 0.105         | 0.032|
| 2000  | 0.089         | 0.023|

Quantitatively, the estimation error is determined by the value $\kappa$, defined by the following formula:

$$\kappa = \left( \int_0^\infty (f'(x) - \hat{f}'(x))^2 dx \right)^{1/2}.$$  \hspace{1cm} (18)

The mean error

$$\bar{\kappa} = M^{-1} \sum_{i=1}^M \kappa_i$$

calculated over $M$ repeated experiments and standard deviations in brackets are represented in Table 3 and Table 4.

| $m$  | rule of thumb | CV     |
|------|---------------|--------|
| 100  | 0.352         | 0.248  |
|      | (0.050)       | (0.070) |
| 1000 | 0.242         | 0.131  |
|      | (0.014)       | (0.014) |
| 2000 | 0.220         | 0.110  |
|      | (0.018)       | (0.020) |

Table 4. Error $\kappa$ for the Weibull distribution

| $m$  | rule of thumb | CV     |
|------|---------------|--------|
| 100  | 1.048         | 0.765  |
|      | (0.144)       | (0.242) |
| 1000 | 0.783         | 0.376  |
|      | (0.044)       | (0.078) |
| 2000 | 0.744         | 0.361  |
|      | (0.024)       | (0.028) |

From the tables and graphics one can see that from two methods of bandwidth selection (the rule of thumb and cross-validation) better results gives the latter one.

5. CONCLUSIONS

Estimation of the probability characteristics of the positive random variables is required in the theory of signal processing, the financial and actuarial mathematics and other important applications. The positivity of the distribution support of the observed random variables results in a significant complication of the models compared to the case of an unbounded support. We present nonparametric kernel estimates of the densities and their derivatives based on the asymmetric gamma kernels. The main part of the paper is devoted to the construction and evaluation of the optimal smoothing parameters (bandwidths) in the kernel density derivative estimates by samples of the independent random variables. For this purpose two well-known methods such as the rule of thumb and the cross-validation are used. It should be noted that in the case of asymmetric support even for the simple rule of thumb method the expression for the smoothing parameter (12) becomes quite cumbersome in comparison with the known corresponding expression based on the Gaussian reference function defined on the whole line. The simulation results show a significant advantage of the cross-validation method for the evaluation of the smoothing parameter. Unfortunately, due to rather low convergence rate ($O(n^{-1/9})$) of the kernel density derivative estimate, a sufficient accuracy requires large sample sizes (several thousands). The estimates of the kernel density derivatives will be used for the nonparametric estimation of logarithmic derivatives of the density determined on the positive real axis. The latter can be used for the problems of unsupervised nonparametric signal processing.