Kernel estimation as alternative to (semi)parametric models in survival analysis

Iveta Selingerová, Stanislav Katina, Ivana Horová

Department of Mathematics and Statistics
Faculty of Science, Masaryk University
Brno

ROBUST 2016
11 - 16 September 2016

Outline

Survival analysis

- Survival time T with distribution function F
- Random right censoring
- Censoring time C with distribution function G
- Observed time $Y = \min(T, C)$ with distribution function L, censoring indicator $\delta = I(T \leq C)$
- Survival function $\overline{F}(t) = P(T \geq t) = 1 - F(t)$
- Hazard function

$$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{\overline{F}(t)}$$

- Cumulative hazard function $\Lambda(t) = \int_0^t \lambda(u) \mathrm{d}u$

Properties

- Survival function of observed time $\overline{L}(t) = \overline{F}(t)\overline{G}(t)$
- Joint density of (Y, δ) is $l(\cdot, \cdot)$ and is composed of two "subdensities" $l(\cdot, 0)$ and $l(\cdot, 1)$
- Subdensity of the event time $r(t) = l(\cdot, 1) = f(t)\overline{G}(t)$
- Hazard function

$$\lambda(t) = \frac{r(t)}{\overline{L}(t)}$$

Nonparametric estimation

- Kaplan-Meier estimator of the survival function

$$\widehat{\overline{F}}_{KM}(t) = \begin{cases} 1 & t \leq Y_{(1)}, \\ \prod_{i:Y_{(i)}<t} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}} & t > Y_{(1)}. \end{cases}$$

- Nelson-Aalen estimator of the cumulative hazard

$$\hat{\Lambda}_{NA}(t) = \begin{cases} 0 & t \leq Y_{(1)}, \\ \sum_{i:Y_{(i)<t}} \frac{\delta_{(i)}}{n-i+1} & t > Y_{(1)}. \end{cases}$$

- Smoothing techniques for estimation of hazard function $\lambda(t)$

Survival models

- $\mathbf{X} = (X_1, X_2, \ldots, X_p)^T$ vector of p covariates
- Conditional hazard function

$$\lambda(t|\mathbf{x}) = \lim_{\Delta t \to 0^+} \frac{P(t \leq T < t + \Delta t | T \geq t, \mathbf{X} = \mathbf{x})}{\Delta t}$$

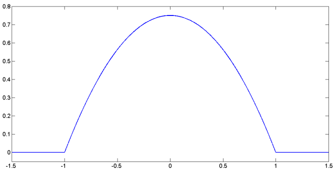
- Parametric survival model
  - Assumption of distribution of survival time
  - Accelerated failure time (AFT) models, parametric proportional hazard (PH) models
- Semiparametric survival model – Cox model
  - Exponential dependence on covariates
  - Assumption proportionality of hazard ratio
  - Nonparametric estimate of baseline hazard
- Nonparametric survival model
  - No assumption of distribution or dependence on covariates
  - Smoothing techniques

Kernel smoothing

- The value of unknown function at a point is estimated as a local weighted average of known observations in the neighbourhood of this point
- Kernel K – real function with support on $[-1,1]$ satisfying

$$\int_{-1}^{1} x^j K(x)\mathrm{d}x = \begin{cases} 1 & j=0, \\ 0 & 0<j<k, \\ b_k \neq 0 & j=k. \end{cases}$$

- Epanechnikov kernel $K(x) = -\frac{3}{4}(x^2-1)I([-1,1])$



Kernel smoothing

- The bandwidth sequence $\{h(n)\}$

$$\lim_{n\to\infty} h(n) = 0, \quad \lim_{n\to\infty} nh(n) = \infty$$

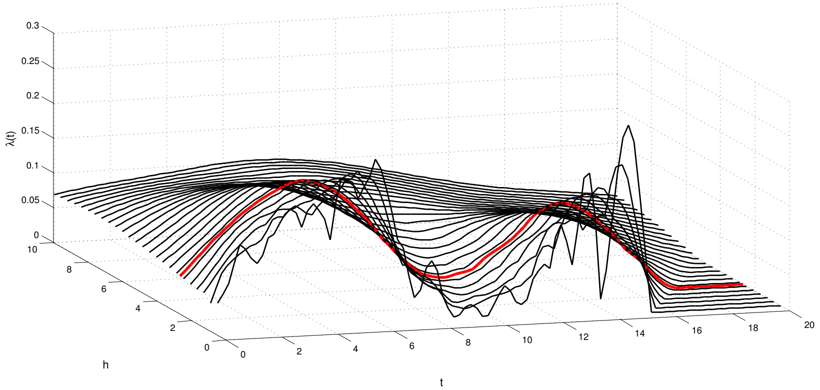
- Statistical properties of an estimator $\hat{\vartheta}(x)$ of a function $\vartheta(x)$
  - Local error of an estimator

$$\mathrm{MSE}\left(\hat{\vartheta}(x)\right) = \mathrm{E}\left(\vartheta(x) - \hat{\vartheta}(x)\right)^2 = \mathrm{Var}\left(\hat{\vartheta}(x)\right) + \mathrm{Bias}^2\left(\hat{\vartheta}(x)\right)$$

  - Global error of an estimator

$$\mathrm{MISE}\left(\hat{\vartheta}\right) = \int \mathrm{MSE}\left(\hat{\vartheta}(x)\right)\omega(x)\mathrm{d}x$$

Bandwidths



Two types of kernel conditional hazard estimates

External estimator

- $\overline{L}(t|x) = \overline{F}(t|x)\overline{G}(t|x)$
  $\lambda(t|x) = \frac{f(t|x)}{\overline{F}(t|x)} = \frac{f(t|x)\overline{G}(t|x)}{\overline{L}(t|x)} = \frac{r(t|x)}{\overline{L}(t|x)}$
- $$\hat{\lambda}_E(t|x) = \frac{\widehat{r}(t|x)}{\widehat{\overline{L}}(t|x)} = \frac{\frac{1}{h_t}\sum_{i=1}^n w_i(x) K\left(\frac{t-Y_i}{h_t}\right)\delta_i}{\sum_{i=1}^n w_i(x) W\left(\frac{Y_i-t}{h_t}\right)}$$

Nadaraya–Watson weights

$$w_i(x) = \frac{K\left(\frac{x-X_i}{h_x}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_x}\right)},\ i = 1, \ldots, n$$

Two types of kernel conditional hazard estimates

Internal estimator

- Beran estimator of conditional cumulative hazard function

$$\hat{\Lambda}(t|x) = \begin{cases} 0 & t \leq Y_{(1)} \\ \sum_{i:Y_{(i)}<t} \frac{\delta_{(i)} w_{(i)}(x)}{1-\sum_{j=1}^{i-1} w_{(j)}(x)} & t > Y_{(1)} \end{cases}$$

-

$$\begin{aligned} \hat{\lambda}_I(t|x) &= \frac{1}{h_t} \int K\left(\frac{t-u}{h_t}\right) \mathrm{d}\hat{\Lambda}(u|x) \\ &= \frac{1}{h_t} \sum_{i=1}^{n} K\left(\frac{t-Y_{(i)}}{h_t}\right) \frac{\delta_{(i)} w_{(i)}(x)}{1-\sum_{j=1}^{i-1} w_{(j)}(x)} \end{aligned}$$

h_t influences smoothing in direction of the time

h_x influences smoothing in direction of the covariate

Cox proportional hazards model

- Cox model
$$\lambda(t|x) = \lambda_0(t)\mathrm{e}^{\beta x}$$
- Properties and assumptions:
    - Hazard ratios are constant over time, i.e.
    $$\frac{\lambda(t|x_1)}{\lambda(t|x_2)} = \mathrm{e}^{\beta(x_1 - x_2)}$$
    - The dependence on covariate is exponential
    - No distribution of the survival time is assumed
- Maximum partial likelihood estimator of the model parameter β

Cox proportional hazards model

- Breslow estimator of baseline hazard function

$$\hat{\Lambda}_0(t) = \sum_{i:t_{(i)}<t} \hat{\lambda}_0(t_{(i)}) = \sum_{i:t_{(i)}<t} \frac{d_i}{\sum_{j\in\mathbb{R}_i} \mathrm{e}^{\beta X_j}}$$

- Kernel estimator of baseline hazard function

$$\begin{aligned}\hat{\lambda}_0(t) &= \frac{1}{h}\int K\left(\frac{t-u}{h}\right)\mathrm{d}\hat{\Lambda}_0(u) \\ &= \frac{1}{h}\sum_{i=1}^{n} K\left(\frac{t-Y_i}{h}\right)\frac{\delta_i}{\sum_{j=1}^{n} I(Y_j \geq Y_i)\mathrm{e}^{\beta X_j}}\end{aligned}$$

Bandwidth selection

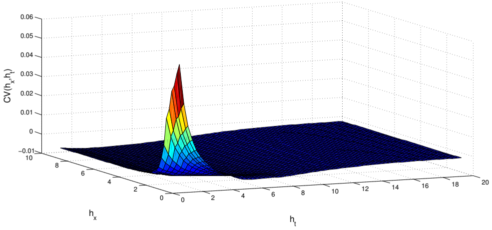
Asymptotically optimal bandwidths

- Minimize MISE
- Theoretical value of bandwidths
- Need to know distribution of survival time, observed time, or covariate

Bandwidth selection

Cross-validation method

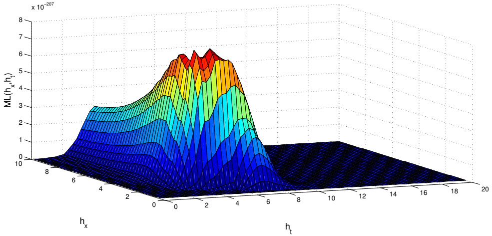
- $$CV(h_x, h_t) = \frac{1}{n}\sum_{i=1}^{n}\int \hat{\lambda}^3(t|X_i)\mathrm{e}^{-\int_0^t \hat{\lambda}(u|X_i)\mathrm{d}u}\mathrm{d}t - \frac{2}{n}\sum_{i=1}^{n}\frac{\hat{\lambda}_{-i}^2(Y_i|X_i)}{\widehat{\overline{L}}(Y_i|X_i)}\mathrm{e}^{-\int_0^{Y_i}\hat{\lambda}_{-i}(u|X_i)\mathrm{d}u}\delta_i$$
- $(\hat{h}_{x,CV}, \hat{h}_{t,CV}) = \arg\min_{(h_x,h_t)} CV(h_x, h_t)$



Bandwidth selection

Maximum likelihood method

- $\mathrm{ML}(h_x, h_t) = \prod_{i=1}^{n} \hat{\lambda}_{-i}^{\delta_i}(Y_i|X_i)\widehat{\overline{F}}_{-i}(Y_i|X_i)$
- $(\hat{h}_{x,ML}, \hat{h}_{t,ML}) = \arg\max_{(h_x,h_t)} \mathrm{ML}(h_x, h_t)$



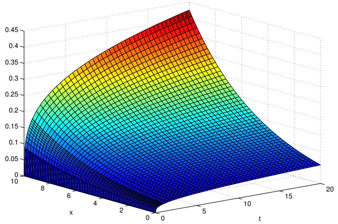
Simulation study

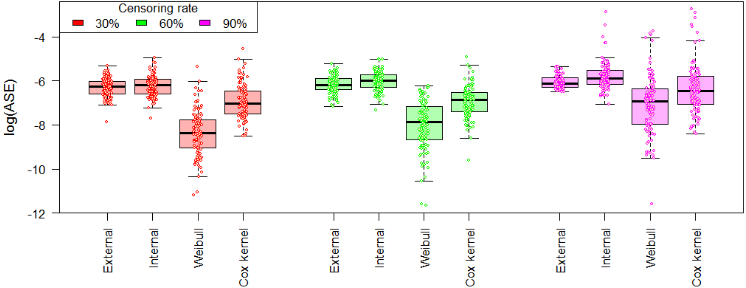
- 200 triples of observed data (X_i, Y_i, δ_i)
- 100 samples
- Specified conditional hazard function $\lambda(t|x)$
- Covariate $X_i \sim \mathcal{U}(0, 10)$
- Survival time T_i obtained as $T_i = F^{-1}(U_i|X_i)$, where $U_i \sim \mathcal{U}(0, 1)$ and $F(t|x) = 1 - \mathrm{e}^{-\int_0^t \lambda(u|x)}\mathrm{d}u$
- Censoring time $C_i \sim \log\mathcal{N}(\mu; 0.2^2)$, where μ influences censoring rate (30%, 60% or 90%)
- Observed time $Y_i = \min(T_i, C_i)$
- Censoring indicator $\delta_i = 1$ for $Y_i = T_i$ and $\delta_i = 0$ for other
- Error between true and estimated conditional hazard function measured by

$$\mathrm{ASE}(\hat{\lambda}) = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\lambda}(Y_i|X_i) - \lambda(Y_i|X_i)\right)^2$$

Weibull model

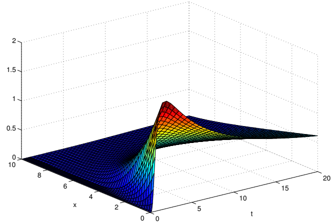
- $\lambda(t|x) = \nu\mu t^{\mu-1}\mathrm{e}^{\beta x}$
- $\nu = 0.018$, $\mu = 1.3$, $\beta = 0.2$
- The interpretation using PH or AFT model

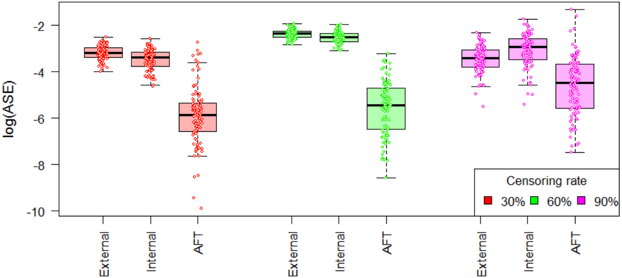




Lognormal model

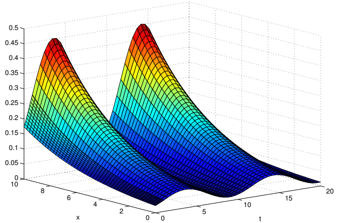
- $\lambda(t|x) = \dfrac{\frac{1}{\sigma t}\phi\left(\frac{\log t - \beta x}{\sigma}\right)}{\Phi\left(\frac{\beta x - \log t}{\sigma}\right)}$
- $\sigma = 0.5,\ \beta = \frac{1}{3}$
- The interpretation using AFT model
  $\ln T_i = \beta X_i + \varepsilon_i$,
  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$

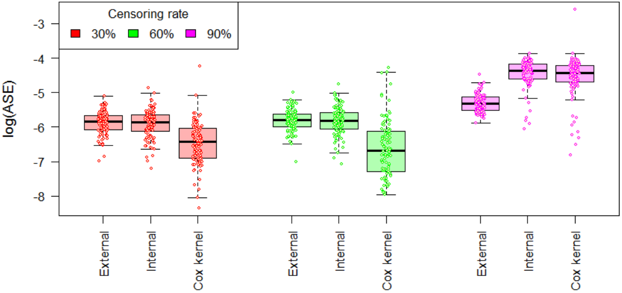




Cox model

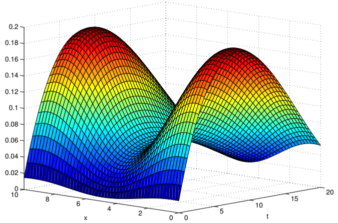
- $\lambda(t|x) = \frac{1}{15\mathrm{e}}\left(1.5 - \cos\left(\frac{t}{1.7} + 1\right)\right)\mathrm{e}^{\beta x}$
- $\beta = 0.2$
- $\lambda_0(t) = \frac{1}{15\mathrm{e}}\left(1.5 - \cos\left(\frac{t}{1.7} + 1\right)\right)$

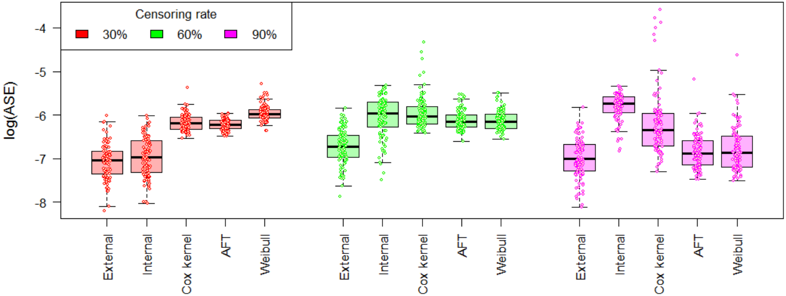




General hazard function

- $\lambda(t|x) = \frac{1}{40000}\left(t\,(t-25)^2 + 200\right) \times \times (\sin(0.7x-5)+2)$



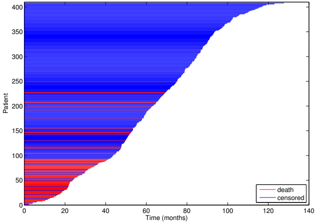


Results of simulation study

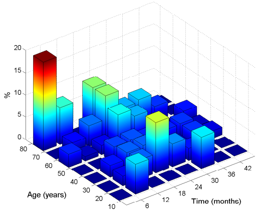
- If assumptions of the (semi)parametric models are satisfied, these models have slightly better results
- This difference between models are diminished with increasing censoring rate
- The differences between kernel estimates and parametric models are more pronounced than between kernel estimates and Cox model. This is due to stronger assumptions of the parametric models.
- The simulation of general conditional hazard function (violated assumptions of the (semi)parametric models) shows kernel estimates as preferable
- In practice, time distribution and dependence on covariates are not known $\Rightarrow$ general conditional hazard is most frequent situation

Triple negative breast cancer

- 408 patients diagnosed and/or treated at Masaryk Memorial Cancer Institute in Brno in the period 2004–2011
- 100 deaths (82 deaths in the first 4 years from diagnosis $\sim$ 80% censoring rate)
- The age at diagnosis varies between 25 and 88 years (the average 55 years)
- Is survival time affected by patient age? (Testing of hypothesis that $\beta = 0$ versus $\beta \neq 0$)

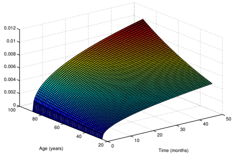


Patients' times



Death rate

Triple negative breast cancer

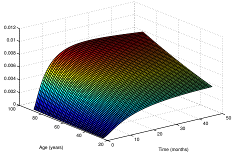


Weibull model p-value=0.129

$\hat{\beta} = 0.0128$, $\hat{\nu} = 0.0005$, $\hat{\mu} = 1.41$

PH interpretation: increase of age about one year increases risk of death 1.0128 times

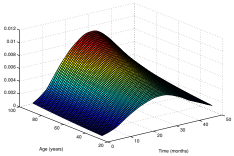
AFT interpretation: increase of age about one year shortens survival time 0.9910 times



Lognormal model p-value=0.045

$\hat{\beta} = -0.0126$, $\hat{\sigma} = 1.290$, $\hat{\mu} = 5.616$

AFT interpretation: increase of age about one year shortens survival time 0.9875 times

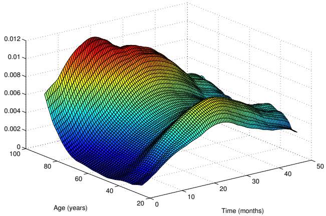


Cox model p-value=0.127

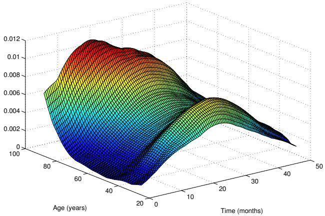
$\hat{\beta} = 0.0126$

PH interpretation: increase of age about one year increases risk of death by 1.0127 times

Triple negative breast cancer



External kernel estimate



Internal kernel estimate

- High risk of death for patients over 70 years and increased risk of death for patients up to 40 years
- Slightly different shape of hazard function for various age groups
- It offers possibility to divide patients into groups with similar risk of death (up to 40, 41-70, over 70 years)

Conclusion

- In practice, the theoretic shape of the conditional hazard function or distribution of the survival time is not known, assumptions of PH model are often violated – using kernel estimates is therefore very helpful alternative
- The kernel estimates of the conditional hazard function can be useful for verifying assumptions of the (semi)parametric models and for finding thresholds of continuous variables
- The kernel estimates are able to capture any changes in the hazard function in direction of covariate and time
- The kernel methods are able to estimate different shapes of the hazard function without any constrains and smooth them out
- The kernel estimation produces functions more useful for presentation

Conclusion

<table>
<tr><th></th><th>Advantage</th><th>Disadvantage</th></tr>
<tr><td>Kernel estimates</td><td>Flexibility
Visualization
No assumptions
Finding groups
with similar risk</td><td>Complexity of calculating
Estimate influenced
by choice of bandwidth</td></tr>
<tr><td>Parametric model</td><td>Easy interpretation
Quality of estimate for
known time distribution</td><td rowspan="2">Inadequate results can
be obtained where
assumptions of models
are violated</td></tr>
<tr><td>Semiparametric model</td><td>Easy interpretation
No distribution
assumption</td></tr>
</table>

References

- I. Selingerová, H. Doleželová, I. Horová, S. Katina, J. Zelinka, *Survival of Patients with Primary Brain Tumors: Comparison of Two Statistical Approaches*, PloS one **11**(2), 2016.
- T. R. Fleming and D. P. Harrington, *Counting processes and survival analysis*, John Wiley & Sons, 2011.
- I. Horová, J. Koláček, J. Zelinka, *Kernel Smoothing in MATLAB: Theory and Practice of Kernel Smoothing*, World Scientific Publishing, 2012.
- L. Spierdijk, *Nonparametric Conditional Hazard Rate Estimation: A Local Linear Approach*, Computational Statistics & Data Analysis **52**, 2008, 2419–2434.
- M. Svoboda et al., *Triple-Negative Breast Cancer: Analysis of Patients Diagnosed and/or Treated at the Masaryk Memorial Cancer Institute between 2004–2009* (in Czech), Clinical Oncology **25**(3), 2012, 188–198.

Thank you for your attention!