Sensitivity Analysis for Correlated Survival Models

Fotios Siannis
MRC Biostatistics Unit, Cambridge
fotios.siannis@mrc–bsu.cam.ac.uk

University of Waterloo
1 December 2006

Outline

- Introduction/Notation
- Parametric Approach
  - Model for sensitivity analysis
  - Example: Lupus Data
  - Proportional Hazards Modelling with Covariates
  - Example: Multiple Myeloma Data (1)
- Semi-Parametric Approach
  - Modified Partial Likelihood
  - Example: Multiple Myeloma Data (2)

Classical Survival Analysis

- Non–Parametric
  - Product–Limit (Kaplan–Meier)
  - Actuarial Method
- Semi–Parametric
  - Cox's Proportional Hazard Model
- Parametric
  - Any convenient known distribution (Weibull, etc)

Common Assumption: Ignorable censoring

Assumptions

(α) Model the dependence between Failure and Censoring Process (Informative Censoring).

(β) The process of censoring times can be seen as a separate failure process, with failures other than the ones of main interest.

- This mean that we introduce the problem within the Competing Risks framework, where the failure of interest and censoring are the only two competing risks, in the absence of random censoring.

(γ) Assess the sensitivity of quantities of interest in the presence of small levels of association between the two processes.

$\longrightarrow$ Define:

- T-**Process** := The failure process.
- C-**Process** := The censoring process, where we observe the censored times as being failure times.

$\longrightarrow$ This research is focused on:

- **Parametric** models,
  where both processes are assumed to follow a known distribution.
- **Semi–parametric** models,
  where we modify the use of Cox's partial likelihood in order to allow for the presence of informative censoring.

Notation

- $h_T(t;\theta)$, $H_T(t;\theta)$, $S_T(t;\theta)$ and $f_T(t;\theta)$ are the marginal hazard, cumulative hazard, survival and density functions of T-process respectively, governed by θ.

- Define $s_T(t;\theta) = \frac{\partial}{\partial\theta}\log f_T(t;\theta)$ and $i_\theta = Var_T\{s_T(T;\theta)\}$.

- Since θ generally depends on covariates $\mathbf{x}$, we write $h_T(t;\mathbf{v}|\mathbf{x})$, $H_T(t;\mathbf{v}|\mathbf{x})$, $S_T(t;\mathbf{v}|\mathbf{x})$ and $f_T(t;\mathbf{v}|\mathbf{x})$, where $\mathbf{v}$ is the vector of regression coefficients.

- Furthermore, $h(T,t|\mathbf{x})$ [or $h(T,t;\theta)$] is the sub-hazard of the T-process, defined

$$h(T,t|\mathbf{x}) = \lim_{\Delta t\to 0^+}\frac{Pr(t\le T<t+\Delta t\big|T\ge t, C\ge t,\mathbf{x})}{\Delta t},$$

  which is the hazard of the failure time process in the presence of censoring.

- Similar notation holds for the censoring process, with $(T,\theta,\mathbf{v}) \leftrightarrow (C,\gamma,\mathbf{u})$.

Parametric Approach

$\longrightarrow$ Propose the Model:

$$P(C = c|T = t) \quad = \quad f_C\big(c; \gamma + \delta i_\gamma^{-\frac{1}{2}} B(t;\theta)\big) \tag{1}$$

where:

- it is assumed that the conditional distribution of C given T has the same form as its marginal distribution $f_C(c;\gamma_i)$, with the only difference being in the location parameter of the distribution which is now allowed to depend on T.

- parameter $\boldsymbol{\delta}$ can be thought of as measuring the $size$ of the dependence between T and C.

- function $B(t;\theta)$, named the `bias function`, can be thought of as measuring the $pattern$ of this dependence.

Notes:

1. No inferences can be drawn for parameter δ, hence we assume that it is known.

2. The aim is to perform sensitivity analysis on quantities of interest (like θ or the estimate of the survival curve) for small levels of association between the two processes (small values of δ). For $\underline{\delta = 0}$ we have independence.

3. Therefore, the joint p.d.f. takes the form

$$f_{T,C}(\mathbf{u}) \simeq f_T(u;\theta) f_C(u;\gamma) \left[1 + \delta \imath_\gamma^{-\frac{1}{2}} s_C(u;\gamma) B(u;\theta)\right].$$

4. Under this parametrization, θ is the main parameter of interest, treating γ as a nuisance parameter.

Log–likelihood when $\delta \neq 0$

In first order approximation we have

$$\log l_\delta \simeq \log l_0 + \delta i_\gamma^{-\frac{1}{2}} \sum_{i=1}^{n} \left[(1 - I_i) s_c(t_i; \gamma) \mu(t_i; \theta) - I_i \frac{\partial H_C(t_i; \gamma)}{\partial \gamma} B(t_i; \theta) \right]$$

where:

- $\log l_0 = \sum_{i=1}^{n} \left[I_i \log h_T(t_i; \theta) + (1 - I_i) \log h_C(t_i; \gamma) - H_T(t_i; \theta) - H_C(t_i; \gamma) \right]$ is the log–likelihood under independence ($\delta = 0$).

- $\mu(t_i; \theta) = \frac{\int_{t_i}^{\infty} B(u;\theta) f_T(u;\theta) du}{S_T(t_i;\theta)}$.

Sensitivity Analysis

By differentiating $\log l_\delta$ with respect to θ we get

$$\hat{\theta}_\delta - \hat{\theta}_0 \simeq \delta i_\gamma^{-\frac{1}{2}} \imath(\theta)^{-1} \sum_{i=1}^{n} \left[(1 - I_i) \frac{\partial \mu(t_i;\theta)}{\partial \theta} s_C(t_i;\gamma) - I_i \frac{\partial B(t_i;\theta)}{\partial \theta} \frac{\partial H_C(t_i;\gamma)}{\partial \gamma} \right],$$

where

$$\imath(\theta) = -\frac{\partial^2 \log l_0}{\partial \theta^2}$$

is the observed information.

- This is what we will name `correlation bias`.
- The term multiplied by δ is called `Sensitivity Index`

The Bias Function $B(t;\theta)$

(i) In order for the joint p.d.f. to provide the correct marginal distributions for T and C we require

$$E_T\big[B(t;\theta)\big] = 0.$$

(ii) $B(t;\theta)$ must have a finite variance, and without any loss of generality

$$Var_T\big[B(t;\theta)\big] = E_T\big[B^2(t;\theta)\big] = 1.$$

(iii) A choice that satisfies the above restrictions is $\boxed{B(t;\theta) = \imath_\theta^{-\frac{1}{2}} s_T(t;\theta)}$, which also provides with a nice symmetry in the joint density

$$f_{T,C}(\mathbf{u}) \simeq f_T(u;\theta) f_C(u;\gamma) \left[1 + \delta \imath_\gamma^{-\frac{1}{2}} \imath_\theta^{-\frac{1}{2}} \; s_T(u;\theta) s_C(u;\gamma)\right].$$

1. Suppose that for given subject, T and C are independent with densities $g_T(t;\theta+\epsilon_T\imath_\theta^{-\frac{1}{2}})$ and $g_C(c;\gamma+\epsilon_C\imath_\gamma^{-\frac{1}{2}})$, where ϵ_T and ϵ_C are, assumed small, random effects with means zero, variances σ_T^2 and σ_C^2, and covariance σ_{TC}. Then

$$f_{T,C}(t,c) \simeq f_T(t,\theta)f_C(c,\gamma)\left[1+\sigma_{TC}(\imath_\theta\imath_\gamma)^{-\frac{1}{2}}s_T(t,\theta)s_C(c,\gamma)\right],$$

   where with the appropriate definition of δ, is the joint density between T and C.

2. Since the form of dependence is completely unknown, our assumptions about it should be as weak as possible so far as the information about θ is concerned. Therefore, the efficient information bound [Bickel et.al.(1992)] is

$$\mathrm{Inf}_B\left\{\imath(\theta)-\delta^2\left(E_T\left[\frac{\partial B(T;\theta)}{\partial\theta}\right]\right)^2\right\},$$

   where the minimization over all functions lead to $B(t;\theta)=\imath_\theta^{-\frac{1}{2}}s_T(t;\theta)$.

Sensitivity indices and confidence intervals

The general form of the bias approximations, up to linear terms in δ, has the form

$$\hat{\theta}_\delta - \hat{\theta}_0 = \delta U + O(\delta^2),$$

where U is the `Sensitivity Index`.

For a confidence interval, the approximations give

$$\{Var(\hat{\theta}_\delta)\}^{\frac{1}{2}} = \{\imath(\theta)\}^{-\frac{1}{2}} + O(\delta^2).$$

Hence, retaining only linear terms in δ, the asymptotic confidence interval for θ is approximately

$$\hat{\theta}_0 - \delta U \pm z_\alpha \{\imath(\theta)\}^{-\frac{1}{2}},$$

where z_α is the appropriate standard normal percentage point.

For a local sensitivity analysis in practice, we fix a maximum, but small, value for δ to give $(-\delta, \delta)$ as a plausible range of dependence parameters that we wish to consider. This leads to

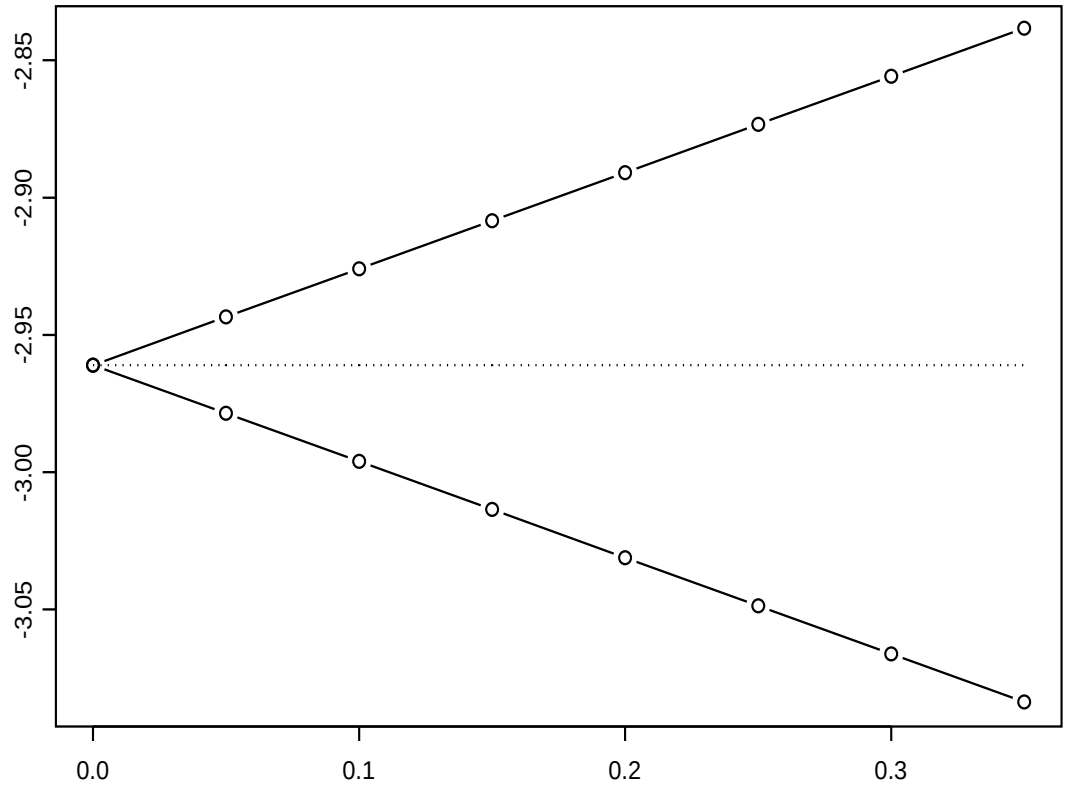
$$\hat{\theta}_\delta = \hat{\theta}_0 \pm \delta U$$

as the plausible range of values of the estimate of θ, and

$$\hat{\theta}_0 \pm \left\{ \delta |U| + z_\alpha \{\imath(\theta)\}^{-\frac{1}{2}} \right\}$$

as the corresponding conservative confidence interval for θ itself.

Sensitivity Analysis on θ



Interpretation of δ

Assume $A(T;\theta)$ and $D(C;\gamma)$ are some functions of T and C respectively, and

$$\rho = \text{Corr}\left(A(T;\theta), D(C;\gamma)\right).$$

Then, generally can be shown that

$$|\rho| \leq |\delta|,$$

which immediately implies that $|Corr(T,C)| \leq |\delta|$.

This provides with a useful interpretation and a plausible range of values for δ.

* Parametric Case

In the simplest of scenarios where both T and C follow exponential distribution, it can be easily shown that

$$\rho = \delta.$$

Upper Bound

Although we do not have the answer to the question "what is the best choice for $B(t;\theta)$", we can still calculate some naive bounds. Straight forward mathematical calculations give

$$E\left[\hat{\theta}_\delta - \hat{\theta}_0\right] \quad \leq \quad |\delta| n \imath_\gamma^{-\frac{1}{2}} \imath(\theta)^{-1} E_T\left[N^2\right]^{\frac{1}{2}},$$

where:

$$N = \left[\frac{\partial \log h_T(t;\theta)}{\partial\theta} - \frac{\partial H_T(t;\theta)}{\partial\theta}\right] \frac{\partial H_C(t;\gamma)}{\partial\gamma} S_C(t;\gamma) + \int_0^t \frac{\partial f_C(u;\gamma)}{\partial\gamma} \frac{\partial H_T(u;\theta)}{\partial\theta} du.$$

The equality is attained only when $B(t;\theta)$ is a linear function of N (not very attractive).

The Ignorability of Censoring at the End of the Study

Assume that censoring at the end of the study is regarded as ignorable. Hence, we observe $Y = \min\{T, C_I, C_E\}$ and the likelihood becomes

$$l' = \prod_{i=1}^{n} Pr\big(T = t_i, T < C_I\big)^{I_i} Pr\big(C_I = t_i, C_I < T\big)^{(1-I_i)Z_i} Pr\big(C_E = t_i\big)^{(1-I_i)(1-Z_i)},$$

where

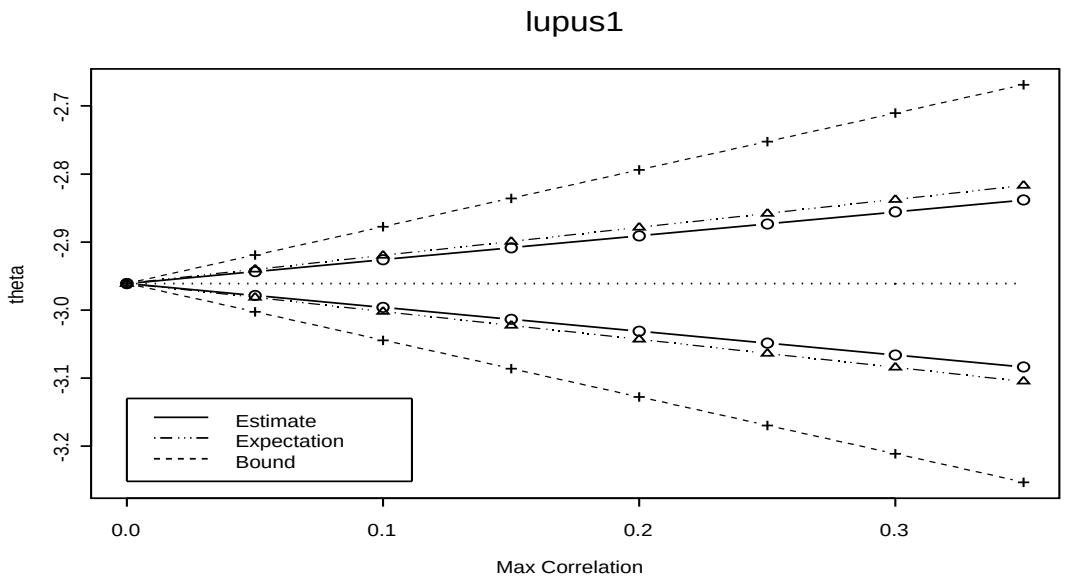
- $I_i = \begin{cases} 1, & \ldots \text{ when failure time} \\ 0, & \ldots \text{ when censored time} \end{cases}$
- $Z_i = \begin{cases} 1, & \ldots \text{ when } Y = C_I \text{ (censored before the end)} \\ 0, & \ldots \text{ when } Y = C_E \text{ (censored at the end)} \end{cases}$
- $Pr(C_E = t) = S_T(t;\theta) S_C(t;\gamma) \left[1 - \delta \frac{\partial H_C(t;\gamma)}{\partial \gamma} \mu(t;\theta)\right]$

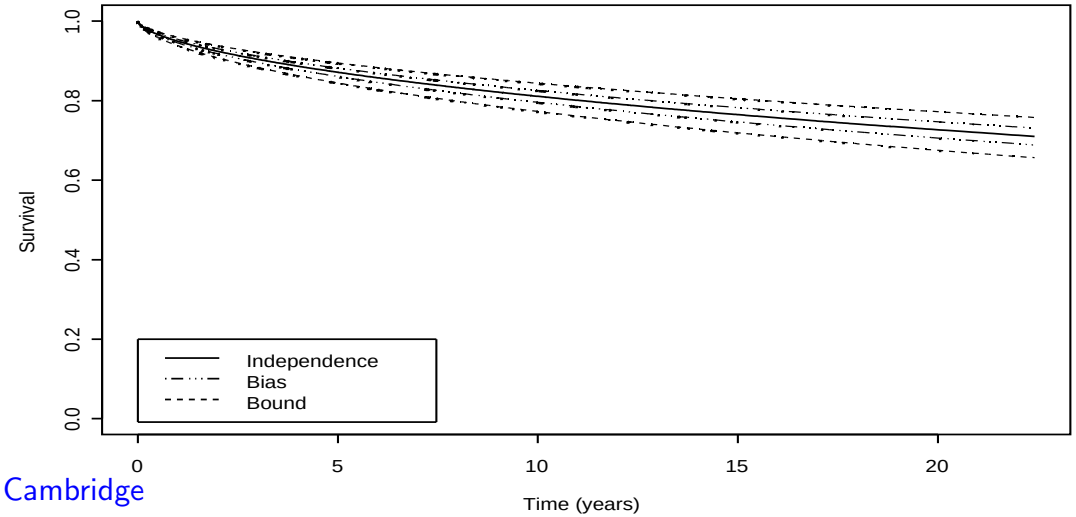
Example: Lupus Data

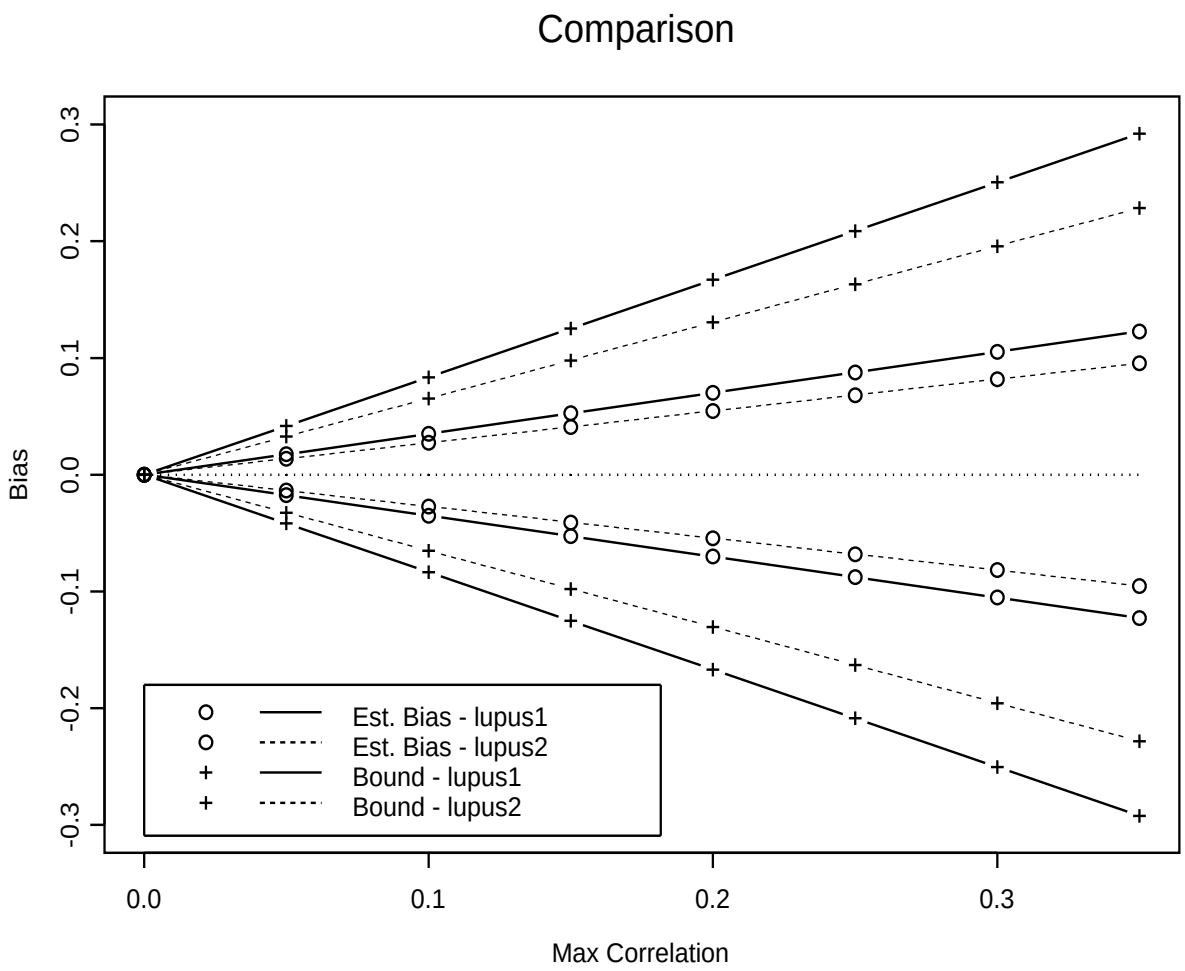
Since 1970, patients with systematic lupus erythematosus (SLE) at the University of Toronto Lupus Clinic have been followed. By September 1991

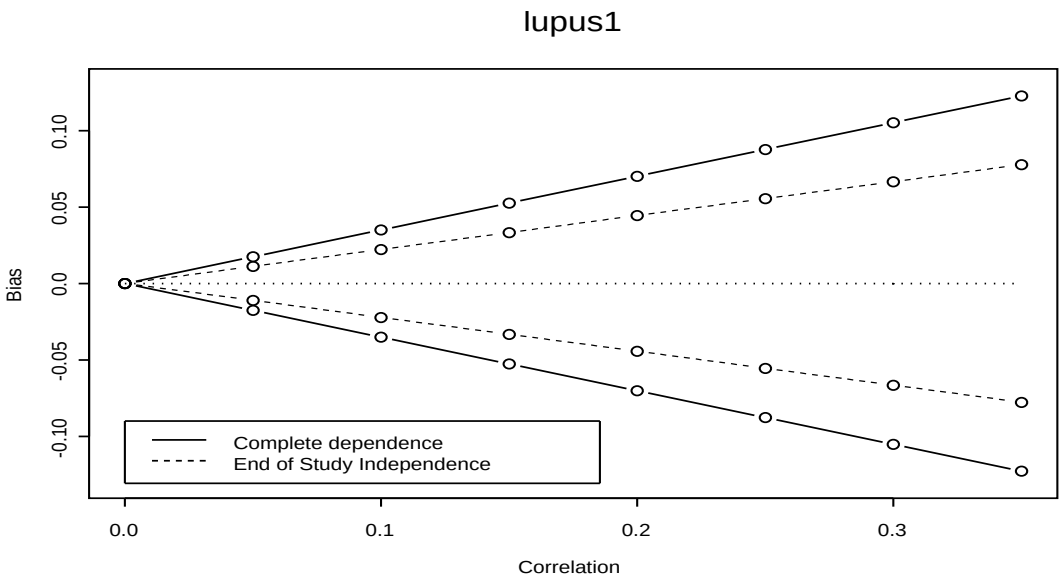
- A total of 576 patients were in the clinic registry.
- 91 died during these years.
- At this time a study of patients who were lost-to-follow-up was undertaken. By December 1992
  - 139 had been traced (out of 248 randomly chosen patients).
  - 22 were reported to have died, increasing the total number of deaths to 113.

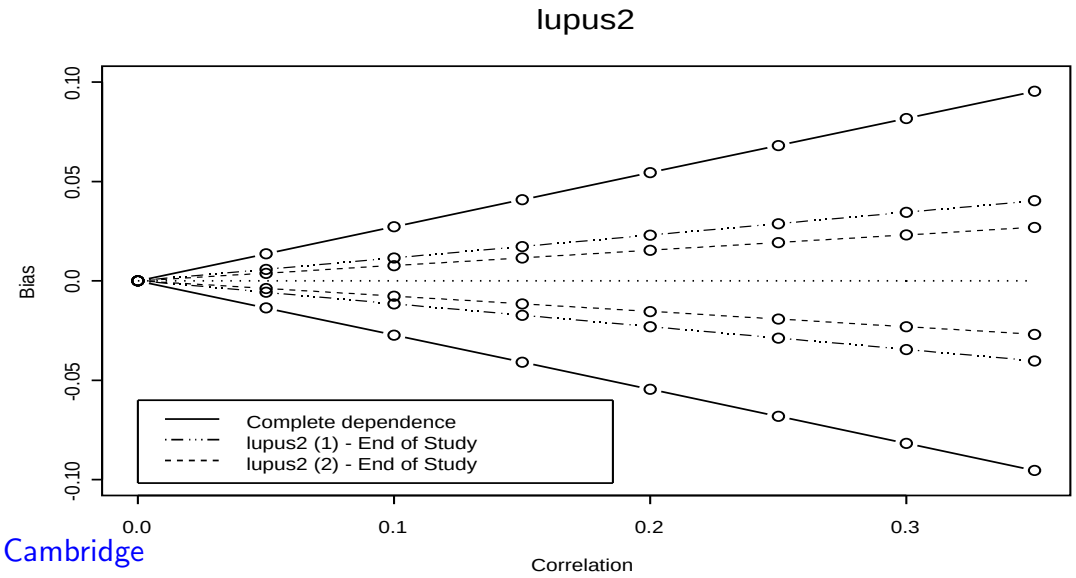
Define `lupus1` to be the original lupus data and `lupus2` to be the data set with the additional observations. A Weibull model was fitted to both the failure and the censoring processes for both `lupus1` and `lupus2`.


lupus1
theta
-2.7
-2.8
-2.9
-3.0
-3.1
-3.2
Estimate
Expectation
Bound
0.0
0.1
0.2
0.3
Max Correlation


Survival
1.0
0.8
0.6
0.4
0.2
0.0
Independence
Bias
Bound
0
5
10
15
20
Time (years)


Comparison
Bias
Max Correlation
0.0
0.1
0.2
0.3
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3
Est. Bias - lupus1
Est. Bias - lupus2
Bound - lupus1
Bound - lupus2


lupus1
Bias
0.10
0.05
0.0
-0.05
-0.10
Complete dependence
End of Study Independence
0.0
0.1
0.2
0.3
Correlation


lupus2
Bias
0.10
0.05
0.0
-0.05
-0.10
Complete dependence
lupus2 (1) - End of Study
lupus2 (2) - End of Study
0.0
0.1
0.2
0.3
Correlation

Proportional Hazards Modelling with Covariates

Assume: $h_T(t;\mathbf{v}|\mathbf{x}) = e^{\mathbf{v}'\mathbf{x}} h_T^*(t)$ and $h_C(c;\mathbf{u}|\mathbf{x}) = e^{\mathbf{u}'\mathbf{x}} h_C^*(c)$

where $\mathbf{v}$ and $\mathbf{u}$ are vectors of parameters of T and C with covariates $\mathbf{x}$.

- $h_T^*(t)$ and $h_C^*(c)$ have known parametric forms (not necessarily the same).
- Vector $\mathbf{x}$ need not be the same for the two processes, hence vectors $\mathbf{v}$ and $\mathbf{u}$ may have different lengths.
- We focus on vector $\mathbf{v}$, treating $\mathbf{u}$ as a vector of nuisance parameters.
- $\hat{\mathbf{v}}_\delta - \hat{\mathbf{v}}_0 \simeq \delta \imath(\mathbf{v})^{-1} \sum_{i=1}^{n} \{\mathbf{x}_i [H_T(t_i;\mathbf{v}|\mathbf{x}) H_C(t_i;\mathbf{u}|\mathbf{x}) - (1-I_i) H_T(t_i;\mathbf{v}|\mathbf{x})]\}$
- Depending on the parametrization, sensitivity analysis can be performed on various quantities of interest (like the median).

Example: Multiple Myeloma Data (1)

Study with 65 multiple myeloma patients (Krall $et.al.$(1975)). We have 4 covariates and we model the hazard functions with a constant baseline hazard function for both processes. Then

$$w_{\mathbf{x}} = \mathbf{v}'\mathbf{x} \qquad \text{and} \qquad z_{\mathbf{x}} = \mathbf{u}'\mathbf{x}$$

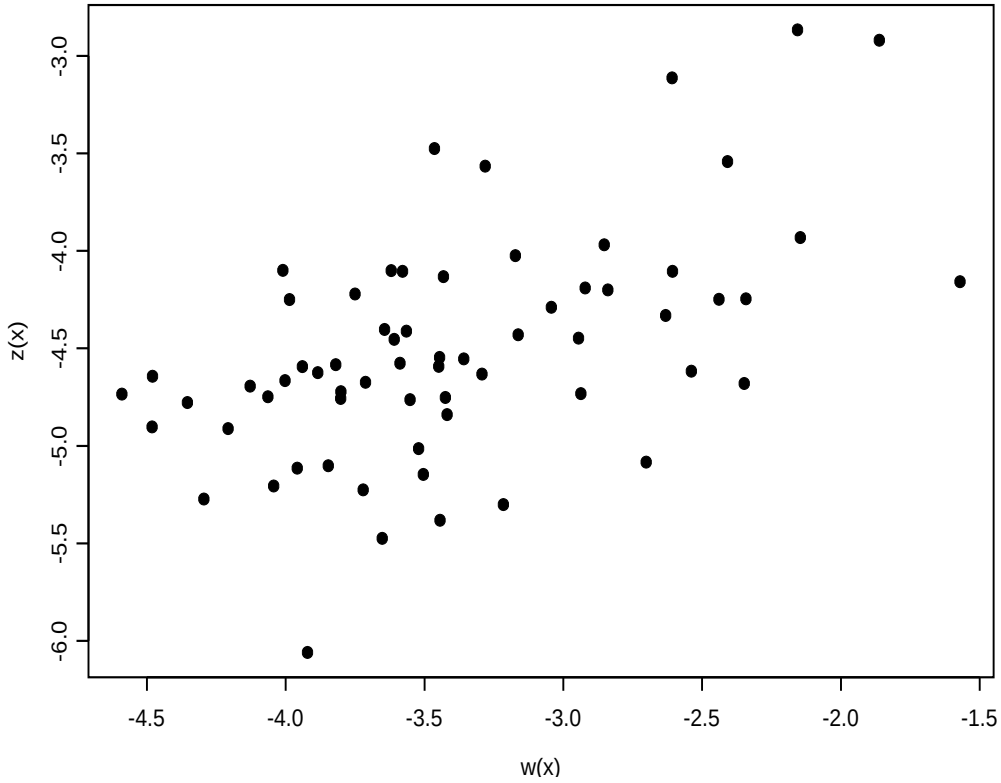
are the log-hazard rates for T and C respectively (including intercept).

Therefore, the sensitivity analysis for the log-hazard rates leads to

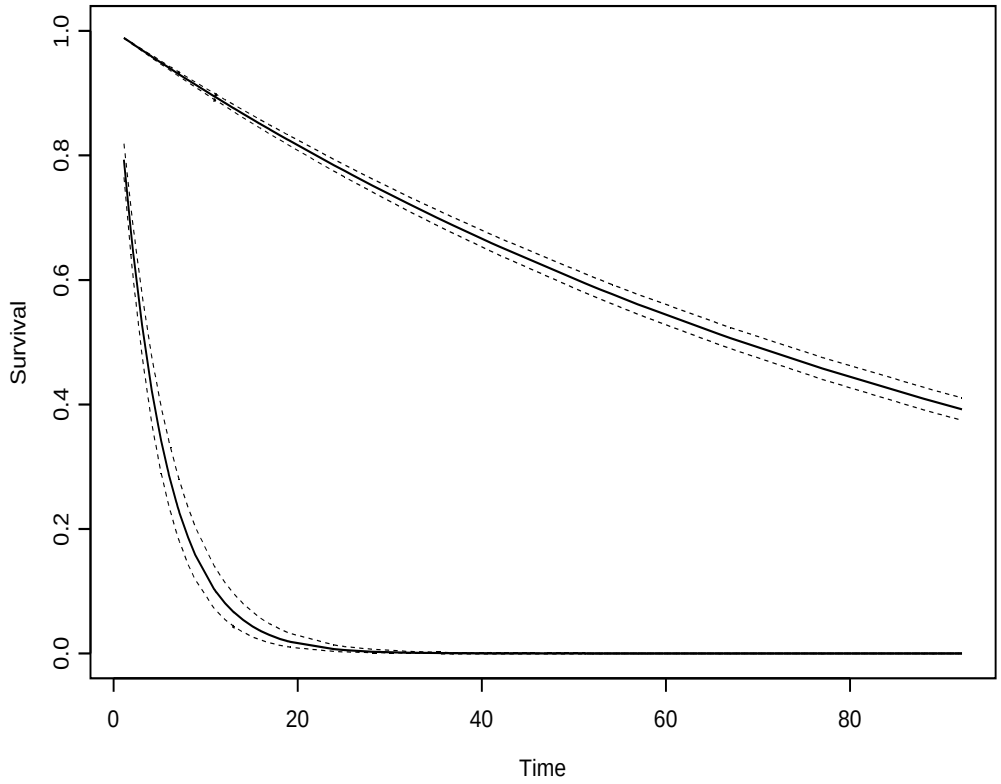
$$\hat{w}^{\delta}_{\mathbf{x}} - \hat{w}^{0}_{\mathbf{x}} \quad \simeq \quad \delta \frac{\sum_{i=1}^{n} \left\{ e^{z_{\mathbf{x}}} t_i^2 - (1 - I_i) t_i \right\}}{\sum_{i=1}^{n} t_i},$$

which is a very simple formula to calculate.

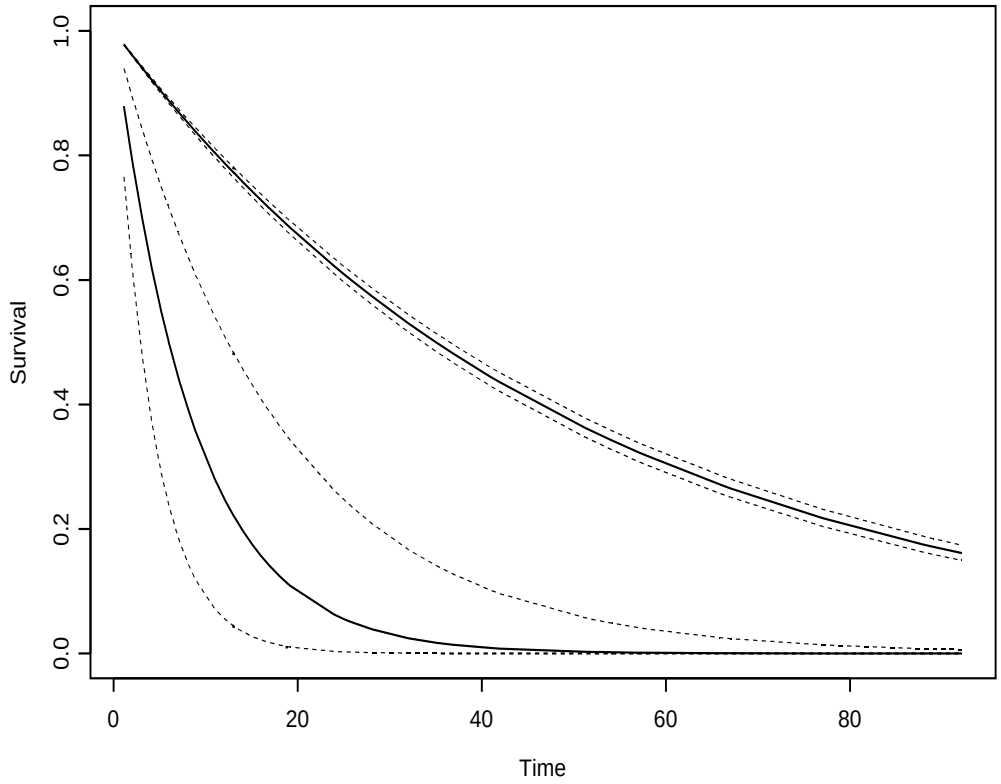
Plot of $w_{\mathbf{x}}$ against $z_{\mathbf{x}}$



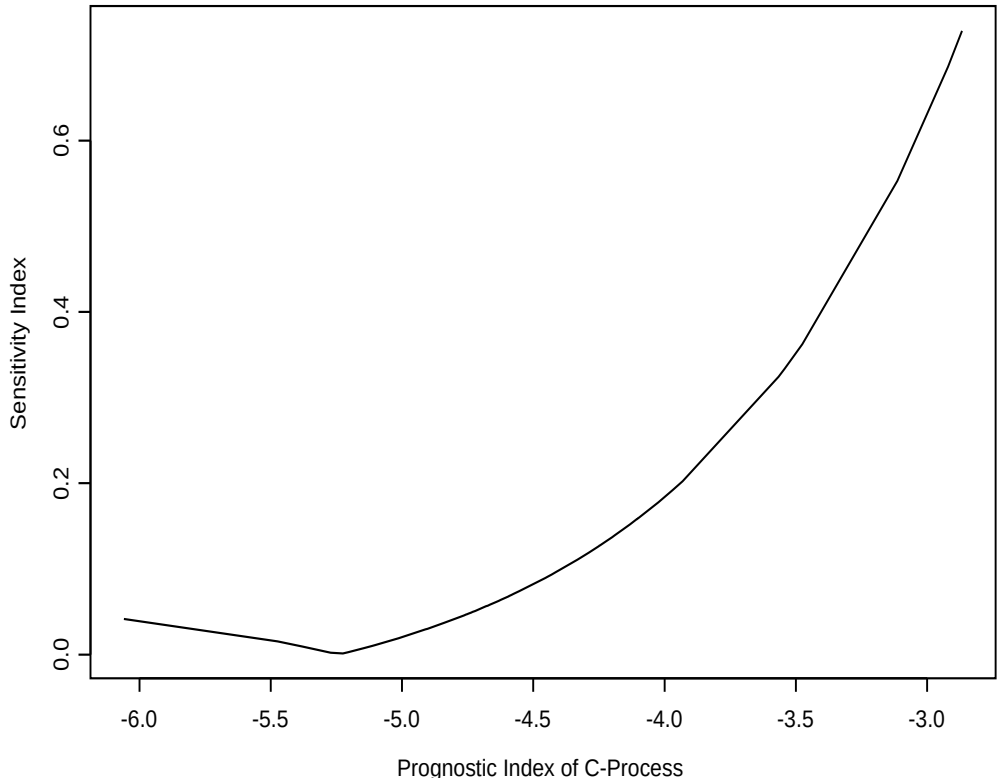
Sensitivity analysis on survival curves for min and max $w_{\mathbf{x}}$ ($\delta = 0.3$)



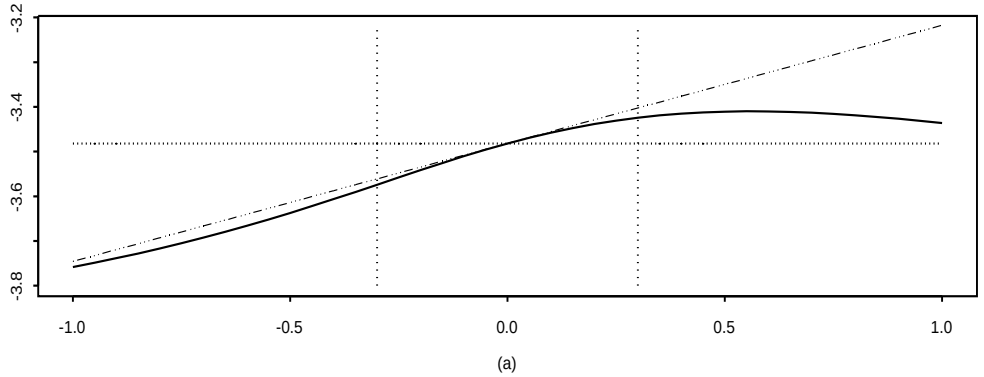
Sensitivity analysis on survival curves for min and max $z_{\mathbf{x}}$ ($\delta = 0.3$)

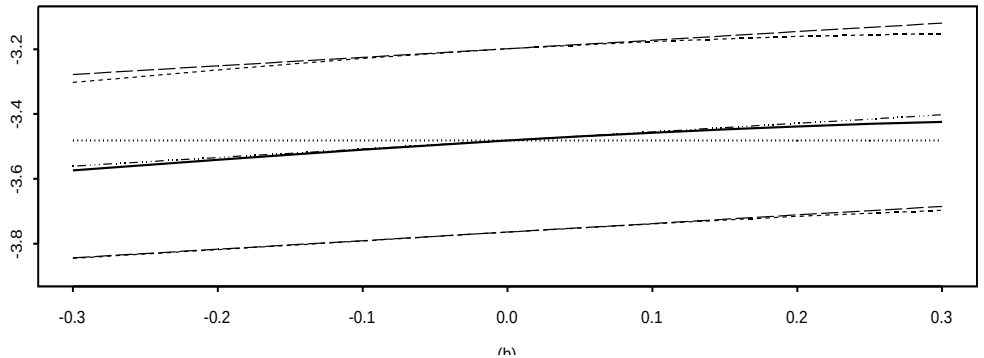


Plot of the absolute value of Sensitivity Index against $z_{\mathbf{x}}$



Explore Approximation





Semi–parametric

$\longrightarrow$ Cox's Partial Likelihood (l_P)

- Cox (1972) proposed that the hazard function is proportional to some other function (baseline hazard function), which depends only on time t

$$h(t;\theta_i) = e^{\theta_i} h_0(t).$$

- Parameter θ_i is a linear combination of a set of explanatory variables.

- Under the assumption of independence between the failure and the censoring times, Cox introduced the Partial Likelihood

$$l_P = \prod_i \frac{h(t;\theta_i)}{\sum_{\ell \in \mathcal{R}_{t_i}} h(t;\theta_\ell)} = \prod_i \frac{e^{\theta_i}}{\sum_{\ell \in \mathcal{R}_{t_i}} e^{\theta_\ell}}$$

$\longrightarrow$ Modified Partial Likelihood (l_M)

An extension to Cox's partial likelihood is the competing risks partial likelihood

$$l_P^* = \prod_i \frac{h(j, t_i|\mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}_{t_i}} h(j, t_i|\mathbf{x}_\ell)},$$

where instead of the marginal hazards the observable sub-hazards for cause j are used. Therefore, considering T and C as acting competing risks we get

$$l_M = \prod_{i=1}^{r} \frac{h(T, t_i|\mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}_{t_i}} h(T, t_i|\mathbf{x}_\ell)} \prod_{j=1}^{k} \frac{h(C, t_j|\mathbf{x}_j)}{\sum_{q \in \mathcal{R}_{t_j}} h(C, t_j|\mathbf{x}_q)},$$

When independence ($\delta = 0$) we have $l_P = l_M$.

- r is the total number of failure and k is the total number of censored observations.

- Sensitivity analysis for informative censoring is performed through model (1)

$$P(C = c|T = t) = f_C\big(c; \gamma + \delta \imath_\gamma^{-\frac{1}{2}} B(t; \theta)\big).$$

- Sub-hazards take the form
  - $h(T, t|\mathbf{x}) \simeq h_T(t; \mathbf{v}|\mathbf{x}) \left[1 + \delta \imath_\gamma^{-\frac{1}{2}} \mu_T(t; \mathbf{v}|\mathbf{x}) \psi(t|\mathbf{x})\right]$
  - $h(C, t|\mathbf{x}) \simeq h_C(t; \mathbf{u}|\mathbf{x}) \left[1 + \delta \imath_\gamma^{-\frac{1}{2}} \mu(t; \mathbf{v}|\mathbf{x})\right],$

  where: $\mu(t; \mathbf{v}|\mathbf{x}) = \frac{\int_t^\infty B(w;\mathbf{v}|\mathbf{x}) f_T(w;\mathbf{v}|\mathbf{x}) dw}{S_T(t;\mathbf{v}|\mathbf{x})}$, $\mu_T(t; \mathbf{v}|\mathbf{x}) = \frac{\partial \mu(t;\mathbf{v}|\mathbf{x})}{\partial T}$

  and $\psi(t|\mathbf{x}) = \frac{H_C(t;\mathbf{u}|\mathbf{x})}{h_T(t;\mathbf{v}|\mathbf{x})}$.

- Under this approach we need to estimate the baseline hazard (Cox or Kalbfleisch & Prentice) to be included in the sensitivity index.

The Ignorability of Censoring at the End of the Study

- Assume that we have two competing risks, the failure T and the informative censoring C_I, where both are subject to ignorable censoring, say C_R.

- Of the k censored observations, w are informative and $k - w$ are not.

- Therefore

$$l_M^I = \prod_{i=1}^{r} \frac{h(T, t_i|\mathbf{x}_i)}{\sum_{\ell \in \mathcal{R}_{t_i}} h(T, t_i|\mathbf{x}_\ell)} \prod_{j=1}^{w} \frac{h(C_I, t_j|\mathbf{x}_j)}{\sum_{q \in \mathcal{R}_{t_j}} h(C_I, t_j|\mathbf{x}_q)},$$

  where the second product is over the w non-ignorable censored observations and not all the censored observations. Baseline for the C_I process is different to the one from the C-process.

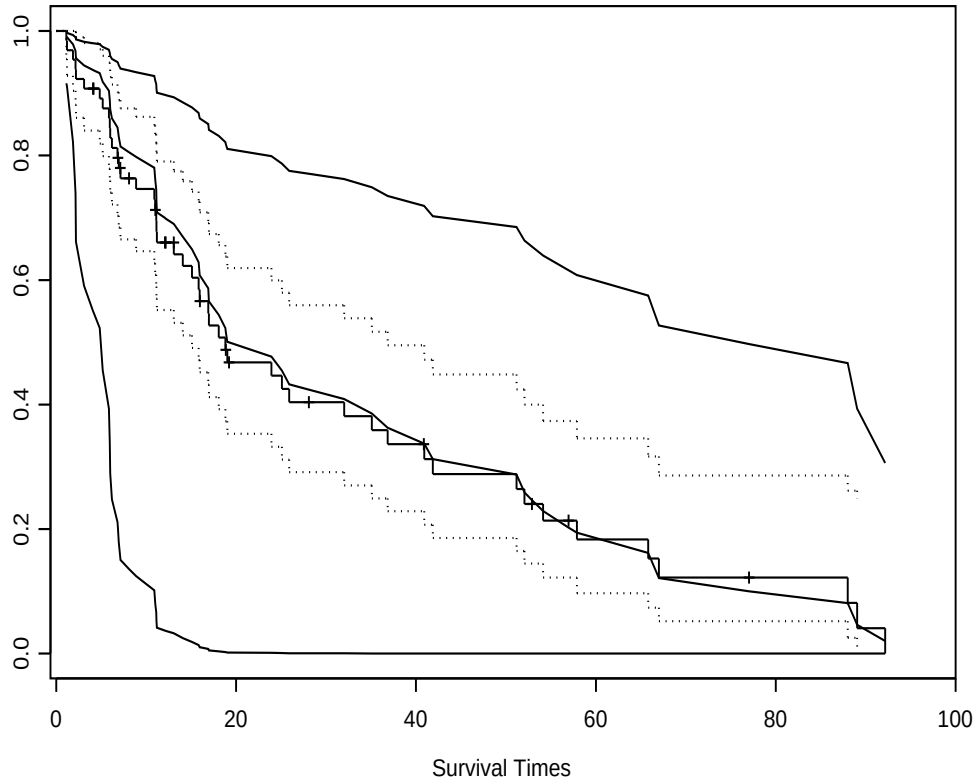
Example: Multiple Myeloma Data (2)

Back to myeloma data. Estimates under Cox's proportional hazard model ($\delta = 0$) are given in the following table

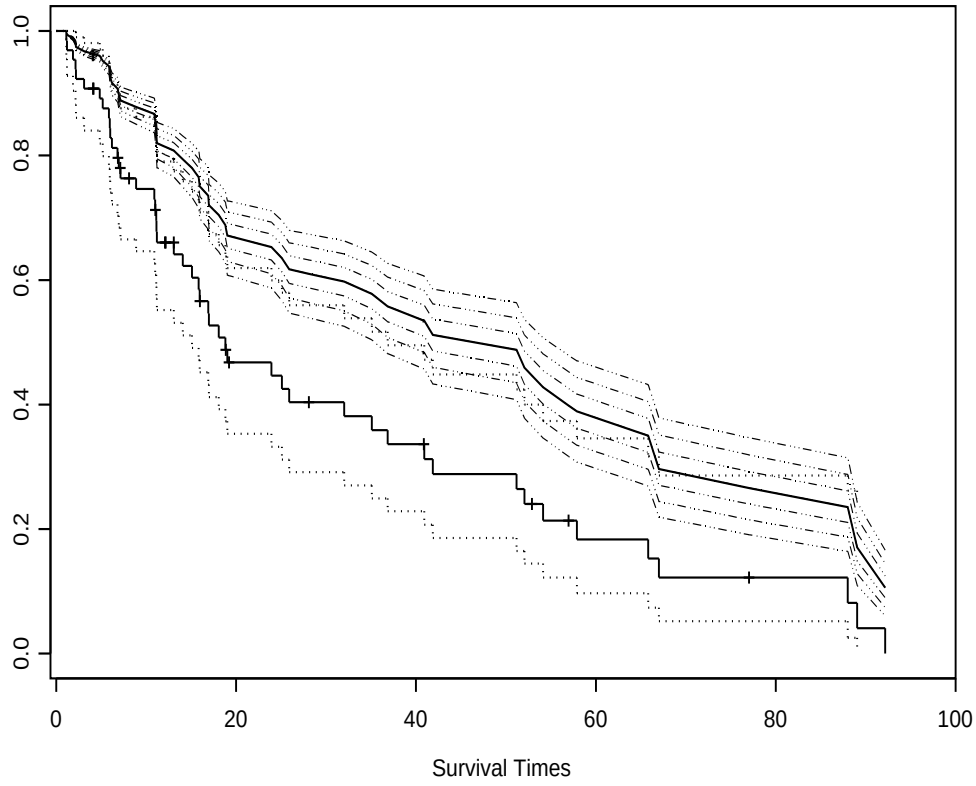
	coef	exp(coef)	se(coef)	z	p
x1	1.832	6.245	0.6476	2.83	0.0047
x2	-0.120	0.887	0.0594	-2.03	0.0430
x9	0.462	1.587	0.4620	1.00	0.3200
x16	0.1397	1.149	0.1000	1.39	0.1600

We can calculate the prognostic index $w_{\mathbf{x}}$ for each individual, and draw the survival curves.

Figure with min and max $w_{\mathbf{x}}$ along with the KM estimate



Changes in the Survival Curve for different values of δ ($w_{\mathbf{x}} = 2.74$)



- If now we allow δ to depart from zero, then the vector of the parameters does not remain the same any more. Using the MPL we perform a sensitivity analysis for values of $\delta \in [-0.3, 0.3]$.

δ	$\mathbf{v}_1$	$\mathbf{v}_2$	$\mathbf{v}_3$	$\mathbf{v}_4$
-0.3	1.7953789	-0.1195844	0.4478341	0.1214729
-0.2	1.8075859	-0.1197229	0.4525561	0.1273153
-0.1	1.8197930	-0.1198615	0.4572780	0.1331576
0	1.832	-0.120	0.462	0.139
0.1	1.8442070	-0.1201385	0.4667220	0.1448424
0.2	1.8564141	-0.1202771	0.4714439	0.1506847
0.3	1.8686211	-0.1204156	0.4761659	0.1565271

Focusing on the parameter of covariate x2 we get

δ	$\mathbf{v}_2$	z
-0.3	-0.1195844	-2.013
-0.2	-0.1197229	-2.016
-0.1	-0.1198615	-2.018
0	-0.120	-2.030
0.1	-0.1201385	-2.022
0.2	-0.122771	-2.025
0.3	-0.1204156	-2.027

where it is clear that v2 is significant for $\delta \in [-0.3, 0.3]$, concluding that correlation does not weakens the role of v2.

- If we remove x9 and x16 (not significant) then we have

	coef	exp(coef)	se(coef)	z	p
x1	1.802	6.062	0.6279	2.87	0.0041
x2	-0.115	0.891	0.0576	-2.00	0.0460

where $\mathbf{v}_2$ is still on the borderline of being significant. If we test again

δ	$\mathbf{v}_2$	z
-0.3	-0.1271504	-2.207
-0.2	-0.1231003	-2.141
-0.1	-0.1190501	-2.070
0	-0.115	-2.000
0.1	-0.1109499	-1.930
0.2	-0.1068997	-1.859
0.3	-0.1028496	-1.786

where $\mathbf{v}_2$ is not significantly different from zero for almost any positive δ.

Acknowledgements:

- Professor John Copas
- Guobing Lu
- Professor Vern Farewell

References:

-Siannis F (2004) 'Applications of a parametric model for informative censoring', *Biometrics* **60**(3),704–714.

-Siannis F, Copas J and Lu G (2005) 'Sensitivity analysis for informative censoring in parametric survival models', *Biostatistics* **6**(1),77–91.

-Siannis F (2006) 'Sensitivity Analysis for Informative Censoring: A Semi–parametric Approach', *Submitted to Biometrics*.