MMR'2000

Bordeaux, July 2000

The Covariate Order Method for Nonparametric Censored Exponential Regression

Bo Lindqvist and Jan Terje Kvaløy

Department of Mathematical Sciences, Norwegian University of Science and Technology,

Trondheim

bo@math.ntnu.no, jtk@math.ntnu.no

http://www.math.ntnu.no/~bo/

CONTENTS

GENERAL SETUP – LIFETIME DATA

Underlying random variables:

- Z = lifetime
- X = covariate
- C = censoring time

Basic assumption:

- Z, C are conditionally independent given X (independent censoring)

Observed:

- $T = \min(Z, C)$
- $\delta = I(Z \leq C)$
- X

Data:

- Independent triplets

  $(T_1, \delta_1, X_1), \cdots, (T_n, \delta_n, X_n)$

COVARIATE ORDERING
(single continuous covariate)

Order data $(T_1, \delta_1, X_1), \cdots, (T_n, \delta_n, X_n)$

with $X_1 \leq X_2 \leq \cdots \leq X_n$

COVARIATE ORDER PROCESS:

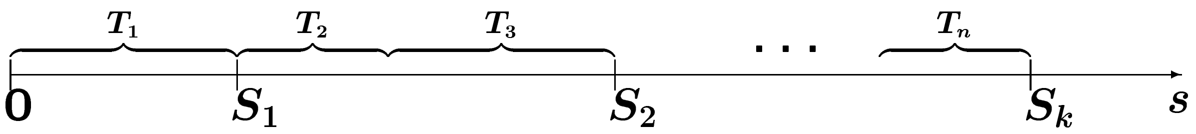
Point process $S_1, S_2, \ldots$ with

$S_i = \sum_{j=1}^{k(i)} T_j$

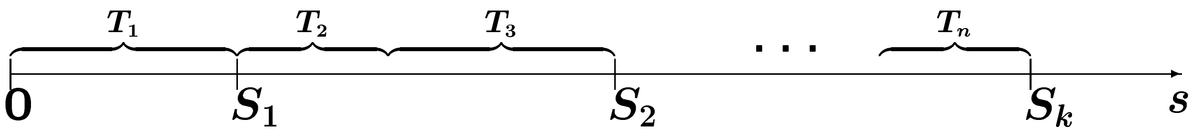
$k(i) = \min\{r | \sum_{j=1}^{r} \delta_j = i\}$.

Example:

$\delta_1 = 1, \delta_2 = 0, \delta_3 = 1, \ldots, , \delta_{n-1} = 0, \delta_n = 1$.



TESTING FOR COVARIATE EFFECT
(Kvaløy 99b)



- NO CENSORING or RANDOM CENSORING (C independent of X)

  $S_1, S_2, \ldots$ is renewal process if lifetime Z is independent of covariate X.

- INDEPENDENT CENSORING (C may depend on X)

  Warning:

  $S_1, S_2, \ldots$ is NOT necessarily a renewal process if lifetime Z independent of covariate X.

  EXCEPTIONAL CASE:

  $Z \sim \exp(\lambda)$, independent of X

  $$\Downarrow$$

  $S_1, S_2, \ldots$ is a homogeneous Poisson process (HPP). under independent censoring.

EXPONENTIAL REGRESSION
(single covariate)

MODEL: Z **given** $X = x$ **is** $\exp(\lambda(x))$;
independent censoring

- $Z|x$ **has density** $f_Z(t|x) = \lambda(x)e^{-\lambda(x)t}$
- $C|x$ **has density** $f_C(t|x)$
- Z, C **independent given** x

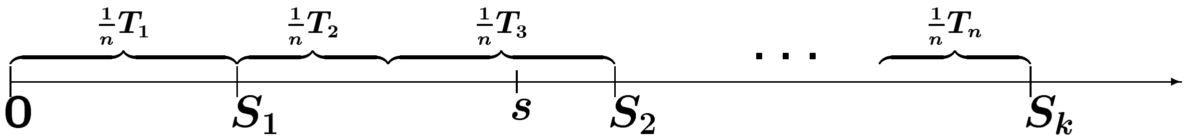
Data:

$(T_1, \delta_1, X_1), \cdots, (T_n, \delta_n, X_n)$

STANDARD NONPARAMETRIC ESTIMATION:

- **Local polynomial regression:**

COVARIATE ORDER METHOD FOR ESTIMATION OF $\lambda(x)$



Define $N(s)$ = number of censored or uncensored events before s

Conditional intensity of process $S_1, S_2, \dots$ at s (see figure) is:

$$\rho(s) = n\lambda(X_{N(s)+1})$$

Idea: Estimate the intensity by kernel estimator ("density estimator") $\hat{\rho}(s)$

Then invert relation

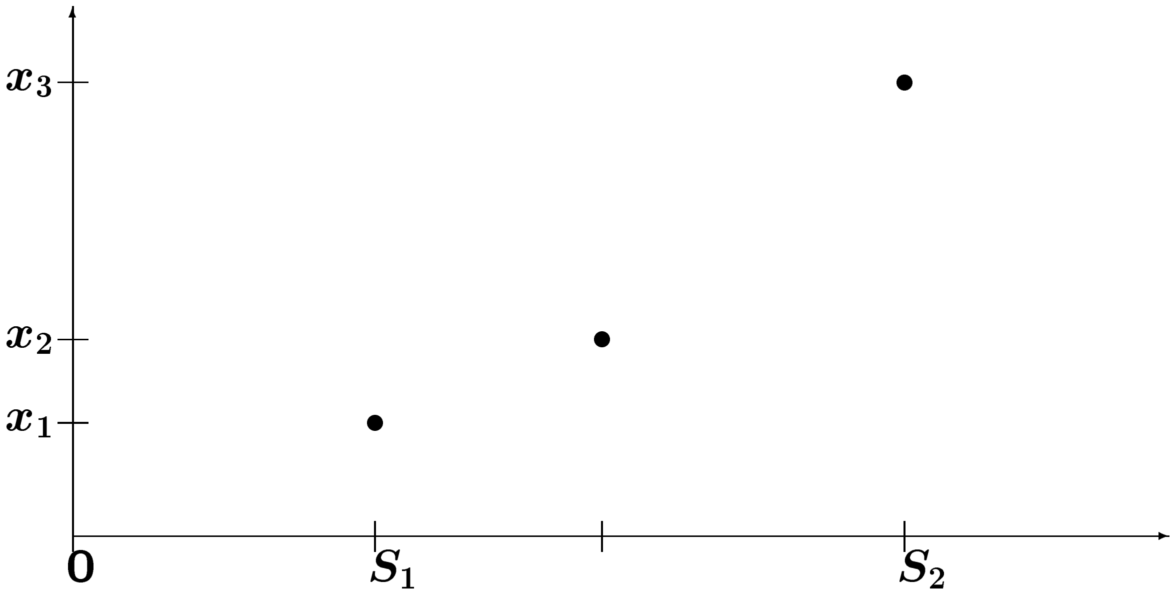
$$\hat{\rho}_n(s) = n\hat{\lambda}(X_{N(s)+1})$$

to obtain $\hat{\lambda}(x)$ at each given x.

Key here: Correspondence between points on s-axis and x-axis.

CORRESPONDENCE FUNCTION

What is the s corresponding to an x?



Step:

$$\bar{s}(x) = (1/n) \sum_{X_i \le x} T_i$$

Piecewise linear:

$$\tilde{s}(x) = \frac{1}{n} \sum_{i=1}^{J-1} T_i + \frac{x - X_{J-1}}{X_J - X_{J-1}} \frac{1}{n} T_J, \quad X_{J-1} < x \le X_J$$

Smooth:

$$\hat{s}(x) = \ldots$$

COVARIATE ORDER ESTIMATOR

$$\hat{\lambda}(x) = \hat{\rho}(\hat{s}(x)) \equiv \frac{1}{nh_s} \sum_{i=1}^{k} K\left(\frac{\hat{s}(x) - S_i}{h_s}\right)$$

is a uniformly consistent estimator for $\lambda(x)$.

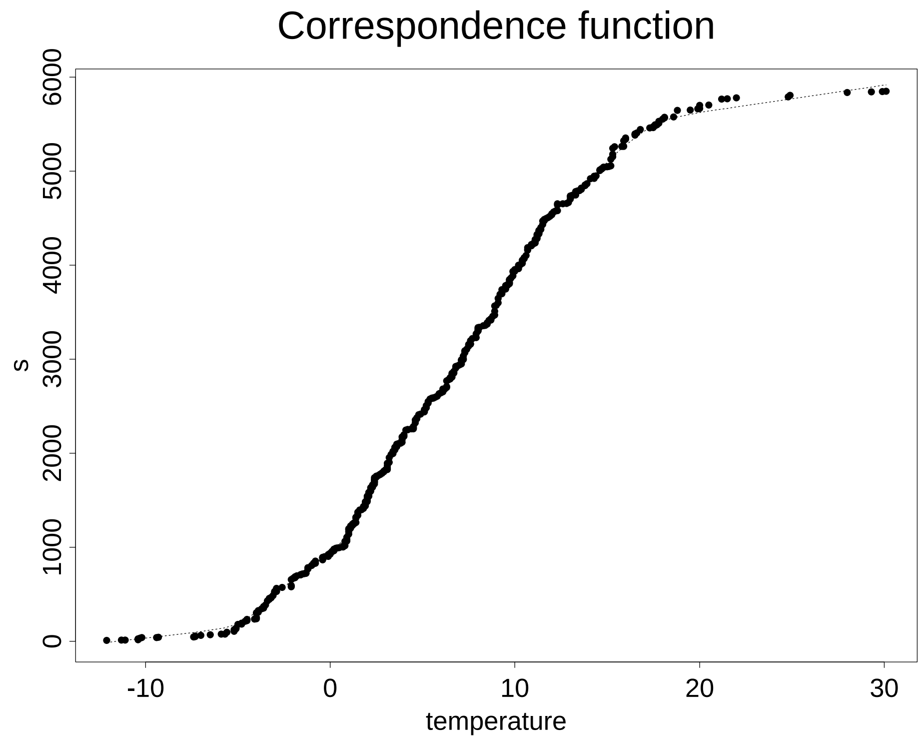
Here:

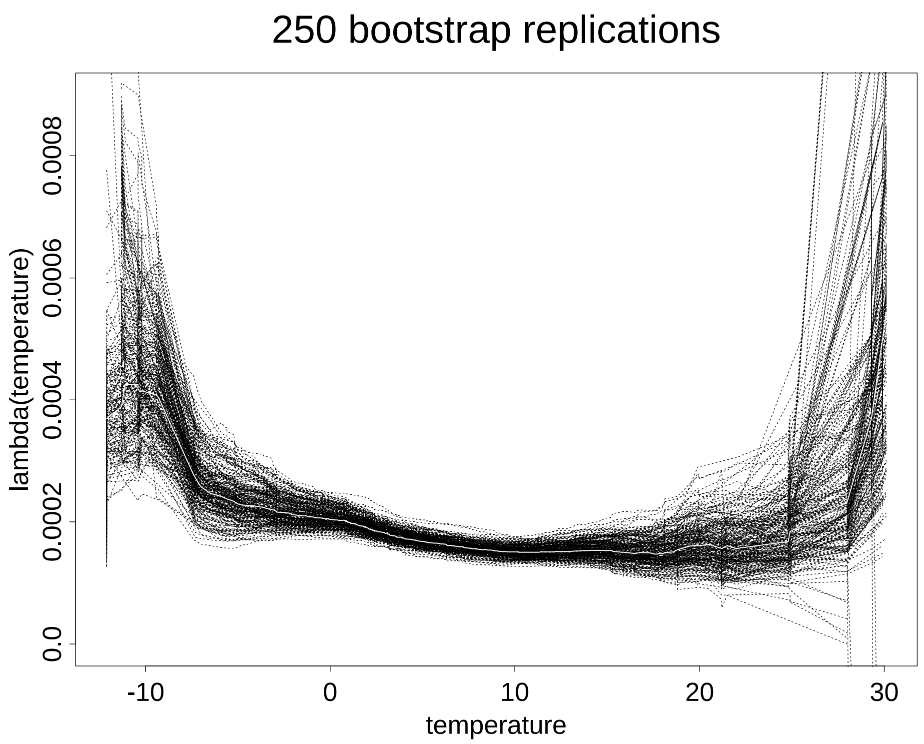
- $K(\cdot)$ is a positive kernel, integral 1
- h_s is a smoothing parameter; may depend on s
- $h_s \to 0$ as $n \to \infty$ for all s.
- There is a sequence h_n such that $h_s \geq h_n$ for all s, n where $nh_n \to \infty$ as $n \to \infty$.

EXAMPLE – CARDIAC ARREST

Times of out-of-hospital cardiac arrests reported to a Norwegian hospital over a 5 years period.

$Z =$ **inter-event times,** $X =$ **temperature.**





SEVERAL COVARIATES

$X = (X_1, \ldots, X_m)$

Local polynomial methods: Generalization in principle immediate

Covariate order method: Generalization less clear without imposing structure

GENERALIZED ADDITIVE MODELS

$$\lambda(x) = e^{\alpha + g_1(x_1) +, \ldots, + g_m(x_m)}$$

$g_1(\cdot), \ldots, g_m(\cdot)$ **estimated iteratively (backfitting)**

Basis:

$Z \sim \exp(e^{\alpha + g_1(x_1) + \ldots + g_m(x_m)})$

$$\Downarrow$$

$Ze^{\alpha + g_1(x_1) + \ldots + g_{j-1}(x_{j-1}) + g_{j+1}(x_{j+1}) + \ldots + g_m(x_m)} \sim \exp(e^{g_j(x_j)})$.

PROPORTIONAL HAZARDS

Basis:

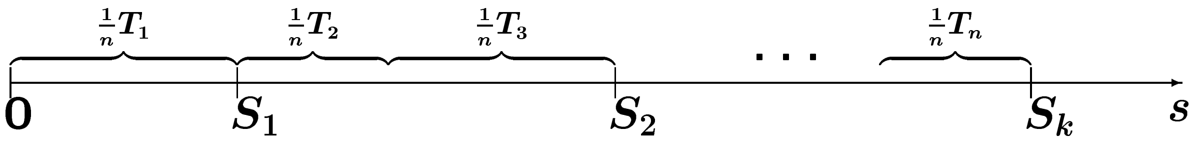
Hazard rate of Z is $\boldsymbol{\lambda}(t|x) = \boldsymbol{\lambda}_0(t)e^{g(\boldsymbol{x})}$

$$\Downarrow$$

$\boldsymbol{\Lambda_0(Z) \sim \exp(e^{g(x)})}$.

CONCLUSIONS

COVARIATE ORDER METHOD FOR NONPARAMETRIC EXPONENTIAL REGRESSION



- Simulations indicate that method is competitive w.r.t. local likelihood methods, in particular for high censoring and few observations.

- Covariate order method is flexible due to "free" choice of density estimation method

- Method is easy to use with existing software

- Method involves informative plots:
  - S-plot contains reconstructed "complete" lifetimes from censored data
  - Correspondence function gives information on covariates and outliers.

- Extension to proportional hazards models is straightforward