Hazard rate function estimation using generalized Birnbaum-Saunders kernel

Sylia Chekkal [1] & Karima Lagha [2] & Nabil zougab [3]

[1] *Laboratory LMA, University of Bejaia, 06000 Bejaia, Algeria. syliachekkal@gmail.com.*
[2] *Laboratory LAMOS, University of Bejaia, 06000 Bejaia, Algeria. karima_lagha@yahoo.com.*
[3] *Laboratory LAMOS, University of Bejaia, 06000 Bejaia, Algeria. nabilzougab@yahoo.fr.*

Abstract. We consider the nonparametric kernel method for the hazard rate function estimation. Since the hazard rate function is positively supported, we use the asymmetric kernels in order to ovoid the problem of high bias in the boundary region. In this work, we consider the class of generalized Birnbaum-Saunders (GBS) kernels because of its flexibility. The asymptotic properties and optimal bandwidth are established for the proposed estimator. Finally we conduct simulation study for sample finite performance.
Keywords. Bandwidth, hazard rate function, kernel method, nonparametric estimation.

Résumé. Dans ce présent papier, nous nous intéressons à l'estimation non paramètrique du taux de défaillance avec la méthode du noyau. Puisque le taux de défaillance est défini sur le support positif $[0, \infty[$, nous utilisons un noyau asymmétrique afin d'éliminer le problème d'effet de bord, qui engendre un biais de plus en plus élevé en se rapprochant du bord. A cet effet nous proposons d'utiliser le noyau GBS associé (Birnbaum-saunders généralisé).
Nous déterminons les propriétés asymptotiques de l'estimateur proposé ainsi que le paramètre de lissage optimal. La performance de l'estimateur est étudiée par simulation des données suivants des lois de fiabilité telles que: lognormale, BS, Gamma.
Mots-clés. Estimation non paramètrique, taux de défaillance, méthode du noyau, effet de bord.

1 Introduction

Nonparametric kernel method using symmetric and asymmetric kernels is widely developed in statistical literature for density, regression and hazard rate functions estimation. In this paper, we focus on kernel estimation of the hazard rate function, based on a random variables $T_1, T_2, ..., T_n$ (independent and identically distributed), which represent the survival times. Since the hazard rate function is positively supported, then it is necessary to use asymmetric kernel instead of the symmetric one, in order to avoid the problem of high bias in the boundary region, which called boundary effect. The hazard rate function is already studied using inverse Gaussian, reciprocal inverse Gaussian (RIG) and Weibull kernels, by Salha (2012), Salha (2012) and Salha and al. (2014), respectively; see also Bouezmarni and al (2011), Bouezmarni and al (2006) for censored data using gamma kernel. The aim of this work is to propose a new kernel estimator of the hazard rate function using the class of generalized Birnbaum-Saunder (GBS kernels), which has as particular cases, BS classical (BS), BS-power exponential (BS-PE) and BS-student (BS-t), see Marchant and al (2013) in the context of the density estimation. The asymptotic properties are established and the optimal bandwidth is obtained using plug-in method. Finally, a simulation study is conducted to test the performance of the proposed GBS kernel estimator.

1.1 Short review on generalized Birbaum Saunders kernels

Let T be a random variable that represent a lifetime of an item, has the probability density function f, distribution function F and survival function R witch are defined on the positive support $[0, \infty[$. We consider a random sample $T_1, T_2, ..., T_n$ from a variable T.

The kernel estimator of the probability density function (pdf), based on GBS kernels is proposed by Marchant and al (2013), and is given as follows

$$\hat{f}_{GBS}(t) \quad = \quad \frac{1}{n}\sum_{i=1}^{n} K_{GBS(h^{\frac{1}{2}},t,g)}(T_i). \tag{1}$$

Where $K_{GBS(h^{\frac{1}{2}},t,g)}$ is the GBS kernel given by

$$K_{GBS(h^{\frac{1}{2}},t,g)}(y) = cg\left(\frac{1}{h}\left(\frac{y}{t}+\frac{t}{y}-2\right)\right)\frac{1}{\sqrt{4h}}\left(\frac{1}{\sqrt{yt}}+\sqrt{\frac{t}{y^3}}\right), \qquad y>0, \quad h>0, \ t>0. \tag{2}$$

Where h is the bandwidth, t is the target, c is the normalization constant and $g = g(u)$ with $u > 0$ is a real function that generates the probability density (pdf) of the random variable Z, such that $f_Z(z) = cg(z^2)$, $\quad z \in R$, where $Z = \frac{1}{\alpha}\left(\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}\right)$; see Marchant and al (2013), Fang and al (1990) for more details.

1.2 GBS kernel estimator of the hazard rate function

The hazard rate function represents the probability that an item with age t will fail in interval (t,t+dt), for small $dt > 0$, defined as

$$\lambda(t) = \lim_{dt \to 0} \frac{P(t \leq T \leq t + dt/T > t)}{dt}, \qquad t > 0.$$

Which can be written as

$$\lambda(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{R(t)}, \qquad t > 0.$$

The proposed estimator of the hazard rate function, using the estimator in (1), is given by

$$\hat{\lambda}_{GBS}(t) = \frac{\hat{f}_{GBS}(t)}{\hat{R}_{GBS}(t)} = \frac{\frac{1}{n}\sum_{i=1}^{n} K_{GBS(h^{\frac{1}{2}},t,g)}(T_i)}{1 - \frac{1}{n}\sum_{i=1}^{n}\int_0^t K_{GBS(h^{\frac{1}{2}},x,g)}(T_i)\ dx}, \qquad t > 0. \tag{3}$$

By replacing the expression of GBS kernel defined in (2), we get

$$\hat{\lambda}_{GBS}(t) = \frac{\sum_{i=1}^{n} g(\frac{1}{h}(\frac{T_i}{t} + \frac{t}{T_i} - 2))(\frac{1}{\sqrt{tT_i}} + \sqrt{\frac{t}{T_i^3}})}{\frac{2n\sqrt{h}}{c} - \sum_{i=1}^{n}\int_0^t g(\frac{1}{h}(\frac{T_i}{x} + \frac{x}{T_i} - 2))(\frac{1}{\sqrt{xT_i}} + \sqrt{\frac{x}{T_i^3}})dx}, \qquad t > 0. \tag{4}$$

Where h is the bandwidth parameter, such that $h = h(n) \to 0$ when $n \to \infty$, c is the normalization constant and g is a generator.

2 Asymptotic properties and bandwidth choice

The bias, variance, mean squared error (MSE) and mean integrated squared error (MISE) of the GBS kernel estimator of the hazard rate function are given in proposition 1 and proposition 2.

Proposition 1 *The bias and the variance of the kernel estimator $\hat{\lambda}_{GBS}$, are given by*

$$Bias(\hat{\lambda}_{GBS}(t)) = \frac{hu_1(g)(tf'(t) + t^2 f''(t))}{R(t)} + o(h). \tag{5}$$

$$Var(\hat{\lambda}_{GBS}(t)) = \frac{n^{-1}h^{-\frac{1}{2}}c^2 t^{-1}\lambda(t)}{c_{g^2}R(t)} + o(n^{-1}h^{-\frac{1}{2}}). \tag{6}$$

Proposition 2 *The mean squared error (MSE) and the mean integrated squared error (MISE) of the kernel estimator $\hat{\lambda}_{GBS}$ are given by*

$$MSE(\hat{\lambda}_{GBS}) = \frac{1}{R^2(t)} MSE(\hat{f}_{GBS}). \tag{7}$$

$$MISE(\lambda\hat{(}t) = \frac{1}{4}h^2 \int_0^\infty \left(\frac{u_1(g)(tf'(t) + t^2 f''(t))}{R(t)}\right)^2 dt + n^{-1}h^{-\frac{1}{2}} \int_0^\infty \frac{c^2 t^{-1} f(t)}{c_{g^2}R^2(t)} dt + o(h^2 + n^{-1}h^{-\frac{1}{2}}). \tag{8}$$

Where f' and f'' are respectively the first and the second derivatives of the density f, c_{g^2} is the normalization constant, such that $\int_R g^2(z^2)dz = \frac{1}{c_{g^2}}$, $z \in R$ and $u_k = u_k(g) = E(U^k)$, with $U = Z^2$ is a random variable following a generalized chi-squared distribution with one degree of freedom denoted by, $U \sim G_{\chi^2}(1, g)$.

The performance of the kernel method depends on the bandwidth h, which controls the smoothness of the estimator. The optimal bandwidth of the proposed estimator $\hat{\lambda}_{GBS}$ is obtained by using the plug-in method and is given by

$$h_{opt} = \left[\frac{c^2 \int_0^\infty \frac{t^{-1}f(t)}{R^2(t)} dt}{c_{g^2} u_1^2(g) \int_0^\infty \left(\frac{tf'(t)+t^2 f''(t)}{R(t)} \right)^2 dt} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \tag{9}$$

3 Simulation study

In this section, we evaluate the performance of the proposed estimator by using MC simulation. We simulate data from the distributions lognormal, BS and gamma.
The experiment are based on 100 random samples of length n=200, n=500, n=1000.
We simulate the samples from the distributions, $BS(2,3)$, $log\mathcal{N}(2,3)$ and $G(3,1/2)$ with pdf given respectively by

$$f_{BS(\alpha,\beta)}(t) = \frac{1}{2\alpha\beta\sqrt{2\pi}} \left[\left(\frac{\beta}{t}\right)^{\frac{1}{2}} + \left(\frac{\beta}{t}\right)^{\frac{3}{2}} \right] exp\left[-\frac{1}{2\alpha^2(\frac{t}{\beta} + \frac{\beta}{t} - 2)} \right], \quad \alpha > 0,\ \beta > 0.$$

$$f_{LN(\mu,\sigma)}(t) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-\frac{(\ln(x)-\mu)^2}{2\sigma^2}), \quad \sigma > 0,\ \mu \in R.$$

$$f_{Gam(\alpha,\beta)}(t) = t^{\alpha-1}\frac{\beta^\alpha \exp(-\beta x)}{\Gamma(\alpha)}, \quad \alpha > 0,\ \beta > 0.$$

After that we estimate the hazard rate function using BS, BS-PE (with v=2) and BS-t (with v=5) kernels, and we determine the estimation error produced by each estimator using the integrated squared error (ISE), given by $ISE(\hat{\lambda}(t)) = \int_0^\infty \left(\lambda(t) - \hat{\lambda}(t)\right)^2 dt$. The results are reported in the Table 1.

Size	Distribution	BS	BS-PE	BS-t
n=200	BS(2,3)	0.00807	0.01700	0.00981
	LN(2,3)	0.02682	0.04774	0.04835
	Gam(3,1/2)	0.01956	0.02740	0.02825
n=500	BS(2,3)	0.00407	0.00740	0.00543
	LN(2,3)	0.02501	0.06187	0.04201
	Gam(3,1/2)	0.01819	0.02035	0.02447
n=1000	BS(2,3)	0.00257	0.00424	0.00436
	LN(2,3)	0.01200	0.04130	0.03491
	Gam(3,1/2)	0.01552	0.01874	0.02301

Table 1: Average ISE

4 Conclusion

In this paper we have proposed a new kernel estimator of the hazard rate function using a class of GBS kernel, in particular BS, BS-PE and BS-t kernels. We have established the asymptotic properties of the estimator and estimated the bandwidth parameter using plug-in method. The simulation study has conducted in order to evaluate the performance of the proposed kernel estimator. The results have showed a good performance of the estimator.

Bibliographie

Marchant, C., Bertin, K., Leiva, V., and Saulo, H. (2013). Generalized Birnbaum-Saunders Kernel Density Estimators and an Analysis of Financial Data, *Computational Statistics and Data Analysis*, 63, pp. 1-15.

Salha, R. (2012). Hazard rate function estimation using inverse Gaussian kernel, *The Islamic university of Gaza journal of Natural and Engineering studies*, 20, pp. 73-84.

Salha, R. (2013). Estimating the density and hazard rate function using the reciprocal inverse gaussian kernel, *15th Applied Stochastic Models and Data Analysis International Conference*, pp. 25-28.

Salha, R. B., Ahmed, H. I. E. S., and Alhoubi, I. M. (2014). Hazard Rate Function Estimation Using Weibull Kernel, *Open Journal of Statistics*, 04, pp. 08-650.

Bouezmarni, T., El Ghouch, A., and Mesfioui, M. (2011). Gamma Kernel Estimators for Density and Hazard Rate of Right-Censored Data, *Journal of Probability and Statistics*, 2011, pp. 1-16.

Bouezmarni, T., and Rombouts, J. V. (2006). Density and hazard rate estimation for censored and a-mixing data using gamma kernels. Available at SSRN 970914.

Fang, K. T., Kotz, S., and Ng, W. K. (1990). Symmetric multivariate and related distributions. London: Chapman and Hall.