Reliability Analysis via Cox-type Regression Models

Uwe Jensen

University of Hohenheim

ESREL 2007, Stavanger

Outline

Axel Gandy, Constanze Lütkebohmert
www.uni-hohenheim.de/∼jensen/spp/index.html

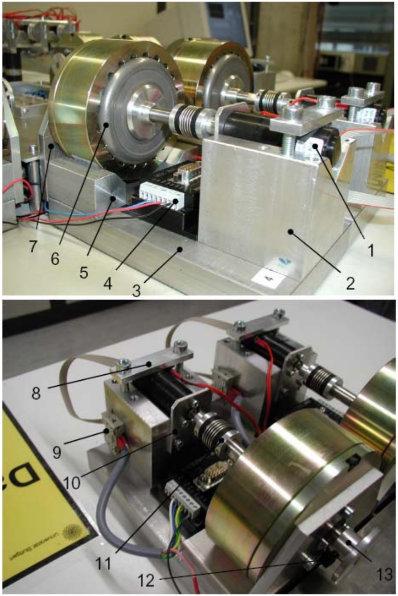
Introductory Example

The electric motor dataset

Dataset: experiments 2003 - 2005 (Prof. Schinköthe/Köder)

- 233 electric motors
- 4 covariates: load, current, voltage, r.p.m.
- 11 percent of the data censored


Achtung
Dauerversuche!
Achtung
Dauerversuche!



Legende:

1 Prüfling
2 Motorbock
3 Grundplatte
4 Ansteuerung
5 Bremsbock
6 Hysteresebremse
7 Rückwand
8 Motorspanner
9 Steckverbinder
10 Verdrehsicherung
11 Metallbalgkupplung
12 Lichtschranke
13 Drehzahlscheibe

Survival Analysis - Event History Analysis

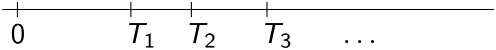
- lifetime $X \sim F$
  hazard rate $\lambda(t) \approx \frac{1}{\Delta t} P(t < X \leq t + \Delta t | X > t)$
  n individuals
  lifetimes $X_1, \ldots, X_n$
  estimate $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t)$

Survival Analysis - Event History Analysis

- lifetime $X \sim F$
  hazard rate $\lambda(t) \approx \frac{1}{\Delta t} P(t < X \leq t + \Delta t | X > t)$
  n individuals
  lifetimes $X_1, \ldots, X_n$
  estimate $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t)$
- right censoring: $\min(X, C) = X \wedge C, \delta = I(X \leq C)$
  Kaplan-Meier estimate $\hat{F}_K(t)$

Survival Analysis - Event History Analysis

- lifetime $X \sim F$
  hazard rate $\lambda(t) \approx \frac{1}{\Delta t} P(t < X \leq t + \Delta t | X > t)$
  n individuals
  lifetimes $X_1, \ldots, X_n$
  estimate $\hat{F}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t)$
- right censoring: $\min(X, C) = X \wedge C, \delta = I(X \leq C)$
  Kaplan-Meier estimate $\hat{F}_K(t)$
- more than one event per individual (object)



$N(t) = \sum_{k=1}^{\infty} I(T_k \leq t)$, $\lambda(t)$ intensity

- additional information due to k observed covariates for individual i: $Y_{ij}, i = 1, \ldots, n, j = 1, \ldots, k$
  examples:
  - medicine: age, sex, blood pressure, ...
  - software reliability: size of source code, programming language, ...
  - reliability: temperature, air pressure, ...
  - insurance: risk premium, insurance sum, ...

Regression Models

multivariate counting process

$$\mathbf{N}(t) = (N_1(t), N_2(t), ..., N_n(t)), 0 \leq t \leq \tau$$

k covariates $Y_{i1}(t), ..., Y_{ik}(t)$ for individual i

$$\mathbf{Y}_i(t) = (Y_{i1}(t), \ldots, Y_{ik}(t))^\top$$

Regression Models

multivariate counting process

$$\mathbf{N}(t) = (N_1(t), N_2(t), ..., N_n(t)), 0 \leq t \leq \tau$$

k covariates $Y_{i1}(t), ..., Y_{ik}(t)$ for individual i

$$\mathbf{Y}_i(t) = (Y_{i1}(t), \ldots, Y_{ik}(t))^\top$$

Cox Model:
$\mathbf{N}(t)$ n-variate counting process with intensity $\boldsymbol{\lambda}(t)^\top = (\lambda_1(t), \ldots, \lambda_n(t))$

$$\lambda_i(t) = \lambda_0(t) R_i(t) \exp\{\boldsymbol{\beta}^\top \mathbf{Y}_i(t)\}$$

λ_0 baseline hazard
R_i at risk indicator
$\boldsymbol{\beta}$ parameter vector

The Aalen Model

Model assumption:

$$\boldsymbol{\lambda}(t) = \mathbf{Y}(t)\boldsymbol{\alpha}(t)$$

$\lambda_i(t) = Y_{i1}(t)\alpha_1(t) + ... + Y_{ik}(t)\alpha_k(t)$

$\boldsymbol{\alpha}(t) = (\alpha_1(t), ..., \alpha_k(t))$ unknown baseline intensities

Interpretation:

$\alpha_j(t)$ mean number of failures per unit time per unit of covariate $Y_{ij}(t)$

Model Checks - The Test

Aim: Model check for classes of regression models of type

$$\lambda_i(t) = f(\mathbf{Y}_i(t), \boldsymbol{\alpha}^c, \boldsymbol{\alpha}^v(t))$$

f:	Continuous function (known)
$\mathbf{Y}_i$:	Vector of predictable stochastic processes (covariates, observable)
$\boldsymbol{\alpha}^c$:	Finite-dimensional regression parameter (unknown)
$\boldsymbol{\alpha}^v$:	Vector of regression functions (unknown)

Model Checks - The Test

Aim: Model check for classes of regression models of type

$$\lambda_i(t) = f(\mathbf{Y}_i(t), \boldsymbol{\alpha}^c, \boldsymbol{\alpha}^v(t))$$

f:	Continuous function (known)
$\mathbf{Y}_i$:	Vector of predictable stochastic processes (covariates, observable)
$\boldsymbol{\alpha}^c$:	Finite-dimensional regression parameter (unknown)
$\boldsymbol{\alpha}^v$:	Vector of regression functions (unknown)

Develop goodness-of-fit tests which can be tuned to be powerful against certain alternatives

Details in Gandy/Je 2005 a,b (SJS, preprint) and
Gandy 2005 (dissertation)
www.uni-hohenheim.de/∼jensen/spp/index.html
http://www.ma.ic.ac.uk/ agandy/

Model Checks

Martingale $\mathbf{M}(t) = \mathbf{N}(t) - \int_0^t \boldsymbol{\lambda}(s)\mathrm{d}s$

Martingale residuals $\hat{\mathbf{M}}(t) = \mathbf{N}(t) - \int_0^t \hat{\boldsymbol{\lambda}}(s)\mathrm{d}s$

Basis for goodness-of-fit test

$$\hat{T}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t c_i(s)\mathrm{d}\hat{M}_i(s) = \frac{1}{\sqrt{n}} \int_0^t \mathbf{c}(s)^\top \mathrm{d}\hat{\mathbf{M}}(s)$$

with predictable weights $\mathbf{c}$

Idea: Choose the weights c such that

- the asymptotic distribution can be derived
- the test is powerful against certain alternatives

Graphical Model Check

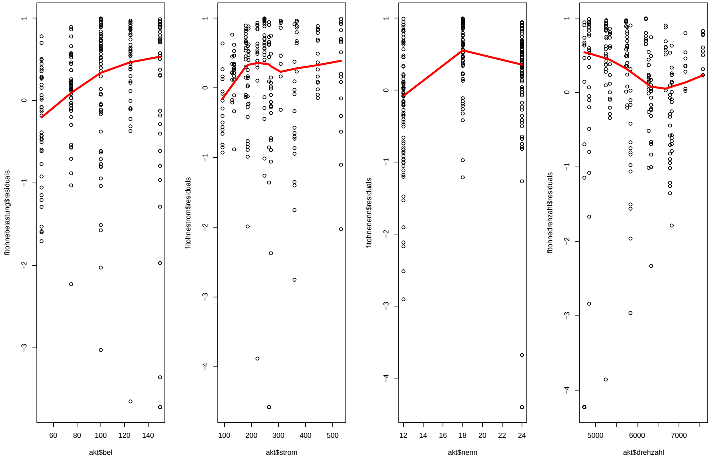
Data analysis of the functional form of the covariates $\mathbf{Y}$ in a Cox model:

Scatterplot of martingale residuals against a covariate

Therneau/Grambsch/Flemming 1990, Gandy/Je/Lütkebohmert 2005

Motor Data Set

Smoothed Martingale Residual Plots



Cox-type Models

Cox's Model with Change Points:
Liang/Self/Liu 90, Luo/Boyett 97, Pons 03

vector of covariates $\mathbf{Y}(t) = (\mathbf{Y_1}(t)^\top, \mathbf{Y_2}(t)^\top)^\top$
$\mathbf{N}(t)$ n-variate counting process with intensity

$$\boldsymbol{\lambda}(t)^\top = (\lambda_1(t), \ldots, \lambda_n(t))$$

$$\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) \exp\left\{\boldsymbol{\beta}_1^\top \mathbf{Y}_{1i}(t) + \boldsymbol{\beta}_2^\top \mathbf{Y}_{2i}(t) + \boldsymbol{\beta}_3^\top (\mathbf{Y}_{2i}(t) - \boldsymbol{\xi})^+\right\},$$

where $\boldsymbol{\theta} = (\boldsymbol{\xi}^\top, \boldsymbol{\beta}^\top)^\top$ with $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top, \boldsymbol{\beta}_3^\top)^\top \in \mathcal{B} \subset \mathbb{R}^{p+2m}$

Change points $\boldsymbol{\xi}$, vector of parameters lying in the known rectangle
$\Xi := [\xi_{11}, \xi_{21}] \times [\xi_{12}, \xi_{22}] \times \cdots \times [\xi_{1m}, \xi_{2m}]$

Short form

$$\lambda_i(t, \boldsymbol{\theta}) = \lambda_0(t) R_i(t) \exp\{\boldsymbol{\beta}^\top \tilde{\mathbf{Y}}_i(t; \boldsymbol{\xi})\},$$

Estimates,Consistency

$\boldsymbol{\theta}_0$ is estimated by the value $\hat{\boldsymbol{\theta}}_n$ that maximizes the logarithm of the partial likelihood

Theorem

Under suitable conditions there exists a neighborhood Θ_0 of $\boldsymbol{\theta}_0$ such that if $\hat{\boldsymbol{\theta}}_n$ lies in Θ_0, it follows that $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\boldsymbol{\theta}_0$ as $n \to \infty$.

Asymptotic Normality

- Standard methods fail,
  the partial likelihood function is **not** differentiable with respect to its parameters
- Results about asymptotic normality of M-estimators,
  key words: criterion function, Lipschitz, limit function with a second order Taylor expansion
  see e.g. van der Vaart: Asymptotic Statistics 1998.

Theorem

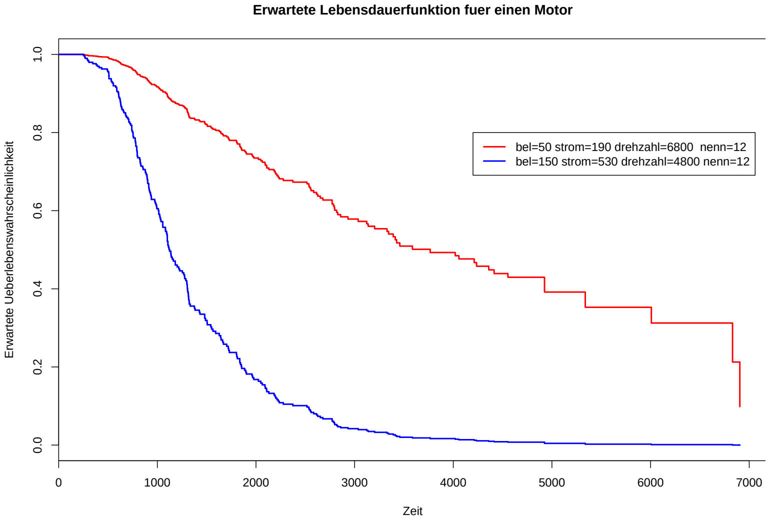
Under suitable technical conditions and under the assumption that $\hat{\boldsymbol{\theta}}_n$ is a consistent sequence of estimators of $\boldsymbol{\theta}_0$, $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is asymptotically normal with mean zero and a covariance matrix which can be estimated consistently.

Applications - Electric Motor Dataset

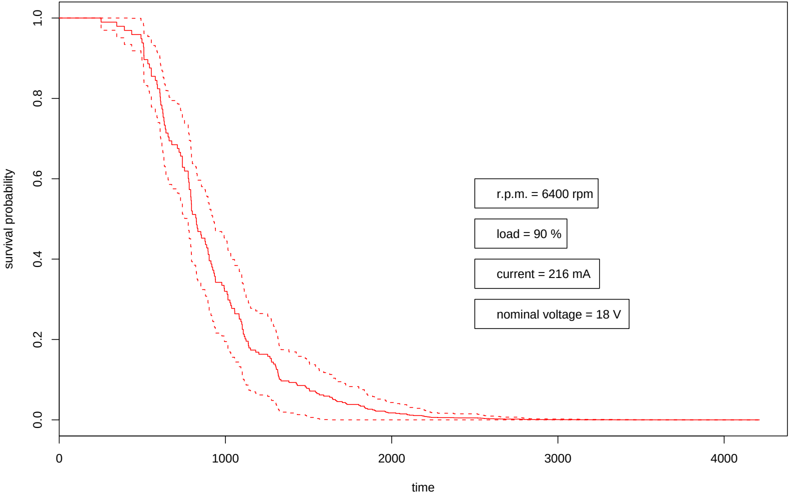
Dataset: experiments 2003 - 2005 (Prof. Schinköthe/Köder)

- 232 electric motors
- 4 covariates: load, current, voltage, r.p.m.
- 11 percent of the data censored

null hypothesis	competing model	p-value
Model O	Model CP in voltage	0.0001



Erwartete Lebensdauerfunktion fuer einen Motor
bel=50 strom=190 drehzahl=6800 nenn=12
bel=150 strom=530 drehzahl=4800 nenn=12
Erwartete Ueberlebenswahrscheinlichkeit
0.0
0.2
0.4
0.6
0.8
1.0
0
1000
2000
3000
4000
5000
6000
7000
Zeit



survival probability
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
time
r.p.m. = 6400 rpm
load = 90 %
current = 216 mA
nominal voltage = 18 V

Summary

- Models in Survival Analysis can be applied in Reliability
- Regression models with covariates
- Model checks by directed goodness of fit tests
- Smoothed martingale residual plots
- Cox model with change points in the covariates
- Consistency and asymptotic normality of estimators can be shown