

Gamma Kernel Estimators for Density and Hazard Rate of Right Censored Data

T. Bouezmarni* A. El Ghouch[†] and M. Mesfioui[‡]

March 28, 2011

Abstract

The nonparametric estimation for the density and hazard rate functions for right censored data using the kernel smoothing techniques is considered. The "classical" fixed symmetric kernel type estimator of these functions performs well in the interior region, but it suffers from the problem of bias in the boundary region. Here, we propose new estimators based on the gamma kernels for the density and the hazard rate functions. The estimators are free of bias and achieve the optimal rate of convergence in terms of integrated mean squared error. The mean integrated squared error, the asymptotic normality and the law of iterated logarithm are studied. A comparison of gamma estimators with the local linear estimator for the density function and with hazard rate estimator proposed by Müller and Wang (1994), which are free from boundary bias, is investigated by simulations.

Key words and phrases. Asymmetric kernels, Kaplan-Meier, density and hazard functions, asymptotic normality, law iterated logarithm.

JEL Classification : C13,C14. *MSC 2000 :* 62G07, 62G08

*Département de mathématiques, Université de Sherbrooke, Canada.

[†]Institut de statistique, biostatistique et sciences actuarielles, Université Catholique de Louvain, Belgium.

[‡]Département de mathématique et d'informatique, Université du Québec à Trois Rivières, Canada,

1 Introduction

Censored data arise in many contexts, for example, in medical follow-up studies in which the occurrence of the event times (called survival) of individuals may be prevented by the previous occurrence of another competing event (called censoring). In such studies, interest focuses on estimating the underlying density and/or hazard rate of the survival time. Non-parametric estimation using kernel smoothing method has received considerable attention in the statistical literature. A popular approach for estimating the density function and the hazard rate function is done using a fixed symmetric kernel density with bounded support and a bandwidth parameter. The kernel determines the shape of the local neighborhood while the bandwidth controls the degree of smoothness. In order to get a reasonable estimator, these two parameters, the kernel and the bandwidth parameter, have to be chosen carefully. For a review about kernel smoothing approaches, we refer the reader to Silverman (1986) and Izenman (1991) for uncensored data and Singpurwalla and Wong (1983), Tanner and Wong (1983), Padgett and McNichols (1984), Mielniczuk (1986) and Lo, Mack, and Wang (1989) in the case of right censoring.

It is well known from the literature that the kernel has less impact than the bandwidth on the resulting estimate. However, when the density function of the data have a bounded support, using the classical kernel leads to an estimator with a large bias near the endpoints. The problem of bias, called also the boundary effect, becomes a serious drawback when a large portion of the sampled data are present in the boundary region. In fact, when we estimate the underlying function near the endpoints, as the support of the kernel exceeds the available range of the data, the bias of the resulting estimator becomes larger. This is especially the case in survival analysis, since the survival time is assumed to be nonnegative variable. So, near zero, the symmetric kernel estimator of the density and the hazard functions underestimates the true ones.

For uncensored data, several methods are available in the literature to overcome this problem. For example, the reflection method of Schuster (1985), the smooth optimum kernel of Müller (1991), the local linear estimator of Lejeune and Sarda (1992), the transformation

approach of Marron and Ruppert (1994) and the boundary kernel of Jones (1993) and Jones and Foster (1996). The local linear estimator is a special case of boundary kernel method. The main idea behind the boundary kernel method is to use an adaptive kernels in the boundary region and to use a fixed symmetric kernel in the interior region. For nonnegative data and in order to overcome the boundary bias problem, Chen (2000) considers the gamma kernel estimator. Simulation results of Jones (1993) and Chen (2000) show that the local linear estimator outperforms the boundary kernel estimator of Müller (1991). For right censored data and to resolve this problem, Müller and Wang (1994) proposes a new class of polynomial boundary kernel estimator for hazard rate function, where the kernel and the bandwidth parameter depend on the point where the estimate is to be evaluated. Hess, Serachitopol, and Brown (1999) show numerically its performance via an extensive simulation study.

In this paper, we adapt the gamma kernel smoothing procedure to estimate the marginal density and the hazard function of positive independent survival data that are subject to right censoring. We show that both estimators are robust against boundary problems. Also, we establish the mean integrated square error, the asymptotic normality and the law of iterated logarithm of the two estimators. Furthermore, via a Monte Carlo study, the finite sample performance of the estimators are investigated under various scenarios.

The paper is organized as follows. Section 2 introduces the gamma kernel estimators for the density and the hazard rate functions for right censored data. In Section 3, we establish the asymptotic properties of the gamma kernel estimators. In Section 4, we investigate the finite sample properties of the gamma kernel estimators. Section 5 provides an application to the classical bone marrow transplant data. The last section is an appendix that gathers the proofs.

2 Methodology

Let T_1, \dots, T_n (survival times) and C_1, \dots, C_n (censoring times) be two i.i.d. nonnegative independent random sequences with distribution functions F and G , respectively. Under the

censoring model, instead of observing T_i , we observe the pair (X_i, δ_i) , where $X_i = \min(T_i, C_i)$ and $\delta_i = I(Y_i \leq C_i)$ with $I(\cdot)$ being the indicator function. We denote by f the density function of F and by $h = f/(1 - F)$ the corresponding hazard function. The nonparametric maximum likelihood approach proposed by Kaplan and Meier (1958) leads to the estimator of F given by

$$\hat{F}(x) = \begin{cases} 1 - \prod_{i:1 \leq i \leq n, X_{(i)} \leq x} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{if } x < X_{(n)} \\ 1 & \text{Otherwise.} \end{cases}$$

where $X_{(1)} \leq \dots \leq X_{(n)}$ and $\delta_{(1)}, \dots, \delta_{(n)}$ are the corresponding δ_i 's. This estimator was studied by many authors. For reference we cite Breslow and Crowley (1974), Wang (1987) and Stute and Wang (1993) among many authors. Lo and Singh (1986) expressed the KM estimator as an i.i.d. mean process with a remainder of negligible order. This result was improved by Lo, Mack, and Wang (1989) and will be useful in this paper to make the connection between the uncensored and censored case.

From now on, we will denote the right endpoint of a given (sub)distribution L by T_L , i.e. $T_L = \sup\{t \geq 0, L(t) < 1\}$. We will also use the notation \bar{L} for the corresponding survival function, i.e. $\bar{L}(\cdot) = 1 - L(\cdot)$. Let H be the distribution function of X , that is $\bar{H} = \bar{F}\bar{G}$. We suppose that $T_G \leq T_F$ or equivalently $T_H = T_F$. In the remainder of this paper, except if mentioned otherwise, all the integrations are taken over the interval $[0, T]$, where $T = T_F$. Based on the smooth kernel technique, we propose to estimate the density by the gamma kernel estimator defined as follows:

$$\hat{f}_b(x) = \int K(x, b)(t) d\hat{F}(t) = \sum_{i=1}^n K(x, b)(X_{(i)}) \mathbb{W}_i \quad (1)$$

where the kernel K is given by

$$K(x, b)(t) = \frac{t^{\rho_b(x)-1} \exp(-t/b)}{b^{\rho_b(x)} \Gamma(\rho_b(x))} \text{ and } \rho_b(x) = \begin{cases} x/b & \text{if } x \geq 2b \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } x \in [0, 2b), \end{cases} \quad (2)$$

the weights \mathbb{W}_i 's are the jumps of \hat{F} at X_i for $i = 2, \dots, n$, and $0 < b \equiv b_n \rightarrow 0$ is the bandwidth parameter.

Naturally the gamma kernel estimator that we propose for the hazard rate h is:

$$h_b(x) = \hat{f}_b(x)/(1 - \hat{F}(x)). \quad (3)$$

As we will see later, those estimators are free of boundary bias, this is due to the fact that the gamma kernel is defined on the positive real and so no weight is assigned outside the support of the underlying density and/or hazard rate functions. The shape of the gamma kernel and the amount of smoothing are not only controlled by the bandwidth parameter, but vary also according to the position where the function is estimated. For uncensored data, Chen (2000) shows that the gamma kernel density estimator achieves the optimal rate of convergence for the mean integrated squared error within the class of non-negative kernel density estimators. Bouezmarni and Scaillet (2005) states the uniform weak consistency for the gamma kernel estimator on any compact set and also the weak convergence in terms of mean integrated absolute error. In the next section, we will prove that even when the data are censored, the gamma kernel estimators perform in the interior and the boundary regions.

3 Asymptotic properties

In this section we state the asymptotic properties of the gamma kernel estimator of the density and the hazard rate functions. We start with the following theorem which will play an important role for the reminder of this section. To be concise, in the following, we will denote by Z either the density or the hazard rate function and \hat{Z}_b either the gamma kernel estimator for the density or the hazard function.

Theorem 1 *Assume that f is twice continuously differentiable. If (i) $\log^2 n/(nb_n^{3/2}) \rightarrow 0$, then the integrated mean squared error of \hat{Z}_b is*

$$IMSE(\hat{Z}_b) = b^2 \int B^2(x)dx + n^{-1}b^{-1/2} \int V(x)dx + o(b^2) + o(n^{-1}b^{-1/2})$$

where for the density function B and V are given by

$$B_f(x) = \frac{1}{2}xf''(x), \quad \text{and} \quad V_f(x) = \frac{1}{2\sqrt{\pi}} \frac{x^{-1/2}f(x)}{\bar{G}(x)},$$

and for the hazard rate function B and V are given by

$$B_h(x) = \frac{B_f(x)}{\bar{F}(x)}, \quad \text{and} \quad V_h(x) = \frac{V_f(x)}{\bar{F}^2(x)}.$$

The optimal bandwidth parameter which minimizes the asymptotic $IMSE(\hat{Z}_b)$ is given by

$$b^* = \left(\frac{1}{4} \frac{\int V(x)dx}{\int B(x)dx} \right)^{2/5} n^{-2/5},$$

and the corresponding optimal asymptotic integrated mean squared error is

$$IMSE(\hat{Z}_b) \sim \frac{5}{4^{4/5}} \frac{(\int V(x)dx)^{4/5}}{(\int B(x)dx)^{1/5}} n^{-4/5}.$$

■

Theorem 1 states that the gamma kernel estimators of the density and hazard rate functions are free of boundary bias and provides the theoretical formula of the optimal bandwidth. However, in real data analysis, to choose an appropriate bandwidth, one needs to use a data driven procedure, for example: the cross-validation, the bootstrap or the method proposed by Bouezmarni and Scaillet (2005). Of course, all those methods need to be carefully adapted to the censoring case. Also, from equation (8) in Appendix, the asymptotic variances of the gamma kernel estimator are of a larger order $O(n^{-1}b^{-1})$ near the boundaries than those $O(n^{-1}b^{-1/2})$ in the interior. However, Theorem 1 shows that the impact of the increased variance near the boundary on the mean integrated squared error is negligible and the optimal rate of convergence in term of integrated mean squared error is achieved by the gamma kernel estimators.

The following proposition deals with the asymptotic normality of the gamma kernel estimators.

Proposition 1 *Under the same conditions in Theorem 1. If (ii) $nb^{5/2} = o(1)$, then for any x such that $f(x) > 0$ and $x \leq T$, we have*

$$\left(n^{1/2} b^{1/4} \frac{\hat{Z}_b(x) - Z(x)}{\sqrt{V^*(x)}} \right) \rightarrow N(0, 1), \text{ in distribution,}$$

where

$$V^*(x) = \begin{cases} V(x) & \text{if } x/b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1)b^{-1/2}}{2^{1+2\kappa}\Gamma^2(\kappa+1)} 2\sqrt{\pi}x^{1/2}V(x) & \text{if } x/b \rightarrow \kappa. \end{cases}$$

■

We establish in the next proposition the law of iterated logarithm of the gamma kernel estimators. Let $\Phi_x(n, b) = 2n^{-1}b^{-1/2}V^*(x) \log \log(n)$, where V^* is defined in Proposition 1.

Proposition 2 *Under the same conditions in Theorem 1. If,*

$$\text{(iii) } \frac{b_n}{b_m} \rightarrow 1, \text{ as } n, m \rightarrow \infty \text{ such that } \frac{n}{m} \rightarrow 1 \quad \text{and} \quad \frac{2 \log^4(n)}{(nb)^2 \Phi_x(n, b)} \rightarrow 0,$$

and

$$\text{(iv) } b_n = o(\log \log(n)/n)^{2/5},$$

then,

$$\limsup_n (\Phi_x(n, b))^{-1/2} |\hat{Z}_b(x) - Z(x)| = 1, \quad a.s. \tag{4}$$

■

4 Simulations results

The finite sample performance of the proposed methodology is studied in this section. Two models are considered:

- **Model A:** the survival time follows an exponential distribution and the censoring times were generated from an uniform distribution $[0, c]$. c was chosen to be the solution of the following equation $pc + \exp(-c) = 1$, where p is the desired percentage of censoring.
- **Model B:** the survival times follow a Weibull distribution with scale parameter $b = 2$ and shape parameter $a = 1.2$ and the censoring times are also generated from a Weibull distribution with shape parameter a and a scale parameter given by $b \left(\frac{1-p}{p}\right)^{1/a}$. This ensures that the degree of censoring is equal to p .

First, we show the results for the density estimator. To evaluate the performance of the gamma kernel estimator for the density function, we compare this later with the local linear estimator of Jones (1993) adapted for the right censoring case. The local linear method is known to be a robust technique against boundary bias problems. We consider different sample sizes, $n = 125, 250, 500, 1000$. As a measure of errors of the estimators we analyze the mean and the standard deviation of the L_2 -norm.

Table 1 and Table 2 provide the results obtained with 1000 replications for the mean and the standard deviation, respectively. Firstly and as expected, when the sample size increases, the mean integrated square error for the two estimators decreases; see Table 1. This is true for both models and all degrees of censoring. For example in model A, we can see that with 10% of censoring and using the gamma kernel estimator, the mean error decreases from 0.0117 to 0.0101 when the sample size goes from 125 to 250. Note that, for the 50% of censoring the rate at which the mean error decreases is much smaller. Secondly, except the 50% censoring case with model A, the gamma kernel estimator outperforms strongly the local linear kernel estimator. Thirdly, for model B, when the degree of censoring increases the mean integrated square error increases as expected. In fact, for model B, the cumulative distribution of the censoring times increases when the degree of censoring p increases. This implies that the mean integrated squared error of the gamma kernel estimators increases. From Table 2, one can see that the variance of both models and both estimators decreases with the sample size, for all the degree of censoring. For model B, the variance of gamma kernel estimator increases with the degree of censoring. Another point to remark is that for model B the

local linear estimator is dominated in terms of variance by gamma kernel estimator in all situations. For model A, the variance of the gamma kernel estimator is smaller than the variance of the local linear estimator for $p=0.1$ and 0.25 , but not for $p=0.5$.

For the hazard function, we compare our gamma kernel estimator with the boundary corrected hazard estimator of Müller and Wang (1994). Table 3 shows the mean and the variance of L_2 errors based on 1000 replications for both $n = 125$ and $n = 250$. The same data generating procedure, as for density function, was used. This results clearly demonstrate that our method is far more efficient than the method proposed by Müller and Wang (1994).

5 Application

To illustrate our approach with real data we consider the classical bone marrow transplant study. The variable of interest is the disease-free survival time, i.e. time to relapse or death. Among a total of 137 observations, there is 83 censored times all of them are caused by the end of the follow up period. The data together with a detailed description of the study can be found in Section 1.3 of Klein and Moeschberger (2003). Figure 1 shows the estimated hazard function using both our method (G) and that proposed by Müller and Wang (1994)(MW) with boundary correction. For the Muller estimator we use the data driven global optimal bandwidth as proposed by the authors. To calculate the MW estimator and its bandwidth parameter we make use of the R package *muhaz*; see Hess and Gentleman (2010). The estimator of Müller and Wang (1994) shown in Figure 1 is based on the bandwidth $b = 204.5$ while for the gamma kernel estimator $b = 153.4$. This is a data driven bandwidth based on the least squared cross-validation method of Marron and Padgett (1987) adapted to our case. Note that this is just a practical choice and may be far from the optimal one. From Figure 1, one can not decide which estimator is better, but given the results of our simulation study, we suspect that the MW method has a global tendency of underestimating the real hazard function and so the real risks after a bone marrow transplant.

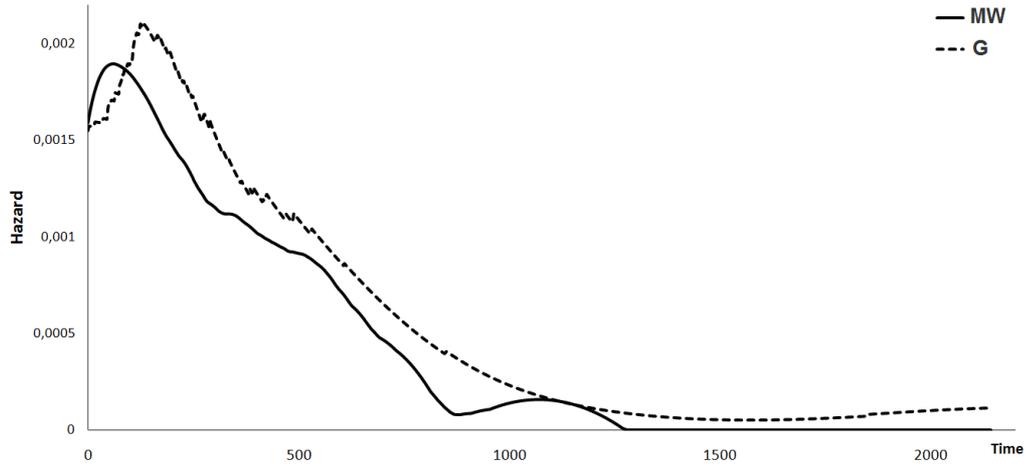


Figure 1: Gamma(G) and Müller and Wang(MW) estimators of the hazard function for the bone marrow transplant data.

Table 1: Mean ($\times 10^{-2}$) of L_2 error for the density function estimators.

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	1.17	1.55	1.01	1.41	0.94	1.34	0.93	1.31
	25	1.48	1.96	1.27	1.83	1.24	1.80	1.18	1.77
	50	3.15	2.01	3.11	1.95	3.09	1.90	3.06	1.85
B	10	1.08	1.44	0.76	1.02	0.58	0.73	0.44	0.52
	25	1.65	3.50	1.56	3.22	1.50	2.99	1.42	2.72
	50	3.08	9.26	3.04	9.16	3.02	9.13	2.91	9.40

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Table 2: Standard deviation ($\times 10^{-2}$) of L_2 error for the density function estimators

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$	
Model	% cens	G	LL	G	LL	G	LL	G	LL
A	10	5.67	8.42	4.12	6.97	3.26	5.80	2.99	5.51
	25	6.81	9.96	4.91	8.11	4.06	7.02	3.18	6.115
	50	8.47	1.14	8.30	1.03	7.99	0.99	7.84	0.93
B	10	5.32	11.23	3.51	8.12	2.81	5.39	2.25	3.125
	25	5.30	21.98	4.52	18.6	3.52	14.87	2.64	10.64
	50	6.79	13.72	5.13	8.82	4.07	7.77	3.09	5.12

G: gamma estimator. LL: local linear estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Table 3: Mean and Standard deviation ($\times 10^{-2}$) of L_2 error for the hazard rate function estimators.

		$n = 125$				$n = 250$			
		Mean		Variance		Mean		Variance	
Model	% cens	G	MW	G	MW	G	MW	G	MW
A	10	6.00	12.46	0.36	0.67	3.20	8.09	0.08	0.26
	25	5.90	14.04	0.16	0.45	4.95	13.09	0.07	0.30
	50	13.20	20.00	0.38	0.64	13.13	19.33	0.22	0.35
B	10	2.32	8.87	0.04	0.27	1.04	5.51	0.02	0.08
	25	2.64	10.04	0.03	0.34	2.27	8.49	0.02	0.09
	50	8.52	11.87	0.10	0.18	8.93	11.99	0.05	0.09

G: gamma estimator. MW: Müller and Wang estimator. Model A: exponential survival times with uniform censoring times. Model B: Weibull survival times with Weibull censoring times. The results are based on 1000 replications.

Appendix

We start this section with some notations. Let

$$g(t) = \int_0^t (1 - H(s))^{-2} dH^1(s),$$

where $H^1(t) := P(X_i \leq t, \delta_i = 1) = \int_0^t (1 - G(s)) dF(s)$ denote the sub-distribution of the uncensored observations. For positive real numbers z and x , and $\delta = 0$ or 1 , let

$$\xi(z, \delta, t) = -g(z \wedge t) + (1 - H)^{-1} I(z \leq t, \delta = 1).$$

We also set $\xi_i(x) = \xi(X_i, \delta_i, x)$. First note that

$$\mathbb{E}(\xi_i(x)) = 0 \text{ and } Cov(\xi_i(t), \xi_i(s)) = g(s \wedge t).$$

The following two lemmas will play an important role in the demonstrations. The first lemma is a du to Lo, Mack, and Wang (1989) which expressed the KM estimator as an i.i.d. mean process with a remainder of negligible order.

Lemma 1 (*Lo, Mack, and Wang (1989)*)

For $t \leq T$

$$\hat{F}(t) - F(t) = n^{-1} \bar{F}(t) \sum_{i=1}^n \xi_i(t) + r_n(t),$$

where

$$\sup_{0 \leq x \leq T} |r_n(x)| = O(\log n/n) \quad a.s.,$$

and for any $\alpha \geq 1$,

$$\sup_{0 \leq x \leq T} \mathbb{E}|r_n(x)|^\alpha = O([\log n/n]^\alpha).$$

■

The next lemma gives a strong approximation of the gamma kernel estimator of the density. This lemma permits us to derive the asymptotic properties of the gamma kernel estimator for density and hazard rate function.

Lemma 2 *The gamma kernel density estimator f_b admits the strong approximation on the interval $[0, T]$:*

$$\hat{f}_b(x) = f(x) + \beta_n(x) + \sigma_n(x) + e_n(x),$$

where

$$\beta_n(x) = \int f(t)K(x, b)(t)dt - f(x), \quad \sigma_n(x) = n^{-1} \sum_n \int \xi_i(t)dK(x, b)(t) \quad (5)$$

and

$$\sup_{0 \leq x \leq T} |e_n(x)| = O(\log n / (nb)) \quad a.s..$$

Also, for any $\alpha \geq 1$,

$$\sup_{0 \leq x \leq T} \mathbb{E}|e_n(x)|^\alpha = O([\log n / (nb)]^\alpha).$$

■

Proof of Lemma 2.

Let us first show that $K(x, b)(T)$ can be made arbitrary small for any $x < T$. From

$$\Gamma(w + 1) \sim \sqrt{2\pi} \exp(w)w^{w+1/2}, \quad \text{as } w \rightarrow \infty$$

it follows that for $s = x - b$,

$$\begin{aligned} K(x, b)(T) &= \frac{T^{s/b} \exp(-T/b)}{b^{s/b+1} \Gamma(s/b + 1)} \\ &\sim \frac{(bs)^{-1/2}}{\sqrt{2\pi}} \exp \left\{ \frac{s}{b} \left(1 - \frac{T}{s} + \log\left(\frac{T}{s}\right) \right) \right\} \end{aligned}$$

Observe that, for any $w > 1$, $1 - w + \log(w) < 0$. So, $K(x, b)(T) = o(b^\alpha)$ for any $\alpha \geq 1$.

Now, using this result, the integration by part and Lemma 1, we obtain

$$\begin{aligned} \hat{f}_b(x) &= \int K(x, b)(t)d\hat{F}(t) \\ &\approx - \int \hat{F}(t)dK(x, b)(t) \\ &= - \int [F(t) + n^{-1} \sum_n \xi_i(t) + r_n(t),]dK(x, b)(t) \\ &\approx \int f(t)K(x, b)(t)dt - n^{-1} \sum_n \int \xi_i(t)dK(x, b)(t) - \int r_n(t)dK(x, b)(t) \\ &= f(x) + \beta_n(x) + \sigma_n(x) + e_n(x). \end{aligned}$$

We deduce the result of Lemma 2 by using the following inequality,

$$\begin{aligned} \int_0^{+\infty} |dK(x, b)(t)| &= b^{-1} \int_0^{+\infty} |K(x, b)(t) - K(x - b, b)(t)| dt \\ &\leq 2b^{-1}. \end{aligned} \quad (6)$$

Proof of Theorem 1. We start with the gamma kernel density estimator. From the asymptotic bias of the gamma kernel estimator for uncensored data (see Chen (2000)) and the fact that for the interior region

$$(\log n/(nb_n))/(n^{-1/2}b^{-1/4}) = (\log n/(n^{1/2}b^{3/4})) \longrightarrow 0,$$

and for the boundary region

$$(\log n/(nb_n))/(n^{-1/2}b^{-1/2}) = (\log n/(n^{1/2}b^{1/2})) \longrightarrow 0,$$

it can be easily verified that, in our case, the bias is given by

$$\mathbb{E}(\hat{f}_b(x)) - f(x) = \begin{cases} \frac{1}{2}xf''(x)b + o(b) + o(n^{-1/2}b^{-1/4}) & \text{if } x \geq 2b \\ \xi_b(x)bf'(x) + o(b) + o(n^{-1/2}b^{-1/2}) & \text{if } x \in [0, 2b). \end{cases} \quad (7)$$

where $\xi_b(x) = (1 - x)(\rho_b(x) - x/b)/(1 + b\rho_b(x) - x)$.

To calculate the asymptotic variance of the gamma kernel estimator, we only need to evaluate the variance of $\sigma_n(x)$ since the other terms are negligible. We start with the fact that

$$Var(\sigma_n(x)) = n^{-1} \int \int \bar{F}(t)\bar{F}(s)g(t \wedge s)dk(t)dk(s).$$

where $k(\cdot) = k(x, b)(\cdot)$. Using integration by parts, the first integral becomes

$$\begin{aligned} \int \bar{F}(t)g(t \wedge s)dk(t) &= \int_0^s \bar{F}(t)g(t)dk(t) + \int_{t \geq s} \bar{F}(t)g(s)dk(t) \\ &= [\bar{F}(t)g(t)k(t)]_0^s - \int_0^s (\bar{F}(t)g(t))'k(t)dt \\ &+ g(s)[\bar{F}(t)k(t)]_{t \geq s} - g(s) \int_{t \geq s} \bar{F}(t)'k(t)dt \\ &= - \int_0^s (\bar{F}(t)g(t))'k(t)dt - g(s) \int_s^\infty \bar{F}(t)'k(t)dt \end{aligned}$$

So that,

$$\begin{aligned}
nVar(\sigma_n(x)) &= - \int \bar{F}(s) \int_0^s (\bar{F}(t)g(t))'k(t)dt dk(s) - \int \bar{F}(s)g(s) \int_{t \geq s} \bar{F}(t)'k(t)dk(s)dt \\
&= - \int (\bar{F}(t)g(t))'k(t) \left(\int_{s \geq t} \bar{F}(s)dk(s) \right) dt - \int \bar{F}(t)'k(t) \left(\int_0^t \bar{F}(s)g(s)dk(s) \right) dt \\
&= - \int (\bar{F}(t)g'(t)k(t) + \bar{F}'(t)g(t)k(t)) \left(\int_{s \geq t} \bar{F}(s)dk(s) \right) dt \\
&\quad - \int \bar{F}(t)'k(t) \int_0^t \bar{F}(s)g(s)dk(s)dt
\end{aligned}$$

Again, using the integration by parts we obtain,

$$\begin{aligned}
\int (\bar{F}(t)g'(t)k(t)) \left(\int_{s \geq t} \bar{F}(s)dk(s) \right) dt &= - \int \bar{F}^2(t)g'(t)k^2(t)dt + O(1), \\
\int (\bar{F}'(t)g(t)k(t)) \left(\int_{s \geq t} \bar{F}(s)dk(s) \right) dt &= - \int \bar{F}'(t)\bar{F}(t)g(t)k^2(t)dt + O(1)
\end{aligned}$$

and

$$\int \bar{F}(t)'k(t) \int_0^t \bar{F}(s)g(s)dk(s)dt = \int \bar{F}'(t)\bar{F}(t)g(t)k^2(t)dt + O(1)$$

Therefore, from $g'(t) = f(t)/[\bar{G}(t)\bar{F}(t)^2]$, we get

$$\begin{aligned}
nVar(\sigma_n(x)) &= \int \bar{F}^2(t)g'(t)k^2(t)dt \\
&= \int f(t)/[\bar{G}(t)]k^2(t)dt \\
&= B_b(x)E(f(\eta_x)\bar{G}^{-1}(\eta_x))
\end{aligned}$$

where η_x is a gamma($2x/b + 1, b/2$) random variable and

$$B_b(x) = \frac{b^{-1}\Gamma(2x/b + 1)}{2^{2x/b+1}\Gamma^2(x/b + 1)}$$

From Chen(2000), we have that for a small value of b ,

$$B_b(x) \sim \begin{cases} \frac{1}{2\sqrt{\pi}}b^{-1/2}x^{-1/2} & \text{if } x/b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}b^{-1} & \text{if } x/b \rightarrow \kappa \end{cases}$$

So after some easy development, we get

$$n\text{Var}(\hat{f}_b(x)) \sim \begin{cases} \frac{1}{2\sqrt{\pi}}b^{-1/2}x^{-1/2}f(x)/\bar{G}(x) & \text{if } x/b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}b^{-1}f(x)/\bar{G}(x) & \text{if } x/b \rightarrow \kappa \end{cases} \quad (8)$$

Finally, we derive the integrated mean squared error from equations (8) and (7).

Now, for the gamma kernel estimator of the hazard function. We start by the following decomposition

$$MSE[h_{b_n}(x)] = \mathbb{E}[h_{b_n}(x) - h(x)]^2 = \mathbb{E}[\text{I} + \text{II} + \text{III}]^2$$

where

$$\text{I} = \hat{f}_{b_n}(x) \left[\frac{\bar{F}(x) - \hat{\bar{F}}(x)}{\bar{F}(x)\hat{\bar{F}}(x)} \right]$$

$$\text{II} = [\hat{f}_{b_n}(x) - \mathbb{E}(\hat{f}_{b_n}(x))] / \bar{F}(x), \text{ and}$$

$$\text{III} = [\mathbb{E}(\hat{f}_{b_n}(x)) - f(x)] / \bar{F}(x).$$

Observe that $\mathbb{E}(\text{II.III}) = \text{III}.\mathbb{E}(\text{II}) = 0$, because the term III is deterministic. Using Schwartz inequity and the boundedness of the gamma kernel density estimator we found that $\mathbb{E}|\text{I}| = O(n^{-1/2})$ and $\mathbb{E}[\text{I}^2] = O(n^{-1})$. Now, from the bias formula of the gamma kernel estimator and the conditions on the bandwidth parameter, we get

$$\mathbb{E}(|\text{I.III}|) = |\text{III}|\mathbb{E}|\text{I}| = \begin{cases} O(n^{-1/2})(O(b) + o(n^{-1/2}b_n^{-1/4})) = o(n^{-1}b_n^{-1/2}) & \text{if } x/b_n \rightarrow \infty, \\ O(n^{-1/2})O(b) + o(n^{-1/2}b_n^{-1/2}) = o(n^{-1}b_n^{-1}) & \text{if } x/b_n \rightarrow \kappa. \end{cases}$$

Again, using the Schwartz's inequality we obtain

$$\mathbb{E}(|\text{I.II}|) \leq [\mathbb{E}(\text{I}^2)]^{1/2} \frac{[\text{Var}(\hat{f}_{b_n}(x))]^{1/2}}{\bar{F}(x)} = \begin{cases} O(n^{-1/2})O(n^{-1/2}b_n^{-1/4}) = o(n^{-1}b_n^{-1/2}) & \text{if } x/b_n \rightarrow \infty, \\ O(n^{-1/2})O(n^{-1/2}b_n^{-1/2}) = o(n^{-1}b_n^{-1}) & \text{if } x/b_n \rightarrow \kappa. \end{cases} \quad (9)$$

Combining these formulas, one can see that

$$\begin{aligned} MSE [h_{b_n}(x)] &= \mathbb{E}(\text{II}^2) + \mathbb{E}(\text{III}^2) + O(n^{-1}) + o((nb_n)^{-1}) \\ &= \mathbb{E}(\text{II}^2) + \mathbb{E}(\text{III}^2) + o((nb_n)^{-1}). \end{aligned} \quad (10)$$

Finally, using the expression of the bias and the variance of the gamma kernel density estimator we derive the desired result of the gamma kernel estimator for the failure rate function. ■

We will only give the proof of the asymptotic normality and the iterated logarithm for the gamma kernel estimator. Thereafter, the result is straightforward for the gamma kernel hazard estimator. The proofs are based on Lemma 2. Indeed, we only need to prove the asymptotic normality and the iterated logarithm for σ_n defined in (5) since the other terms are negligible.

Proof of Proposition 1.

By (i), see Theorem 1, we have $n^{1/2}b^{1/4}e_n(x) = o(1)$ and under the conditions on the bandwidth parameter we have $n^{1/2}b^{1/4}\beta_n(x) = o(1)$.

Therefore, we need to state that

$$\left(n^{1/2}b^{1/4} \frac{\sigma_n(x)}{\sqrt{V^*(x)}} \right) \rightarrow N(0, 1), \text{ in distribution}$$

But since

$$\sigma_n(x) = \sum_n W_i(x, b) \text{ where } W_i(x, b) = n^{-1} \int \xi_i(t) dK(x, b)(t)$$

it suffice to prove that

$$\frac{\sum_{i=1}^n \mathbb{E}(|W_i|^3)}{(n \text{ var}(W_1))^{3/2}} \rightarrow 0.$$

In fact, using inequality (6),

$$\sup_x |W_i(x, b)| = O\left(\frac{1}{nb}\right)$$

Therefore, from the variance of $\sigma_n(x)$,

$$\begin{aligned} \frac{\sum_{i=1}^n \mathbb{E}(|W_i|^3)}{(n \operatorname{var}(W_1))^{3/2}} &\leq O\left(\frac{1}{nb}\right) (n \operatorname{var}(W_1))^{-1/2} = \begin{cases} O((nb)^{-3/2}) & \text{if } x \geq 2b \\ O(n^{-3/2}n^{-4/5}) & \text{if } x \in [0, 2b). \end{cases} \\ &= o(1) \quad (\text{since } b = o(n^{-2/5})). \end{aligned}$$

.

■

Proof of Proposition 2.

Condition (iv) ensures that $n^{1/2}b^{1/4}\beta_n(x) = o(1)$ and $n^{1/2}b^{1/4}e_n(x) = o(1)$.

On the other hand, we apply Theorem 1 in Hall (1981) to $S_n = \sum_{i=1}^n W_i$, where W_i is defined as in the proof of Proposition 1, we get under condition (iii) on the bandwidth parameter

$$\limsup_n (\Phi_x(n, b))^{-1/2} |\sigma_n| = 1, \quad \text{a.s.}$$

which concludes the proof of the theorem.

■

References

- BOUEZMARNI, T., AND O. SCAILLET (2005): “Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data,” *Econometric Theory*, 21, 390–412.
- BRESLOW, N., AND J. CROWLEY (1974): “A Large Sample Study of the Life Table and Product-Limit Estimates Under Random Censorship,” *Annals of Statistics*, 2, 437–443.
- CHEN, S. (2000): “Probability Density Functions Estimation Using Gamma Kernels,” *Annals of the Institute of Statistical Mathematics*, 52, 471–480.
- HALL, P. (1981): “Laws of the Iterated Logarithm for Nonparametric Density Estimators,” *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 56, 47–61.

- HESS, K., AND R. GENTLEMAN (2010): *Muhaz: Hazard Function Estimation in Survival Analysis, R package version 1.2.5*.
- HESS, K., D. SERACHITOPOL, AND B. BROWN (1999): “Hazard Function Estimators: a Simulation Study,” *Statistics in Medicine*, 18, 3075–3088.
- IZENMAN, A. J. (1991): “Recent Developments in Nonparametric Density Estimation.,” *Journal of the American Statistical Association*, 86, 205–224.
- JONES, M. (1993): “Simple Boundary Correction for Kernel Density Estimation,” *Statistical Computing*, 3, 135–146.
- JONES, M., AND P. FOSTER (1996): “A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation,” *Statistica Sinica*, 6, 1005–1013.
- KAPLAN, E., AND P. MEIER (1958): “Nonparametric Estimation from Incomplete Observations,” *Journal of the American Statistical Association*, 53, 457–481.
- KLEIN, J. P., AND M. L. MOESCHBERGER (2003): *Survival analysis. Techniques for censored and truncated data. 2nd ed.* Statistics for Biology and Health. New York, NY: Springer. xv, 536 p.
- LEJEUNE, M., AND P. SARDA (1992): “Smooth Estimators of Distribution and Density Functions,” *Computational Statistics and Data Analysis*, 14, 457–471.
- LO, S., Y. MACK, AND J. WANG (1989): “Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator,” *Probability Theory and Related Fields*, 80, 473–473.
- LO, S., AND K. SINGH (1986): “The Product-Limit Estimator and the Bootstrap: Some Asymptotic Representations,” *Probability Theory and Related Fields*, 71, 318–348.
- MARRON, J., AND D. RUPPERT (1994): “Transformations to Reduce Boundary Bias in Kernel Density Estimation,” *Journal of the Royal Statistical Society, Series B*, 56, 653–671.

- MARRON, J. S., AND W. J. PADGETT (1987): “Asymptotically Optimal Bandwidth Selection for Kernel Density Estimators from Randomly Right-Censored Samples,” *The Annals of Statistics*, 15, pp. 1520–1535.
- MIELNICZUK, J. (1986): “Some Asymptotic Properties of Kernel Estimators of a Density Function in Case of Censored Data,” *Annals of Statistics*, 14, 766–773.
- MÜLLER, H. (1991): “Smooth Optimum Kernel Estimators near Endpoints,” *Biometrika*, 78, 521–530.
- MÜLLER, H., AND J. WANG (1994): “Hazard Rate Estimation Under Random Censoring with Varying Kernels and Bandwidth,” *Biometrics*, 50, 61–76.
- PADGETT, W., AND D. MCNICHOLS (1984): “Nonparametric density estimation from censored data.,” *Communications in Statistics, Theory and Methods*, 13, 1581–1611.
- SCHUSTER, E. (1985): “Incorporating Support Constraints into Nonparametric Estimators of Densities,” *Communications in Statistics - Theory and Methods*, 14, 1123–1136.
- SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- SINGPURWALLA, N., AND M.-Y. WONG (1983): “Kernel Estimators of the Failure rate Function and Density Function: An Analogy,” *Journal of the American Statistical Association*, 78, 478–481.
- STUTE, W., AND J. WANG (1993): “A Strong Law under Random Censorship,” *Annals of Statistics*, 21, 1591–1607.
- TANNER, M., AND W. WONG (1983): “The Estimation of the Hazard Function from Randomly Censored Data by the Kernel Method,” *Annals of Statistics*, 11, 989–993.
- WANG, J. (1987): “A Note on the Uniform Consistency of the Kaplan-Meier Estimator,” *Annals of Statistics*, 15, 1313–1316.