Hazard Rate Function Estimation

Using Erlang Kernel

Raid B. Salha

Department of Mathematics
The Islamic University of Gaza
Palestine

Hazem I. El Shekh Ahmed

Department of Mathematics
Al Quds Open University
Palestine

Iyad M. Alhoubi

University College of Science and Technology
Palestine

Abstract

In this paper, we define the Erlang kernel and use it to nonparametically estimation of the probability density function (pdf) and the hazard rate function for independent and identically distributed (iid) data.The bias, variance and the optimal bandwidth of the proposed estimator are investigated. Moreover, the asymptotic normality of the proposed estimator is investigated. The performance of the proposed estimator is tested using simulation study and real data.

Keywords: Erlang kernel, hazard rate function, kernel estimation, asymptotic normality

1. Introduction

Hazard rate functions can be used for several statistical analyses in medicine, engineering and economics. For instance, they are commonly used when presenting results in clinical trials involving survival data.

Several methods for hazard function estimation have been considered in the literature. Hazard function estimation by nonparametric methods has an advantage in flexibility because no formal assumptions are made about the mechanism that generates the sample order than the randomness.
Estimators of the hazard function based on kernel smoothing have been studied extensively. For instance, see Salha [8, 7], O. Scaillet [5], Watson and Leadbetter [12] and Rice and Rosenblatt [6].

The performance off the estimator at boundary points differs from the interior points due to so-called "boundary effects" that occur in nonparametric curve estimation problems. More specifically, the bias off the estimator at boundary points. To remove those boundary effects in kernel density estimation, a variety of methods have been developed in the literature. Some well-known methods are summarized below:
(i) The reflection method (Cline and Hart [3]; Schuster [10]; Silverman, [11].
(ii) The local linear method (Cheng [2]; Zhang and Karunamuni [13]).

Chen [1] solved this problem by replacing the symmetric kernels by asymmetric Gamma kernel which never assigns weight outside the support.

A lot of people who care of estimation do many specific distributions, such as normal, log-normal, gamma and inverse-gamma distributions, and this made us pose a question which is: Is there any other distributions which can be used as a kernel in the estimation and then give us acceptable results?

In this paper, we propose the Erlang kernel which also never assigns weight outside the support. The Erlang distribution is a continuous probability distribution with wide applicability primarily due to its relation to the exponential and Gamma distributions. The Erlang distribution was developed by A. K. Erlang to examine the number of telephone calls which might be made at the same time to the operators of the switching stations.

This paper is organised in six sections. In the first section, we present some information about kernel smoothing, hazard functions and the proposed kernel. In the second section, we introduce some definitions and relations, and state the conditions under which the results of the paper will be proved. In the third section, we investigate the bias, variance and optimal bandwidth of the Erlang kernel estimator. In the fourth section, we investigate the asymptotic

normality of the Erlang kernel estimator of the pdf and of the hazard function estimator. In the fifth section, the performance of the proposed estimator will be tested via two applications, simulated and real life data. In the sixth section, we introduce comments and the conclusion.

2. PRELIMINARIES

In this section, we state the conditions under which the results of the paper will be proved. Also, we mention the definition of gamma function and some relations related to it, then we will define Erlang kernel.

Conditions
  (1) Let $\{x_i\}_{i=1}^n$ be a random sample from a distribution of unknown probability density function f defined on $[0, \infty)$ such that f has a continuous second derivative.
  (2) $0 < \int x^4 (\acute{f})^2 (x) dx < \infty$ and $\int \frac{f(x)}{x} dx < \infty$
  (3) h is smoothing parameter satisfying $h + \frac{1}{nh^5} \longrightarrow 0$ as $n \to \infty$.

The Gamma function is defined to be the improper integral:

$$\Gamma(x) = \int_0^\infty t^{x-1} exp(-t) dt, \ x \geq 1$$

The Taylor expansion about zero of $\frac{1}{\Gamma(x)}$ is available and given by:

$$\frac{1}{\Gamma(x)} = x + \gamma x^2 + \tfrac{1}{2}(\gamma^2 - \tfrac{\pi^2}{6})x^3 + o(x^3)$$
$$\approx x.$$

Also, through this paper we will approximate the folloing:
The beta function

$$B(\frac{1}{h}, \frac{1}{h}) = \frac{(\Gamma(\frac{1}{h}))^2}{\Gamma(\frac{2}{h})} \approx 4h^3,$$

$$\left(\frac{1}{2}\right)^{2\frac{h+1}{h}} \approx \frac{h}{4},$$

$$h^{1+\frac{1}{h}} \approx h^2.$$

This approximation because we concern with values of h close to zero.

In this paper we consider the following Erlang kernel function:

$$K_E(x, \frac{1}{h})(t) = \frac{1}{\Gamma(1+\frac{1}{h})} \left[\frac{1}{x}(1+\frac{1}{h})\right]^{\frac{h+1}{h}} t^{\frac{1}{h}} exp(-\frac{t}{x}(1+\frac{1}{h})), \ t \geq 0$$

If a random variable Y has a pdf $K_E(x, \frac{1}{h})(t)$, then $E(Y) = x$, and the variance is $var(Y) = \frac{x^2 h}{1+h}$.
We propose the following estimator of the probability density function $f(.)$,

the Erlang estimator

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_E(x, \frac{1}{h})(X_i)$$

3. BIAS, VARIANCE AND OPTIMAL BANDWIDTH

In this section, the bias and variance will be evaluated. Then we will discuss the optimal bandwidth of the Erlang kernel estimator.

Proposition 1.
The bias of the proposed estimator is given by:

$$Bias(\hat{f}(x)) = \frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x) + o(1) \tag{1}$$

Proof:

$$E(\hat{f}(x)) = \int_0^\infty K_E(x, \frac{1}{h})(y)f(y) \, dy = E(f(\xi_x))$$

where ξ_x follows an Erlang distribution with mean $\mu_x = E(\xi_x) = x$ and variance, $V_x = Var(\xi_x) = \frac{x^2 h}{1+h}$.
The Taylor expansion about μ_x for $f(\xi_x)$ is:

$f(\xi_x) = f(\mu_x) + \acute{f}(\mu_x)(\xi_x - \mu_x) + \frac{1}{2}\acute{f}(\mu_x)(\xi_x - \mu_x)^2 + o(1)$

So, $E(\hat{f}(\xi_x)) = f(x) + \frac{1}{2}\acute{f}(x)var(\xi_x) + o(1)$

$$E(\hat{f}(\xi_x)) = f(x) + \frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x) + o(1) \tag{2}$$

Hence,

$$Bias(\hat{f}(x)) = \frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x) + o(1)$$

Proposition 2.
The variance of the proposed estimator is given by:

$$Var(\hat{f}(x)) = \frac{(2+h)(\frac{1}{2})^{2\frac{h+1}{h}}}{2nxh^3}f(x) + o((nh^2)^{-1}) \tag{3}$$

Proof:
$Var(\hat{f}(x)) = \frac{1}{n}Var\left[K_E(x, \frac{1}{h})(t)\right] = \frac{1}{n}\left[E\left(K_E^2(x, \frac{1}{h})(t)\right)\right] - \left[E(K_E(x, \frac{1}{h})(t))\right]^2$
where
$$K_E^2(x, \frac{1}{h})(t) = \frac{1}{\left[\Gamma(1+\frac{1}{h})\right]^2}\left[\frac{1}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{\frac{2}{h}}exp(-2\frac{t}{x}(1+\frac{1}{h}))$$

$$= \frac{1}{\left[\Gamma(1+\frac{1}{h})\right]^2}\left(\frac{1}{2}\right)^{2\frac{h+1}{h}} t^{-1}\Gamma(2\frac{h+1}{h}).\left[\frac{1}{\Gamma(2\frac{h+1}{h})}\left[\frac{2}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{1+\frac{2}{h}}exp(-2\frac{t}{x}(1+\frac{1}{h}))\right]$$

$$= R_{x,h}t^{-1}.\left[\frac{1}{\Gamma(2\frac{h+1}{h})}\left[\frac{2}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{1+\frac{2}{h}}exp(-2\frac{t}{x}(1+\frac{1}{h}))\right]$$

where $R_{x,h} = \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2h^3}$.

Therefore, $E[K_E^2(x,\frac{1}{h})(t)] = E[R_{x,h}\eta_x^{-1}f(\eta_x)]$

where η_x follows as Erlang distribution with mean $\mu_x = E(\eta_x) = x$ and variance $V_x = Var(\eta_x) = \frac{hx^2}{2(h+1)}$.

The Taylor expansion of $\eta_x^{-1}f(\eta_x)$ about μ_x is as follows:

$\eta_x^{-1}f(\eta_x) = \mu_x^{-1}f(\mu_x) + \left[\frac{-1}{\mu_x^2}f(\mu_x) + \frac{1}{\mu_x}\grave{f}(\mu_x)\right](\eta_x - \mu_x) + \frac{1}{2}\left[\frac{2}{\mu_x^3}f(\mu_x) - 2\frac{1}{\mu_x^2}\grave{f}(\mu_x) + \frac{1}{\mu_x}\acute{f}(\mu_x)\right]$
$(\eta_x - \mu_x)^2 + o(1)$

So,

$E\left[\eta_x^{-1}f(\eta_x)\right] = x^{-1}f(x) + \frac{1}{2}\left[\frac{2}{x^3}f(x) - \frac{2}{x^2}\grave{f}(x) + \frac{1}{x}\acute{f}(x)\right]\frac{hx^2}{2(h+1)} + o(1)$

This implies that

$E\left[R_{x,h}\eta_x^{-1}f(\eta_x)\right] = \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2h^3x}f(x) + \frac{(2+h)hx^2\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{8(h+1)h^3}\left[\frac{2}{x^3}f(x) - \frac{2}{x^2}\grave{f}(x) + \frac{1}{x}\acute{f}(x)\right] +$
$o(h^{-2})$

Therefore,

$$Var(\hat{f}(x)) = \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nxh^3}f(x) + o(n^{-1}h^{-2}) - \frac{1}{n}(f(x) + o(1))^2$$
$$= \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nxh^3}f(x) + o((nh^2)^{-1})$$

Optimal Bandwidth:

First of all, we will define MSE and Mean Integrated Squared Error (MISE) as follows:

$$MSE(\hat{f}) = \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nxh^3}f(x) + o((nh^2)^{-1}) + \left[\frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x) + o(1)\right]^2$$
$$= \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nxh^3}f(x) + o((nh^2)^{-1}) + \frac{1}{4}x^4\frac{h^2}{(1+h)^2}\left[\acute{f}(x)\right]^2 + o(1)$$
$$\approx \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nxh^3}f(x) + \frac{1}{4}x^4\frac{h^2}{(1+h)^2}\left[\acute{f}(x)\right]^2$$

Therefore,

$$MISE(\hat{f}) = \frac{(2+h)\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}}{2nh^3}I_1 + \frac{1}{4}\frac{h^2}{(1+h)^2}I_2 \tag{4}$$

where, $I_1 = \int \frac{f(x)}{x}dx$ and $I_2 = \int x^4\left[\acute{f}(x)\right]^2 dx$

Therefore, we can approximate MISE to be:

$$MISE(\hat{f}) = \frac{I_1}{4nh^2} + \frac{h^2}{4}I_2 \tag{5}$$

We will now find the optimal bandwidth by minimizing (5) with respect to h, so we have

$$\frac{d\ MISE(\hat{f})}{dh} = \frac{-I_1}{2nh^3} + \frac{h}{2}I_2, \tag{6}$$

$$\frac{d^2\ MISE(\hat{f})}{dh^2} = \frac{3I_1}{2nh^4} + \frac{1}{2}I_2 > 0. \tag{7}$$

Setting (6) equal zero yields the optimal bandwidth h_{opt} for the given pdf and kernel:

$$h_{opt} = \left(\frac{I_1}{nI_2}\right)^{\frac{1}{4}} \tag{8}$$

In addition, (7) proves that this value minimize (5).
Substituting (8) for h in (5) gives the minimum MISE for the given pdf and kerne which is given by:

$$MISE_{opt} = \left(\frac{I_1 I_2}{n}\right)^{\frac{1}{2}} \tag{9}$$

Note that h_{opt} depend on the sample size n, the kernel and on the unknown pdf.

4. ASYMPTOTIC NORMALITY

In this section, we state the main two theorems talking about the asymptotic normality for the proposed estimator and an important lemma which we will use in the second main theorem.

Definition 1. *A hazard rate function is defined as the probability of an event happening in a short time interval. More precisely, it is defined as:*

$$r(x) = \lim_{\Delta x \to 0} \frac{P(X \le x + \Delta x | X > x)}{\Delta x}, \quad x > 0.$$

The hazard rate function can be written as the ratio between the pdf $f(.)$ and the survivor function $S(.) = 1 - F(.)$ as follows:

$$r(x) = \frac{f(x)}{S(x)}$$

The kernel estimator of $S(.)$ is $\hat{S}(x) = 1 - \hat{F}(x)$ where,

$$\hat{F}(x) = \int_0^x \hat{f}(u)du = \frac{1}{n}\sum_{i=1}^n \int_0^x K_E(u, \frac{1}{h})(X_i)du$$

Definition 2. *we defined the proposed estimator for the hazard rate function to be:*

$$\hat{r}(x) = \frac{\hat{f}(x)}{\hat{S}(x)}$$

Theorem 1. *Under conditions 1, 2, and 3, the following holds*

$$2^{\frac{1}{h}}\sqrt{\frac{nh^5}{h+2}}(\hat{f}(x) - f(x)) \xrightarrow{d} N\left(0, \frac{f(x)}{8x}\right) \qquad (10)$$

Proof:

Let $V_{ni} = K_E(x, \frac{1}{h})(X_i)$, $i = 1, 2, ..., n$. Then $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} V_{ni}$, where V_{ni}, $i = 1, 2, ..., n$ are independent and identically distributions (iid) and $\sigma_{ni}^2 < \infty$.

We show now that the Liapounov condition is satisfied, that is for some $\delta > 0$,

$$\lim_{n \to \infty} \frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)} = 0$$

First of all, we have:

$$|V_n|^{2+\delta} = \frac{1}{[\Gamma(1+\frac{1}{h})]^{2+\delta}}\left[\frac{1}{x}(1+\frac{1}{h})\right]^{(2+\delta)\frac{h+1}{h}} t^{\frac{2+\delta}{h}} exp(-(2+\delta)\frac{t}{x}(1+\frac{1}{h})), \ t \geq 0$$

$$= \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{\delta\frac{h+1}{h}} t^{\frac{\delta}{h}} exp(-\delta\frac{t}{x}(1+\frac{1}{h}))\Gamma(2\frac{h+1}{h})}{[\Gamma(1+\frac{1}{h})]^{2+\delta}} \cdot \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{\frac{2}{h}} exp(-2\frac{t}{x}(1+\frac{1}{h}))}{\Gamma(2\frac{h+1}{h})}$$

$$= \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{\delta\frac{h+1}{h}} t^{\frac{\delta}{h}} exp(-\delta\frac{t}{x}(1+\frac{1}{h}))t^{-1}\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}\Gamma(2\frac{h+1}{h})}{[\Gamma(1+\frac{1}{h})]^{2+\delta}} \cdot \frac{\left[\frac{2}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{1+\frac{2}{h}} exp(-2\frac{t}{x}(1+\frac{1}{h}))}{\Gamma(2\frac{h+1}{h})}$$

$$= \alpha t^{\frac{\delta}{2}-1} \cdot \frac{\left[\frac{2}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{1+\frac{2}{h}} exp(-2\frac{t}{x}(1+\frac{1}{h}))}{\Gamma(2\frac{h+1}{h})}$$

where, $\alpha = \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{\delta\frac{h+1}{h}}\left(\frac{1}{2}\right)^{2\frac{h+1}{h}}\Gamma(2\frac{h+1}{h})exp(-\delta\frac{t}{x}(1+\frac{1}{h}))}{[\Gamma(1+\frac{1}{h})]^{2+\delta}}$. Then we have:

$$E(|V_n|^{2+\delta}) = \int_0^\infty \alpha y^{\frac{\delta}{h}-1}\left[\frac{\left[\frac{2}{x}(1+\frac{1}{h})\right]^{2\frac{h+1}{h}} t^{1+\frac{2}{h}} exp(-2\frac{t}{x}(1+\frac{1}{h}))}{\Gamma(2\frac{h+1}{h})}\right] f(y)dy = E(\alpha \xi_x^{\frac{\delta}{h}-1}f(\xi_x)),$$

where, ξ_x follows an Erlang distribution with mean $\mu_x = E(\xi_x) = x$ and variance $var(\xi_x) = \frac{x^2 h}{2(1+h)}$

The Taylor expansion of $\xi_x^{\frac{\delta}{h}-1}f(\xi_x)$ about the mean $\mu_x = E(\xi_x) = x$ as follows:

$$\xi_x^{\frac{\delta}{h}-1}f(\xi_x) = x^{\frac{\delta}{h}-1}f(x) + \left[(\frac{\delta}{h}-1)x^{\frac{\delta}{h}-2}f(x) + x^{\frac{\delta}{h}-1}\acute{f}(x)\right](\xi_x - x) +$$

$$\left[(\frac{\delta}{h}-1)(\frac{\delta}{h}-2)x^{\frac{\delta}{h}-3}f(x) + 2(\frac{\delta}{h}-1)x^{\frac{\delta}{h}-2}\acute{f}(x) + x^{\frac{\delta}{h}-1}\acute{f}(x)\right](\xi_x - x)^2 + o(h^{-2})$$

$$E(\xi_x^{\frac{\delta}{h}-1}f(\xi_x)) = x^{\frac{\delta}{h}-1}f(x) +$$
$$h^{-2}\left[(\delta-h)(\delta-2h)x^{\frac{\delta}{h}-3}f(x) + 2h(\delta-h)x^{\frac{\delta}{h}-2}\acute{f}(x) + h^2 x^{\frac{\delta}{h}-1}\acute{f}(x)\right]\frac{hx^2}{2(h+1)} + o(h^{-2})$$

$$= x^{\frac{\delta}{h}-1}f(x) + \frac{h^{-1}x^{\frac{\delta}{h}-1}}{2(h+1)}\left[(\delta-h)(\delta-2h)f(x) + 2h(\delta-h)x\acute{f}(x) + h^2 x^2\acute{f}(x)\right] + o(h^{-2})$$

Therefore,

$$E(|V_n|^{2+\delta}) = \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{\delta\frac{h+1}{h}} \left(\frac{1}{2}\right)^{1+\frac{2}{h}} (h+2)exp(-\delta\frac{t}{x}(1+\frac{1}{h}))}{4h^5 \left[\Gamma(1+\frac{1}{h})\right]^{\delta}} x^{\frac{\delta}{h}-1} f(x) + o(h^{-7})$$

(11)

Now substituting $\delta = 0$, the following is hold:

$$Var(V_n) = \frac{\left(\frac{1}{2}\right)^{1+\frac{2}{h}} (h+2)}{4h^5} x^{-1} f(x) + o(h^{-7}).$$

Hence,

$$\frac{E|V_n - E(V_n)|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)} \leq \frac{E|V_n|^{2+\delta}}{n^{\frac{\delta}{2}}\sigma^{2+\delta}(V_n)}$$

$$\rightarrow \frac{\left[\frac{1}{x}(1+\frac{1}{h})\right]^{\delta\frac{h+1}{h}} \left(\frac{1}{2}\right)^{1+\frac{2}{h}} (h+2)exp(-\delta\frac{t}{x}(1+\frac{1}{h}))x^{\frac{\delta}{h}-1}f(x)}{n^{\frac{\delta}{2}}\left[\Gamma(1+\frac{1}{h})\right]^{\delta}4h^5\left[\frac{\left(\frac{1}{2}\right)^{1+\frac{2}{h}}(h+2)}{4h^5}x^{-1}f(x)\right]^{\frac{2+\delta}{2}}}$$

$$= \frac{x^{-\delta-1}\left[1+\frac{1}{h}\right]^{\delta\frac{h+1}{h}} \left(\frac{1}{2}\right)^{1-\frac{\delta}{h}} exp(-\frac{t\delta}{x}(1+\frac{1}{h}))(h+2)f(x)}{n^{\frac{\delta}{2}}\left[\Gamma(1+\frac{1}{h})\right]^{\delta}[4h^5]^{-\frac{\delta}{2}}\left[\frac{1}{2}(h+2)x^{-1}f(x)\right]^{\frac{\delta+2}{2}}}$$

$$= \frac{x^{-\delta-1}[1+h]^{\delta\frac{h+1}{h}} [4h^5]^{\frac{\delta}{2}} exp(-\frac{t\delta}{x}(1+\frac{1}{h}))\frac{1}{2}(h+2)f(x)}{n^{\frac{\delta}{2}}\left(\frac{1}{4}\right)^{\frac{\delta}{2h}}[h]^{\delta\frac{h+1}{h}}\left[\Gamma(1+\frac{1}{h})\right]^{\delta}\left[\frac{1}{2}(h+2)x^{-1}f(x)\right]^{\frac{\delta+2}{2}}}$$

$$\rightarrow 0$$

The last term by condition 3 vanishes as $n \rightarrow 0$. Also, the remaining component of the last term are bounded.

Lemma 1. *Under conditions 1, 2 and 3 the following holds*

$$\sqrt{nh}|\hat{F}(x) - F(x)| \xrightarrow{p} 0 \qquad (12)$$

Proof:

First of all, we have from the definition of $\hat{F}(x)$ the following relations:

$$\begin{aligned}
E\hat{F}(x) &= \int_0^\infty \int_0^x k_E(u,\frac{1}{h})(y)du f(y)dy \\
&= \int_0^x \int_0^\infty k_E(u,\frac{1}{h})(y)f(y)dydu \\
&= \int_0^x E(f(\xi_x))du \\
&= \int_0^x (f(u) + \frac{1}{2}u^2\frac{h}{h+1}\acute{f}(u))du + o(1) \\
&= F(x) + o(1)
\end{aligned}$$

where ξ_x is defined as in Proposition 1. This implies that, $E\hat{F}(x) - F(x) = o(1)$
Therefore,

$$(nh)^{\frac{1}{2}}|E\hat{F}(x) - F(x)| = o((nh)^{\frac{1}{2}}) \xrightarrow{p} 0 \qquad (13)$$

On the other hand, $\hat{F}(x)$ can be written as follows:
$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^n \int_0^x k_E(u,\frac{1}{h})(y)du = \frac{1}{n}\sum_{i=1}^n W_i(x)$ where $W_i(x) = \int_0^x k_E(u,\frac{1}{h})(y)du$

Now, given $\varepsilon > 0$, $\delta > 0$, then we have:

$$
\begin{aligned}
P\left[(nh)^{\frac{1}{2}}|\hat{F}(x) - E\hat{F}(x)| > \varepsilon\right] &\leq \frac{E|\hat{F}(x) - E\hat{F}(x)|^{2+2\delta}}{\varepsilon^{2+2\delta}(nh)^{-(1+\delta)}} \\
&= \varepsilon^{-2-2\delta}(nh)^{1+\delta}E|\hat{F}(x) - E\hat{F}(x)|^{2+2\delta} \\
&= \varepsilon^{-2-2\delta}(nh)^{1+\delta}n^{-2-2\delta}E|\sum_{i=1}^{n}W_i(x) - EW_i(x)|^{2+2\delta} \\
&\leq \varepsilon^{-2-2\delta}(n^{-1}h)^{1+\delta}\sum_{i=1}^{n}E|W_i(x)|^{2+2\delta} + \\
&\quad \varepsilon^{-2-2\delta}(n^{-1}h)^{1+\delta}\sum_{i=1}^{n}|EW_i(x)|^{2+2\delta}
\end{aligned}
$$

The second term vanishes as $n \longrightarrow \infty$ and $h \longrightarrow 0$, since from 2 we have

$$
(n^{-1}h)^{1+\delta}\sum_{i=1}^{n}|EW_i(x)|^{2+2\delta} = o((n^{-1}h)^{1+\delta}),
$$

Further, by replacing δ with 2δ in (11) and assuming that

$$
\tau_{h,\delta} = \frac{[(1+h)]^{2\delta\frac{h+1}{h}}\left(\frac{1}{2}\right)^{1+\frac{2}{h}}(h+2)}{\left[\Gamma(1+\frac{1}{h})\right]^{2\delta}}
$$

we have:

$$
\begin{aligned}
P\left[(nh)^{\frac{1}{2}}|\hat{F}(x) - E\hat{F}(x)| > \varepsilon\right] &\leq \varepsilon^{2+2\delta}(n^{-1}h)^{1+\delta}\sum_{i=1}^{n}E|W_i(x)|^{2+2\delta} \\
&= \varepsilon^{-2-2\delta}n^{-1-\delta}h^{1+\delta}n\int_0^{\infty}\int_0^{x}\left[k_E(u,\tfrac{1}{h})(y)\right]^{2+2\delta}duf(y)dy \\
&= \varepsilon^{-2-2\delta}n^{-\delta}h^{1+\delta}\int_0^{x}\int_0^{\infty}\left[k_E(u,\tfrac{1}{h})(y)\right]^{2+2\delta}f(y)dydu \\
&= \varepsilon^{-2-2\delta}\int_0^{x}\left[\frac{\tau_{h,\delta}exp(-2\delta\frac{t}{u}(1+\frac{1}{h}))}{4n^{\delta}h^{\delta+\frac{2\delta}{h}+4}}u^{-2\delta-1}f(u)\right]du \\
&\quad +o((n^{-1}h)^{\delta}h^{-6}) \to 0
\end{aligned}
$$

$$
(nh)^{\frac{1}{2}}|\hat{F}(x) - E\hat{F}(x)| \xrightarrow{p} 0. \tag{14}
$$

Since $|\hat{F}(x) - F(x)| \leq |\hat{F}(x) - E\hat{F}(x)| + |E\hat{F}(x) - F(x)|$, then by (13) and (14), we have:

$$(nh)^{\frac{1}{2}}|\hat{F}(x) - F(x)| \leq (nh)^{\frac{1}{2}}|\hat{F}(x) - E\hat{F}(x)| + (nh)^{\frac{1}{2}}|E\hat{F}(x) - F(x)| \xrightarrow{p} 0$$

This complete the proof of the lemma..

Theorem 2. *Under conditions 1, 2, and 3, the following holds*

$$
2^{\frac{1}{h}}\sqrt{\frac{nh^5}{h+2}}(\hat{r}(x) - r(x)) \xrightarrow{d} N(0, \frac{r(x)}{8xS(x)}) \tag{15}
$$

Proof:

$$
\begin{aligned}
2^{\frac{1}{h}}\sqrt{\tfrac{nh5}{h+2}}(\hat{r}(x)-r(x)) \;=\; & 2^{\frac{1}{h}}\sqrt{\tfrac{nh5}{h+2}}\big(\tfrac{\hat{f}(x)}{\hat{S}(x)}-\tfrac{f(x)}{S(x)}\big)\\[4pt]
=\;& 2^{\frac{1}{h}}\sqrt{\tfrac{nh5}{h+2}}\big(\tfrac{\hat{f}(x)}{\hat{S}(x)}-\tfrac{f(x)}{\hat{S}(x)}-\tfrac{f(x)}{S(x)}+\tfrac{f(x)}{\hat{S}(x)}\big)\\[4pt]
=\;& \tfrac{2^{\frac{1}{h}}}{\hat{S}(x)}\sqrt{\tfrac{nh5}{h+2}}\big(\hat{f}(x)-f(x)-\tfrac{f(x)\hat{S}(x)}{S(x)}+f(x)\big)\\[4pt]
=\;& \tfrac{2^{\frac{1}{h}}}{\hat{S}(x)}\sqrt{\tfrac{nh5}{h+2}}\big(\hat{f}(x)-f(x)\big)\\[4pt]
& -\tfrac{2^{\frac{1}{h}}f(x)}{S(x)\hat{S}(x)}\sqrt{\tfrac{nh5}{h+2}}\big(\hat{S}(x)-S(x)\big)\\[4pt]
\longrightarrow^{d}\;& N\big(0,\tfrac{r(x)}{8xS(x)}\big)
\end{aligned}
$$

Note that the second term vanishes by (12), and the first term is asymptotically normally distributed by (10).

Moreover, from (10) and (15) too we have:

$$
\begin{aligned}
E(\hat{r}(x)) \;=\;& \tfrac{E((\hat{f})(x))}{E((\hat{S})(x))}\\[4pt]
=\;& \tfrac{f(x)+\frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x)}{S(x)}+o(1)\\[4pt]
=\;& r(x)+\tfrac{\frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x)}{S(x)}+o(1)
\end{aligned}
$$

Therefore,

$$
Bias(\hat{r}(x)) = \frac{\frac{1}{2}x^2\frac{h}{1+h}\acute{f}(x)}{S(x)}+o(1)
$$

and,

$$
Var(\hat{r}(x)) = \frac{(2+h)(\frac{1}{2})^2 2^{\frac{h+1}{h}}}{2nxh^3}\frac{r(x)}{S(x)}+o((nh^2)^{-1})
$$

5. Applications

In this section, the performance of the proposed estimator in estimating the pdf and hazard rate function is tested upon two applications using a simulated and real life data.

5.1. **A Simulation Study.** A sample of size 200 from the exponential distribution with pdf $f(x)=exp(-x)$ is simulated. We computed the bandwidth using the relation

$$
h_{opt}=0.79Rn^{-\frac{1}{5}}, \tag{16}
$$

see [11] page 47 and it equals (0.36658).

The density and the hazard functions were estimated using the Erlang estimator. The estimated values and the true exponential pdf are plotted in Figures 1(a), this figure shows that the performance of the Erlang estimator is acceptable at the boundary near the zero. In the interior the behavior of the pdf estimator becomes more similar for large values. Also figure 1(b)

shows that the performance of the Erlang estimator of the hazard function is acceptable at the boundary near the zero which we concern on. The mean squared error (MSE) of proposed estimator of the density function is equal to 0.000266937 and for the hazard function is evaluated for the interval $[0, 0.5]$ -because we concern about closest values to zero- and is equal to 0.02289262.

5.2. **Real Data.** In this subsection, we used the suicide data given in Silverman [11], to exhibit the practical performance of the Erlang estimator. The data gives the lengths of the treatment spells (in days) of control patients in suicide study. We used the logarithm of the data to draw figures 2(a) and 2(b)using bandwidth equals 0.480411 which computed by (16), these figures exhibit the two estimated functions of the probability density and hazard rate functions, respectively.
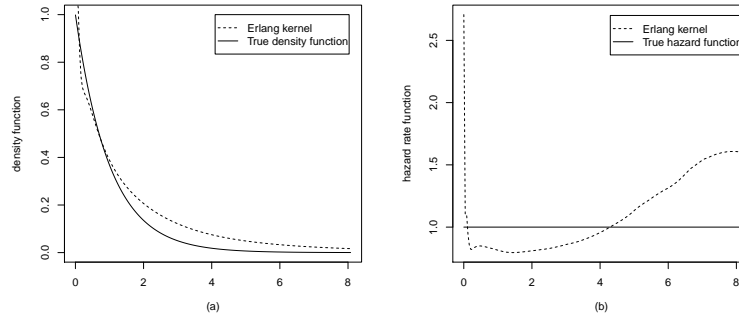

FIGURE 1. (a) The Erlang kernel estimator of the density function, (b) the hazard rate function for the simulated data of the exponential distribution
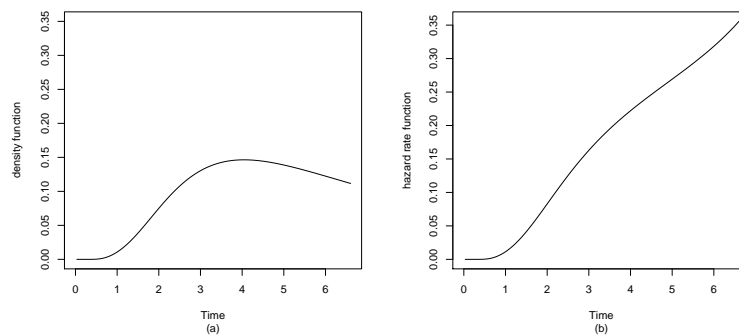

FIGURE 2. (a) The Erlang kernel estimator of the density function, (b) the hazard rate function for the suicide data.

6. Comments and Conclusion

In this paper, we have proposed a new kernel estimator of the hazard rate function for (iid) data based on the Erlang kernel with nonnegative support, we show that the bias is depends on the smoothing parameter h and the estimated point x, and it goes to zero as $h \longrightarrow 0$, also it gets small for the values of x closed to zero. The variance is investigated and we noticed that it depends also on h and x. On the other hand, it goes to zero as $h \longrightarrow 0$, and gets large at the values of x close to zero. Moreover, the optimal bandwidth and the asymptotic normality were investigated.

In addition, the performance of the proposed estimator is tested in two applications. In a simulation study using exponential sample we noticed that the performance of the proposed estimator is acceptable, and gives a small MSE. Using real data, we exhibited the practical performance of the Erlang estimator.

References

[1] Chen, S. X. (2000). Probability density function estimation using gamma kernels. Ann. Inst. Stat. Math., 52, 471- 480.
[2] Cheng, M.Y., Fan, J. and Marron, J.S. (1997). On Automatic Boundary Corrections. The Annals of Statistics, 25, 1691-1708.
[3] Cline, D.B.H and Hart, J.D. (1991). Kernel Estimation of Densities of Discontinuous Deriva-tives. Statistics, 22, 69-84.
[4] N. D. Singpurwalla and M. Y. Wong (1983). Kernel estimation of failure rate function and density estimation: an analogy. J. Amer. Statist. Assoc. 78(3), 478-481.
[5] O. Scaillet (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric Statistics, 16, 217-226.
[6] Rice, J. and Rosenblatt, M. (1976). Estimation of the log survivor function and hazard function. Sankhya Series A 38, 60-78.
[7] Salha, R. (2013). Estimating the density and hazard rate functions using the reciprocal inverse Gaussian kernel. 15th Applied Stochastic Models and Data Analysis International Conference, Mataro (Barcelona) Spain, June 25-28, 2013.
[8] Salha, R. (2012). Hazard rate function estimation using inverse Gaussian kernel. The Islamic university of Gaza journal of Natural and Engineering studies, 20(1), 73-84.
[9] Salha, R. (2009). Adaptive kernel estimation of the hazard rate function. The Islamic univercity - Gaza Journal, 17(1), 71-81.
[10] Schuster, E.F. (1985). Incorporating Support Constraints Into Nonparametric Estimators of Densities.Communications in Statistics, Part A - Theory and Methods, 14, 1123-1136.
[11] Silverman, B. W. (1986). Density estimation for statisticas and data analysis. Chapman and Hall, London.
[12] Watson, G. and Leadbetter, M. (1964). Hazard analysis I, Biometrika, 51, 175-184.
[13] Zhang, S. and Karunamuni, R.J. (1998). On Kernel Density Estimation Near Endpoints. Journal of Statistical Planning and Inference, 70, 301-316.

Received: June 11, 2014