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
• Introduction
• Extension to censored data
• Asymptotic normality
• Simulations
**Aim**: density estimation

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

**Parametric approach**

- $f \in \{f(x, \theta) : \theta \in \Theta\}, \ \theta \in \Theta \subset \mathbb{R}^p$
- 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))
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

\[ A\text{Bias} \tilde{f}(x) = \frac{1}{2} h^2 f''(x) \int u^2 K(u) du \]

\[ A\text{Var} \tilde{f}(x) = (nh)^{-1} f(x) \int K^2(u) du - \frac{f(x)^2}{n} \]
**Objective:** Reduce the bias for kernel estimator

**Idea:** Use information from a parametric guide \( f(x, \theta) \), \( \theta \in \Theta \subset \mathbb{R}^p \) unknown

**Multiplicative correction:**

\[
    f(x) = f(x, \theta) r(x, \theta)
\]

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

\[
    r(x, \theta) = \frac{f(x)}{f(x, \theta)}
\]
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:

\[
\hat{f}(x, \hat{\theta}) = f(x, \hat{\theta}) \hat{r}(x, \hat{\theta})
\]
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

\[f(x, \theta) = constant \rightarrow \text{the classical kernel estimator}\]
Statistical properties: $\theta_*$ the least false value of $\theta$ according to the Kullback-Leibler distance.

$A\text{Bias } \hat{f}(x, \hat{\theta}) = \frac{1}{2} h^2 f(x, \theta_*) r''(x, \theta_*) \int u^2 K(u) du$

$A\text{Var } \hat{f}(x, \hat{\theta}) = (nh)^{-1} f(x) \int K^2(u) du - \frac{f(x)^2}{n}$

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:

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

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

\[
\hat{f}(x, \hat{\theta}) = f(x, \hat{\theta})\hat{r}(x, \hat{\theta})
\]
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

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

and \(\bar{F}_n(x) = 1 - F_n(x)\)
Extension to censored data

Guided kernel hazard rate’s estimator:

Define the hazard rate

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

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

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 \( \theta^* \) is the least false value of \( \theta \) according to some distance
Asymptotic normality

The estimator with a fixed guide $f_0(.)$

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

**Proposition :** Under some assumptions, we have,

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

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 $$\xi_i(x) = \int_0^{X_i \land x} (\bar{L}(s))^{-2} dL_1(s) + (\bar{L}(X_i))^{-1}.I\{X_i \leq x, \delta_i = 1\}$$, $$\bar{L}(x) = P(T_i > x, C_i > x), L_1(x) = P(X_i \leq x, \delta_i = 1)$$ and $$\sup_{x \leq T} |R_n(x)| = O_p (\log n/nh)$$.
The estimator with a fixed guide $f_0(.)$

**Theorem**

*Under some assumptions, we have,*

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

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

The estimator with an estimated guide $f(., \hat{\theta})$

$\hat{\theta}$ : the maximum likelihood estimator
$\theta_* :$ the least false value of $\theta$ 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:
  \[ |f(x, \theta_\ast) r''(x, \theta_\ast)| < |f''(x)| \]

- Variance: does not change

Censored case vs uncensored case:

- Bias: Not affected
- Variance: increases with censoring
Asymptotic normality

Extension to hazard function $\lambda(.)$

**Theorem**

*Under some regularity assumptions, we have,*

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

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

Normal density with a student’s t non estimated guide

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

Student's guide with different df

\[ f_{\{0\}}(x) \]

- \( df=\infty \)
- \( df=10 \)
- \( df=3 \)
- \( df=1 \)
Simulations:

Normal density with a student’s t non estimated guide

<table>
<thead>
<tr>
<th>Method</th>
<th>$IBias^2$</th>
<th>$IVar$</th>
<th>IMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TKE</td>
<td>1.129</td>
<td>1.467</td>
<td>2.596</td>
</tr>
<tr>
<td>GKE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$df$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\infty$</td>
<td>0.597</td>
<td>1.256</td>
<td>1.853</td>
</tr>
<tr>
<td>10</td>
<td>0.691</td>
<td>1.169</td>
<td>1.860</td>
</tr>
<tr>
<td>3</td>
<td>0.866</td>
<td>1.116</td>
<td>1.982</td>
</tr>
<tr>
<td>1</td>
<td>1.057</td>
<td>1.364</td>
<td>2.421</td>
</tr>
</tbody>
</table>
Simulations:

Weibull density with estimated exponential guide

Data:

\[ T \sim \text{weibull}(a, b) \]
\[ C \sim \text{weibull}(a, d), \quad d = b((1 - p)/p)^{1/a} \]

\( p \) is the censoring rate

Parametric guide:

\[ f(x, \hat{\theta}) = \hat{\theta} \exp(-\hat{\theta} x), \quad \hat{\theta} \text{ is the MLE} \]
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

Weibull(1,4)

Weibull(3,4)

Weibull(6,4)

Weibull(9,4)
Simulations:

Weibull density with estimated exponential guide

<table>
<thead>
<tr>
<th>Censoring rate</th>
<th>10%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\times 10^3$</td>
<td>$\times 10^3$</td>
</tr>
<tr>
<td>$a$</td>
<td>Method</td>
<td>$IBias^2$</td>
</tr>
<tr>
<td>1</td>
<td>GKE</td>
<td>5.449</td>
</tr>
<tr>
<td></td>
<td>TKE</td>
<td>10.721</td>
</tr>
<tr>
<td>3</td>
<td>GKE</td>
<td>6.933</td>
</tr>
<tr>
<td></td>
<td>TKE</td>
<td>11.840</td>
</tr>
<tr>
<td>6</td>
<td>GKE</td>
<td>19.203</td>
</tr>
<tr>
<td></td>
<td>TKE</td>
<td>30.135</td>
</tr>
<tr>
<td>9</td>
<td>GKE</td>
<td>43.626</td>
</tr>
<tr>
<td></td>
<td>TKE</td>
<td>41.844</td>
</tr>
</tbody>
</table>
Further research

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

