IUG Journal of Natural and Engineering Studies
Vol.20, No.1, pp 73-84 **2012**, ISSN 1726-6807, http://www.iugaza.edu.ps/ar/periodical/

Hazard Rate Function Estimation Using Inverse Gaussian Kernel

Raid B. Salha
Department of Mathematics, the Islamic University of Gaza
Email: rbsalha@iugaza.edu.ps

Abstract: In this paper, we consider the nonparametric estimation of the hazard rate function for independent identically distributed (iid) data using kernel estimation techniques. Since survival times are positive with potentially a high concentration at zero, one has to take into account the bias problem when the hazard rate function is estimated in the boundary region. To overcome the boundary bias problem, we use the Inverse Gaussian (IG) kernel, since it has a positive support.
The asymptotic mean squared error (AMSE) and the asymptotic normality of the proposed estimator are investigated. Also, the selection of an optimal bandwidth is discussed since it plays an important role in the kernel estimation.

Keywords: Inverse Gaussian kernel, hazard rate function, kernel estimation, asymptotic mean square error, boundary bias.

تقدير دالة معدل المخاطرة باستخدام نواة دالة جاوس العكسية

ملخص: في هذا البحث ندرس التقدير اللامعلمي لدالة المخاطرة باستخدام طرق تقدير النواة. ولأن تحليل البقاء يعتمد على بيانات ايجابية متركزة بالقرب من الصفر فيجب أن يؤخذ بعين الاعتبار مشكلة التحيز عند تقدير دالة المخاطرة في منطقة الحدود بالقرب من الصفر. وللتغلب على هذه المشكلة فإننا نستخدم تقدير معكوس جاوس والذي يتميز بأنه يمتلك دعم ايجابيا. كما تم في هذا البحث دراسة تقارب الخطأ التربيعي للتقدير المقترح وتقاربه لتوزيع طبيعي. بالإضافة لذلك تم مناقشة كيفية اختيار اتساع النافذ حيث إنها تلعب دورا مهما في تقدير النواة .

1. Introduction

In medical trails, reliability, survival analysis and many other fields, the occurrence of the event of interest called lifetime (or time to failure) forms the modeling basis, although often these times are not completely observed. Hazard rate estimation for the lifetime event is a basic tool for processing survival analysis.

Many methods for hazard estimation have been considered in the literature, and in particular nonparametric ones have known an important recent development. Estimation of the hazard rate by nonparametric methods has the advantage of flexibility because no formal assumptions are made about the mechanism that generates the sample order than the randomness.

Estimators of hazard function based on kernel smooth estimation have been studied extensively in literature. For example, see Watson and Leadbetter (1964), Rice and Rosenblatt (1976), Singpurwalla and Wong (1983) and

Salha (2009). However, when the support of the curve under estimation is bounded, many nonparametric estimators appear to be biased more than the usual in regions near the endpoints. To solve this problem, boundary kernels are used only within the boundary region. This is an efficient way to correct boundary bias but it requires complicated adjustment to the estimator.

For positive data, a natural way to overcome the boundary bias problem when estimating a density nonparametric ally is to consider kernels with positive support. Recently, Chen (2000) has proposed a nice way to circumvent the well known boundary bias or edge effect that appears in standard kernel density estimation. Boundary bias is due to weight allocation by the fixed symmetric kernel outside the density support when smoothing is carried out near the boundary. The remedy consists in replacing symmetric kernels by asymmetric Gamma kernel which never assigns weight outside the support, in addition to nice asymptotic features. Scaillet (2004) has used this idea and proposed two new classes of density estimators, rely on the use of inverse Gaussian (IG) and reciprocal inverse Gaussian (RIG) probability density function as kernels in place of the Gamma density function.

In this paper, we consider the nonparametric estimation of the hazard rate function for iid data using the Inverse Gaussian (IG) kernel estimation. As gamma kernel estimators, the IG kernel estimator is free of boundary bias, always non-negative, and achieves the optimal rate of convergence for the mean integrated squared error (MISE) within the class of nonnegative kernel density estimators. Furthermore its variance reduces s the position where the smoothing is made moves away from the boundary. In contrast with the gamma kernel estimators, the IG kernel estimator avoids the presence of the first derivatives of the probability density function in its bias, see Scaillet (2004). The asymptotic mean squared error (AMSE) and the asymptotic normality of the proposed estimator are investigated. Also, the selection of an optimal bandwidth is discussed in the last section, since it plays an important role.

2. Preliminaries

In this section, we state the conditions under which the results of the paper will be proved. Also, two important propositions from Scaillet (2004) are stated in Lemma 1.

Conditions

1. Let $X_1, X_2, ..., X_n$ be a random sample from a distribution with an unknown probability density function f defined on $[0,\infty)$, such that f is twice continuously differentiable, and $\int_0^\infty \left(x^3 f''(x)\right)^2 dx < \infty.$

2. h is a smoothing parameter satisfying $h + \frac{1}{nh} \to 0$, and $nh^{\frac{5}{2}} \to 0$ as $n \to \infty$.

Scaillet (2004) considered the following Inverse Gaussian kernel

$$K_{IG}(x, \frac{1}{h})(u) = \frac{1}{\sqrt{2p\, hu^3}} \exp\left(-\frac{1}{2hx}(\frac{u}{x} - 2 + \frac{x}{u}) \right), \tag{1}$$

If a random variable Y has a probability density functions $K_{IG}(x, \frac{1}{h})$ then $E(Y) = x$, and $Var(Y) = x^3 h$.

Scaillet (2004) has proposed the following estimator of the probability density function $f(\cdot)$, the Inverse Gaussian estimator,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{IG}(x, \frac{1}{h})(X_i). \tag{2}$$

Definition of the proposed estimator

The hazard rate function is defined as the instantaneous probability that duration X will end in the next time instant. More precisely, the hazard rate function is defined as

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

It can be shown that the hazard rate function can be written as the ratio of the density function $f(\cdot)$ and the survivor function $S(\cdot) = 1 - F(\cdot)$ of X, i.e.

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

The kernel estimator for the survivor function $S(\cdot)$ is constructed using the kernel density estimator in Equation (2)

$$\hat{S}(x) = 1 - \hat{F}(x),$$

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

Now, the proposed estimator for the hazard rate function is given by

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

Lemma 1.

Under the conditions (1) and (2), the following hold

(i) $Bias\left(\hat{f}(x)\right) = \frac{1}{2} x^3 f''(x) h + o(h).$

(ii) $Var\left(\hat{f}(x)\right) = \frac{1}{2n\sqrt{p h}} x^{-\frac{3}{2}} f(x) + o(n^{-1} h^{-\frac{1}{2}}).$

Proof. See Proposition 1 and 2 in Scaillet (2004).

3. Main Results

In this section, the two main results, Theorem 1 and Theorem 2 of this paper are stated and proved.

The first main result in this paper is stated in Theorem 1 which gives the asymptotic normality of the Inverse Gaussian kernel estimator of the probability density function.

Theorem 1.

Under the conditions (1) and (2), the following holds

$$\sqrt{nh^{\frac{1}{2}}}(\hat{f}(x) - f(x)) \xrightarrow{d} N\left(0, \frac{1}{2\sqrt{p}} x^{-\frac{3}{2}} f(x)\right).$$

Proof

Let $V_{ni} = K_{IG}(x, \frac{1}{h})(X_i),\ i = 1, 2, ..., n,$ then $\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} V_{ni}$, where $V_{ni},\ i = 1, 2, ..., n$ are independent identically distributed random variables.

Now, we show that Liapounov condition is satisfied, that is for some $d > 0$,

$$\lim_{n\to\infty} \frac{E\,|V_n - E(V_n)|^{2+d}}{n^{\frac{d}{2}} s^{2+d}(V_n)} = 0.$$

Let h_x be an Inverse Gaussian, $IG(x, \frac{2+d}{h}),$ distributed random variable.

Hence $m_x = E(h_x) = x$ and $T_x = Var(h_x) = \frac{x^3 h}{2+d}.$

$$E\,|V_n|^{2+d} = E\left[\left(\frac{1}{\sqrt{2\pi h y^3}}\right)^{2+d}\exp\left(-\frac{(2+d)}{2hx}\right)\left(\frac{y}{x}-2+\frac{x}{y}\right)\right]$$

$$=\frac{1}{(2\pi h)^{1+\frac{d}{2}}}\int_0^\infty y^{-3(1+\frac{d}{2})}\left[\exp\left(-\frac{(2+d)}{2hx}\right)\left(\frac{y}{x}-2+\frac{x}{y}\right)\right]f(y)\,dy$$

$$=\frac{\sqrt{2\pi h}}{\sqrt{2+d}\,(2\pi h)^{1+\frac{d}{2}}}\int_0^\infty y^{-\frac{3}{2}(1+d)}K_{IG}(x,\frac{2+d}{h})(y)f(y)\,dy$$

$$=\frac{\sqrt{2\pi h}}{\sqrt{2+d}\,(2\pi h)^{1+\frac{d}{2}}}E\left(h_x^{-\frac{3}{2}(1+d)}f(h_x)\right).$$

By using the Taylor's series to expand $f(h_x)$ about m_x, we obtain

$$h_x^{-\frac{3}{2}(1+d)}f(h_x)=m_x^{-\frac{3}{2}(1+d)}f(m_x)+(m_x^{-\frac{3}{2}(1+d)}f'(m_x)-\frac{3}{2}m_x^{-\frac{5}{2}(1+d)}f(m_x))(h_x-m_x)$$

$$+\frac{1}{2}(m_x^{-\frac{3}{2}(1+d)}f''(m_x)-\frac{3}{2}m_x^{-\frac{5}{2}(1+d)}f'(m_x)-\frac{3}{2}m_x^{-\frac{5}{2}(1+d)}f'(m_x)+\frac{15}{4}m_x^{-\frac{7}{2}(1+d)}f(m_x))(h_x-m_x)^2+o(h).$$

$$E\left(h_x^{-\frac{3}{2}(1+d)}f(h_x)\right)=x^{-\frac{3}{2}(1+d)}f(x)+\frac{1}{2}(x^{-\frac{3}{2}(1+d)}f''(x)-\frac{3}{2}x^{-\frac{5}{2}(1+d)}f'(x)$$

$$-\frac{3}{2}x^{-\frac{5}{2}(1+d)}f'(x)+\frac{15}{4}x^{-\frac{7}{2}(1+d)}f(x))T_x+o(h)$$

$$=x^{-\frac{3}{2}(1+d)}f(x)+\frac{1}{2(2+d)}(x^{-\frac{3}{2}(1+d)}f''(x)-\frac{3}{2}x^{-\frac{5}{2}(1+d)}f'(x)$$

$$-\frac{3}{2}x^{-\frac{5}{2}(1+d)}f'(x)+\frac{15}{4}x^{-\frac{7}{2}(1+d)}f(x))x^3h+o(h)$$

$$=x^{-\frac{3}{2}(1+d)}f(x)+o(h).$$

This implies that

$$\mathrm{E}|\mathrm{V_n}|^{2+d}=\frac{1}{\sqrt{2+d}\,(2\pi \mathrm{h})^{\frac{1+d}{2}}}\;x^{-\frac{3}{2}(1+d)}f(x)+o\left(h^{-\frac{(1+d)}{2}}\right).$$

Now, substituting $d = 0,$ the following holds

$$Var(V_n) = \frac{1}{2\sqrt{p}} h^{-\frac{1}{2}} x^{-\frac{3}{2}} f(x) + o(h^{-\frac{1}{2}}).$$

$$\frac{E\,|V_n - E(V_n)|^{2+d}}{n^{\frac{d}{2}} s^{2+d}(V_n)} \le \frac{E\,|V_n|^{2+d}}{n^{\frac{d}{2}}\left(\frac{1}{2\sqrt{p}} h^{-\frac{1}{2}} x^{-\frac{3}{2}} f(x)\right)^{\frac{2+d}{2}}} \to \frac{\frac{1}{\sqrt{2+d}\,(2p\mathrm{h})^{\frac{1+d}{2}}}\, x^{-\frac{3}{2}(1+d)} f(x)}{n^{\frac{d}{2}}\left(\frac{1}{2\sqrt{p}} h^{-\frac{1}{2}} x^{-\frac{3}{2}} f(x)\right)^{\frac{2+d}{2}}}$$

$$= \frac{\frac{1}{\sqrt{2+d}\,(2p)^{\frac{1+d}{2}}}\, x^{-\frac{3}{2}(1+d)} f(x)}{n^{\frac{d}{2}} h^{\frac{d}{4}}\left(\frac{1}{2\sqrt{p}} x^{-\frac{3}{2}} f(x)\right)^{\frac{2+d}{2}}} = \frac{\frac{1}{\sqrt{2+d}\,(2p)^{\frac{1+d}{2}}}\, x^{-\frac{3}{2}(1+d)} f(x)}{\left(nh^{\frac{1}{2}}\right)^{\frac{d}{2}}\left(\frac{1}{2\sqrt{p}} x^{-\frac{3}{2}} f(x)\right)^{\frac{2+d}{2}}} \to 0,$$

The last term vanishes as $n \to \infty$, since Condition (2) implies that $h \to 0$ and $nh \to \infty$, then $h^{\frac{1}{2}}$ goes to zero slower than h and this implies that $nh^{\frac{1}{2}} \to \infty.$ On the other hand, the remaining components of the last term are bounded by Condition (1).

This completes the proof of the theorem.

Now, Lemma 2 is stated and proved. Lemma 2 is important to derive the second main result in this paper. In this lemma it will be shown that the error in estimating the cumulative density function vanishes with probability.

Lemma 2.

Under the conditions (1) and (2), the following holds

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

Proof.

Firstly, from the definition of $\hat{F}(x)$, the following two facts in Relations (3) and (4) hold.

$$E\hat{F}(x)=\int_0^\infty\int_0^x K_{IG}(u,\frac{1}{h})(y)du\,f(y)dy=\int_0^x\int_0^\infty K_{IG}(u,\frac{1}{h})(y)f(y)dydu=\int_0^x E(f(x_u))du$$
$$=\int_0^x (f(u)+\frac{1}{2}u^3 f''(u)h)du+o(h)=F(x)+\frac{h}{2}\left[\int_0^x u^3 f''(u)du\right]+o(h)=F(x)+o(h).$$
$$E\hat{F}(x)-F(x)=o(h).$$

This implies that,

$$\sqrt{nh^{\frac{1}{2}}}\,|E\hat{F}(x)-F(x)|=o((nh^{\frac{5}{2}})^{\frac{1}{2}})\to 0. \qquad (3)$$

Now, $\hat{F}(x)$ can be written in the following form

$$\hat{F}(x)=\frac{1}{n}\sum_{i=1}^{n}\int_0^x K_{IG}(u,\frac{1}{h})(X_i)du=\frac{1}{n}\sum_{i=1}^{n}W_i(x),\ \text{where } W_i(x)=\int_0^x K_{IG}(u,\frac{1}{h})(X_i)du.$$

Let $e>0, d>0$ be given.

$$P\left[(nh^{\frac{1}{2}})^{\frac{1}{2}}|\hat{F}(x)-EF(x)|>e\right]\le e^{-2-2d}(nh^{\frac{1}{2}})^{1+d}E\,|\frac{1}{n}\sum_{i=1}^{n}[W_i(x)-EW_i(x)]|^{2+d}$$
$$=e^{-2-2d}h^{\frac{1+d}{2}}n^{-1-d}E\,|\sum_{i=1}^{n}[W_i(x)-EW_i(x)]|^{2+d}\le 2^{1+d}e^{-2-2d}(n^{-1}h^{\frac{1}{2}})^{1+d}\sum_{i=1}^{n}E\,|W_i(x)|^{2+2d}$$
$$+2^{1+d}e^{-2-2d}(n^{-1}h^{\frac{1}{2}})^{1+d}\sum_{i=1}^{n}|EW_i(x)|^{2+2d}.$$

$$(n^{-1}h^{\frac{1}{2}})^{1+d}\sum_{i=1}^{n}E\,|W_i(x)|^{2+2d}=(n^{-1}h^{\frac{1}{2}})^{1+d}n\int_0^\infty\int_0^x\left[K_{IG}(u,\frac{1}{h})(y)\right]^{2+2d}du\,f(y)dy$$
$$=n^{-d}h^{\frac{1+d}{2}}\int_0^x\int_0^\infty\left[K_{IG}(u,\frac{1}{h})(y)\right]^{2+2d}f(y)dydu$$

$$= n^{-d} h^{\frac{1+d}{2}} \int_0^x \frac{1}{\sqrt{2+2d}\,(2p\,h)^{\frac{1+2d}{2}}} u^{-\frac{3}{2}(1+2d)} f(u)du + o(h^{-\frac{1+2d}{2}}) \le C\, n^{-d} h^{\frac{1+d}{2}} h^{-\frac{1+2d}{2}}$$

$$= C\, n^{-d} h^{-\frac{d}{2}} = C\,(nh^{\frac{1}{2}})^{-d} \to 0.$$

This implies that,

$$\sqrt{nh^{\frac{1}{2}}}\, |\hat{F}(x) - EF(x)| \xrightarrow{p} 0. \qquad (4)$$

Secondly, using Relations (3), (4) and the following fact

$$|\hat{F}(x) - F(x)| \le |\hat{F}(x) - E\hat{F}(x)| + |E\hat{F}(x) - F(x)|,$$

we obtain that

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

This completes the proof of the lemma

The second main result in this paper is stated in Theorem 2 which gives the asymptotic normality of the Inverse Gaussian kernel estimator of the hazard rate function.

Theorem 2.

Under the conditions (1) and (2), the following holds

$$\sqrt{nh^{\frac{1}{2}}}\left(\hat{r}(x) - r(x)\right) \xrightarrow{d} N\left(0, \frac{1}{2\sqrt{p}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)}\right).$$

Proof.

$$\sqrt{nh^{\frac{1}{2}}}\left(\hat{r}(x)-r(x)\right)=\sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}(x)}{\hat{S}(x)}-\frac{f(x)}{S(x)}\right)=\sqrt{nh^{\frac{1}{2}}}\left(\frac{\hat{f}(x)}{\hat{S}(x)}-\frac{f(x)}{\hat{S}(x)}-\frac{f(x)}{S(x)}+\frac{f(x)}{\hat{S}(x)}\right)$$

$$=\frac{\sqrt{nh^{\frac{1}{2}}}}{\hat{S}(x)}\left(\hat{f}(x)-f(x)-\frac{f(x)\hat{f}(x)}{S(x)}+f(x)\right)$$

$$=\frac{\sqrt{nh^{\frac{1}{2}}}}{\hat{S}(x)}\left[\hat{f}(x)-f(x)\right]+\frac{\sqrt{nh^{\frac{1}{2}}}f(x)}{S(x)\hat{S}(x)}\left[\hat{S}(x)-S(x)\right]. \qquad (5)$$

The proof of the theorem is completed by a combination of Theorem 1, Lemma 2 and Equation (5). Since by Theorem 1, the first term in Equation (5) is asymptotically normally distributed and the second term vanishes by Lemma 2.

From Theorem 1 and 2, we get that

$$E\left(\hat{r}(x)\right)=\frac{E\left(\hat{f}(x)\right)}{E\left(\hat{S}(x)\right)}=\frac{f(x)+\frac{1}{2}x^{3}f''(x)h}{S(x)}+o(h)=r(x)+\frac{\frac{1}{2}x^{3}f''(x)h}{S(x)}+o(h),$$

$$Bias\left(\hat{r}(x)\right)=\frac{\frac{1}{2}x^{3}f''(x)h}{S(x)}+o(h)$$

and

$$Var\left(\hat{r}(x)\right)=\frac{1}{2n\sqrt{p\,h}}x^{-\frac{3}{2}}\frac{r(x)}{S(x)}+o(n^{-1}h^{-\frac{1}{2}}).$$

3. Bandwidth selection

The selection of the bandwidth in kernel estimation plays an important role. It depends on choosing a value of the bandwidth that minimizing the asymptotic mean squared error. Using the same techniques of Scaillet (2004), we obtain the following:

The asymptotic mean square error (AMSE) is given by

$$\text{AMSE} = \left(\frac{\frac{1}{2} x^{3} f''(x) h}{S(x)} \right)^{2} + \frac{1}{2n\sqrt{p\,h}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)}. \tag{6}$$

Note that the AMSE of the proposed estimator for points near zero is smaller than that of the Gaussian kernel estimator.

$$\text{AMSE(for Gaussian Kernel)} = \left(\frac{\frac{1}{2} f''(x) h^{2}}{S(x)} \right)^{2} + \frac{1}{2nh\sqrt{p}} x^{-\frac{3}{2}} \frac{r(x)}{S(x)}.$$

Now, $f(x)$ and not on x. Note that the AMSE depends only on
then equating it to zero and solving h differentiate the AMSE with respect to
, we obtain h for

$$h = \left(\frac{1}{2\sqrt{p}} \frac{f(x)}{(f''(x))^{2}} \right)^{\frac{2}{5}} x^{-3} n^{-\frac{2}{5}}. \tag{7}$$

Substitute Equation (7) in Equation (6), we get

$$\text{AMSE} = \left(\frac{x^{3} f''(x)}{2S(x)} \left(\frac{f(x)}{2\sqrt{p}\,(f''(x))^{2}} \right)^{\frac{2}{5}} x^{-3} n^{-\frac{2}{5}} \right)^{2}$$

$$+ \frac{x^{-\frac{3}{2}} r(x)}{2n\sqrt{p}\,S(x)} \left(\left(\frac{f(x)}{2\sqrt{p}\,(f''(x))^{2}} \right)^{\frac{2}{5}} x^{-3} n^{-\frac{2}{5}} \right)^{-\frac{1}{2}}$$

$$= \left(\frac{(f(x))^{\frac{2}{5}} (f''(x))^{\frac{1}{5}}}{2\left(\sqrt{p}\right)^{\frac{2}{5}} S(x)} \right)^{2} n^{-\frac{4}{5}} + \frac{r(x)\left(f(x)\right)^{-\frac{1}{5}} n^{-\frac{4}{5}}}{\left(2\sqrt{p}\right)^{\frac{4}{5}} S^{2}(x) \left(f''(x)\right)^{-\frac{2}{5}}}$$

$$= \frac{5}{4}\left(\frac{f(x)}{2\sqrt{p}}\right)^{\frac{4}{5}} \frac{\left(f''(x)\right)^{\frac{2}{5}}}{S^2(x)} n^{-\frac{4}{5}}.$$

Regarding global properties, the optimal bandwidth h^* and AMISE are

$$h^* = \left(\frac{\frac{1}{2\sqrt{p}} \int_0^{\infty} x^{-3/2} f(x)\,dx}{\int_0^{\infty} (x^3 f''(x))^2 dx} \right)^{\frac{2}{5}} n^{-\frac{2}{5}},$$

$$\text{AMISE} = \frac{5}{4}\left(\frac{1}{2\sqrt{p}} \int_0^{\infty} x^{-3/2} f(x)\,dx \right)^{\frac{4}{5}} \frac{\left(\int_0^{\infty} (x^3 f''(x))^2 dx\right)^{\frac{2}{5}}}{S^2(x)} n^{-\frac{4}{5}}.$$

In practice, the bandwidth selection can be done by using the rule in Equation (8) which was proposed by Scalliet (2004)

$$h^{**} = \left(\frac{16 s^5 \exp(1/8\,(7s^2 - 20m))}{12 + 68 s^2 + 225 s^4} \right)^{\frac{2}{5}} n^{-\frac{2}{5}}, \qquad (8)$$

where the unknown parameters m and s^2 are estimated by the arithmetic mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and the sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ respectively.

4. Conclusion

This paper makes use of the Inverse Gaussian kernel to estimate nonparametrically the marginal density and the hazard rate function. The estimator use adaptive weights depending on the point in which we estimate the functions. Also, the new proposed estimator can be modified but considering a variable bandwidth depending on the point in which we estimate the function. We derive the asymptotic mean square error, the asymptotic normality and the strong consistency of the proposed estimator.

The AMSE of the proposed estimator is smaller than that of the Gaussian kernel estimator for points near zero.

References

[1] Chen, S. X. (2000). Probability density function estimation using gamma kernels. Ann. Inst. Stat. Math., 52, 471- 480.

[2] J. Rice and M. Rosenblatt (1976). Estimation of the log survivor function and hazard function. Sankhya Series A 38, 60-78.

[3] N. D. Singpurwalla and M. Y. Wong (1983). Kernel estimation of the failure rate function and density estimation: an analogy. J. Amer. Statist. Assoc. 78(3), 478-481.

[4] O. Scaillet (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels. Nonparametric Statistics, 16, 217-226.

[5] Salha, R (2009). Adaptive kernel estimation of the hazard rate function. The Islamic university – Gaza Journal, 17(1), 71-81.

[6] Silverman, B. W. (1986). Density estimation for statistics and data analysis. Chapman and Hall, London.

[7] Wand, M. P. and Jones, M. C. (1995). Kernel smoothing, Chapman and Hall.

[8] Watson, G. and Leadbetter, M. (1964). Hazard analysis I, Biometrika, 51, 175-184.