Session 3. Time-dependent covariates

Content
< simulation study
< example CML trial
< irregularly observed time-dependent covariate
< predicting by landmarking
< predicting by retrospective modeling
Simulation study

< Time-dependent covariate $X(t)$

Wiener process with

$$E[X(t)]' = 0, \ cov(X(t), X(s))' = s^2 e^{|t-s|}$$

with $s' = 1$ and $? = 1$

simple stationary autoregressive model

no measurement error

< hazard function $h(t)' = h_0 e^{\beta X(t)}$ with $h_0' = 0.1$ and $\beta' = 2$

< discretized in steps of $? = \frac{1}{12}$

< simulated on 2000 patients with follow-up till $t=6$
The first 5 simulated patients
Resulting survival function
The corresponding hazard looks like

It is not constant and much larger than 0.1 !!!!
Explanation:

\[ h(t)' \cdot E[h_0 \exp(\beta X(t)|T^t)] \]

at \( t=0 \)

\[ h(0)' \cdot h_0 \exp(\frac{1}{2} \beta^2)' \cdot 0.1 \cdot \exp(2)' \cdot 0.73 \]

(not quite so in the data, due to discretization)
Distribution of $X(t)$ changes over time

< selection on survival $\tilde{N}$

< regression to the mean $\dot{E}$
Histogram at $t=6$

Mean below zero and slightly skewed to the left.
No explicit solution for equilibrium distribution
Since death and censoring are typically Missing at Random the model can completely be recovered from the data. Model easy to fit because of monotone drop-out
<X-model by regression on the past
<hazard model by logistic regression in risk sets

Big question: how to predict.
< Need only the last observation.
< Consider prediction based on first X-value $X_0$
Formally \( h(t|X_0) \) \( E[h_0 \exp(\beta X(t)|X(0), X_0, T$t] \\
No explicit solution as in the Manton-Woodbury model \\
Approximation based on ignoring \( T$t condition \\
\( X(t)|X(0) \sim N(e^{\sigma t}, 1 \sigma e^\sigma 2t) \)
\( E[\exp(\beta X(t) X_0]) \exp(\beta e^\sigma X_0 \sigma /2 \beta^2 (1 \sigma e^\sigma 2t)) \)
\( \ln(h(t|X_0)) \ln(h_0) \sigma X_0 \beta \sigma /2 (1 \sigma e^\sigma 2t) \beta^2 \)
Approximation is not more than approximation.
Modelling by exponentials .....
Conclusion: long-range hazard hard to model because of unpredictable behaviour of hazard
Retrospective time
Boxplots

remaining time

Heidelberg, November 2000, session 3, page 14
Histogram of last measurement before death
Some correlation with time of last observation

Model: \[ 1.06 \times 0.56 \times \exp(4.55 \times t_{last}) \times s^{0.73} \]
Retrospective model is purely Markovian, but not stationary.

Autoregressive model per month

\[ X_t' \& 0.055% 0.879( X_{t+1} ' 0.385( \& 0.84)) \text{ (limiting } s' 0.84) \].

Graph shows mean and st.dev as function of time remaining
Purely Markovian model makes that prediction depends only on last observation.

\[
f(t | X(t_{\text{last}}))' = \frac{f(t) f(X(t_{\text{last}}) | T' t)}{\int_{t_{\text{last}}} f(t) f(X(t_{\text{last}}) | T' t) dt}
\]

with

\[
m_{t_{\text{last}}} 
\]

\[
f(X(t_{\text{last}}) | T' t) \sim N(\mu_{t_{\text{last}}} , \sigma^2_{t_{\text{last}}})
\]

Not identical with forward prediction, but close (study going on)
CML-trial

- 190 patients
- variables at baseline
  - age
  - sokal (scoring system for overall prognosis)
- WBC ($10^9/l$) measured during follow-up
  transformed to $10 \log(WBC) - 0.95$
- endpoint survival
Overall Kaplan-Meier

![Graph showing time to death (years) vs. S(t)](image-url)
Table 1: Summary of the $WBC$ measurements in the data.

<table>
<thead>
<tr>
<th>Number of $WBC$-measurements</th>
<th>Number of patients</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 10</td>
<td>23</td>
<td>12.1</td>
</tr>
<tr>
<td>11 – 15</td>
<td>24</td>
<td>12.6</td>
</tr>
<tr>
<td>16 – 20</td>
<td>17</td>
<td>8.9</td>
</tr>
<tr>
<td>21 – 30</td>
<td>48</td>
<td>25.3</td>
</tr>
<tr>
<td>31 – 40</td>
<td>28</td>
<td>14.7</td>
</tr>
<tr>
<td>41 – 50</td>
<td>27</td>
<td>14.2</td>
</tr>
<tr>
<td>51 – 60</td>
<td>12</td>
<td>6.3</td>
</tr>
<tr>
<td>≥ 60</td>
<td>11</td>
<td>5.8</td>
</tr>
<tr>
<td>Total</td>
<td>190</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 2: Summary of the data.

<table>
<thead>
<tr>
<th>Time (years)</th>
<th>Total patients at start of the year</th>
<th>Mean number of patients in that year</th>
<th>Number of WBC observations per year</th>
<th>Events</th>
<th>Observation rate</th>
<th>Event rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1</td>
<td>190</td>
<td>188.5</td>
<td>2384</td>
<td>7</td>
<td>12.6</td>
<td>0.04</td>
</tr>
<tr>
<td>1-2</td>
<td>183</td>
<td>174.2</td>
<td>1179</td>
<td>18</td>
<td>6.8</td>
<td>0.10</td>
</tr>
<tr>
<td>2-3</td>
<td>165</td>
<td>149.5</td>
<td>856</td>
<td>30</td>
<td>5.7</td>
<td>0.20</td>
</tr>
<tr>
<td>3-4</td>
<td>134</td>
<td>122.5</td>
<td>548</td>
<td>17</td>
<td>4.5</td>
<td>0.14</td>
</tr>
<tr>
<td>4-5</td>
<td>102</td>
<td>85.5</td>
<td>336</td>
<td>15</td>
<td>3.9</td>
<td>0.18</td>
</tr>
<tr>
<td>5-6</td>
<td>72</td>
<td>53.0</td>
<td>186</td>
<td>13</td>
<td>3.5</td>
<td>0.25</td>
</tr>
<tr>
<td>6-7</td>
<td>42</td>
<td>31.4</td>
<td>105</td>
<td>6</td>
<td>3.4</td>
<td>0.19</td>
</tr>
<tr>
<td>7-end</td>
<td>17</td>
<td>9.6</td>
<td>40</td>
<td>3</td>
<td>4.2</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Some typical LWBC patterns
Two time-dependent covariates

- \( X_t \) last observed value of LWBC
- \( TEL_t \) time elapsed since last observation

Model, inspired by the simulation study

\[
\ln(h(T|X_t, TEL_t, Z) | \ln(h_0(t)) | Z \beta | X_t | (TEL_t) | d(TEL_t) | (TEL) | \exp(\alpha_1 TEL) | d(TEL) | dexp(\alpha_2(TEL))
\]
Result

Nonparametric vs parametric

Heidelberg, November 2000, session 3, page 25
Table 3: Model with $TEL$ and last observed $LWBC$.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Simple model</th>
<th></th>
<th>Extended model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\beta}$</td>
<td>s.e.(\hat{\beta})</td>
<td>$\hat{\beta}$</td>
<td>s.e.(\hat{\beta})</td>
</tr>
<tr>
<td>age</td>
<td>0.018</td>
<td>0.008</td>
<td>0.023</td>
<td>0.008</td>
</tr>
<tr>
<td>Sokal score</td>
<td>0.453</td>
<td>0.180</td>
<td>0.345</td>
<td>0.192</td>
</tr>
<tr>
<td>$LWBC$</td>
<td>1.534</td>
<td>0.214</td>
<td>0.438</td>
<td>0.325</td>
</tr>
<tr>
<td>$LWBC e^{-2TEL}$</td>
<td></td>
<td></td>
<td>2.510</td>
<td>0.636</td>
</tr>
<tr>
<td>$1-e^{-6TEL}$</td>
<td></td>
<td></td>
<td>2.901</td>
<td>0.429</td>
</tr>
<tr>
<td>deviance</td>
<td>61.1</td>
<td></td>
<td>123.0</td>
<td></td>
</tr>
</tbody>
</table>
Refinement

Exponentially weighted linear extrapolation of $X$ as estimate of the current value of $X(t)$, truncated at -1 or +1
Table 4: Model with Kernel smoother with trend for \( LWBC \), with \( h = 0.067 \).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Simple model</th>
<th></th>
<th>Extended model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{\beta} )</td>
<td>s.e.((\hat{\beta}))</td>
<td>( \hat{\beta} )</td>
<td>s.e.((\hat{\beta}))</td>
</tr>
<tr>
<td>age</td>
<td>0.020</td>
<td>0.008</td>
<td>0.023</td>
<td>0.008</td>
</tr>
<tr>
<td>Sokal score</td>
<td>0.444</td>
<td>0.185</td>
<td>0.415</td>
<td>0.190</td>
</tr>
<tr>
<td>( LWBC_{wls} )</td>
<td>1.059</td>
<td>0.173</td>
<td>0.021</td>
<td>0.179</td>
</tr>
<tr>
<td>( LWBC_{wls}e^{-2TEL} )</td>
<td></td>
<td></td>
<td>3.204</td>
<td>0.505</td>
</tr>
<tr>
<td>( 1-e^{-6TEL} )</td>
<td></td>
<td></td>
<td>2.988</td>
<td>0.426</td>
</tr>
<tr>
<td>deviance</td>
<td>55.3</td>
<td></td>
<td>134.6</td>
<td></td>
</tr>
</tbody>
</table>
Predicting from the last model, starting at $t_0$

$$\ln(\hat{?}(t|Z,X_t^{(i)},TEL_0)^{\prime}) \ln(\hat{?}_0(t))\%Z\beta\%X_t^{(i)}(TEL_0%&t_0)%d(TEL_0)$$

$X_t^{(i)}$ prediction of $X$

$TEL_0$ age of observation at time of prediction

Choice between $d(TEL_0)$ and $d(TEL_0%&t_0)$ depending on interpretation of $TEL$-effect
Figure 4: The latent multi-state model with two different states (monitored and not-monitored) with one-way transition intensity $\kappa$ and corresponding hazard rates $\lambda_a$ and $\lambda_b$. 
Predictions using fixed (genuine) TEL-effect lead to

Figure 5: Predictive use of the model, (a) Cumulative baseline hazard; (b) LWBC observations and projections for two patients; (c) Predicted survival curves for the same patients.
Predicting by landmarking

Landmark idea:

*If you want to predict at some $t_0$, use all data available at that landmark point, throw away all time-dependent covariate information after $t_0$ and make a predictive model.*

Different situations

- continuous observations
- irregular observations
Continuous observations

- $t_0$ landmark point
- $X(t_0)$ last observations
- $Z$ fixed covariate
- $s$ time since $t_0$ ($t' t_0 s$)

Model

$$h(t_0 | X(T_0), Z) = h_0(s) e^{\beta(s) X(t_0) Z}$$

$$\beta(s) = \beta_0 + \delta_{as}$$

(everything may depend on $t_0$)
Application to CML-trial data with $X(t_0)$ last observation before $t_0$. Results for $t_0=1,2,3$ and 4 years

Table 1: Parameter estimates and standard deviations using the separate landmark analyses for the landmarks $t_0 = 1, 2, 3, 4$ year.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Landmark 1</th>
<th>Landmark 2</th>
<th>Landmark 3</th>
<th>Landmark 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>0.17 (0.08)</td>
<td>0.09 (0.09)</td>
<td>0.11 (0.11)</td>
<td>0.08 (0.14)</td>
</tr>
<tr>
<td>Sokal score</td>
<td>0.53 (0.18)</td>
<td>0.68 (0.22)</td>
<td>0.21 (0.37)</td>
<td>0.36 (0.43)</td>
</tr>
<tr>
<td>LWBC($t_0$)</td>
<td>0.73 (0.36)</td>
<td>0.75 (0.43)</td>
<td>0.18 (0.58)</td>
<td>0.42 (0.38)</td>
</tr>
<tr>
<td>LWBC($t_0$)$e^{-\alpha}$</td>
<td>3.45 (1.42)</td>
<td>1.97 (1.70)</td>
<td>3.72 (1.50)</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>2.5</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Wanted: model for all $t_0$.

Simple procedure:

- create prediction data-sets for k landmarks on the interval of interest (the interval $[1,4]$ in the example)
- stack the data-sets
- fit the model

$$ h(t_0) = h_0(s) e^{\beta(s)X(t_0) + Z}\ dt_0 $$

(or more complicated if you think this is too simple)
Table 2: Parameter estimates and bootstrapped s.e.'s of the multiple landmark model for a different number $k$ of equidistant chosen landmarks at [1,4] year. The sum of the bootstrapped variances $SBV$ is the criterion to decide on the number of landmarks.

<table>
<thead>
<tr>
<th>$k$</th>
<th>age</th>
<th>Sokal score</th>
<th>$LWBC(t_0)$</th>
<th>$LWBC(t_0)e^{-2s}$</th>
<th>$t_0$</th>
<th>$SBV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.15</td>
<td>0.51 (0.16)</td>
<td>0.57 (0.31)</td>
<td>1.39 (1.06)</td>
<td>0.12</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>0.12</td>
<td>0.52 (0.19)</td>
<td>0.72 (0.32)</td>
<td>0.42 (0.80)</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>4</td>
<td>0.12</td>
<td>0.53 (0.18)</td>
<td>0.62 (0.34)</td>
<td>1.73 (0.72)</td>
<td>0.10</td>
<td>0.06</td>
</tr>
<tr>
<td>5</td>
<td>0.12</td>
<td>0.54 (0.19)</td>
<td>0.70 (0.31)</td>
<td>1.65 (0.65)</td>
<td>0.10</td>
<td>0.06</td>
</tr>
<tr>
<td>6</td>
<td>0.11</td>
<td>0.59 (0.19)</td>
<td>0.78 (0.32)</td>
<td>1.32 (0.62)</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>7</td>
<td>0.12</td>
<td>0.56 (0.20)</td>
<td>0.64 (0.35)</td>
<td>1.31 (0.62)</td>
<td>0.10</td>
<td>0.06</td>
</tr>
</tbody>
</table>
Questions that arise:
• What exactly are you doing?
• What are the standard errors of your coefficients?
• How many landmarks should you take.

Approaches
• mathematical via
  S pseudo-likelihood
  S sandwich-estimators
• bootstrapping
Bootstrapping can produce
• covariance matrix of estimated coefficients (st. errors)
• cov. matrix of baseline hazard (not done)
• SBV=trace(cov. matrix) as overall criterion for precision

_Sum of Bootstrapped Variances_

General impression:
• about five landmarks suffice
Prediction is straightforward from the model, provided we have an estimate of $h_0(s)$. Rather straightforward in stacked data-set.

Consider patient(s) with

- age=60
- sokal=0.84
- $t_0=2.5$ and $LWBC(t_0)=-1,0,1$
- $t_0=1,3,5$ and $LWBC(t_0)=0.5$
Irregular observations

Last observation \( X(t_0) \) at \( t' \), \( t_0 \& TEL \).

Prediction model

\[
    h(t_0 | X(t_0), Z, t_0, TEL) \ h_0(s) \ \beta(TEL,s) X(t_0) \ \beta(TEL,s) h_0(s) \ \beta(TEL,s) \ e \ \beta(TEL,s) \ \beta(TEL,s) \ e \ \beta(TEL,s) \ e \ \beta(TEL,s) \ e
\]

\[
    g_1(TEL, s)' (1 \ & TEL) \ \text{or} \ g_2(TEL, s)' (1 \ & (s \ & TEL))
\]
Table 3: Results for the $SBV$ criterion and the bootstrapped deviance. These results are given for the extended prediction model with both a fixed and time-dependent $g(TEL, s)$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$SBV$</th>
<th>$BDEV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TEL fixed</td>
<td>TEL time-dependent</td>
</tr>
<tr>
<td>2</td>
<td>3.66</td>
<td>5.13</td>
</tr>
<tr>
<td>3</td>
<td>2.59</td>
<td>3.41</td>
</tr>
<tr>
<td>4</td>
<td>1.52</td>
<td>1.98</td>
</tr>
<tr>
<td>5</td>
<td>1.58</td>
<td>2.11</td>
</tr>
<tr>
<td>6</td>
<td>1.47</td>
<td>1.87</td>
</tr>
<tr>
<td>7</td>
<td>1.44</td>
<td>1.80</td>
</tr>
</tbody>
</table>
Models compared on the basis of bootstrapped deviance \( BDEV = \hat{\beta} \sum_{\text{Boot}} (\hat{\beta})^{-1} \hat{\beta} \)

Fixed model \( g_1 \) seems to fit slightly better.

Final model

<table>
<thead>
<tr>
<th>age</th>
<th>Sokal score</th>
<th>LWBC((t_0))</th>
<th>LWBC((t_0))</th>
<th>(1 - e^{-TEL})</th>
<th>(e^{-2(s+TEL)})</th>
<th>(t_0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.14 (0.09)</td>
<td>0.52 (0.20)</td>
<td>0.58 (0.34)</td>
<td>3.74 (1.10)</td>
<td>1.11 (0.39)</td>
<td>0.01 (0.07)</td>
<td></td>
</tr>
</tbody>
</table>

Making predictive model for an irregularly observed time-dependent covariate is much harder than for a continuous one.
Interesting question: **coherence?**

- last observation at $t_{obs}$
- prediction wanted at $t_{pred}$
- could use any landmark $t_{obs} \#_0 \# t_{pred}$

Would that matter? Yes, but not very much.
Example:
age=60  Sokal=0.84,
$LWBC(t_{obs})=-1, 0$ or $1$
$t_{obs}'=2, t_{pred}'=3$ and
$t_0=2$ or $3$
Retrospective modeling

- $\ T$  survival time
- $\ X$  time-dependent covariate (process)
- $\ Z$  fixed covariate

Prospective model

$$f(X,T|Z)\ f(X|Z)f(T|X,Z)$$

Retrospective modeling (pattern-mixture à la Hogan & Laird)

$$f(X,T|Z)\ f(T|Z)f(X|T,Z)$$
Type of models

- \( f(T|Z) \)  Cox-model or parametric survival model
- \( f(X|T,Z) \)  Generalized Linear Mixed Model

Censored observations handled by either

- using observed T’s only in modeling \( f(X|T,Z) \) (HL)
- multiple imputation (discussed here) (PMDA)
- full ML (combined with EM) (not discussed)

HL  Hogan & Laird

PMDA  Poor man’s Data Augmentation
Prediction of survival after $t_0$ using **all relevant**

**information** before $t_0$ by means of Bayes Theorem

$$f(T|X(\ldots t_0),Z,T\$t_0)' \quad \frac{f(X(\ldots t_0)|T,Z)f(T)}{\int_{t_0}^{\infty} f(X(\ldots t_0)|T,Z)f(T)dT}$$

$X(\ldots t_0)$ up to $t_0$
Application to example

Table 1: Loglikelihood of several parametric survival functions for the initial estimation of $f(T|Z)$, including the fixed covariates age and Sokal score.

<table>
<thead>
<tr>
<th>Parametric survival function</th>
<th>Loglikelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>-280.0</td>
</tr>
<tr>
<td>Weibull</td>
<td>-276.3</td>
</tr>
<tr>
<td>Gaussian</td>
<td>-296.8</td>
</tr>
<tr>
<td>Logistic</td>
<td>-300.8</td>
</tr>
<tr>
<td>Log logistic</td>
<td>-278.4</td>
</tr>
<tr>
<td>Log normal</td>
<td>-281.9</td>
</tr>
</tbody>
</table>
Model for covariate process given survival time formulated in retrospective time $s^1 T & t$
Figure 2: Smoothing splines for the \textit{LWBC} courses in retrospective time for 6 randomly drawn patients with an event.
Favorite model

\[ X_i(s) \beta_0 \circ \beta_1 \circ (s) e^{s \circ \varepsilon_i(s)} \]

- fixed parameters \( \beta_0, \beta_1 \) and \( \varepsilon \)
- \( (b_0, b_1) \sim N(0, S_b) \)
- \( e(t) \) autoregressive, mean 0, variance \( s^2 \), autocorrelation \( \gamma^t \)
Table 2: Results of the parameter estimates for $f(Y|T)$ and $f(T|Z)$. The s.e.’s are given between brackets. In the third column, the results using the procedure of Hogan and Laird (HL) are given, whereas in the fourth column, the results using the Poor Man’s Data Augmentation algorithm (PMDA) are given.

<table>
<thead>
<tr>
<th>density</th>
<th>parameter</th>
<th>HL</th>
<th>PMDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(Y</td>
<td>T)$</td>
<td>$\alpha_0$</td>
<td>-0.007 (0.020)</td>
</tr>
<tr>
<td></td>
<td>$\alpha_1$</td>
<td>1.837 (0.227)</td>
<td>1.986 (0.303)</td>
</tr>
<tr>
<td></td>
<td>$\gamma$</td>
<td>2.073 (0.146)</td>
<td>2.335 (0.356)</td>
</tr>
<tr>
<td></td>
<td>$\rho$</td>
<td>0.316</td>
<td>0.345</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2$</td>
<td>0.091</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>$\Sigma_{11}$</td>
<td>0.027</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>$\Sigma_{12}$</td>
<td>-0.031</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>$\Sigma_{22}$</td>
<td>1.384</td>
<td>1.448</td>
</tr>
</tbody>
</table>

| $f(T|Z)$| intercept | 2.54 (0.20) | 2.50 (0.20) |
|         | age       | -0.01 (0.01) | -0.01 (0.01) |
|         | Sokal score | -0.31 (0.10) | -0.33 (0.09) |
|         | scale     | 0.58 (0.08) | 0.57 (0.07) |
The graph shows the relationship between LWBC and Time to event (years). The equations for the two curves are:

- \[ E(Y|T) = 0.02 + 1.99 \exp(-2.34t), \text{ PMDA} \]
- \[ E(Y|T) = 0.01 + 1.84 \exp(-2.07t), \text{ HL} \]
Prediction example: data of some patient that dies after nearly 7 years