Resampling methods in clinical research

Laboratoire de Statistique Théorique et Appliquée, UPMC, Paris 6

9 Avril 2015

John O'Quigley and Michel Broniatowski

Outline

General

- $X_1, ..., X_n$
- $F(x) = \Pr(X \leq x)$
- $F_n(x) = \sum_{i=1}^{n} n^{-1} I(X_i \leq x)$
- $\mu = E(X; F) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} x dF(x)$
- $\bar{x} = E(X; F_n) = \int_{-\infty}^{+\infty} x dF_n(x)$
- Variance, median, coefficient of variation: $\sigma^2(F)$, $m(F)$, $CV(F)$

Bootstrap resampling

- To obtain more accurate inference, in particular more accurate confidence intervals
- To facilitate inference for parameter estimators in complex situations.

A broad discussion including several challenging applications is provided by Politis (1998).

Bootstrap resampling

Write $\theta = \theta(F)$ and $\tilde{\theta} = \theta(F_n)$ as an estimator for $\theta(F)$.

- Infinitely many i.i.d. samples, each of size n, from F provides exact sampling properties of any estimator $\tilde{\theta} = \theta(F_n)$.
- Take a very large number, say B, of samples of size n from F provides approximations to the sampling properties of $\tilde{\theta}$,
- Replace F by F_n

Bootstrap resampling

Each of the B samples is viewed as an i.i.d. sample from $F_n(t)$.

- The i th resample of size n can be written $X^*_{1i}, X^*_{2i}, ..., X^*_{ni}$ and has empirical distribution $F_n^{*i}(t)$.
- $\theta(F)$ is the population quantity of interest.
- $\theta(F_n)$ is this same quantity defined with respect to the empirical distribution F_n
- $\theta(F_n^{*i})$ is again the same quantity defined with respect to the i th empirical distribution of the resamples $X^*_{1i}, X^*_{2i}, ..., X^*_{ni}$.
- Finally, $F_B(\theta)$ is the bootstrap distribution of $\theta(F_n^{*i})$, i.e., the empirical distribution of $\theta(F_n^{*i})$ $(i = 1, \ldots, B)$.

Bootstrap resampling: large sample theory

- As $B \to \infty$, $\int u dF_B(u) \xrightarrow{p} \theta(F_n)$
- As $n \to \infty$, $\theta(F_n) \xrightarrow{p} \theta(F)$.
- F_B deals with the distribution of $\theta(F_n^{*i})$ $(i = 1, \ldots, n)$ and therefore, when our focus of interest changes from one parameter to another, from say θ_1 to θ_2 the function F_B would be generally quite different.
- This is not the case for F, F_n and F_n^{*i}

Bootstrap resampling: confidence intervals

- Var $\{\theta(F_n)\} = \sigma_B^2 = \int u^2 dF_B(u) - \left(\int u dF_B(u)\right)^2$
- Var$(\cdot|F_B)$ is wrt the distribution $F_B(x)$.
- Var $\{\theta(F_n)|F_B\}$, can be used as an estimator of Var $\{\theta(F_n)|F\}$
- Var $\theta(F_n) = \sigma^2\,, \quad Q_\alpha = \sigma z_\alpha\,, \quad \Phi(z_\alpha) = \alpha.$
- $I_{1-\alpha}(\theta) = \{\theta(F_n) - Q_{1-\alpha/2},\ \theta(F_n) - Q_{\alpha/2}\}$

Bootstrap resampling: confidence intervals

- Instead of $\sigma^2(F)$ use σ_B^2.
- Instead of normal approximation can define $F_B(Q_\alpha) = \alpha$.
- If no exact solution use nearest approximation $F_B^{-1}(\alpha)$ as Q_α.
- The intervals are called bootstrap “root” intervals

Bootstrap resampling: confidence intervals

- Use $I_{1-\alpha}(\theta) = \{\theta(F_n) + Q_{\alpha/2},\ \ \theta(F_n) + Q_{1-\alpha/2}\}$ where $Q_{\alpha/2}$ and $Q_{1-\alpha/2}$ are percentiles of F_B
- The intervals are called bootstrap percentile intervals
- When F_B symmetric, $Q_{\alpha/2} + Q_{1-\alpha/2} = 0$, and percentile intervals and root intervals coincide
- They coincide, in particular, under normal approximation

Bootstrap resampling: accuracy of confidence intervals

- Edgeworth expansions show accuracy of all 3 types of interval to be the same
- Studentized methods lead to slight gains in accuracy For the i th bootstrap sample our estimate of the variance is Var $\{\theta(F_n^{*i})\} = \sigma_{*i}^2 = \int u^2 dF_n^{*i}(u) - \left(\int u dF_n^{*i}(u)\right)^2$ and we then consider the standardized distribution of the quantity, $\theta(F_n^{*i})/\sigma_{*i}$.
- Bias-corrected, accelerated intervals, called BC_a intervals

Closely related methods

- Jacknife
- Cross-validation
- Permutation tests, eg. Fisher's “exact” test
- Subsampling
- Wild, block, cluster, smooth, parametric bootstrap

Main limitations and difficulties

- Small samples
- High dimensional problems
- Maintaining correlation structure in complex models
- Misspecified models, when $X(F_n) \not\to X(F)$

ACR (analytic conditional resampling)

- If $F(x) = \Pr(X \leq x)$ and $Y = F(X)$ then $\Pr(Y < y) = y$
- Let $Y_{(1)}, ..., Y_{(n)}$ be order statistics for $Y_1, ..., Y_n$ and let $W_i = Y_{(i)} - Y_{(i-1)}$ for $i = 1, ..., n$.

Main Theorem

Let V_i , $i = 1, ..., n$, be i.i.d. where $\Pr(V_i \leq v) = 1 - \exp(-v)$, and let $W_i^* = V_i / \sum_{j=1}^{n} V_j$ then, $W_i \cong W_i^*$.

ACR (analytic conditional resampling)

- We condition on the data $X_1, ..., X_n$
- Estimating equation is of the form $\int U(s)dG(s) = 0$
- Obtain “exact” analytic conditional solutions based on W^*
- Numerical approximations based on Cornish-Fisher or Saddlepoint methods

Inference for the mean

- $X_1, ..., X_n \sim F(x)$ with $E(X) = \mu$, Var $(X) = \sigma^2$
- $\bar{x} = \hat{\mu} = n^{-1} \sum_i X_i$
- $\bar{x} = \sum_i W_i X_i \,, \quad \sum_{i=1}^{n} W_i = 1$
- Replace W_i by W_i^*
- Replace μ by $\bar{x}$ and σ^2 by usual empirical variance; then the two approaches agree in the first two moments.

Fisher's exact test

Table: Conditional test of two proportions

	Group 1	Group 2	Total
+ ve	4	3	7
- ve	9	18	27
Total	13	21	34

- $T = \sum_{i=1}^{34} W_i Y_i$
- Base test on det$(A|H_0)$ where A is above matrix.

Analytic conditional resampling in survival analysis

When conditionning on the risk sets,

- $\mathcal{Z}(t_i)$ is the covariate of the failing subject at time t_i,
- $\mathcal{E}_{\beta_0}(Z|t_i)$ and $\mathcal{V}_{\beta_0}(Z|t_i)$ are the expectation and variance of $\mathcal{Z}(t_i)$,
- $\mathcal{Z}(t_i)$ is independent of $\mathcal{Z}(t_j)$, $i \neq j$
  (Cox 1975, Andersen and Gill 1982).

Analytic conditional resampling in survival analysis

When conditionning on the risk sets,

- $\mathcal{Z}(t_i)$ is the covariate of the failing subject at time t_i,
- $\mathcal{E}_{\beta_0}(Z|t_i)$ and $\mathcal{V}_{\beta_0}(Z|t_i)$ are the expectation and variance of $\mathcal{Z}(t_i)$,
- $\mathcal{Z}(t_i)$ is independent of $\mathcal{Z}(t_j)$, $i \neq j$
  (Cox 1975, Andersen and Gill 1982).

For $j = 1, ..., k_n$, the standardized score process is

$$\begin{aligned} U^*(\beta_0, t_j) &= \frac{1}{\sqrt{k_n}} \sum_{i=1}^{j} \mathcal{V}_{\beta_0}(Z|t_i)^{-1/2} \Big(\mathcal{Z}(t_i) - \mathcal{E}_{\beta_0}(Z|t_i) \Big) \\ &= \frac{1}{\sqrt{k_n}} \int_0^{t_j} \mathcal{V}_{\beta_0}(Z|s)^{-1/2} \Big(\mathcal{Z}(s) - \mathcal{E}_{\beta_0}(Z|s) \Big) d\bar{N}(s). \end{aligned}$$

For $j = 1, ..., k_n$, the standardized score process is

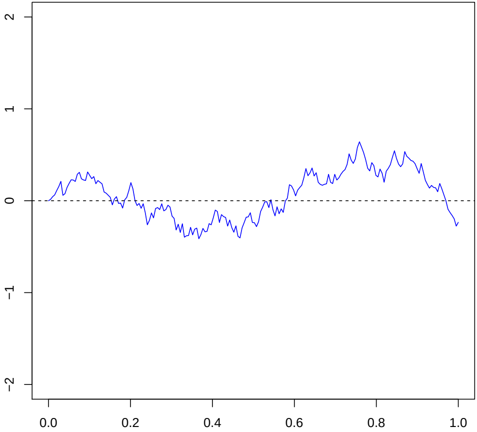
$$U^*(\beta_0, t_j) = \frac{1}{\sqrt{k_n}} \sum_{i=1}^{j} \mathcal{V}_{\beta_0}(Z|t_i)^{-1/2} \Big(\mathcal{Z}(t_i) - \mathcal{E}_{\beta_0}(Z|t_i) \Big).$$

Theorem

$\forall\, t \in [0;1]$, *under the Cox model of parameter* β_0,

$$U^*(\beta_0, t) \xrightarrow[n\to\infty]{D} W(t),$$

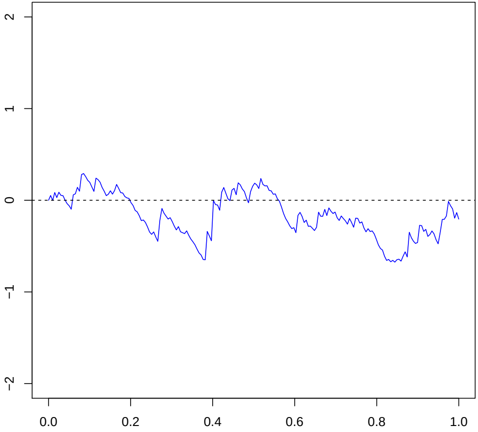
where W *is a standard Brownian motion process.*



$$\beta = \log(4)$$

$$n = 200$$

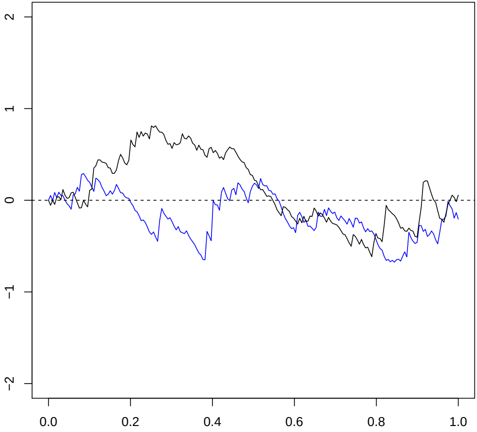
$$Z \sim \mathcal{E}(0.5)$$



$\beta = 0$

$n = 200$

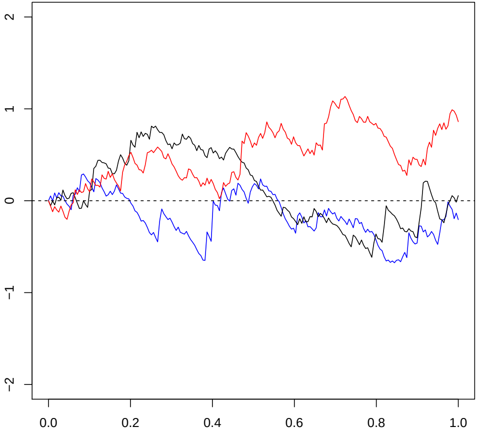
$Z \sim \mathcal{E}(0.5)$



$\beta = 0$

$n = 200$

$Z \sim \mathcal{E}(0.5)$



$\beta = 0$

$n = 200$

$Z \sim \mathcal{E}(0.5)$

Theorem

$\forall t > 0, \forall \beta_0$, *under the Cox model of parameter* β,

$$U^*(\beta_0, t) - \sqrt{n}(\beta - \beta_0)A_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

where $A_n(t)$ *is an explicit process converging to a deterministic known function* $A(t)$ *with probability 1.*

Theorem

$\forall t > 0, \forall \beta_0$, *under the Cox model of parameter* β*,*

$$U^*(\beta_0, t) - \sqrt{n}(\beta - \beta_0)A_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

where $A_n(t)$ *is an explicit process converging to a deterministic known function* $A(t)$ *with probability 1.*

For a fixed n, for all $t > 0$,

$$U^*(\beta_0, t) \quad \overset{D}{\approx} \quad W(t) + \sqrt{n}(\beta - \beta_0)A(t)$$

Theorem

$\forall t > 0, \forall \beta_0$, *under the Cox model of parameter* β,

$$U^*(\beta_0, t) - \sqrt{n}(\beta - \beta_0)A_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

where $A_n(t)$ *is an explicit process converging to a deterministic known function* $A(t)$ *with probability 1.*

For a fixed n, for all $t > 0$,

$$U^*(\beta_0, t) \overset{D}{\approx} W(t) + \sqrt{n}(\beta - \beta_0)A(t)$$
$$U^*(\beta_0, t) \overset{D}{\approx} W(t) + C\sqrt{n}(\beta - \beta_0)t$$

In this situation, $U^*(\beta_0, t)$ can be approximated by a standard brownian motion with linear drift.

Theorem

$\forall t > 0, \forall \beta_0$, under the Cox model of parameter β,

$$U^*(\beta_0, t) - \sqrt{n}(\beta - \beta_0)A_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

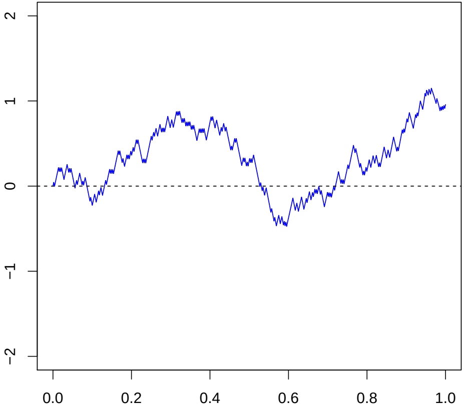
where $A_n(t)$ is an explicit process converging to a deterministic known function $A(t)$ with probability 1.

For a fixed n, for all $t > 0$,

$$U^*(\beta_0, t) \overset{D}{\approx} W(t) + \sqrt{n}(\beta - \beta_0)A(t)$$
$$U^*(\beta_0, t) \overset{D}{\approx} W(t) + C\sqrt{n}(\beta - \beta_0)t$$

In this situation, $U^*(\beta_0, t)$ can be approximated by a standard brownian motion with linear drift.

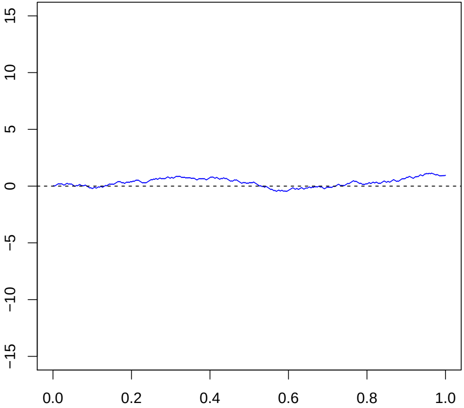
Application: In practice, the process U^* is assessed in $\beta_0 = 0$.



$$\beta = 0$$

$$n = 500$$

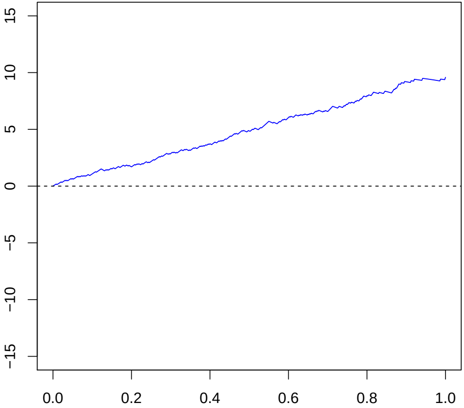
$$Z \sim \mathcal{B}(0.5)$$



$\beta = 0$

$n = 500$

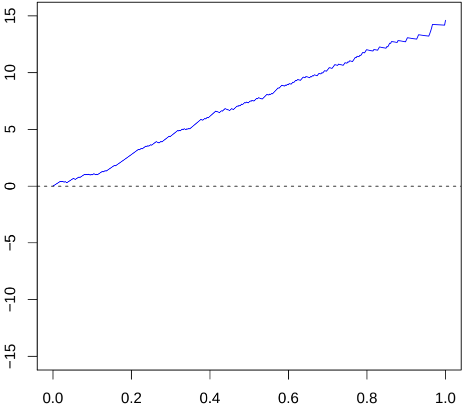
$Z \sim \mathcal{B}(0.5)$



$\beta = \log(2)$

$n = 500$

$Z \sim \mathcal{B}(0.5)$



$\beta = \log(4)$

$n = 500$

$Z \sim \mathcal{B}(0.5)$

Non-proportional hazards model

Define the class of non-proportional hazards model as follows

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta(t)Z),$$

where $\lambda_0(t)$ is a baseline hazard, $\beta(t)$ a time-dependent regression parameter and Z a covariate.

Non-proportional hazards model

Define the class of non-proportional hazards model as follows

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta(t)Z),$$

where $\lambda_0(t)$ is a baseline hazard, $\beta(t)$ a time-dependent regression parameter and Z a covariate.

When $\beta(t) = \beta_0$, the model satisfies the proportional hazards assumption (Cox 1972).

Non-proportional hazards model

Define the class of non-proportional hazards model as follows

$$\lambda(t|Z) = \lambda_0(t)\exp(\beta(t)Z),$$

where $\lambda_0(t)$ is a baseline hazard, $\beta(t)$ a time-dependent regression parameter and Z a covariate.

When $\beta(t) = \beta_0$, the model satisfies the proportional hazards assumption (Cox 1972).

What is the shape of the standardized score process if the "true" coefficient is not constant over time ?

Theorem

$\forall t \in [0;1]$, *under the non-proportional hazards model of parameter* $\beta(t)$,

$$U^*(0,t) - \sqrt{n}\beta(t)B_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

where $B_n(t)$ *is an explicit process converging to a deterministic known function* $B(t)$, *with probability 1.*

Theorem

$\forall t \in [0;1]$, *under the non-proportional hazards model of parameter* $\beta(t)$,

$$U^*(0,t) - \sqrt{n}\beta(t)B_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

where $B_n(t)$ *is an explicit process converging to a deterministic known function* $B(t)$*, with probability 1.*

For a fixed n, for all $t > 0$,

$$U^*(0,t) \quad \overset{D}{\approx} \quad W(t) + \sqrt{n}B(t)\beta(t)$$

Theorem

$\forall t \in [0;1]$, *under the non-proportional hazards model of parameter* $\beta(t)$,

$$U^*(0,t) - \sqrt{n}\beta(t)B_n(t) \xrightarrow[n\to\infty]{D} W(t),$$

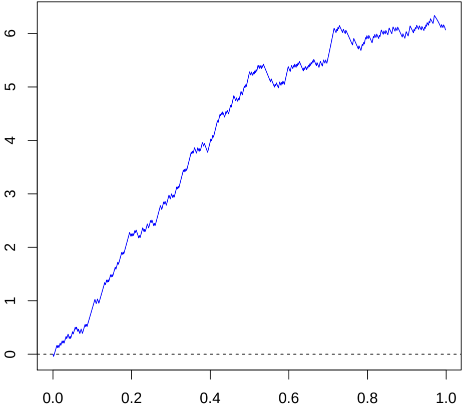
where $B_n(t)$ *is an explicit process converging to a deterministic known function* $B(t)$*, with probability 1.*

For a fixed n, for all $t > 0$,

$$\begin{aligned} U^*(0,t) &\overset{D}{\approx} W(t) + \sqrt{n}B(t)\beta(t) \\ U^*(0,t) &\overset{D}{\approx} W(t) + \sqrt{n}Kt\beta(t) \end{aligned}$$

In this case, the shape of the drift reflects the shape of $\beta(t)$.

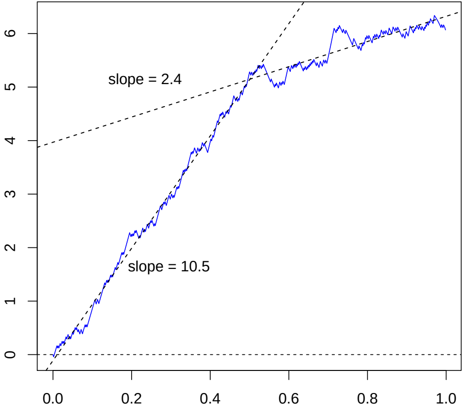
Non-proportional hazards



n=600

$\beta(t)$=$\mathbf{1}_{t\leq 0.5}$+$0.2\mathbf{1}_{t>0.5}$

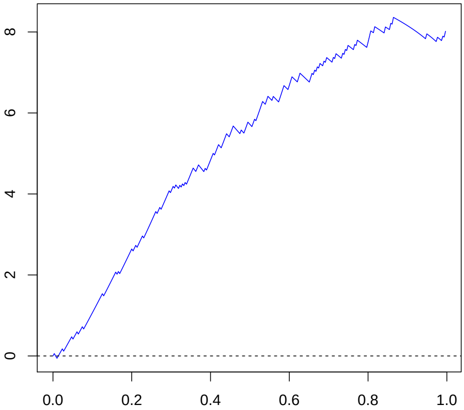
Non-proportional hazards



n=600

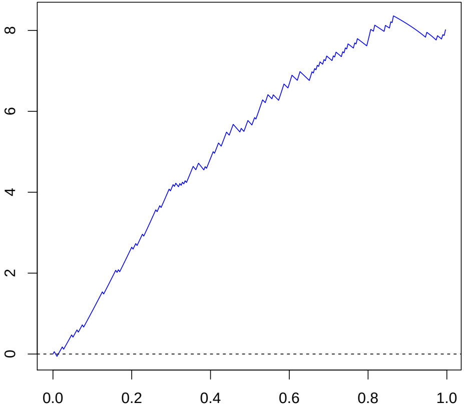
$\beta(t)$=$\mathbf{1}_{t\leq 0.5}$+$0.2\mathbf{1}_{t>0.5}$

ratio of slopes
=$2.4/10.5 \approx 0.22$.



n=300

$\beta(t)$=$2(1-t)$



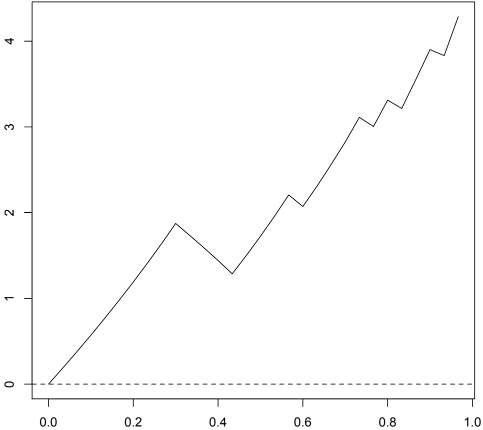
n=300

$\beta(t)$=$2(1-t)$

$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	1.01	0.22
$\beta_0(1-t)$	1.99	**0.40**
$\beta_0(1-t)^2$	2.63	0.38
$\beta_0(1-t^2)$	1.52	0.34
$\beta_0 \log(t)$	-0.86	0.26

Freireich dataset

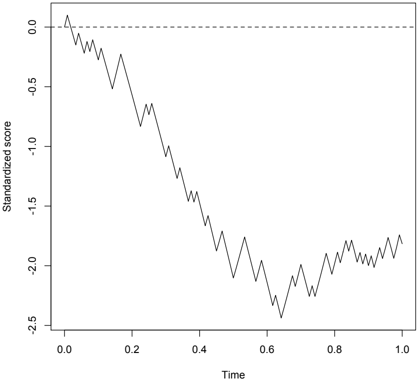
42 patients, 29% of censoring.



$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	1.6	**0.41**
$\beta_0 t$	2.43	0.32
$\beta_0 t^2$	3.00	0.28
$\beta_0 t^3$	3.48	0.25
$\beta_0(1-t)$	2.7	0.3
$\beta_0(1-t)^2$	3.68	0.25
$\beta_0(1-t^2)$	2.03	0.32

Advanced lung cancer- Karnofsky score

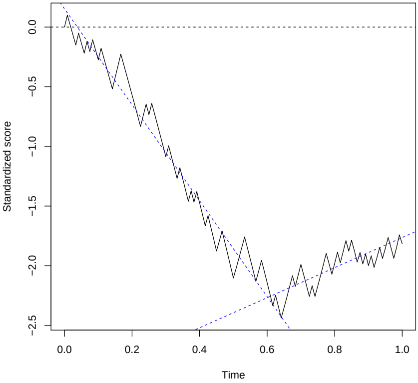
167 patients, 23% of censoring.



$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	-0.33	0.03
$\beta_0(\mathbf{1}_{t\leq 0.6} -0.31.\mathbf{1}_{t\geq 0.6})$	-0.58	0.06
$\beta_0\mathbf{1}_{t\leq 0.5}$	-0.82	**0.08**
$\beta_0(1-t)$	-0.83	0.05
$\beta_0(1-t)^2$	-1.06	0.05
$\beta_0(1-t^2)$	-0.64	0.05

Advanced lung cancer- Karnofsky score

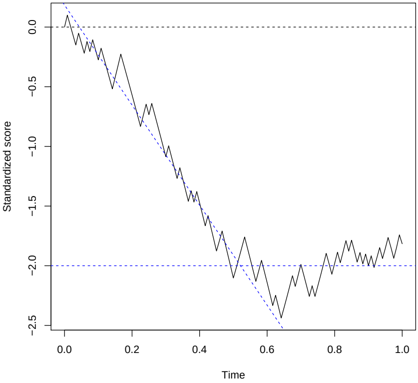
167 patients, 23% of censoring.



$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	-0.33	0.03
$\beta_0(\mathbf{1}_{t\leq 0.6}$ $-0.31.\mathbf{1}_{t\geq 0.6})$	-0.58	0.06
$\beta_0\mathbf{1}_{t\leq 0.5}$	-0.82	**0.08**
$\beta_0(1-t)$	-0.83	0.05
$\beta_0(1-t)^2$	-1.06	0.05
$\beta_0(1-t^2)$	-0.64	0.05

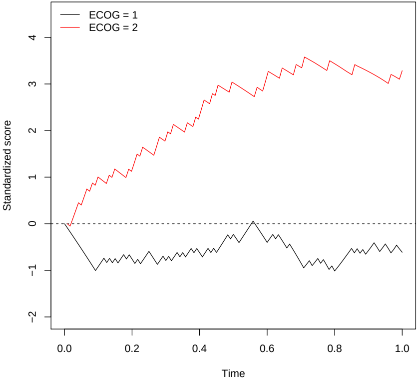
Advanced lung cancer- Karnofsky score

167 patients, 23% of censoring.



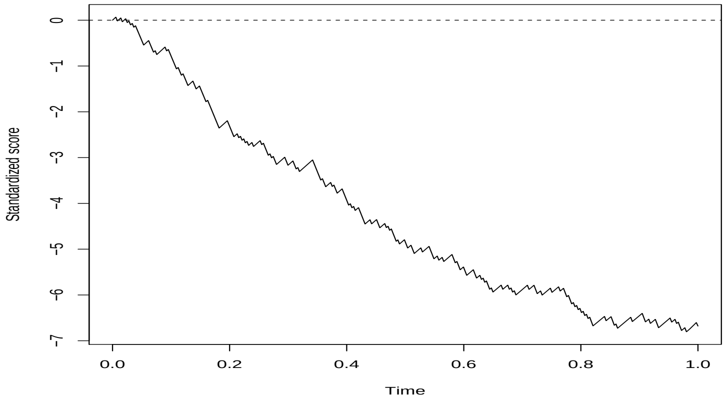
$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	-0.33	0.03
$\beta_0(\mathbf{1}_{t\leq 0.6} - 0.31.\mathbf{1}_{t\geq 0.6})$	-0.58	0.06
$\beta_0\mathbf{1}_{t\leq 0.5}$	-0.82	**0.08**
$\beta_0(1-t)$	-0.83	0.05
$\beta_0(1-t)^2$	-1.06	0.05
$\beta_0(1-t^2)$	-0.64	0.05

Advanced lung cancer - ECOG

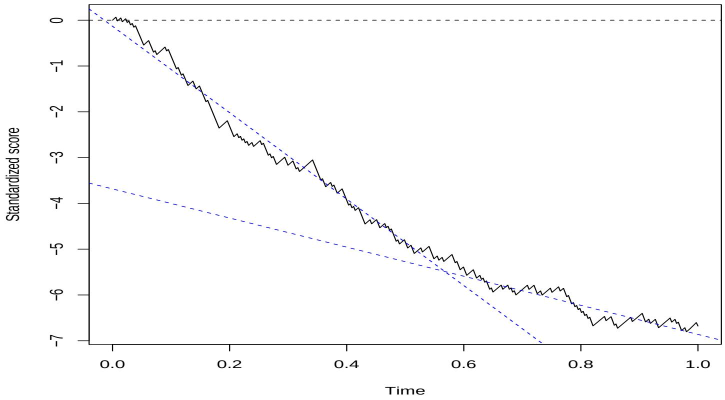


$\beta(t)$	$\hat{\beta}_0$	R^2
β_0	0.72	0.09
$\beta_0 \mathbf{1}_{t \leq 0.5}$	1.10	**0.13**
$\beta_0(1-t)$	1.36	0.12
$\beta_0(1-t)^2$	1.69	0.11
$\beta_0(1-t^2)$	1.07	0.11

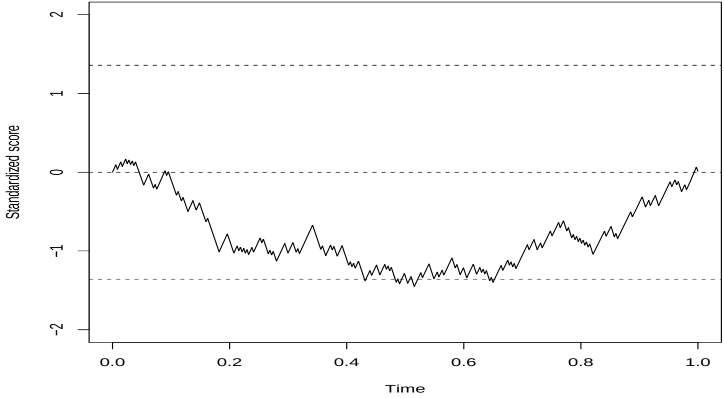
How to model NPH problems



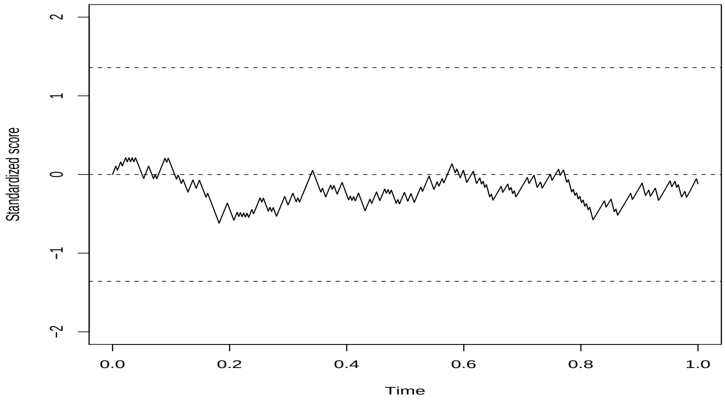
How to model NPH problems



How to model NPH problems



How to model NPH problems



Distance from origin test (log-rank)

$H_0 : \beta = \beta_0$ against $H_1 : \beta \neq \beta_0$.

Lemma

Set z^α such that $\alpha = 2(1 - \Phi(z^\alpha))$. The distance from origin test rejects H_0 with a type I error α if $\left|U^*(\beta_0, t)/\sqrt{t}\right| \geq z^\alpha$.

The test p-value is given by $2\left[1 - \Phi\left(\left|U^*(\beta_0, t)\right|/\sqrt{t}\right)\right]$. At time t this is then a good test for absence of effect

Integrated Brownian motion test

Area under the curve is given by

$$J(\beta_0, t) = \int_0^t U^*(\beta_0, u)du$$

Lemma

$J(\beta_0, t)$ converges in distribution to integrated brownian motion, i.e., a Gaussian process with mean zero and covariance process

$$\text{Cov}\ \{J(\beta_0, s), J(\beta_0, t)\} = s^2 \left(\frac{t}{2} - \frac{s}{6}\right) \quad (s < t). \tag{1}$$

p-value corresponding to the null hypothesis obtains from

$$\text{Pr}\ \left\{\sqrt{3}|J(\beta_0, t)/t\sqrt{t}| > z\right\} = 2(1 - \Phi(z)).$$

Combining AUC and log-rank

Lemma

*Under the model proportional hazards model (**??**) with parameter $\beta(t) = \beta_0$, the covariance function of $J(\beta_0, t)$ and $U^*(\beta_0, t)$, converges in probability to $t^2/2$.*

Distance from origin test most powerful under PH alternatives while AUC can be more powerful under non-PH alternatives. Combinations good in both situations.

$$D(\theta,\beta_0,t) = \theta U^*(\beta_0,t) + (1-\theta)J(\beta_0,t)\,,\ \ 0 \leq \theta \leq 1.$$

Therefore, under the hypothesis $H_0 : \beta = \beta_0$, the following corollary is obtained

Corollary

Under $H_0 : \beta = \beta_0$, $D(\theta,\beta_0,t)$ converges in distribution to a centered normal law with variance equals to $t\theta^2 + \frac{t^3}{3}(1-\theta)^2 + \frac{t^2}{2}\theta(1-\theta)$.

$$M(\beta_0, t) = \max\left\{|U^*(\beta_0, t)|, \sqrt{3}|J(\beta_0, t)|\right\}. \quad (2)$$

The gaussian limit distribution of the vector of which components are the distance from origin and the area under the curve statistics enables the following corollary

Corollary

$$\forall q > 0, \quad P\left(M(\beta_0, t) \geq q\right) \underset{n \to \infty}{\longrightarrow} 1 - 2\int_0^q \int_0^q \phi(u, v; 0, \Sigma(t)) du dv, \quad (3)$$

where $\phi(u, v; 0, \Sigma(t))$ is the density of the centered normal distribution in $\mathbb{R}^2$ with $\Sigma(t) = \begin{pmatrix} t & \sqrt{3}t^2/2 \\ \sqrt{3}t^2/2 & t^3 \end{pmatrix}$ as a covariance matrix, assessed in (u, v).

Example in marrow transplantation

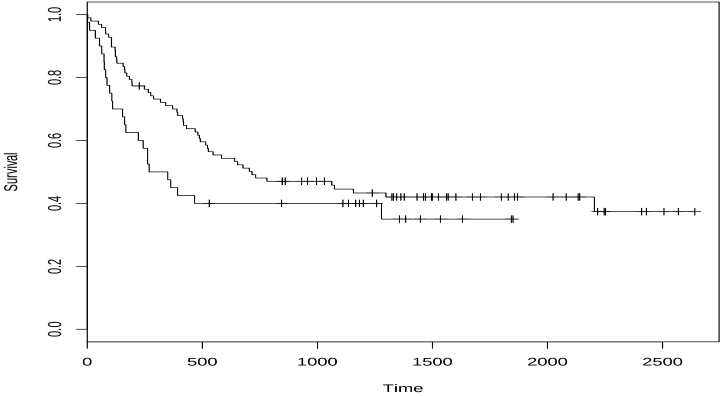


Figure: Kaplan-Meier estimates of survival for marrow transplant data

Process for marrow transplantation data

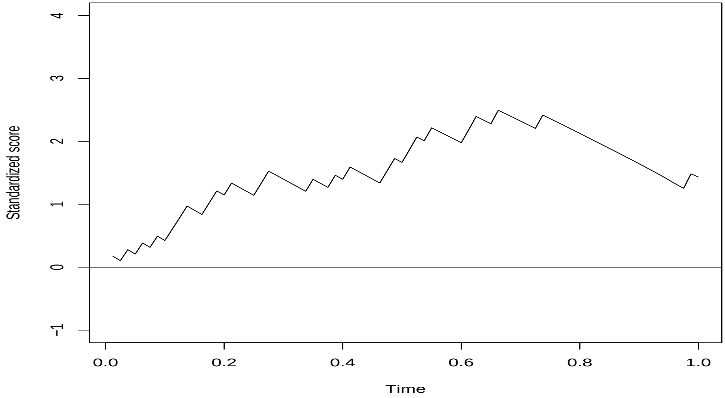


Figure: log-rank p-value = 0.15, AUC p-value = 0.008, Adaptive p-value = 0.01

(Figure **??**).