Parametrically guided kernel density and hazard functions estimation with censored data

M. Talamakrouni
I. Van Keilegom, A. El Ghouch

Université Catholique de Louvain - ISBA

Young Researchers Day - September 20, 2013

Outline

- Introduction
- Extension to censored data
- Asymptotic normality
- Simulations

Introduction

Aim: density estimation

$X_1, \ldots, X_n$ are iid from **unknown** density f

Parametric approach

- $f \in \{f(x,\theta) : \theta \in \Theta\},\ \theta \in \Theta \subset \Re^p$
- Plug-in approach : $\hat{f}(x) = f(x, \hat{\theta})$, $\hat{\theta}$ is an estimator
- $\sqrt{n}$ rate of convergence
- **Strong** parametric assumptions

Introduction

Nonparametric approach

- No assumptions on parametric form
- $\sqrt{nh}$ rate of convergence (**slower**)

Ideally

To combine the advantages of both parametric and nonparametric approaches

How ?

Parametrically guided nonparametric approach (Hjort and Glad (1995))

Introduction

Traditional kernel estimator

$$\tilde{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)$$

K: kernel function, h: bandwidth

Statistical properties

$$\begin{aligned} ABias\,\tilde{f}(x) &\doteq \frac{1}{2} h^2 f''(x) \int u^2 K(u)du \\ AVar\,\tilde{f}(x) &\doteq (nh)^{-1} f(x) \int K^2(u)du - \frac{f(x)^2}{n} \end{aligned}$$

Introduction

Objective: Reduce the bias for kernel estimator

Idea: Use information from a parametric guide $f(x,\theta)$, $\theta \in \Theta \subset \Re^p$ unknown

Multiplicative correction:

$$f(x) = f(x,\theta) r(x,\theta)$$

$r(x,\theta)$ is the correction factor:

$$r(x,\theta) = \frac{f(x)}{f(x,\theta)}$$

Introduction

The two steps estimation procedure:

1. Parametric estimator $f(x, \hat{\theta})$
2. Nonparametric estimator of $r(x, \hat{\theta}) = f(x)/f(x, \hat{\theta})$

$$\hat{r}(x, \hat{\theta}) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) \frac{1}{f(X_i, \hat{\theta})}$$

The ensuing estimator:

$$\hat{f}(x, \hat{\theta}) \quad = \quad f(x, \hat{\theta}) \hat{r}(x, \hat{\theta})$$

Guided kernel estimator (Hjort and Glad (1995))

$$\hat{f}(x,\hat{\theta}) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)\frac{f(x,\hat{\theta})}{f(X_i,\hat{\theta})}$$

Special case: Uniform guide

$f(x,\theta) = constant \rightarrow$ the classical kernel estimator

Introduction

Statistical properties: θ_* the least false value of θ according to the Kullback-Leibler distance.

$$
\begin{aligned}
ABias\,\hat{f}(x,\hat{\theta}) &\doteq \frac{1}{2}h^2 f(x,\theta_*)r''(x,\theta_*)\int u^2K(u)du \\
AVar\,\hat{f}(x,\hat{\theta}) &\doteq (nh)^{-1}f(x)\int K^2(u)du - \frac{f(x)^2}{n}
\end{aligned}
$$

Advantage:

Bias reduction in all cases where:

$$|f(x,\theta_*)r''(x,\theta_*)| < |f''(x)|$$

No change in variance

Asymptotic normality: has not been investigated

Guided estimator will:

- **Converge** to the true density function "no matter" whether the parametric part is correct or not.
- Have **similar asymptotic behavior** to the kernel estimates if the parametric density is not correct.
- **Adapt automatically** to the parametric guide when the latter is locally close to the true density.

Extension to censored data

Random right censoring:

T: variable of interest with d.f F and density f
C: censoring variable with d.f G
$X = min(T,C)$: observed time
$\delta = I(T \leq C)$: censoring indicator

Assumption:

T and C are independent

Extension to censored data

Kernel estimator for density f (Blum and Susarla (1980)):

$$f_n(x) = \frac{1}{h}\int K\left(\frac{x-u}{h}\right)dF_n(u)$$

Kaplan Meier estimator (1958):

$$F_n(x) = \prod_{i:\, X_{(i)}\leq x}\left(1-\frac{1}{n-i+1}\right)^{\delta_{(i)}}$$

Statistical properties:

$$ABias\, f_n(x) \doteq \frac{1}{2}h^2 f''(x)\int u^2K(u)du$$

$$AVar\, f_n(x) \doteq (nh)^{-1}[f(x)/(1-G(x))]\int K^2(u)du$$

Extension to censored data

The two steps estimation procedure:

1. Parametric estimator $f(x, \hat{\theta})$
2. Nonparametric estimator of $r(x, \hat{\theta}) = f(x)/f(x, \hat{\theta})$

$$\hat{r}(x, \hat{\theta}) = \frac{1}{h} \int K\left(\frac{x-u}{h}\right) \frac{1}{f(u, \hat{\theta})} dF_n(u)$$

The ensuing estimator:

$$\hat{f}(x, \hat{\theta}) \quad = \quad f(x, \hat{\theta}) \hat{r}(x, \hat{\theta})$$

Extension to censored data

Guided kernel estimator:

$$\hat{f}(x,\hat{\theta}) = \frac{1}{h}\int K\left(\frac{x-u}{h}\right)\frac{f(x,\hat{\theta})}{f(u,\hat{\theta})}dF_n(u)$$

In practice:

$$\hat{f}(x,\hat{\theta}) = \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right)\frac{f(x,\hat{\theta})}{f(X_i,\hat{\theta})}W_i$$

where

$$W_i = \begin{cases} \bar{F}_n(X_i) - \bar{F}_n(X_{i+1}), & i = 1,\dots,n-1, \\ \bar{F}_n(X_n), & i = n. \end{cases}$$

and $\bar{F}_n(x) = 1 - F_n(x)$

Extension to censored data

Guided kernel hazard rate's estimator:

Define the hazard rate

$$\lambda(x) = f(x)/\bar{F}(x)$$

$$\hat{\lambda}(x,\hat{\theta}) = \hat{f}(x,\hat{\theta})/\bar{F}_n(x)$$

where $\bar{F}_n(.)$ is the Kaplan-Meier estimator

Asymptotic normality

Two steps:

- Asymptotic normality of $\hat{f}(.)$ the guided kernel estimator using non random guide f_0
- Provide sufficient conditions for the asymptotic equivalence of $\hat{f}(x, \hat{\theta})$ and $\hat{f}(x, \theta^*)$ where θ^* is the least false value of θ according to some distance

$f_0(.)$ is a known function that approximate $f(.)$

Proposition : Under some assumptions, we have,

1.
$$\sup_{x\leq T}|\hat{f}(x)-f(x)| = O_p((nh)^{-1/2}(\log n)^{1/2})$$

2.
$$\hat{f}(x)-f(x) = \frac{1}{nh}\sum_i \int_{-1}^{1} \xi_i(x-uh)K'(u)du + R_n(x)$$

where $\xi_i(x) = \int_0^{X_i\wedge x}(\bar{L}(s))^{-2}dL_1(s) + (\bar{L}(X_i))^{-1}.I\{X_i \leq x, \delta_i = 1\}$,
$\bar{L}(x) = P(T_i > x, C_i > x), L_1(x) = P(X_i \leq x, \delta_i = 1)$ and
$\sup_{x\leq T}|R_n(x)| = O_p(\log n/nh)$.

Theorem

Under some assumptions, we have,

$$\sqrt{nh}\left(\hat{f}(x) - f(x) - \frac{1}{2}h^2\mu_K^2 r''(x) f_0(x) + o_p(h^2)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(x)}{1-G(x)} \int K^2(u)du\right)$$

where $r(x) = f(x)/f_0(x)$ *and* $G(x) = P(C \leq x)$

$\hat{\theta}$: the maximum likelihood estimator
θ_* : the least false value of θ according to the Kullback-Leibler distance

Theorem

Under some regularity assumptions, we have,

$$\sqrt{nh}\left(\hat{f}(x,\hat{\theta}) - f(x) - \frac{1}{2}h^2\mu_K^2 r''(x,\theta_*)f(x,\theta_*) + o_p(h^2)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{f(x)}{1-G(x)}\int K^2(u)du\right)$$

$G(x) = P(C \leq x)$

Asymptotic normality

Guided kernel estimator vs traditional kernel estimator:

- Bias: reduction in all cases where:

$$|f(x,\theta_*)r''(x,\theta_*)| < |f''(x)|$$

- Variance: does not change

Censored case vs uncensored case :

- Biais: Not affected
- Variance: increases with censoring

Theorem

Under some regularity assumptions, we have,

$$\sqrt{nh}\left(\hat{\lambda}(x,\hat{\theta}) - \lambda(x) - \frac{1}{2}h^2\mu_K^2 r''(x,\theta_*)f(x,\theta_*)/\bar{F}(x) + o_p(h^2)\right)$$
$$\xrightarrow{d} \mathcal{N}\left(0, \frac{\lambda(x)}{1-H(x)}\int K^2(u)du\right)$$

$H(x) = P(T \leq x)$

Data:

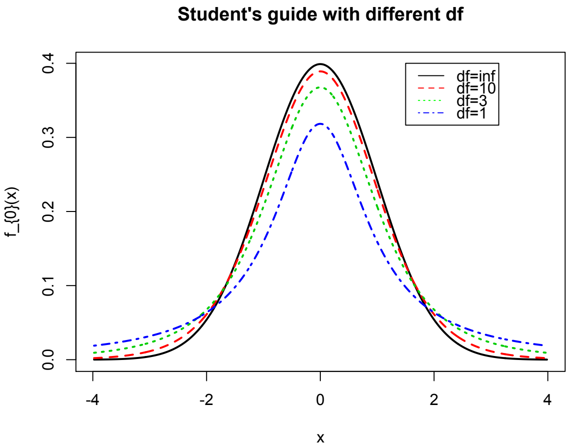
- $T \sim N(0,1)$
- $C \sim N(0,1)$
- Censoring rate: 50%
- Gaussian kernel
- Replications $N = 500$, observations $n = 100$
- Bandwidths are selected via a grid search

Non estimated guide:

- Student's t-distribution
- Four values of df : $\infty, 10, 3, 1$

Simulations :

Normal density with a student's t non estimated guide



Simulations:

Normal density with a student's t non estimated guide

		$\times 10^3$		
Method		$IBias^2$	$IVar$	$IMSE$
TKE		1.129	1.467	2.596
GKE	df			
	∞	0.597	1.256	1.853
	10	0.691	1.169	1.860
	3	0.866	1.116	1.982
	1	1.057	1.364	2.421

Weibull density with estimated exponential guide

Data:

$T \sim weibull(a, b)$
$C \sim weibull(a, d)$, $d = b((1-p)/p)^{1/a}$
p is the censoring rate

Parametric guide:

$f(x, \hat{\theta}) = \hat{\theta} \exp(-\hat{\theta} x)$, $\hat{\theta}$ is the MLE

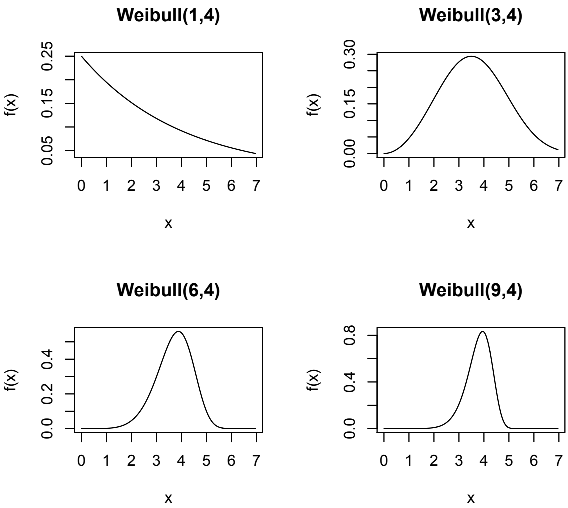
Simulations:

Weibull density with estimated exponential guide

- Two censoring rates 10% and 40%
- Replications $N = 500$, observations $n = 200$
- Four values of a : $1, 3, 6, 9$, $b = 4$
- Epanechnikov kernel function
- Bandwidths are selected via a grid search

Simulations :

Weibull density with estimated exponential guide



Simulations:

Weibull density with estimated exponential guide

Censoring rate		10%			40%		
			$\times 10^3$			$\times 10^3$	
a	Method	$IBias^2$	IVar	IMSE	$IBias^2$	IVar	IMSE
1	GKE	5.449	0.030	5.479	5.456	0.045	5.501
	TKE	10.721	0.005	10.726	10.719	0.007	10.726
3	GKE	6.933	0.025	6.958	7.432	0.043	7.475
	TKE	11.840	0.041	11.881	11.839	0.053	11.892
6	GKE	19.203	0.529	19.731	23.097	0.736	23.833
	TKE	30.135	0.149	30.284	30.350	0.186	30.536
9	GKE	43.626	0.427	44.053	44.787	0.295	45.082
	TKE	41.844	0.571	42.415	41.895	0.687	42.582

Further research

- Extension to multivariate case
- Extension to conditional density
- Application to real data

References

- Blum, J.R., and Susarla, V. (1980). Maximal deviation theory of density and failure rate estimates based on censored data. *Multivariate Analysis* **5**, 213-222.
- Hjort, N.L. and Glad, I.K. (1995). Nonparametric density estimation with parametric start. *Ann. Statist.* **23**, 882- 904.
- Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457-481.