Multivariate survival analysis

Luc Duchateau, Ghent University
Paul Janssen, Hasselt University
Multivariate survival analysis

Overview of course material
Lecture 1: Multivariate survival data examples

- Univariate survival: independent event times
  \[ T_1, T_2, \ldots, T_n \]

- Multivariate survival data: clustered event times
  \[ (T_{11}, T_{12}, \ldots, T_{1n_1}), \ldots, (T_{s1}, T_{s2}, \ldots, T_{sn_s}) \]
Lecture 2: The different analysis approaches

- Ignore dependence: basic survival analysis
  \[ h_{ij}(t) = h_0(t) \exp \left( x_{ij}^t \beta \right) \]

- The marginal model
  \[ h_{ij}(t) = h_0(t) \exp \left( x_{ij}^t \beta \right) \]

- The fixed effects model
  \[ h_{ij}(t) = h_0(t) \exp \left( x_{ij}^t \beta + c_i \right) \]
The stratified model

\[ h_{ij}(t) = h_{i0}(t) \exp(x^t_{ij}\beta) \]

The copula model

\[ S_p(t_1, t_2) = C_\theta(S_{1,p}(t_1), S_{2,p}(t_2)) \]

The frailty model

\[ h_{ij}(t) = h_0(t) w_i \exp(x^t_{ij}\beta) \]
Lecture 3: Frailties: past, present and future

- First introduction of longevity factor to better model population mortality
  = modeling overdispersion in univariate survival data

- Populations consisting of subpopulations and the requirement for different frailties
Lecture 4: Basics of the parametric gamma frailty model

\[ h_{ij}(t) = h_0(t) u_i \exp(x_{ij}^t \beta) \]

- Analytical solution: marginal likelihood by integrating out frailties
- Characteristics of this model through its Laplace transform
Lecture 5: Parametric frailty models with other frailty densities

\[ h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t \beta) \]

- Parametric
- Positive stable
- Log normal

- Analytical solution exists also for positive stable
- No analytical solution exists for log normal, the frailties are integrated out numerically
Lecture 6: The semi-parametric gamma frailty model through EM and PPL

\[ h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t \beta) \]

The marginal likelihood contains unknown baseline hazard.

Solutions can be found through the EM-algorithm or through the penalised partial likelihood approach.

Multivariate survival data
Overview of course material
Lecture 7: Bayesian analysis of the semi-parametric gamma frailty model

\[ h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t\beta) \]

Clayton developed an approach to fit this same model using MCMC algorithms.

\[ f_U(u) = \frac{u^{1/\theta-1}\exp(-u/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} \]
Lecture 8: Multifrailty and multilevel models

- Multifrailty: two frailties in one cluster

\[ h_{ij}(t) = h_0(t) \exp \left( w_{0i} + (\beta_1 + w_{1i}) x_{ij1} + x_{ij(-1)}^t \beta_{(-1)} \right) \]

Solution based on Bayesian approach. The posterior densities are obtained by Laplacian integration.
Multilevel: two or more cluster sizes, one clustering level nested in the other

\[ h_{ijk}(t) = h_0(t) u_i z_{ij} \exp(x_{ijk}^t \beta) \]

Solution based on frequentist approach by integrating out one level analytically and the other level numerically.
Multivariate survival data
Examples
Multivariate survival data types

Criteria to categorise
- Cluster size: 1, 2, 3, 4, >4
- Hierarchy: 1 or 2 nesting levels
- Event ordering: none or ordered in space/time
Univariate survival data

- Clusters of fixed size 1
- Example 1: East Coast fever transmission dynamics

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Time to ECF contact</th>
<th>Status</th>
<th>Breed</th>
<th>Month of birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>309</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>240</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>126</td>
<td>1</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>365</td>
<td>0</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>62</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>232</td>
<td>365</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Bivariate survival data

- Clusters of fixed size 2
- Example 2: diagnosis of fracture healing

<table>
<thead>
<tr>
<th>Dogid</th>
<th>Time to diagnosis</th>
<th>Status</th>
<th>Method</th>
<th>Fracture type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>1</td>
<td>RX</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>1</td>
<td>US</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>1</td>
<td>RX</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>1</td>
<td>US</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>35</td>
<td>0</td>
<td>RX</td>
<td>1</td>
</tr>
<tr>
<td>106</td>
<td>35</td>
<td>1</td>
<td>US</td>
<td>1</td>
</tr>
</tbody>
</table>
## Example 3: Udder quarter reconstitution

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Time to reconstitution</th>
<th>Status</th>
<th>Drug</th>
<th>Heifer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.50</td>
<td>0</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>6.50</td>
<td>0</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6.50</td>
<td>0</td>
<td>A</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.93</td>
<td>1</td>
<td>P</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1.90</td>
<td>1</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>0.41</td>
<td>1</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>3.80</td>
<td>1</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>99</td>
<td>6.50</td>
<td>0</td>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>4.93</td>
<td>1</td>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>100</td>
<td>6.50</td>
<td>0</td>
<td>P</td>
<td>1</td>
</tr>
</tbody>
</table>
Quadruples of correlated event times

- Cluster of fixed size 4
- Example 4: Correlated infection times in 4 udder quarters

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Lower</th>
<th>Upper</th>
<th>Midpoint</th>
<th>Status</th>
<th>Heifer</th>
<th>Quarter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>261</td>
<td>297</td>
<td>279</td>
<td>0</td>
<td>1</td>
<td>LF</td>
</tr>
<tr>
<td>1</td>
<td>261</td>
<td>297</td>
<td>279</td>
<td>1</td>
<td>1</td>
<td>LR</td>
</tr>
<tr>
<td>1</td>
<td>48</td>
<td>78</td>
<td>63</td>
<td>1</td>
<td>1</td>
<td>RF</td>
</tr>
<tr>
<td>1</td>
<td>141</td>
<td>165</td>
<td>153</td>
<td>1</td>
<td>1</td>
<td>RR</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>93</td>
<td>76.5</td>
<td>1</td>
<td>1</td>
<td>LF</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>93</td>
<td>76.5</td>
<td>1</td>
<td>1</td>
<td>LR</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>93</td>
<td>76.5</td>
<td>1</td>
<td>1</td>
<td>RF</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>93</td>
<td>76.5</td>
<td>1</td>
<td>1</td>
<td>RR</td>
</tr>
</tbody>
</table>
Event times in clusters of varying size
Few clusters of large size

- One level of clustering, but varying cluster size
- Example 5: Peri-operative breast cancer treatment

<table>
<thead>
<tr>
<th>Patid</th>
<th>Time to death/recurrence</th>
<th>Status</th>
<th>Institute</th>
<th>Country</th>
<th>Periop</th>
<th>Nodal status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>230</td>
<td>1</td>
<td>1</td>
<td>F</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>3100</td>
<td>0</td>
<td>1</td>
<td>F</td>
<td>Y</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>560</td>
<td>1</td>
<td>1</td>
<td>F</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2793</td>
<td>1800</td>
<td>0</td>
<td>14</td>
<td>N</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>2794</td>
<td>130</td>
<td>1</td>
<td>14</td>
<td>N</td>
<td>N</td>
<td>0</td>
</tr>
<tr>
<td>2795</td>
<td>1900</td>
<td>0</td>
<td>14</td>
<td>N</td>
<td>Y</td>
<td>0</td>
</tr>
</tbody>
</table>
Event times in clusters of varying size
Many small clusters

- One level of clustering, but varying cluster size
- Example 6: Breast conserving therapy for Ductal Carcinoma in Situ (DCIS)

<table>
<thead>
<tr>
<th>Patid</th>
<th>Time to local recurrence</th>
<th>Status</th>
<th>Institute</th>
<th>Radiotherapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1996</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1114</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2535</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1009</td>
<td>3056</td>
<td>0</td>
<td>46</td>
<td>0</td>
</tr>
<tr>
<td>1010</td>
<td>893</td>
<td>0</td>
<td>46</td>
<td>1</td>
</tr>
</tbody>
</table>
Event times in clusters of varying size
Many large clusters

- One level of clustering, but varying cluster size
- Example 7: Time to culling of heifer cows as a function of early somatic cell count

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Time to death</th>
<th>Status</th>
<th>Herd</th>
<th>Measurement day</th>
<th>log(SCC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>230</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>6.2</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>5.3</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>0</td>
<td>1</td>
<td>13</td>
<td>4.1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14244</td>
<td>158</td>
<td>0</td>
<td>140</td>
<td>9</td>
<td>4.4</td>
</tr>
<tr>
<td>14245</td>
<td>65</td>
<td>1</td>
<td>140</td>
<td>6</td>
<td>5.9</td>
</tr>
<tr>
<td>14246</td>
<td>88</td>
<td>1</td>
<td>140</td>
<td>12</td>
<td>8.0</td>
</tr>
</tbody>
</table>
Example 8: Time to first insemination of heifers as function of time-varying milk protein concentration

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Start</th>
<th>End</th>
<th>Status</th>
<th>Herd</th>
<th>Ureum (%)</th>
<th>Protein (%)</th>
<th>Parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>3.12</td>
<td>3.22</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>3.23</td>
<td>3.13</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>27</td>
<td>0</td>
<td>1</td>
<td>3.34</td>
<td>3.05</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>28</td>
<td>33</td>
<td>0</td>
<td>1</td>
<td>3.44</td>
<td>2.97</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>34</td>
<td>42</td>
<td>0</td>
<td>1</td>
<td>3.51</td>
<td>2.92</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>43</td>
<td>50</td>
<td>0</td>
<td>1</td>
<td>3.57</td>
<td>3.01</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>51</td>
<td>58</td>
<td>0</td>
<td>1</td>
<td>3.62</td>
<td>3.09</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>59</td>
<td>63</td>
<td>0</td>
<td>1</td>
<td>3.68</td>
<td>3.17</td>
<td>6</td>
</tr>
<tr>
<td>1</td>
<td>64</td>
<td>69</td>
<td>1</td>
<td>1</td>
<td>3.71</td>
<td>3.22</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>6.60</td>
<td>3.02</td>
<td>5</td>
</tr>
</tbody>
</table>

...
Example 9: Time to first insemination of heifer cows as a function of initial milk protein concentration

<table>
<thead>
<tr>
<th>Cowid</th>
<th>Time</th>
<th>Status</th>
<th>Herd</th>
<th>Ureum (%)</th>
<th>Protein (%)</th>
<th>Parity</th>
<th>Heifer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>69</td>
<td>1</td>
<td>1</td>
<td>3.12</td>
<td>3.22</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>51</td>
<td>71</td>
<td>1</td>
<td>1</td>
<td>3.61</td>
<td>3.42</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>52</td>
<td>34</td>
<td>1</td>
<td>2</td>
<td>2.33</td>
<td>2.87</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10503</td>
<td>37</td>
<td>0</td>
<td>181</td>
<td>2.43</td>
<td>3.40</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>10504</td>
<td>27</td>
<td>1</td>
<td>181</td>
<td>2.44</td>
<td>3.11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10512</td>
<td>78</td>
<td>1</td>
<td>181</td>
<td>3.53</td>
<td>3.52</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10513</td>
<td>156</td>
<td>1</td>
<td>181</td>
<td>1.67</td>
<td>3.09</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Example 10: Prognostic index evaluation in bladder cancer

<table>
<thead>
<tr>
<th>Patid</th>
<th>Time</th>
<th>Status</th>
<th>Centre</th>
<th>PI</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>310</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1954</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1758</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>119</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>84</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2501</td>
<td>392</td>
<td>1</td>
<td>63</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Clustered event times with ordering

- Event times in a cluster have a certain ordering
- Example 11: Recurrent asthma attacks in children

<table>
<thead>
<tr>
<th>Patid</th>
<th>Start</th>
<th>End</th>
<th>Status</th>
<th>Drug</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>168</td>
<td>1</td>
<td>P</td>
</tr>
<tr>
<td>1</td>
<td>172</td>
<td>420</td>
<td>0</td>
<td>P</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>56</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>57</td>
<td>134</td>
<td>1</td>
<td>D</td>
</tr>
<tr>
<td>2</td>
<td>148</td>
<td>325</td>
<td>0</td>
<td>D</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>0</td>
<td>465</td>
<td>0</td>
<td>D</td>
</tr>
</tbody>
</table>
Event times in 2 nested clustering levels

- Smaller clusters of event times are nested in a larger cluster
- Example 12: Infant mortality in Ethiopia

<table>
<thead>
<tr>
<th>Childid</th>
<th>Time</th>
<th>Status</th>
<th>District</th>
<th>Village</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>363</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>361</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>233</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>361</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>841</td>
<td>361</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>842</td>
<td>360</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8162</td>
<td>240</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Modeling multivariate survival data.
The different approaches.
Overview

- Basic quantities/likelihood in survival
- The different approaches
  - The marginal model
  - The fixed effects model
  - The stratified model
  - The copula model
  - The frailty model
- Efficiency comparisons
Basic quantities in survival

- The probability density function of event time $T$
  \[ f(t) \]
- The cumulative distribution function
  \[ F(t) = P(T \leq t) = \int_{0}^{t} f(u) du \]
- The survival function
  \[ S(t) = 1 - F(t) = P(T > t) \]
The hazard function

\[ h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \]

The cumulative hazard function

\[ H(t) = \int_0^t h(u) \, du \]

\[ h(t) = -\frac{d}{dt} \log(S(t)) \]

\[ S(t) = \exp(-H(t)) \]
We consider the survival likelihood for right censored data \((y_i, \delta_i)\) assuming uninformative censoring.

\[
L = \prod_{i=1}^{n} (f(y_i))^\delta_i (S(y_i))^{1-\delta_i}
\]

\[
= \prod_{i=1}^{n} (h(y_i))^\delta_i S(y_i)
\]
Survival likelihood maximisation

- Likelihood maximisation leads to ML estimates, for instance, constant hazard
  \[ h_i(t) = \lambda \]
  with likelihood and loglikelihood
  \[
  L = \prod_{i=1}^{n} \lambda^{\delta_i} \exp(-\lambda y_i) \quad l = d \log(\lambda) - \lambda \sum_{i=1}^{n} y_i
  \]

- Equating first derivative to zero gives
  \[
  \frac{dl}{d\lambda} = \frac{d}{\lambda} - \sum_{i=1}^{n} y_i \quad \hat{\lambda} = \frac{d}{\sum_{i=1}^{n} y_i}
  \]
Survival likelihood maximisation: Example

- Time to diagnosis using US

\[ d = 106, \quad \sum_{i=1}^{n} y_i = 180.433 \]

\[ \hat{\lambda} = \frac{d}{\sum_{i=1}^{n} y_i} = \frac{106}{180.433} = 0.587 \]
Survival likelihood maximisation: Example in R – explicit solution

- Time to diagnosis using US

```r
library(survival)
timetodiag <- read.table("c:\\docs\\bookfrailty\\data\\diag.csv",
header = T, sep=";")
t1 <- timetodiag$t1/30; t2 <- timetodiag$t2/30
c1 <- timetodiag$c1; c2 <- timetodiag$c2

# Explicit solution for lambda
sumy <- -sum(t1); d <- -sum(c1);
lambda1 <- -d/sumy
sumy; d; lambda1
```

Basic quantities/likelihood in survival
Survival likelihood maximisation: Example in R – iterative solution

```r
#likelihood maximisation
survlik<-function(p){d*log(p[1])+p[1]*sumy}
nlm(survlik,0.5)

$minimum [1] 162.3838
$estimate[1] 0.5874742
$gradient[1] 4.447998e+505
$code[1] 1
$iterations[1] 4
```

Basic quantities/likelihood in survival 35
Maximisation and Newton-Raphson

- Closed form solutions exist only in exceptional cases
- In most cases iterative procedures have to be used for maximisation
- Newton Raphson (NR) procedure finds a point $\alpha$ for which $f(\alpha) = 0$
- Thus we can find the estimate that maximises the likelihood by using the NR procedure on the first derivative of the (log) likelihood
Newton-Raphson procedure

- Aim: find point $\alpha$ of function with $f(\alpha) = 0$
- We start with point $x^{(n)}$ and use Taylor series expansion

$$f(\alpha) = f(x^{(n)}) + (\alpha - x^{(n)})f'(x^{(n)}) + (\alpha - x^{(n)})^2 f''(x^{(n)}) + \ldots$$

- We only take first two terms and find

$$\alpha \approx x^{(n)} - \frac{f(x^{(n)})}{f'(x^{(n)})} = x^{(n+1)}$$

- We take the right hand side as new estimate $x^{(n+1)}$ and proceed iteratively
Newton-Raphson procedure: Illustration

- Iterative procedure: 
  \[ x_{(n+1)} = x_{(n)} - \frac{f(x_{(n)})}{f'(x_{(n)})} \]
- Example graphically:

\[ y = f(x_{(n)}) + f'(x_{(n)})(x - x_{(n)}) \]

\[ f(x) = x^3 - 1 \]

\[ x_{(n)} = 1.5 \]

\[ x_{(n+1)} = 1.5 - \frac{(1.5)^3 - 1}{3(1.5)^2} \]

\[ = 1.148 \]
Newton-Raphson procedure Example

- Estimating $\lambda$ though NR:

$$y = \frac{dl(\lambda_{(n)})}{d\lambda} + \frac{d^2l(\lambda_{(n)})}{d\lambda^2}(\lambda - \lambda_{(n)})$$

$$\frac{dl}{d\lambda} = d/\lambda - \sum_{i=1}^{n} y_i$$

$$\frac{d^2l}{d\lambda^2} = -d\lambda^{-2}$$

$\lambda_{(n)} = 0.52$

$\lambda_{(n+1)} = 0.5797$

$\lambda_{(n+2)} = ?$

Basic quantities/likelihood in survival
Newton-Raphson procedure: Second iteration for the example

\[ \lambda_{(n+2)} = \lambda_{(n+1)} - \frac{f(\lambda_{(n+1)})}{f'(\lambda_{(n+1)})} \]

\[ \lambda_{(n+2)} = \lambda_{(n+1)} - \frac{d/\lambda_{(n+1)} - \sum_{i=1}^{n} y_i}{-d\lambda^{-2}_{(n+1)}} \]

\[ \lambda_{(n+2)} = 0.5797248 - \frac{106/0.5797248 - 180.433}{-106 \times (0.5797248)^{-2}} = 0.5874746 \]
Variance estimate from likelihood

- An estimate of the variance of $\hat{\lambda}$ is given by the inverse of minus the second derivative
  \[
  \frac{d^2 l}{d\lambda^2} = -d\lambda^{-2}
  \]
  evaluated at $\hat{\lambda}$

- Thus we have
  \[
  \hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}^2}{d} = \frac{0.5874^2}{106} = 0.00325
  \]
Variance from likelihood in R

- Minus the second derivative(s) of the log likelihood is obtained when using option ‘hessian=T’

```r
survlik <- function(p) {-d*log(p[1])+p[1]*sumy}
results <- nlm(survlik, 0.5, hessian=T)
1/results$hessian

[,1]
[1,] 0.003257014
```

Basic quantities/likelihood in survival
Multidimensional NR procedure

- The NR procedure for one parameter
  \[ \lambda_{(n+1)} = \lambda_{(n)} - \frac{f(\lambda_{(n)})}{f'(\lambda_{(n)})} \]
  \[ \lambda_{(n+1)} = \lambda_{(n)} - \left( \frac{d^2l(\lambda_{(n)})}{d\lambda^2} \right)^{-1} \left( \frac{dl(\lambda_{(n)})}{d\lambda} \right) \]

  can be extended to multidimensional version
  \[ \eta_{(n+1)} = \eta_{(n)} + I^{-1}(\hat{\eta}) S(\hat{\eta}) \]

  \( I^{-1}(\hat{\eta}) \) the observed information matrix and
  \( S(\hat{\eta}) \) the score vector
Overview

- Basic quantities/likelihood in survival
- The different approaches
  - The marginal model
  - The fixed effects model
  - The stratified model
  - The copula model
  - The frailty model
- Efficiency comparisons
The marginal model

- The marginal model approach consists of two stages
  - Stage 1: Fit the model without taking into account the clustering
  - Stage 2: Adjust for the clustering in the data
Consistency of marginal model parameter estimates

- The ML estimate $\hat{\beta}$ from the Independence Working Model (IWM)
  \[ h_{ij}(t) = h_0(t) \exp(x_{ij}^t \beta) \]
  is a consistent estimator for $\beta$ (Huster, 1989)

- More generally, the ML estimate $\hat{\eta}$ ($\hat{\beta}$ and possibly baseline parameters) from the IWM is also a consistent estimator for $\eta$

- Parameter $\beta$ refers to the whole population
Adjusting the variance of IWM estimates

- The variance estimate based on the inverse of the information matrix of $\eta$ is an inconsistent estimator of $\text{Var}(\hat{\eta})$
- One possible solution: jackknife estimation
- General expression of jackknife estimator for iid data (Wu, 1986)

$$
\left( \frac{N - a}{N} \right) \sum_{i=1}^{N} \left( \hat{\eta}_i - \hat{\eta} \right) \left( \hat{\eta}_i - \hat{\eta} \right)^t
$$

with $N$ the number of observations and $a$ the number of parameters
The grouped jackknife estimator

- For clustered observations: grouped jackknife estimator

\[
\left( \frac{s - a}{s} \right) \sum_{i=1}^{s} (\hat{\eta}_i - \hat{\eta}) (\hat{\eta}_i - \hat{\eta})^t
\]

with \( s \) the number of clusters
Sandwich estimator of variance of IWM estimates

- Implication: fit model $s$ times, computer-intensive!
- Lipsitz and Parzen (1996) propose the following approximation
- Likelihood maximisation is based on Newton-Raphson iteration with each step based on the Taylor series approximation

\[
\hat{\eta}^{(k+1)} = \hat{\eta}^{(k)} + \left( \sum_{i=1}^{s} I_i \left( y_{i1}, y_{i2} \mid \hat{\eta}^{(k)} \right) \right)^{-1} \sum_{i=1}^{s} S_i \left( y_{i1}, y_{i2} \mid \hat{\eta}^{(k)} \right)
\]
A good starting value is $\hat{\eta}$, and the first Newton-Raphson iteration step is then

$$\hat{\eta}_{-i}^{(1)} = \hat{\eta} + \left( \sum_{j=1,j\neq i}^{s} \mathbf{I}_j (y_{i1}, y_{i2} \mid \hat{\eta}) \right)^{-1} \sum_{j=1,j\neq i}^{s} \mathbf{S}_j (y_{i1}, y_{i2} \mid \hat{\eta})$$

In this expression, however, we have

$$\sum_{j=1,j\neq i}^{s} \mathbf{S}_j (y_{i1}, y_{i2} \mid \hat{\eta}) = \left( \sum_{j=1}^{s} \mathbf{S}_j (y_{i1}, y_{i2} \mid \hat{\eta}) \right) - \mathbf{S}_i (y_{i1}, y_{i2} \mid \hat{\eta})$$

$$= 0 - \mathbf{S}_i (y_{i1}, y_{i2} \mid \hat{\eta})$$
... and can thus be rewritten as

\[
\left( \hat{\eta}_{-i}^{(1)} - \hat{\eta} \right) = - \left( \sum_{j=1, j \neq i}^{s} I_j (y_{i1}, y_{i2} \mid \hat{\eta}) \right)^{-1} S_i (y_{i1}, y_{i2} \mid \hat{\eta})
\]

Plugging this into the grouped jackknife:

\[
\left( \frac{s - a}{s} \right) \sum_{i=1}^{s} \left[ \left( \sum_{j=1, j \neq i}^{s} I_j (y_{i1}, y_{i2} \mid \hat{\eta}) \right)^{-1} S_i (y_{i1}, y_{i2} \mid \hat{\eta}) \right]
\]

\[
S_i^t (y_{i1}, y_{i2} \mid \hat{\eta}) \left( \sum_{j=1, j \neq i}^{s} I_j (y_{i1}, y_{i2} \mid \hat{\eta}) \right)^{-1}
\]
Further assume \( s \gg a \) and always using the same information matrix, we obtain the robust variance estimator

\[
\mathbf{I}^{-1} (\hat{\eta}) \mathbf{S} (\hat{\eta}) \mathbf{S}^t (\hat{\eta}) \mathbf{I}^{-1} (\hat{\eta})
\]

with \( \mathbf{I} (\hat{\eta}) \) the information matrix and \( \mathbf{S} (\hat{\eta}) \) the score vector.

This approximation corresponds to the sandwich estimator obtained by White (1982) starting from a different viewpoint.
Example marginal model with jackknife estimator

- Example 3: Blood-milk barrier reconstitution
- Estimates from IWM model with time-constant hazard rate assumption are given by
  \[ \hat{\lambda} = 0.216 \quad \hat{\beta} = 0.176 \quad (se = 0.162) \]
- Grouped jackknife = approximation
  \[ \hat{\beta} = 0.176 \quad (se = 0.153) \]
The jackknife estimator in R

- Reading the data

```r
reconstitution <- read.table("c://docs//presentationsfrailty//Rotterdam//data\ //reconstitution.csepv",header=T,sep="",""")

#Create 5 column vectors, five different variables
cowid <- reconstitution$cowid
timerec <- reconstitution$timerec
stat <- reconstitution$stat
trt <- reconstitution$trt
heifer <- reconstitution$heifer
```
Fitting the unadjusted model

```r
res.unadjust <- survreg(Surv(timerec, stat) ~ trt, dist = "exponential", data = reconstitution)
b.unadjust <- -res.unadjust$coef[2]
stdb.unadjust <- sqrt(res.unadjust$var[2, 2])
l.unadjust <- exp(-res.unadjust$coef[1])
```
Obtaining the grouped jackknife estimator

dat <- data.frame(timerec = timerec, trt = trt, stat = stat, cowid = cowid)
res <- survreg(Surv(timerec, stat) ~ trt, data = dat, dist = "exp")
init <- c(res$coeff[1], res$coeff[2])
ncows <- length(levels(as.factor(cowid)))
bdel <- matrix(NA, nrow = ncows, ncol = 2)
for (i in 1:ncows){
temp <- reconstitution[reconstitution$cowid != i,]
coeff <- survreg(Surv(timerec, stat) ~ trt,  
              data = temp, dist = "exponential")$coeff
bdel[i,1] <- -coeff[2]; bdel[i,2] <- exp(-coeff[1])
}
sqrt(0.98 * sum((bdel[,1] - b.unadjust)^2))

The different approaches - the marginal model
Obtaining the grouped 1-step jackknife estimator

```r
bdel <- matrix(NA, nrow = ncows, ncol = 2)
for (i in 1:ncows) {
  temp <- reconstitution[reconstitution$cowid != i,]
  coeff <- survreg(Surv(timerec, stat) ~ trt, data = temp,
                   init = init, maxiter = 1, dist = "exponential")$coeff
  bdel[i, 1] <- -coeff[2]
  bdel[i, 2] <- exp(-coeff[1])
}
sqrt(0.98 * sum((bdel[, 1] - b.unadjust)^2))
```

The different approaches - the marginal model
The jackknife estimator in R

Problems

- The grouped jackknife estimator can be obtained right away by the `cluster` command
  - For Weibull, for instance, we have
    ```r
    res.adjust <- survreg(Surv(timerec, stat) ~ trt + cluster(cowid),
                         data = reconstitution)
    ```
  - For exponential, this does not seem to work ...

The different approaches - the marginal model
Jackknife estimator-simulations

- In the example, the jackknife estimate of the variance is SMALLER than the unadjusted variance!!
- Is the jackknife estimate always smaller than estimate from unadjusted model, or does this depend on the data?
- Generate data from the frailty model of time to reconstitution data with

\[
\lambda = 0.23, \beta = 0.18, \theta = 0.3
\]
We generate 2000 datasets, each of 100 pairs of two subjects for the settings

1. Matched clusters, no censoring
2. 20% of clusters 2 treated or untreated subjects, no censoring
3. Matched clusters, 20% censoring

Matched clusters: the two animals sharing the clusters have different covariate levels
Simulation results

<table>
<thead>
<tr>
<th>Model</th>
<th>First setting</th>
<th>Second setting</th>
<th>Third setting</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>IWM</td>
<td>IWM</td>
<td>IWM</td>
</tr>
<tr>
<td></td>
<td>JC</td>
<td>JC</td>
<td>JC</td>
</tr>
<tr>
<td>( \hat{\beta} ) Median (5%-95% quantile)</td>
<td>0.1757 (−0.1067; 0.4803)</td>
<td>0.1817 (−0.1299; 0.5079)</td>
<td>0.1666 (−0.0928; 0.4204)</td>
</tr>
<tr>
<td>s.e. (( \hat{\beta} )) Median</td>
<td>0.1414</td>
<td>0.1414</td>
<td>0.1655</td>
</tr>
<tr>
<td>Coverage 95% CI</td>
<td>0.881</td>
<td>0.849</td>
<td>0.964</td>
</tr>
</tbody>
</table>

The different approaches - the marginal model
The fixed effects model

- The fixed effects model is given by

\[ h_{ij}(t) = h_0(t) \exp(x_{ij}^t \beta + c_i) \]

with \( c_i \) the fixed effect for cluster \( i \), \( c_1 = 0 \)

- Assume for simplicity \( h_0(t) = \lambda \) and

\[ \lambda^t = (\lambda_1, \ldots, \lambda_s) \text{ with } \lambda_i = \lambda \exp(c_i), i = 1, \ldots, s \]

\[ h_{ij}(t) = \lambda_i \exp(x_{ij}^t \beta) \]
The fixed effects model: ML solution

- General survival likelihood expression

\[ L = \prod_{i=1}^{s} \prod_{j=1}^{n_i} (h_{ij}(y_{ij}))^{\delta_{ij}} \exp \left( - \int_{0}^{y_{ij}} h_{ij}(t) \right) \]

- For fixed effects model with constant hazard

\[ L_{fe}(\lambda, \beta) = \prod_{i=1}^{s} \lambda_{i}^{d_{i}} \exp \left( \beta \sum_{j=1}^{n_i} \delta_{ij} x_{ij} - \lambda_{i} \sum_{j=1}^{n_i} y_{ij} \exp (x_{ij}\beta) \right) \]
Fixed effects and loglinear model

- Software often uses loglinear model

\[ \log T_{ij} = \mu + x_{ij}^t \alpha + k_i + \sigma E_{ij} \]

with, for \( h_0(t) = \lambda \), \( E_{ij} \sim f_E(e) = \exp(e - \exp(e)) \)

- Comparison can be based on \( S(t) \)

\[ S_{ij}(t) = P(T_{ij} > t) = P(\log T_{ij} > \log t) \]

\[ = P(\mu + x_{ij}^t \alpha + k_i + \sigma E_{ij} > \log t) \]

\[ = P(E_{ij} > (\log t - \mu - x_{ij}^t \alpha - k_i) / \sigma) \]

\[ = P[\exp(E_{ij}) > \exp((\log t - \mu - x_{ij}^t \alpha - k_i) / \sigma)] \]
Survival function loglinear model (1)

- We look for

\[ E_{ij} \sim f_E(e) = \exp(e - \exp(e)) \implies \exp(E_{ij}) \sim ???

- General rule

\[ y = g(x) \rightarrow f_Y(y) = f_X(g^{-1}(y)) \mid \frac{dg^{-1}(y)}{dy} \mid

- Applied to

\[ y = \exp(E_{ij}), \quad g^{-1}(y) = \log(E_{ij})\]

\[ \exp(E_{ij}) \sim \exp(\log(e) - \exp(\log(e)))e^{-1} = \exp(e)\]
Equivalence fixed effects model and loglinear model

- Therefore, we have for loglinear model

\[ S_{ij}(t) = P \left[ \exp \left( E_{ij} \right) > \exp \left( (\log t - \mu - x_{ij}^t \alpha - k_i) / \sigma \right) \right] \]

\[ = \exp \left[ - \exp \left( (\log t - \mu - x_{ij}^t \alpha - k_i) / \sigma \right) \right] \]

**VS**

\[ S_{ij}(t) = \exp \left[ -\lambda t \exp \left( x_{ij}^t \beta + c_i \right) \right] \]

\[ \hat{\beta} = -\hat{\alpha}, \hat{c}_i = -\hat{k}_i, \hat{\lambda} = \exp(-\hat{\mu}) \quad \sigma = 1 \]

\[ V(\hat{\beta}) = V(\hat{\alpha}), V(\hat{c}_i) = V(\hat{k}_i), V(\hat{\lambda}) = \left( \exp(-\hat{\mu}) \right)^2 V(\hat{\mu}) \]
The delta method - general

- Original parameters $\zeta^t = (\zeta_1, \ldots, \zeta_k)$
- Interest in univariate continuous function $g(\zeta)$
- Use one term Taylor expansion of $g(\hat{\zeta})$

$$g(\hat{\zeta}) \approx g(\zeta) + \gamma^t (\hat{\zeta} - \zeta)$$

with

$$\gamma^t = \left( \frac{\partial g(\zeta)}{\partial \zeta_1}, \ldots, \frac{\partial g(\zeta)}{\partial \zeta_k} \right)$$

$$\text{Var}(g(\hat{\zeta})) \approx \gamma^t \text{Var}(\hat{\zeta} - \zeta) \gamma \approx \gamma^t \text{Var}(\hat{\zeta}) \gamma$$
The delta method - specific

- Interest in univariate cont. function
  \[ \hat{\lambda} = \exp(-\hat{\mu}) \]

- The one term Taylor expansion of
  \[ \text{Var}(\exp(-\hat{\mu})) \approx \gamma^t \text{Var}(\hat{\mu}) \gamma \]

  with \[ \gamma = \frac{\partial \exp(-\mu)}{\partial \mu} = -\exp(-\mu) \]

  \[ \text{Var}(\exp(-\hat{\mu})) \approx (\exp(-\mu))^2 \text{Var}(\hat{\mu})) \approx (\exp(-\hat{\mu}))^2 \text{Var}(\hat{\mu})) \]
Example: within cluster covariate

- Treatment effect for reconstitution data using R-function `survreg` (loglin. model): the effect of the drug

```r
res.fixed.trt <- survreg(Surv(timerec, stat) ~ as.factor(cowid) + as.factor(trt), dist = "exponential", data = reconstitution)
summary(res.fixed.trt)
```
## Parameter estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>21.0</td>
<td>3331.69</td>
</tr>
<tr>
<td>trt</td>
<td>-0.185</td>
<td>0.19</td>
</tr>
<tr>
<td>cowid2</td>
<td>-18.8</td>
<td>3331.69</td>
</tr>
<tr>
<td>(\ldots)</td>
<td>(2 \times 10^{-6}\times \hat{c}_i = -\hat{k}_i \approx 0)</td>
<td>(\hat{\beta} = -\hat{\alpha} = 0.185)</td>
</tr>
<tr>
<td>cowid65</td>
<td>(2 \times 10^{-6}\times \hat{c}_i = -\hat{k}_i \approx 0)</td>
<td>(\hat{\beta} = -\hat{\alpha} = 0.185)</td>
</tr>
<tr>
<td>(\ldots)</td>
<td>(2 \times 10^{-6}\times \hat{c}_i = -\hat{k}_i \approx 0)</td>
<td>(\hat{\beta} = -\hat{\alpha} = 0.185)</td>
</tr>
<tr>
<td>cowid100</td>
<td>-18.6</td>
<td>3331.69</td>
</tr>
</tbody>
</table>

\[ \text{s.e. } \left( \hat{\lambda} \right) = \exp ( -\hat{\mu} ) \text{ s.e. } (\hat{\mu}) = 2 \times 10^{-6} \]
Parameter interpretation

\[ \hat{\lambda} = \hat{\lambda}_1 \] corresponds to constant hazard of untreated udder quarter of cow 1

\[ \exp\left(-\hat{\mu} - \hat{\kappa}_i\right) \] corresponds to constant hazard of untreated udder quarter of cow i

- Cowid65 \(\approx 0\)
- Cowid100 \(\exp(-21+18.8)=0.11\)

Treatment effect: HR=\(\exp(0.185)=1.203\) with 95% CI \([0.83;1.75]\)
Example: between cluster covariate

- Heifer effect for reconstitution data introducing heifer first in the model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20.90</td>
<td>4723</td>
</tr>
<tr>
<td>heifer</td>
<td>-20.10</td>
<td>6680</td>
</tr>
<tr>
<td>cowid2</td>
<td>1.21</td>
<td>4723</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cowid100</td>
<td>-18.50</td>
<td>4723</td>
</tr>
</tbody>
</table>

Hazard ratio \( \exp(-\hat{\alpha}) \) impossibly high
Heifer effect for reconstitution data introducing cowids first in the model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20.90</td>
<td>4723</td>
</tr>
<tr>
<td>cowid2</td>
<td>-18.90</td>
<td>4723</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cowid100</td>
<td>-18.50</td>
<td>4723</td>
</tr>
<tr>
<td>heifer</td>
<td>0.00</td>
<td>6680</td>
</tr>
</tbody>
</table>

Hazard ratio equal to 1
Heifer effect for reconstitution data

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>20.90</td>
<td>4723</td>
</tr>
<tr>
<td>heifer</td>
<td>−20.10</td>
<td>6680</td>
</tr>
<tr>
<td>heifer</td>
<td>0.00</td>
<td>6680</td>
</tr>
</tbody>
</table>

Two estimates for heifer effect, which one (if any) is correct? Write down incidence matrix for intercept, cowid and heifer for 6 cows, with first 3 cows being heifers.
### Incidence matrix heifer – cowid

<table>
<thead>
<tr>
<th>Intercept</th>
<th>ci2</th>
<th>ci3</th>
<th>ci4</th>
<th>ci5</th>
<th>ci6</th>
<th>heifer</th>
<th>parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The incidence matrix is as follows:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 2 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 4 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 4 \\
\end{bmatrix}
\]

\[C7 = C1 - C4 - C5 - C6\]

**COMPLETE CONFOUNDING**

The different approaches - the fixed effects model
Incidence matrix heifer – cowid

<table>
<thead>
<tr>
<th>Intercept</th>
<th>ci2</th>
<th>ci3</th>
<th>ci4</th>
<th>ci5</th>
<th>ci6</th>
<th>heifer</th>
<th>parity</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C4</td>
<td>C5</td>
<td>C6</td>
<td>C7</td>
<td>C8</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>

$C_8 = C_1 + C_4 + C_5 + 3C_6$ → COMPLETE CONFOUNDING

The different approaches - the fixed effects model
The stratified model

- The stratified model is given by

\[ h_{ij}(t) = h_{i0}(t) \exp(x_{ij}^t \beta) \]

- Maximisation of partial likelihood,

\[
\prod_{i=1}^{s} \prod_{j=1}^{n_i} \left( \frac{\exp(x_{ij}^t \beta)}{\sum_{l \in R_i(y_{ij})} \exp(x_{il}^t \beta)} \right)^{\delta_{ij}}
\]

with \( R_i(y_{ij}) = \{ l : y_{il} \geq y_{ij} \} \) the risk set for cluster \( i \)
Example for bivariate data

- Consider the case of bivariate data, e.g. the reconstitution data.
- Write down a simplified version of the general expression.

\[ \prod_{i=1}^{s} \prod_{j=1}^{n_i} \left( \frac{\exp(x_{ij}^t \beta)}{\sum_{l \in R_i(y_{ij})} \exp(x_{il}^t \beta)} \right)^{\delta_{ij}} \]

- When does a cluster contribute to the likelihood?
The partial likelihood for reconstitution data

\[ L(\beta) = \prod_{i=1}^{s} \left[ \left( \frac{\exp(x_{i1}/\beta) \ I(y_{i1} < y_{i2})}{\exp(x_{i1}/\beta) + \exp(x_{i2}/\beta)} \right)^{\delta_{i1}} + \left( \frac{\exp(x_{i2}/\beta) \ I(y_{i2} < y_{i1})}{\exp(x_{i1}/\beta) + \exp(x_{i2}/\beta)} \right)^{\delta_{i2}} \right] \]

Estimates

\[ \hat{\beta} = 0.131, \text{ s.e. } \left( \hat{\beta} \right) = 0.209 \]
The stratified model in R

- Use the coxph function for time to reconstution data
  
  ```R
  > res.strat.trt<-coxph(Surv(timerec,stat)~as.factor(trt) + strata(cowid),data=reconstitution)
  > summary(res.strat.trt)
  
  coef      exp(coef)   se(coef)      z     p
  as.factor(trt)1  0.131      1.14         0.209   0.625   0.53
  
  exp(coef)  exp(-coef)   lower .95 upper .95
  as.factor(trt)1   1.14           0.878         0.757        1.72
  
  Likelihood ratio test= 0.39  on 1 df,  p=0.531
  Wald test         = 0.39  on 1 df,  p=0.532
  Score (logrank) test = 0.39  on 1 df,  p=0.532
  ```
The copula model

- If all clusters have the same size, another useful approach is the copula model.
- We consider bivariate data in which the first (second) subject in each cluster has the same covariate information, and the covariate information differs between the two subjects in the same cluster.
- Time to diagnosis of being healed:

<table>
<thead>
<tr>
<th>Dogid</th>
<th>RX</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>
A copula model is often fitted using a two-stage model:

- First specify the population (marginal) survival functions for the first (second) subject in a cluster and obtain estimates for them.
- We then generate the joint survival function by linking the population survival functions through the survival copula function (Frees et al., 1996).
Example of copula model

- Time to diagnosis of being healed

<table>
<thead>
<tr>
<th>Dogid</th>
<th>RX</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>83</td>
<td>83</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>35</td>
<td>35</td>
</tr>
</tbody>
</table>

\[ S_p(t_1, t_2) = C_\theta(S_{1,p}(t_1), S_{2,p}(t_2)) \]

\[ S_{1,p}(t) \quad S_{2,p}(t) \]

\[ \lambda_1, \rho_1 \quad \lambda_2, \rho_2 \]
Bivariate copula model likelihood

- Four different possible contributions of a cluster

\[
\prod_{i=1}^{s} (f_p(y_{i1}, y_{i2}))^{\delta_i_1 \delta_i_2} \left( - \frac{\partial S_p(y_{i1}, y_{i2})}{\partial y_{i1}} \right)^{\delta_i_1(1-\delta_i_2)} \times \left( - \frac{\partial S_p(y_{i1}, y_{i2})}{\partial y_{i2}} \right)^{(1-\delta_i_1)\delta_i_2} (S_p(y_{i1}, y_{i2}))^{(1-\delta_i_1)(1-\delta_i_2)}
\]

- Estimated population survival functions are inserted, only copula parameters unknown

The different approaches - the copula model
The Clayton copula

- The Clayton copula (Clayton, 1978) is
  \[ S_p(t_1, t_2) = (S_{1,p}^{-\theta}(t_1) + S_{2,p}^{-\theta}(t_2) - 1)^{-1/\theta} \]

- The Clayton copula corresponds to the family of Archimedean copulas, i.e.,
  \[ C(v, u) = p(q(v) + q(u)), \text{ with } p(s) = q^{-1}(s) \]

  with in the Clayton copula case
  \[ p(s) = (1 + \theta s)^{-1/\theta}, \quad q(s) = (s^{-\theta} - 1) / \theta \]
Clayton copula likelihood for parametric marginals

- Two censored observations
  \[ S_p(y_{i1}, y_{i2}) = \left( S_{1,p}^{-\theta}(y_{i1}) + S_{2,p}^{-\theta}(y_{i2}) - 1 \right)^{-1/\theta} \]

- Observation \( j \) censored
  \[ \frac{\partial S_p(y_{i1}, y_{i2})}{\partial y_{ij}} = \left( S_{1,p}^{-\theta}(y_{i1}) + S_{2,p}^{-\theta}(y_{i2}) - 1 \right)^{-1/\theta-1} \times S_{j,p}^{-\theta-1}(y_{ij}) f_{j,p}(y_{ij}) \]

- No observations censored
  \[ f_p(y_{i1}, y_{i2}) = (1 + \theta) \left( S_{1,p}^{-\theta}(y_{i1}) + S_{2,p}^{-\theta}(y_{i2}) - 1 \right)^{-1/\theta-2} \]
  \[ S_{1,p}^{-\theta-1}(y_{i1}) S_{2,p}^{-\theta-1}(y_{i2}) f_{1,p}(y_{i1}) f_{2,p}(y_{i2}) \]
Example Clayton copula

- For diagnosis of being healed data, first fit separate models for RX and US technique
- For instance, separate parametric models

\[
\begin{align*}
\text{RX}: h_{1,p}(t) &= \lambda_1 \rho_1 t^{\rho_1-1} & \text{RX}: S_{1,p}(t) &= \exp(-\lambda_1 t^{\rho_1}) \\
\text{US}: h_{2,p}(t) &= \lambda_2 \rho_2 t^{\rho_2-1} & \text{US}: S_{2,p}(t) &= \exp(-\lambda_2 t^{\rho_2}) \\
\text{RX}: f_{1,p}(t) &= \lambda_1 \rho_1 t^{\rho_1-1} \exp(-\lambda_1 t^{\rho_1}) \\
\text{US}: f_{2,p}(t) &= \lambda_2 \rho_2 t^{\rho_2-1} \exp(-\lambda_2 t^{\rho_2})
\end{align*}
\]
Estimates for marginal models are

\[ \hat{\lambda}_1 = 0.106, \quad \hat{\rho}_1 = 2.539, \quad \hat{\lambda}_2 = 0.219, \quad \text{and} \quad \hat{\rho}_2 = 2.323 \]

Based on these estimates we obtain

\[ \hat{S}_{1,p}(y_{i1}), \quad \hat{S}_{2,p}(y_{i2}), \quad \hat{f}_{1,p}(y_{i1}), \quad \text{and} \quad \hat{f}_{2,p}(y_{i2}) \]

which can be inserted in the likelihood expression which is then maximised for \( \theta \)

\[ \hat{\theta} = 0.89 \quad \rightarrow \quad \hat{r} = \frac{\hat{\theta}}{\hat{\theta} + 2} = 0.308 \]
For parametric marginal models, the likelihood can also be maximised simultaneously for all parameters leading to

\[ \hat{\lambda}_1 = 0.145, \; \hat{\rho}_1 = 2.341, \; \hat{\lambda}_2 = 0.233, \; \text{and} \; \hat{\rho}_2 = 2.212 \]

\[ \hat{\theta} = 1.066 \]

Thus, for small sample sizes, the two-stage approach can differ substantially from the one-stage approach.
Alternatives can be used for marginal models

- Nonparametric models
- Semiparametric models

\[ RX: h_{1,p}(t) = \lambda \rho t^{\rho-1} \]

\[ US: h_{2,p}(t) = \lambda \exp(\beta) \rho t^{\rho-1} \]

leading to

\[ \hat{\theta} = 0.87 \]
Clayton copula in R

timetodiag <- read.table("c:\\docs\\bookfrailty\\data\\diag.csv", header = T, sep=";")

t1<-timetodiag$t1/30;t2<-timetodiag$t2/30
c1<-timetodiag$c1;c2<-timetodiag$c2

surv1<-survreg(Surv(t1,c1)~1); surv2<-survreg(Surv(t2,c2)~1)
l1<- exp(-surv1$coeff/surv1$scale); r1<-(1/surv1$scale)
l2<-exp(-surv2$coeff/surv2$scale); r2<-(1/surv2$scale)
s1<-exp(-l1*t1^(r1)); f1<-s1*r1*l1*t1^(r1-1)
s2<-exp(-l2*t2^(r2)); f2<-s2*r2*l2*t2^(r2-1)
R-function

loglikcon.gamma.twostage<-function(theta){
P<-s1^(-theta)+ s2^(-theta)-1
loglik<- -(1-c1)*(1-c2)*(1/theta)*log(P)
  +c1*(1-c2)*(-(1+1/theta)*log(P)-(theta+1)*log(s1)+log(f1))
  +c2*(1-c1)*(-(1+1/theta)*log(P)-(theta+1)*log(s2)+log(f2))
  +c1*c2*(log(1+theta)-(2+1/theta)*log(P)-(theta+1)*log(s1)+log(f1)-(theta+1)*log(s2)+log(f2))
-sum(loglik)
}
nlm(loglikcon.gamma.twostage,c(0.5))
Results of R-function

> nlm(loglikcon.gamma.twostage,c(0.5))
$minimum
[1] 233.0512

$estimate
[1] 0.9659607

$gradient
[1] 4.197886e+05

$code
[1] 1

$iterations
[1] 4
The positive stable copula

The positive stable copula is given by

\[ S_p(t_1, t_2) = \exp \left\{ - \left[ (\log S_{1,p}(t_1))^{1/\theta} + (\log S_{2,p}(t_2))^{1/\theta} \right]^\theta \right\} \]

Is this an Archimedean copula?
The positive stable copula

The positive stable copula is given by

\[ S_p(t_1, t_2) = \exp \left\{ - \left[ (-\log S_{1,p}(t_1))^{1/\theta} + (-\log S_{2,p}(t_2))^{1/\theta} \right]^{\theta} \right\} \]

Is this an Archimedean copula?

\[ S_p(t_1, t_2) = p \left[ q(S_{1,p}(t_1)) + q(S_{2,p}(t_2)) \right] \]

Yes, it is, take

\[ p(s) = \exp(-s^\theta) \quad q(s) = (-\log s)^{1/\theta} \]

\[ p(q(s)) = \exp(-((-\log s)^{1/\theta}))^{\theta} = s \]
The positive stable copula and likelihood contributions

The positive stable copula is given by

\[ S_p(t_1, t_2) = \exp \left\{ - \left[ \left( - \log S_{1,p}(t_1) \right)^{1/\theta} + \left( - \log S_{2,p}(t_2) \right)^{1/\theta} \right]^\theta \right\} \]

Write down the likelihood assuming a semiparametric model for the population survival functions
The positive stable copula is given by

\[
S_p(t_1, t_2) = \exp \left\{ - \left[ (- \log S_{1,p}(t_1))^{1/\theta} + (- \log S_{2,p}(t_2))^{1/\theta} \right]^\theta \right\}
\]

Write down the likelihood assuming a nonparametric model for the population survival functions

- No events:  
  \[
  S_p(t_1, t_2) = - \exp (-z^\theta)
  \]

  \[
z = \left[ - \log (S_{1,p}(t_1)) \right]^{1/\theta} + \left[ - \log (S_{2,p}(t_2)) \right]^{1/\theta}
  \]
- **Likelihood contributions**
  - No events
    
    \[ S_p(t_1, t_2) = -\exp(-z^\theta) \]
    
    \[ z = \left[ -\log(S_{1,p}(t_1)) \right]^{1/\theta} + \left[ -\log(S_{2,p}(t_2)) \right]^{1/\theta} \]
  
  - One event
    
    \[ \frac{\partial}{\partial t_j} S_p(t_1, t_2) = -\exp(-z^\theta) z^{\theta-1} \frac{(-\log S_{j,p}(t_j))^{1/\theta-1}}{S_{j,p}(t_j)} \]
  
  - Two events
    
    \[ \frac{\partial^2}{\partial t_1 \partial t_2} S_p(t_1, t_2) = \exp(-z^\theta) \frac{(-\log S_{1,p}(t_1))^{1/\theta-1}}{S_{1,p}(t_1)} \frac{(-\log S_{2,p}(t_2))^{1/\theta-1}}{S_{2,p}(t_2)} \left[ z^{2(\theta-1)} + (1/\theta - 1)z^{\theta-2} \right] \]
The frailty model

- The ‘shared’ frailty model is given by

\[ h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t \beta) \]

with the frailty \( u_i \sim f_U(u) \)

- An alternative formulation is given by

\[ h_{ij}(t) = h_0(t) \exp(x_{ij}^t \beta + w_i) \]

with \( u_i = \exp(w_i) \)
The gamma frailty model

- Gamma frailty distribution is easiest choice

\[ f_U(u) = \frac{u^{1/\theta-1} \exp(-u/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)} \]

with \( E(U) = 1 \) and \( \text{Var}(U) = \theta \)
Marginal likelihood for the gamma frailty model

- Start from conditional (on frailty) likelihood

\[
L_i(\xi, \beta \mid u_i) = \prod_{j=1}^{n_i} (h_0(y_{ij})u_i \exp(x_{ij}^t \beta))^{\delta_{ij}} \exp(-H_0(y_{ij})u_i \exp(x_{ij}^t \beta))
\]

with $\xi$ containing the baseline hazard parameters, e.g., for Weibull $\xi = (\lambda, \rho)$
Marginal likelihood: integrating out the frailties...

- Integrate out frailties using distribution

\[ L_{\text{marg},i}(\zeta) = \int_0^\infty \prod_{j=1}^{n_i} (h_0(y_{ij})u \exp(x_{ij}^t, \beta))^{\delta_{ij}} \]

\[ \exp(-H_0(y_{ij})u \exp(x_{ij}^t, \beta)) \times \frac{u^{1/\theta-1}}{\theta^{1/\theta} \Gamma(1/\theta)} \exp(-u/\theta) \, du \]

with \( \zeta = (\xi, \theta, \beta) \)
Closed form expression for marginal likelihood

- Integration leads to

\[ L_{\text{marg},i}(\zeta) = \frac{\Gamma(d_i + 1/\theta) \prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(x_{ij}^{t}\beta)) \delta_{ij}}{\left(1/\theta + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(x_{ij}^{t}\beta)\right)^{1/\theta+d_i} \theta^{1/\theta}\Gamma(1/\theta)} \]

\[ l_{\text{marg}}(\zeta) = \sum_{i=1}^{s} \left[ d_i \log \theta - \log \Gamma(1/\theta) + \log \Gamma(1/\theta + d_i) \right. \]

\[ -\left(1/\theta + d_i\right) \log \left(1 + \theta \sum_{j=1}^{n_i} H_{ij,c}(y_{ij})\right) + \sum_{j=1}^{n_i} \delta_{ij} \left(x_{ij}^{t}\beta + \log h_0(y_{ij})\right) \]
Maximisation of marginal likelihood leads to estimates

- Marginal likelihood no longer contains frailties. By maximisation estimates of \( \zeta = (\xi, \theta, \beta) \) are obtained.

- Furthermore, the asymptotic variance-covariance matrix can be obtained as the inverse of the observed information matrix \( I(\zeta) = -H(\zeta) \) with \( H(\zeta) \) the Hessian matrix with entries

\[
\frac{\partial^2}{\partial \zeta_i \partial \zeta_j} l_{\text{marg}}(\zeta)
\]
Efficiency comparisons in the reconstitution data example

- Estimates (se) for reconstitution data

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\beta}$ (s.e.)</th>
<th>$\hat{\lambda}$ (s.e.)</th>
<th>$\hat{\theta}$ (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unadjusted</td>
<td>0.176 (0.162)</td>
<td>0.216 (0.025)</td>
<td></td>
</tr>
<tr>
<td>Marginal</td>
<td>0.176 (0.153)</td>
<td>0.216 (0.025)</td>
<td></td>
</tr>
<tr>
<td>Fixed effects</td>
<td>0.185 (0.190)</td>
<td>$7 \times 10^{-10}$ ($2 \times 10^{-6}$)</td>
<td></td>
</tr>
<tr>
<td>Frailty</td>
<td>0.171 (0.168)</td>
<td>0.256 (0.038)</td>
<td>0.286 (0.141)</td>
</tr>
<tr>
<td>Stratified</td>
<td>0.131 (0.209)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Efficiency comparisons-theory

For parametric model with constant hazard, we can obtain the expected information for $\beta$ (Wild, 1983)

$$\mathcal{I}(\beta) = -E\left(\frac{\partial^2 \log L}{\partial \beta^2}\right)$$

Due to asymptotic independence of parameters, the asymptotic variance for $\hat{\beta}$ is given by

$$(\mathcal{I}(\beta))^{-1}$$
- **Unadjusted model** \( \mathcal{I}_1(\beta) = \sum_{i=1}^{s} (x_{i1}^2 + x_{i2}^2) \)

- **Fixed effects model**

\[
\mathcal{I}_2(\beta) = 2 \sum_{i=1}^{s} \frac{(x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2}{3}
\]

- **Frailty model**

\[
\mathcal{I}_3(\beta) = \sum_{i=1}^{s} \left[ \frac{1/\theta (x_{i1}^2 + x_{i2}^2)}{1/\theta + 3} + \frac{2 \left( (x_{i1} - \bar{x}_i)^2 + (x_{i2} - \bar{x}_i)^2 \right)}{1/\theta + 3} \right]
\]

\[
\mathcal{I}_3(\beta) = \frac{1/\theta}{1/\theta + 3} \mathcal{I}_1(\beta) + \frac{3}{1/\theta + 3} \mathcal{I}_2(\beta)
\]
Efficiency comparisons-simulations

- Generate data from the frailty model with

\[ \lambda = 0.23, \beta = 0.18, \theta = 0.3 \]

- We generate 2000 datasets, each of 100 pairs of two subjects for the settings
  1. Matched clusters, no censoring
  2. 20% of clusters 2 treated or untreated subjects, no censoring
  3. Matched clusters, 20% censoring
## Simulation results

<table>
<thead>
<tr>
<th>Model</th>
<th>( \hat{\beta} ) Median (5%-95% quantile)</th>
<th>( \text{s.e.}(\hat{\beta}) ) Median</th>
<th>Coverage 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda \exp (\beta x_{ij}) )</td>
<td>0.1757 (−0.1067; 0.4803)</td>
<td>0.1414</td>
<td>0.881</td>
</tr>
<tr>
<td>( \lambda_i \exp (\beta x_{ij}) )</td>
<td>0.1846 (−0.1001; 0.4626)</td>
<td>0.1730</td>
<td>0.950</td>
</tr>
<tr>
<td>( \lambda u_i \exp (\beta x_{ij}) )</td>
<td>0.1788 (−0.0691; 0.4376)</td>
<td>0.1536</td>
<td>0.944</td>
</tr>
<tr>
<td>( h_{i0}(t) \exp (\beta x_{ij}) )</td>
<td>0.1603 (−0.1201; 0.4895)</td>
<td>0.2010</td>
<td>0.962</td>
</tr>
<tr>
<td><strong>Second setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda \exp (\beta x_{ij}) )</td>
<td>0.1817 (−0.1299; 0.5079)</td>
<td>0.1414</td>
<td>0.849</td>
</tr>
<tr>
<td>( \lambda_i \exp (\beta x_{ij}) )</td>
<td>0.1871 (−0.1268; 0.5117)</td>
<td>0.1930</td>
<td>0.949</td>
</tr>
<tr>
<td>( \lambda u_i \exp (\beta x_{ij}) )</td>
<td>0.1836 (−0.0723; 0.4540)</td>
<td>0.1601</td>
<td>0.947</td>
</tr>
<tr>
<td>( h_{i0}(t) \exp (\beta x_{ij}) )</td>
<td>0.2007 (−0.2007; 0.5645)</td>
<td>0.2247</td>
<td>0.952</td>
</tr>
<tr>
<td><strong>Third setting</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda \exp (\beta x_{ij}) )</td>
<td>0.1666 (−0.0928; 0.4204)</td>
<td>0.1655</td>
<td>0.964</td>
</tr>
<tr>
<td>( \lambda_i \exp (\beta x_{ij}) )</td>
<td>0.2222 (−0.1523; 0.6007)</td>
<td>0.1975</td>
<td>0.906</td>
</tr>
<tr>
<td>( \lambda u_i \exp (\beta x_{ij}) )</td>
<td>0.1858 (−0.1004; 0.4663)</td>
<td>0.1724</td>
<td>0.951</td>
</tr>
<tr>
<td>( h_{i0}(t) \exp (\beta x_{ij}) )</td>
<td>0.1744 (−0.1691; 0.5288)</td>
<td>0.2132</td>
<td>0.952</td>
</tr>
</tbody>
</table>
Semantics and History of the term frailty
Semantics of term frailty

- **Medical field**: gerontology
  - Frail people higher morbidity/mortality risk
  - Determine frailty of a person (e.g. Get-up and Go test)
  - Frailty: fixed effect, time varying, surrogate

- **Modelling**: statistics
  - Frailty often at higher aggregation level (e.g. hospital in multicenter clinical trial)
  - Frailty: random effect, time constant, estimable
History of term frailty - Beard

- Introduced by Beard (1959) in univariate setting to improve population mortality modelling by allowing heterogeneity.
- Beard (1959) starts from Makeham’s law (1868)

\[ h(t) = \alpha + \beta \exp(\lambda t) \]

with \( \alpha (\alpha > 0) \) the constant hazard and with \( \beta > 0, \lambda > 0 \) the hazard increases with time.
- Longevity factor is added to model

\[ h_i(t) = \alpha + u_i \beta \exp(\lambda t) \quad u_i \sim f_U \quad \text{with support } [0, \infty) \]
- Beard’s model \( h_i(t) = \alpha + u_i \beta \exp(\lambda t) \)
- Population survival function
  \[
  S_f(t) = \int_0^\infty \exp(-\alpha t) \exp \left( -u \beta \int_0^t \exp(\lambda v) dv \right) f_U(u) du
  \]
- Population hazard function
  \[
  h_f(t) = \frac{-d \log S_f(t)}{dt}
  = \frac{\int_0^\infty (\alpha + u \beta \exp(\lambda t)) \exp(-\alpha t) \exp \left( -u \beta \int_0^t \exp(\lambda v) dv \right) f_U(u) du}{\int_0^\infty \exp(-\alpha t) \exp \left( -u \beta \int_0^t \exp(\lambda v) dv \right) f_U(u) du}
  \]
With one parameter gamma distribution as frailty distribution, we obtain a Perk’s logistic curve

\[ S_f(t) = \exp(-\alpha t) \left[ 1 + \frac{\beta \theta}{\lambda} (\exp(\lambda t) - 1) \right]^{-1/\theta} \]

\[ h_f(t) = -\frac{d \log S_f(t)}{dt} = \alpha + \frac{B \exp(\lambda t)}{1 + C \exp(\lambda t)} \]

\[ B = \frac{\beta \lambda \theta^{-1}}{\lambda \theta^{-1} - \beta} \quad \text{and} \quad C = \frac{\beta}{\lambda \theta^{-1} - \beta} \]
With one parameter gamma distribution as frailty distribution, we obtain a Perk’s logistic curve

\[
h_f(t) = -\frac{d \log S_f(t)}{dt} = \alpha + \frac{B \exp(\lambda t)}{1 + C \exp(\lambda t)}
\]

\[
B = \frac{\beta \lambda \theta^{-1}}{\lambda \theta^{-1} - \beta} \quad \text{and} \quad C = \frac{\beta}{\lambda \theta^{-1} - \beta}
\]

To what value does the hazard function goes asymptotically for time to infinity?
The hazard function of Perk’s logistic curve starts at $\alpha$ and goes asymptotically to $\alpha + B/C = \alpha + \lambda/\theta$

Example: $\alpha = 0.003$, $\lambda = 0.1$, $\theta = 0.167$, $\beta = 0.01$
History of term frailty - Vaupel

- Term frailty first introduced by Vaupel (1979) in univariate setting to obtain individual mortality curve from population mortality curve

- For the case of no covariates

\[
\frac{h_i(t)}{h_j(t)} = \frac{u_i}{u_j} \quad h_i(t) = u_i h_0(t)
\]
Vaupel assumed gamma frailties and thus

\[ h_i(t) = u_i h_0(t) \]

\[ h_i(t) = u_i h_f(t) S_f^{−\theta}(t) \]

\[ h_i(t) = \frac{u_i h_f(t)}{\mathbb{E}(U | T > t)} \]

Conditional mean decreases with time,
Individual hazard increases relatively with time
From hazard to conditional probability

\[ P_i(t) = P(t \leq T_i < t + \Delta t \mid T_i \geq t) = \frac{S_i(t) - S_i(t + \Delta t)}{S_i(t)} \]

\[ S_i(t) = \exp \left( - \int_0^t u_i h_f(v) S_f^{-\theta}(v) dv \right) \]

\[ = \exp \left[ u_i \int_0^t \left( \frac{d}{dv} \log S_f(v) \right) S_f^{-\theta}(v) dv \right] \]

\[ = \exp \left[ -\frac{1}{\theta} u_i \left( S_f^{-\theta}(t) - 1 \right) \right] \]
Example Swedish population two centuries

\[ P_i(t) = 1 - \exp \left[ -\frac{u_i}{\theta} \left( S_f^{\theta}(t + \Delta t) - S_f^{\theta}(t) \right) \right] \]

Frailty semantics
Example Swedish population two centuries

\[ P_i(t) = 1 - \exp\left[ -\frac{u_i}{\theta} \left( S_f^{-\theta}(t + \Delta t) - S_f^{-\theta}(t) \right) \right] \]
Frailty – two subpopulations

- Vaupel and Yashin (1985) studied heterogeneity due to two subpopulations
  - Population 1: $S_{1,p}(t) \quad \pi(0)$
  - Population 2: $S_{2,p}(t) \quad 1 - \pi(0)$

\[
\pi(t) = \frac{\pi(0)S_{1,p}(t)}{\pi(0)S_{1,p}(t) + (1 - \pi(0))S_{2,p}(t)}
\]

\[
h_p(t) = \pi(t)h_{1,p}(t) + (1 - \pi(t))h_{2,p}(t)
\]
Smokers: high and low recidivism rate

\[ h_{1,p}(t) = 0.06, \ h_{2,p}(t) = 0.08 \]
\[ \pi(0) = 0.8 \]

Is the population hazard constant, decreasing or increasing over time?

Calculate the population hazard at time 0 and at 20 years.
Smokers: high and low recidivism rate

\[ h_{1,p}(t) = 0.06, \ h_{2,p}(t) = 0.08 \quad \pi(0) = 0.8 \]

The population hazard

- At time 0: \[ h_p(0) = \pi(0)h_{1,p}(0) + (1 - \pi(0))h_{2,p}(0) \]
  \[ = 0.8 \times 0.06 + 0.2 \times 0.08 = 0.064 \]

- At 20 years:
  \[ S_{1,p}(20) = \exp -0.06 \times 20 = 0.3012 \]
  \[ S_{2,p}(20) = \exp -0.08 \times 20 = 0.2019 \]

\[ \pi(20) = \frac{0.8 \times 0.3012}{0.8 \times 0.3012 + 0.2 \times 0.2019} = 0.8564 \]

\[ h_p(20) = 0.8564 \times 0.06 + 0.1435 \times 0.08 = 0.0628 \]
Smokers: high and low recidivism rate: pictures

\[ h_{1,p}(t) = 0.06, \quad h_{2,p}(t) = 0.08 \]

\[ \pi(0) = 0.8 \]
Reliability engineering

\begin{align*}
    h_{1,p}(t) &= 0.14 \\
    h_{2,p}(t) &= 0.001 + 0.0015t \\
    \pi(0) &= 0.5
\end{align*}
Two hazards increasing at different rates

\[ h_{1,p}(t) = 0.0001 \exp(0.2t) \]
\[ h_{2,p}(t) = 0.0001 \exp(0.1t) \]
\[ \pi(0) = 0.5 \]
Two parallel hazards (at log scale)

\[ h_{1,p}(t) = 0.01 \exp(0.04t) \]
\[ h_{2,p}(t) = 0.02 \exp(0.04t) \]
\[ \pi(0) = 0.8 \]
Basics of parametric gamma frailty model
Overview

- The gamma (frailty) density
- The marginal likelihood for the parametric gamma frailty model
- An example: time to first insemination with heifer as covariate
- Laplace transform
- From the joint survival function to the marginal likelihood
- Functions at the population level
- An example: udder quarter infection data
The gamma frailty density

- Two-parameter gamma density

\[ f_U(u) = \frac{\gamma^\delta u^{\delta-1} \exp(-\gamma u)}{\Gamma(\delta)} \]

- One-parameter gamma density: \( \delta = \gamma \)

\[ f_U(u) = \frac{u^{1/\theta-1} \exp(-u/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}} \]

E\( (U) = \frac{\delta}{\gamma} \)

Var\( (U) = \frac{\delta}{\gamma^2} \)

E\( (U) = 1 \)

Var\( (U) = \theta = 1/\gamma \)
One-parameter gamma density: pictures
The marginal likelihood for the parametric gamma frailty model

- Shared gamma frailty model (a conditional hazards model)
  \[ h_{ij}(t) = h_0(t) \exp(x_{ij}^t \beta + w_i) \]
  \[ h_{ij}(t) = h_0(t) u_i \exp(x_{ij}^t \beta) \]
  with \( U_1, \ldots, U_s \) a random sample from one-parameter gamma density

- The conditional likelihood for cluster \( i \) is
  \[
  L_i(\xi, \beta \mid u_i) = \prod_{j=1}^{n_i} (h_0(y_{ij}) u_i \exp(x_{ij}^t \beta))^{\delta_{ij}} \exp(-H_0(y_{ij}) u_i \exp(x_{ij}^t \beta))
  \]
The marginal likelihood for cluster $i$ is, with $\zeta = (\xi, \theta, \beta)$ and $d_i = \sum_{j=1}^{n_i} \delta_{ij}$,

$$L_{marg,i}(\zeta) = \int_0^\infty \prod_{j=1}^{n_i} (h_0(y_{ij}) u \exp(x_{ij}^t \beta)) \delta_{ij} \exp(-H_0(y_{ij}) u \exp(x_{ij}^t \beta)) \times \frac{u^{1/\theta - 1}}{\theta^{1/\theta} \Gamma(1/\theta)} \exp(-u/\theta) \, du$$

which has the following explicit solution

$$L_{marg,i}(\zeta) = \frac{\Gamma(d_i + 1/\theta) \prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(x_{ij}^t \beta)) \delta_{ij}}{\left(1/\theta + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(x_{ij}^t \beta)\right)^{1/\theta + d_i} \theta^{1/\theta} \Gamma(1/\theta)}$$
The marginal loglikelihood is

\[ l_{\text{marg}}(\zeta) = \sum_{i=1}^{s} \left[ d_i \log \theta - \log \Gamma(1/\theta) + \log \Gamma(1/\theta + d_i) - (1/\theta + d_i) \log \left( 1 + \theta \sum_{j=1}^{n_i} H_{ij,c}(y_{ij}) \right) + \sum_{j=1}^{n_i} \delta_{ij} (x_{ij}^t \beta + \log h_0(y_{ij})) \right] \]

where \( H_{ij,c}(y_{ij}) = H_0(y_{ij}) \exp(x_{ij}^t \beta) \)

- ML estimates \( \hat{\zeta} \) for the components of \( \zeta = (\xi, \theta, \beta) \) can be found by maximising this loglikelihood.
- The estimated variance-covariance matrix is obtained as the inverse of the observed information matrix evaluated at \( \hat{\zeta} \), \( \mathbf{I}(\hat{\zeta}) = -\mathbf{H}(\hat{\zeta}) \)
An example: time to first insemination with heifer as covariate

- See Example 9
- The hazard for cow $j$ from herd $i$ is
  \[ h_{ij}(t) = h_0(t)u_i \exp(x_{ij}\beta) \]
  with $x_{ij}$ the heifer covariate ($x_{ij} = 0$ for multiparous, $x_{ij} = 1$ for heifer), $\beta$ the heifer effect and $u_i$ the frailty term for herd $i$
- Assume $h_0(t)$ Weibull, then
  \[ S_{ij}(t) = \exp(-\lambda u_i \exp(x_{ij}\beta)t^\rho) \]
  i.e., given the value of $u_i$ the event times follow a Weibull distribution with parameters $\lambda u_i \exp(x_{ij}\beta)$ and $\rho$
Maximising the marginal loglikelihood (with time in months – for convergence reasons) we obtain

\[ \hat{\lambda} = 0.174 \quad (se = 0.009) \]
\[ \hat{\rho} = 1.769 \quad (se = 0.014) \]
\[ \hat{\theta} = 0.394 \quad (se = 0.041) \]
\[ \hat{\beta} = -0.153 \quad (se = 0.023) \]

For herd with \( w_i = 1 \) the hazard ratio is \( \exp(-0.153) = 0.858 \). The 95% CI is [0.820,0.897], i.e., the hazard of first insemination for heifers is significantly lower.
Relation between months and days

\[ \lambda u_i \exp \left( x_{ij} \beta \right) t^\rho = \lambda_d u_i \exp \left( x_{ij} \beta \right) \left( t \times \frac{365.25}{12} \right)^\rho \]

(the cumulative hazard ratio at any time is required to be the same)

\[ \lambda_d = \lambda \times \left( \frac{365.25}{12} \right)^{-\rho} \]
Hazard functions for time to first insemination

Hazard functions for time to first insemination for (a) multiparous cows and for (b) heifers, with corresponding cumulative distribution functions for (c) multiparous cows and for (d) heifers. The full line corresponds to the hazard function in herds with frailty equal to one, the lower and upper lines correspond to the hazard functions of herds with frailty equal to the 5th and the 95th percentiles of the gamma distribution with mean one and variance $\theta = 0.394$. 

An example: time to first insemination with heifer as covariate
Median time to event for herd i in heifers, resp. multiparous cows ($x_{ij} = 1$, resp. $x_{ij} = 0$): $(M_{i1},$ resp. $M_{i0})$.

$$S_{ij}(M_{i1}) = 0.5$$

$$\exp(-\lambda du_i \exp(\beta) M_{i1}^\rho) = 0.5$$

$$\lambda du_i \exp(\beta) M_{i1}^\rho = \log 2$$

$$M_{i1} = g(U_i) = \left( \frac{\log 2}{\lambda du_i \exp(\beta)} \right)^{1/\rho}$$

$$f_{M_{i1}}(m) = \rho \left( \frac{\log 2}{\theta \lambda d \exp(\beta)} \right)^{1/\theta} \frac{1}{\Gamma(1/\theta)} \left( \frac{1}{m} \right)^{1+\rho/\theta}$$

$$\exp\left( -\frac{\log 2}{\theta m^\rho \lambda d \exp(\beta)} \right)$$

An example: time to first insemination with heifer as covariate
Median time to first insemination

Fig. 2.2. Density functions for the median time to first insemination in multiparous cows and heifers.

An example: time to first insemination with heifer as covariate
Maximising marginal likelihood in R

- First read in time to first insemination data

```r
# Read data
insemfix <- read.table("c://docs//presentationsfrailty//Rotterdam//data//insemination.datc", header=T, sep="","")
# Create four column vectors, four different variables
herd <- insemfix$herdnr; timeto <- (insemfix$end*12/365.25)
stat <- insemfix$score; heifer <- insemfix$par2
# Derive some values
n <- length(levels(as.factor(herd)))
di <- aggregate(stat, by=list(herd), FUN=sum)[,2]; r <- sum(di)
```
Define loglikelihood function for transformed parameters

# Observable likelihood weibull with transformed variables
# l = exp(p[1]), theta = exp(p[2]), beta = p[3], rho = exp(p[4])
# r = No events, di = number of events by herd
likelihood.weibul <- function(p){
cumhaz <- -exp(heifer*p[3])*(timeto^(exp(p[4])))*exp(p[1])
cumhaz <- aggregate(cumhaz, by=list(herd), FUN=sum)[,2]
lnhaz <- stat*(heifer*p[3]+log((exp(p[4])*timeto^(exp(p[4])-1))*exp(p[1])))
lnhaz <- aggregate(lnhaz, by=list(herd), FUN=sum)[,2]
lik <- -r*log(exp(p[2]))-
    sum((di+1/exp(p[2]))*log(1+cumhaz*exp(p[2]))) + sum(lnhaz) + sum(sapply(di,function(x)
        ifelse(x==0,0,log(prod(x+1/exp(p[2])-seq(1,x))))))
-lik}

Maximising marginal likelihood in R
• Maximise the function for the transformed parameters and return parameter estimates

\[
\text{res<-nlm(likelihood.weibul, c(log(0.174), log(0.39), -0.15, log(1.76)))}
\]

\[
\text{lambda<-exp(res$estimate[1])}
\]

\[
\text{theta<-exp(res$estimate[2])}
\]

\[
\text{beta<-res$estimate[3]}
\]

\[
\text{rho<-exp(res$estimate[4])}
\]
- Obtain standard errors for parameter estimates from Hessian matrix on original variables

# Observable likelihood weibull with original variables to obtain the Hessian
likelihood.weibul.natural <- function(p) {
  cumhaz <- exp(heifer*p[3])*(timeto^(p[4]))*p[1]
  cumhaz <- aggregate(cumhaz, by=list(herd), FUN=sum)[,2]
  lnhaz <- stat*(heifer*p[3]+log(p[4]*timeto^(p[4]-1)*p[1]))
  lnhaz <- aggregate(lnhaz, by=list(herd), FUN=sum)[,2]
  lik <- -r*log(p[2]) -
      sum((di+1/p[2])*log(1+cumhaz*p[2]))+sum(lnhaz)+
      sum(sapply(di, function(x) log(prod(x+1/p[2]-seq(1,x)))))
      -lik
}
res.nat <- nlm(likelihood.weibul.natural, 
               c(lambda,theta,beta,rho), hessian=T,iterlim=1)
stderrs <- sqrt(diag(solve(res.nat$hessian)))
Laplace transform

- Characteristic function

\[ \phi(s) = E(\exp(isX)) \]

- Moment generating function

\[ E(\exp(iX)) \]

- Laplace transform for positive r.v.

\[ \mathcal{L}(s) = E(\exp(-sX)) = \int_{0}^{\infty} \exp(-sx)f(x)\,dx \]
Generate \( n^{\text{th}} \) moment

- Use \( n^{\text{th}} \) derivative of Laplace transform

\[
\mathcal{L}^{(n)}(s) = (-1)^n \int_0^\infty x^n \exp(-sx) f(x) \, dx
\]

- Evaluate at \( s=0 \)

\[
\mathcal{L}^{(n)}(0) = (-1)^n \int_0^\infty x^n f(x) \, dx = (-1)^n \mathbb{E}(X^n)
\]

- The Laplace transform for the one-parameter gamma density is

\[
\mathcal{L}(s) = (1 + \theta s)^{-1/\theta}
\]
From the joint survival and density functions to the marginal likelihood

- Joint survival function in conditional model

\[ S_i(t_{ni}) = \exp \left[ -u_i \left( H_0(t_1) \exp(x_{i1}^t \beta) + \ldots + H_0(t_{ni}) \exp(x_{ini}^t \beta) \right) \right] \]

- Use as notation

\[ H_{ij,c}(t) = H_0(t) \exp(x_{ij}^t \beta) \quad H_{i,,c}(t_{ni}) = \sum_{j=1}^{n_i} H_{ij,c}(t_j) \]

- For cluster of size \( n \) with covariates \( \mathbf{x} = (x_1^t, \ldots, x_n^t)^t \)

\[
S_{\mathbf{x},f}(t_n) = \int_0^\infty \exp \left( -uH_{\mathbf{x},c}(t_n) \right) f_U(u)du \\
= \mathbb{E} \left[ \exp \left( -UH_{\mathbf{x},c}(t_n) \right) \right] = \mathcal{L}(H_{\mathbf{x},c}(t_n))
\]
For cluster of size $n$ with covariates $\mathbf{x} = (x_1^t, \ldots, x_n^t)^t$
the joint density is

$$f_{\mathbf{x}, f}(t_n) = (-1)^n \frac{\partial^n}{\partial t_1 \ldots \partial t_n} S_{\mathbf{x}, f}(t_1, \ldots, t_n)$$

$$= (-1)^n \int_0^\infty \frac{\partial^n}{\partial t_1 \ldots \partial t_n} \exp \left[ -u \left( H_{x_1, c}(t_1) + \ldots + H_{x_n, c}(t_n) \right) \right] f_U(u) du$$

$$= (-1)^n \prod_{j=1}^n h_{x_j, c}(t_j) \int_0^\infty \exp \left( -uH_{x, c}(t_n) \right) (-u)^n f_U(u) du$$

$$= (-1)^n \prod_{j=1}^n h_{x_j, c}(t_j) \mathcal{L}^{(n)}(H_{x, c}(t_n))$$

From the joint survival and density functions to the marginal likelihood
Applied to Laplace transform of gamma frailty we obtain

\[ S_{x,f}(t_n) = (1 + \theta H_{x,c}(t_n))^{-1/\theta} \]

\[ \mathcal{L}^{(k)}(s) = (-1)^k (1 + \theta s)^{-1/\theta - k} \prod_{l=0}^{k-1} (1 + l\theta) \]

\[ f_{x,f}(t_n) = (1 + \theta H_{x,c}(t_n))^{-1/\theta - n} \prod_{l=0}^{n-1} (1 + l\theta) \prod_{j=1}^{n} h_{x_j,c}(t_j) \]

\[ L_{\text{marg},i}(\zeta) = \prod_{j=1}^{n_i} h_{x_{ij},c}(y_{ij}) \left( 1 + \theta \sum_{j=1}^{n_i} H_{x_{ij},c}(y_{ij}) \right)^{-1/\theta - d_i} \prod_{l=0}^{d_i-1} (1 + l\theta) \]

\[ = \left( \prod_{j=1}^{n_i} h_{x_{ij},c}(y_{ij}) \right)^{-d_i} \mathcal{L}^{(d_i)} \left( \sum_{j=1}^{n_i} H_{x_{ij},c}(y_{ij}) \right) \]

From the joint survival and density functions to the marginal likelihood.
Note. Same as the $L_{\text{marg},i}(\zeta)$, but now from more general Laplace point of view.
Functions at the population level

- Population survival function (integrate out the frailty from the conditional survival function)

\[ S_{x,f}(t) = \int_{0}^{\infty} \exp(-uH_{x,c}(t)) f_{U}(u)du = \mathcal{L}(H_{x,c}(t)) \]

- Population density function

\[ f_{x,f}(t) = \frac{d(1 - S_{x,f}(t))}{dt} = -\mathcal{L}^{(1)}(H_{x,c}(t)) h_{x,c}(t) \]

- Population hazard function

\[ h_{x,f}(t) = \frac{-\mathcal{L}^{(1)}(H_{x,c}(t))}{\mathcal{L}(H_{x,c}(t))} h_{x,c}(t) \]
Ratio of population and conditional hazard

\[
\frac{h_{X,f}(t)}{h_{X,c}(t)} = \frac{-\mathcal{L}^{(1)}(H_{X,c}(t))}{\mathcal{L}(H_{X,c}(t))}
\]

Applied to the gamma frailty we have

\[
S_{X,f}(t) = \mathcal{L}(H_{X,c}(t)) = (1 + \theta H_{X,c}(t))^{-1/\theta}
\]

\[
f_{X,f}(t) = (1 + \theta H_{X,c}(t))^{-1/\theta-1} h_{X,c}(t)
\]

\[
h_{X,f}(t) = (1 + \theta H_{X,c}(t))^{-1} h_{X,c}(t) = S_{X,f}(t) h_{X,c}(t)
\]

\[
\frac{h_{X,f}(t)}{h_{X,c}(t)} = S_{X,f}^{\theta}(t)
\]
Ratio of population and conditional hazard for the gamma frailty

**Fig. 4.2.** The ratio of the population and the conditional hazard function for the gamma frailty distribution for different values of $\theta$. 

Functions at the population level
Population hazard ratio for a binary (0-1) covariate

\[ HR_p(t) = \frac{\mathcal{L}(1)(H_0(t) \exp(\beta))}{\mathcal{L}(H_0(t) \exp(\beta))} \frac{\mathcal{L}(H_0(t))}{\mathcal{L}(1)(H_0(t))} \exp(\beta) \]

Applied to the gamma frailty we have

\[ HR_p(t) = \frac{(1 + \theta H_0(t) \exp(\beta))^{-1} h_0(t) \exp(\beta)}{(1 + \theta H_0(t))^{-1} h_0(t)} \]

\[ = \frac{1 + \theta H_0(t)}{1 + \theta H_0(t) \exp(\beta)} \exp(\beta) \]
Population hazard ratio for the gamma frailty

Fig. 4.3. The evolution of the population hazard ratio as a function of the population survival function for the conditional hazard ratio equal to two for the gamma frailty distribution for different values of $\theta$. 

Functions at the population level
In general we have

\[ f_U(u \mid T > t) = \frac{P(T > t \mid u)}{P(T > t)} = \exp\left(-uH_0(t)\right) f_U(u) \frac{\mathcal{L}(H_0(t))}{\mathcal{L}(H_0(t))} \]

Applied to the gamma frailty we have

\[ f_U(u \mid T > t) = \frac{\exp\left(-uH_0(t)\right) f_U(u)}{(1 + \theta H_0(t))^{-1/\theta}} \]

\[ = \frac{(1/\theta + H_0(t))^{1/\theta} u^{1/\theta - 1} \exp\left[-(1/\theta + H_0(t))u\right]}{\Gamma\left(1/\theta\right)} \]

\[ \sim \text{Gamma}\left(1/\theta, 1/\theta + H_0(t)\right) \]
with conditional mean

$$E(U \mid T > t) = \frac{1/\theta}{1/\theta + H_0(t)} = (1 + \theta H_0(t))^{-1} = S^\theta_f(t) \leq 1$$

and conditional variance

$$\text{Var}(U \mid T > t) = \frac{1/\theta}{(1/\theta + H_0(t))^2}$$
Fig. 4.4. The conditional variance of $U$ given the survival time exceeds $t$ for the gamma frailty distribution.
An example: udder quarter infection data

- See Example 4
- Binary covariate: heifer/multiparous ($\beta$ is the heifer effect)
  - Weibull baseline hazard
  - Gamma frailty density
- Maximising the marginal loglikelihood we obtain

$$\hat{\lambda} = 0.838 \quad (se = 0.214)$$
$$\hat{\rho} = 1.979 \quad (se = 0.106)$$
$$\hat{\beta} = 0.317 \quad (se = 0.328)$$
$$\hat{\theta} = 1.793 \quad (se = 0.304)$$
Fig. 4.5. The conditional and population hazard functions for the udder quarter infection data for (a) primiparous and (b) multiparous cows for the gamma frailty distribution with $\hat{\theta} = 1.793$. 
Fig. 4.6. The conditional hazard ratio (constant) and the population hazard ratio of primiparous versus multiparous cows for the udder quarter infection data for the gamma frailty distribution with $\hat{\theta} = 1.793$. 
Mean and variance of conditional frailty density

Fig. 4.7. The conditional mean and variance of $U$ given the survival time exceeds $t$ for the udder quarter infection data with the gamma frailty distribution with $\hat{\theta} = 1.793$.  

Updating (no covariates, gamma frailty)
The parametric gamma frailty model: variations on the theme

Overview

- A piecewise constant baseline hazard
- Recurrent events
A piecewise constant baseline hazard

- See Example 7
- Event time: time to culling
  Covariate: logarithm of somatic cell count (SCC)
- The model

\[ h_{ij}(t) = (\lambda_1 I(t \leq 70) + \lambda_2 I(70 < t \leq 300) + \lambda_3 I(t > 300)) \times \exp(\beta \log(SCC_{ij})) \]

- The MLE are given by
  \[ \hat{\lambda}_1 = 0.007 \quad (se = 0.001) \]
  \[ \hat{\lambda}_2 = 0.011 \quad (se = 0.001) \quad \hat{\lambda}_3 = 0.097 \quad (se = 0.008) \]
  \[ \hat{\beta} = 0.088 \quad (se = 0.016) \quad \hat{\theta} = 0.144 \quad (se = 0.0029) \]
- LR test for constant baseline hazard versus piecewise constant baseline hazard

\[ H_0 : \lambda_1 = \lambda_2 = \lambda_3 \]

versus

\[ H_a : \lambda_i \neq \lambda_j \quad \text{for at least one pair } (i, j) \]

with \( i \neq j \) and \( i, j \in \{1, 2, 3\} \)

\[ LR = 2(-15125.74 + 13608.25) = 3034 \]

The p-value (using a \( \chi^2 \)-distribution with 2 degrees of freedom) is smaller than 0.00001
Time to culling: piecewise constant hazard frailty model in R

- First read in time to culling data
  
  ```
  # Read data
  culling <- read.table('c:\culling.datc', header=T, sep=";")
  culltime <- culling$timetocul*12/365.25
  logSCC <- culling$logSCC
  herd <- as.factor(culling$herdnr)
  status <- culling$status
  endp1 <- 70*12/365.25; endp2 <- 300*12/365.25
  n <- length(levels(as.factor(herd)))
  di <- aggregate(status, by=list(herd), FUN=sum)[,2]
  r <- sum(di)
  ```

A piecewise constant baseline hazard
- Fit model with linear effect of log(SCC), constant hazard

# parametric exponential covariate lnscc1
likelihood.exp<-function(p){
cumhaz<-exp(logSCC*p[1])*exp(p[3])*culltime
cumhaz<-aggregate(cumhaz,by=list(herd),FUN=sum)[,2]
lnhaz<-status* (logSCC*p[1]+log(exp(p[3])))
lnhaz<-aggregate(lnhaz,by=list(herd),FUN=sum)[,2]
lik<- -r*log(exp(p[2]))-sum((di+1/exp(p[2]))
   *log(1+cumhaz*exp(p[2]))) +sum(lnhaz)-
   n*log(gamma(1/exp(p[2])))
   +sum(log(gamma(di+1/exp(p[2])))))
-lik}
initial<-c(0,log(0.2),log(0.01))
t<-nlm(likelihood.exp,initial,print.level=2)
beta<-t$estimate[1];theta<-exp(t$estimate[2])
lambda<-exp(t$estimate[3]);likelihood.linear<-t$minimum
- Results model with linear effect of log(SCC), constant hazard

\[
\begin{align*}
\beta & = -t\text{estimate}[1] \\
\theta & = -\exp(t\text{estimate}[2]) \\
\lambda & = -\exp(t\text{estimate}[3]) \\
\text{likelihood.linear} & = -t\text{minimum}
\end{align*}
\]

> beta
[1] 0.08813477

> theta
[1] 0.06230903

> lambda
[1] 0.01606552

> likelihood.linear
[1] 15079.36
Fit model with linear effect of log(SCC); piecewise constant hazard

likelihood.pieceexp<-function(p){
cumhaz<-exp(logSCC*p[1])*exp(p[3])*pmin(culltime,endp1)
cumhaz<-cumhaz+exp(logSCC*p[1])
   *exp(p[4])*pmax(0,pmin(endp2-endp1,culltime-endp1))
cumhaz<-cumhaz+exp(logSCC*p[1])*exp(p[5]) *pmax(0,culltime-endp2)
cumhaz<-aggregate(cumhaz,by=list(herd),FUN=sum)[,2]
lnhaz<-status*(logSCC*p[1]+log(as.numeric(culltime<endp1))
   *exp(p[3])+as.numeric(endp1<=culltime)
   *as.numeric(culltime<endp2)*exp(p[4])
   +as.numeric(culltime>=endp2)*exp(p[5])))
lnhaz<-aggregate(lnhaz,by=list(herd),FUN=sum)[,2]
lik<-r*log(exp(p[2]))-
    sum((di+1/exp(p[2]))*log(1+cumhaz*exp(p[2]))) + sum(lnhaz)-
    n*log(gamma(1/exp(p[2]))) +sum(log(gamma(di+1/exp(p[2]))))
    -lik}
initial <- c(0, log(0.2), log(0.01), log(0.01), log(0.01))
t <- nlm(likelihood.piecexp, initial, print.level = 2)

beta <- t$estimate[1]; theta <- exp(t$estimate[2])
lambda1 <- exp(t$estimate[3]); lambda2 <- exp(t$estimate[4])
lambda3 <- exp(t$estimate[5])
likelihood.piecewise <- -t$minimum

> beta
[1] 0.08831314

> theta
[1] 0.1440544

> lambda1
[1] 0.006735632

> lambda2
[1] 0.01151625

> lambda3
[1] 0.09857648

> likelihood.piecewise
[1] 13562.13

A piecewise constant baseline hazard
Comparing constant and piecewise constant hazard function models

> LR<-2*(-likelihood.piecewise+likelihood.linear)
> 1-pchisq(LR,2)
[1] 0
Recurrent events

- See Example 11
- Time to asthma attack is the event time
  Covariate: placebo versus drug
  \[ x_i = 0 \text{ versus } x_i = 1 \]
- Patient \( i \) has \( n_i \) at risk periods

\[ \left( y_{i11}, y_{i12}, \delta_{i1} \right), \ldots, \left( y_{in_i1}, y_{in_i2}, \delta_{in_i} \right) \]

which can be represented in a graphical way (first two patients)
Different timescales can be considered

- Calendar time

- Gap time
We give four conditional hazard models and we use the marginal likelihood to estimate the model parameters.

For all four models the marginal likelihood that needs to be maximised is

\[
    l_{marg}(\zeta) = \sum_{i=1}^{s} \left[ d_i \log \theta - \log \Gamma(1/\theta) + \log \Gamma(1/\theta + d_i) \right]
\]

\[-(1/\theta + d_i) \log \left( 1 + \theta \sum_{j=1}^{n_i} H_{ij,c}(y_{ij}) \right) + \sum_{j=1}^{n_i} \delta_{ij} \log h_{ij,c}(y_{ij})\]

where \( \zeta = (\lambda, \rho, \theta, \beta) \) represents drug effect, heterogeneity between patients, and Weibull baseline hazard.

For each of the four models we need to specify the precise meaning of \( h_{ij,c}(\cdot) \) and \( H_{ij,c}(\cdot) \). Recurrent events
Calendar time model

\[ h_{ij}(t) = \begin{cases} 
  h_0(t)u_i \exp(\beta x_i) & \text{for } y_{ij1} \leq t \leq y_{ij2}, j = 1, \ldots, n_i \\
  0 & \text{otherwise}
\end{cases} \]

\[ H_{ij}(y_{ij}) = \int_{y_{ij1}}^{y_{ij2}} h_{ij}(t)\,dt \]

Weibull baseline hazard

\[ h_0(t) = \lambda \rho t^{\rho - 1} \]
Gap time model. Information used is

\[(y_{i12} - y_{i11}, \delta_{i1}), \ldots, (y_{in_i2} - y_{in_i1}, \delta_{in_i})\]

\[h_{ij}(t) = \begin{cases} 
    h_0(t - y_{ij1})u_i \exp(\beta x_i) & \text{for } y_{ij1} \leq t \leq y_{ij2}, j = 1, \ldots, n_i \\
    0 & \text{otherwise}
\end{cases}\]

\[H_{ij}(y_{ij}) = \int_{0}^{y_{ij2} - y_{ij1}} h_{ij}(t)dt\]

Weibull baseline hazard \[h_0(t) = \lambda \rho t^{\rho - 1}\]
- Gap time model with hazard for first event different and constant

\[
h_{ij}(t) = \begin{cases} 
\lambda_f u_i \exp(\beta x_i) & \text{for } 0 \leq t \leq y_{i12} \\
\lambda \rho(t - y_{ij1})^{p-1} u_i \exp(\beta x_i) & \text{for } y_{ij1} \leq t \leq y_{ij2}, j = 2, \ldots, n_i \\
0 & \text{otherwise}
\end{cases}
\]

- Combining gap and calendar time

\[
h_{ij}(t) = \begin{cases} 
\lambda_f \rho_f t^{\rho_f-1} u_i \exp(\beta x_i) & \text{for } 0 \leq t \leq y_{i12} \\
(\lambda_f \rho_f t^{\rho_f-1} + \lambda \rho(t - y_{ij1})^{p-1}) u_i \exp(\beta x_i) & \text{for } y_{ij1} \leq t \leq y_{ij2}, j = 2, \ldots, n_i \\
0 & \text{otherwise}
\end{cases}
\]
### Overview of ML estimates for the four models

<table>
<thead>
<tr>
<th>Model</th>
<th>( \beta )</th>
<th>( \theta )</th>
<th>( \lambda )</th>
<th>( \rho )</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1:</td>
<td>-0.300</td>
<td>0.574</td>
<td>0.230</td>
<td>1.029</td>
<td>3906.8</td>
</tr>
<tr>
<td>Weibull-calendar</td>
<td>(0.0152)</td>
<td>(0.0055)</td>
<td>(0.0009)</td>
<td>(0.0013)</td>
<td></td>
</tr>
<tr>
<td>Model 2:</td>
<td>-0.254</td>
<td>0.402</td>
<td>0.316</td>
<td>0.829</td>
<td>3863.0</td>
</tr>
<tr>
<td>Weibull-gap</td>
<td>(0.0121)</td>
<td>(0.0041)</td>
<td>(0.0007)</td>
<td>(0.0006)</td>
<td></td>
</tr>
<tr>
<td>Model 3:</td>
<td>-0.251</td>
<td>0.372</td>
<td>0.346</td>
<td>0.762</td>
<td>3838.4</td>
</tr>
<tr>
<td>Weibull-gap but first event exp</td>
<td>( \hat{\lambda}_f = 0.217 )</td>
<td>( (0.0005) )</td>
<td></td>
<td>( \hat{\rho}_f = 0.946 )</td>
<td>(0.0008)</td>
</tr>
<tr>
<td>Model 4:</td>
<td>-0.245</td>
<td>0.421</td>
<td>0.134</td>
<td>0.687</td>
<td>3885.4</td>
</tr>
<tr>
<td>Weibull-gap calendar</td>
<td>(0.0126)</td>
<td>(0.0045)</td>
<td>(0.0007)</td>
<td>(0.0020)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( \hat{\lambda}_f = 0.210 )</td>
<td>( (0.0020) )</td>
<td>( \hat{\rho}_f = 0.946 )</td>
<td>( (0.0008) )</td>
<td></td>
</tr>
</tbody>
</table>

Recurrent events
Fig. 2.5. The hazard as a function of time for a subject with event history depicted at the bottom of the picture according to (a) Model 1: calendar, (b) Model 2: gap, (c) Model 3: gap but first event exp, and (d) Model 4: gap calendar.
First read in asthma data

```r
# Read data
w <- read.table("c:/docs/bookfrailty/data/asthma.dat", head=T)
di <- aggregate(w$st.w, by=list(w$id.w), FUN=sum)[,2]
r <- sum(di); n <- length(di)
trt <- aggregate(w$trt.w, by=list(w$id.w), FUN=min)[,2]
begin <- 12*w$start.w/365.25
end <- 12*w$stop.w/365.25
ptno <- w$id.w; status <- w$st.w
gap <- end - begin; fevent <- w$fevent
```
# Weibull-calendar-no specific hazard first event likelihood
likelihood.weibull.cp.nof<-function(p){
cumhaz<-p[3]*(end^p[4]-begin^p[4])
cumhaz<-aggregate(cumhaz,by=list(ptno),FUN=sum)[,2]
lnhaz<-status*(log(p[3]*p[4]*end^(p[4]-1)))
lnhaz<-aggregate(lnhaz,by=list(ptno),FUN=sum)[,2]
lik<-r*log(p[2])-n*log(gamma(1/p[2]))+sum(log(gamma(di+1/p[2])))-
sum((di+1/p[2])*log(1+p[2]*exp(p[1]*trt)*cumhaz))+
sum(di*p[1]*trt)+sum(lnhaz)-lik
}

Recurrent events 182
# Weibull-gap-no specific hazard first event likelihood
likelihood.weibull.gap.nof<-function(p){
cumhaz<-p[3]*((end-begin)^p[4])
cumhaz<-aggregate(cumhaz,by=list(ptno),FUN=sum)[,2]
lnhaz<-status*(log(p[3]*p[4])^((end-begin)^p[4]-1)))
lnhaz<-aggregate(lnhaz,by=list(ptno),FUN=sum)[,2]
lik<-r*log(p[2])-
   n*log(gamma(1/p[2]))+sum(log(gamma(di+1/p[2])))-
   sum((di+1/p[2])*log(1+p[2]*exp(p[1]*trt)*cumhaz))-
   sum(di*p[1]*trt)+sum(lnhaz)
-lik
}

Recurrent events 183
# Weibull-gap- first event exponential likelihood
likelihood.weibull.gap.fev.exp <- function(p) {
  cumhaz <- aggregate(cumhaz, by = list(ptno), FUN = sum)[, 2]
  lnhaz <- aggregate(lnhaz, by = list(ptno), FUN = sum)[, 2]
  lik <- -r * log(p[2]) -
        n * log(gamma(1/p[2])) + sum(log(gamma(di + 1/p[2]))) -
        sum(di * p[1] * trt) + sum(lnhaz)
  -lik
}

Recurrent events
# Weibull-gap-calendar-nofirstevent
likelihood.weibull.all <- function(p) {
cumhaz <- aggregate(cumhaz, by = list(ptno), FUN = sum)[, 2]
lnhaz <- aggregate(lnhaz, by = list(ptno), FUN = sum)[, 2]
lik <- -r*log(p[2]) - 
      n*log(gamma(1/p[2])) + sum(log(gamma(di + 1/p[2]))) - 
      sum(di * p[1] * trt) + sum(lnhaz) 
- lik
}

Recurrent events
Dependence measures
Overview

- Setting, for $i=1, \ldots, s$
  - binary data: $(T_{i1}, T_{i2})$
  - covariate information: $\mathbf{x}_i = (x_{i1}, x_{i2}) = (x_1, x_2) = \mathbf{x}$

- Kendall’s $\tau$: an overall measure of dependence
- The cross ratio function: a local measure of dependence
- Other local dependence measures
- An example: the gamma frailty for $\theta = 2/3$
Kendall’s $\tau$: an overall measure of dependence

$\tau = E [\text{sign}((T_{i_1} - T_{k_1})(T_{i_2} - T_{k_2}))]$  
\text{sign}(x) = -1, 0, 1 \text{ for } x < 0, x = 0, x > 0$

$\iff$

$\tau = P((T_{i_1} - T_{k_1})(T_{i_2} - T_{k_2}) > 0) - P((T_{i_1} - T_{k_1})(T_{i_2} - T_{k_2}) < 0)$

-1 \leq \tau \leq 1

$\tau = 2P((T_{i_1} - T_{k_1})(T_{i_2} - T_{k_2}) > 0) - 1$

= 2p - 1
Express $p$ in terms of joint survival and joint density function

$$p = \Pr((T_{i1} - T_{k1})(T_{i2} - T_{k2}) > 0)$$

$$= \int_0^{\infty} \int_0^{\infty} \Pr((t_1 - T_{k1})(t_2 - T_{k2}) > 0) f_f(t_1, t_2) dt_1 dt_2$$

$$= \int_0^{\infty} \int_0^{\infty} \Pr(T_{k1} > t_1, T_{k2} > t_2) f_f(t_1, t_2) dt_1 dt_2$$

$$+ \int_0^{\infty} \int_0^{\infty} \Pr(T_{k1} < t_1, T_{k2} < t_2) f_f(t_1, t_2) dt_1 dt_2$$

$$= \int_0^{\infty} \int_0^{\infty} S_f(t_1, t_2) f_f(t_1, t_2) dt_1 dt_2 + \int_0^{\infty} \int_0^{\infty} F_f(t_1, t_2) f_f(t_1, t_2) dt_1 dt_2$$

$$= 2 \int_0^{\infty} \int_0^{\infty} S_f(t_1, t_2) f_f(t_1, t_2) dt_1 dt_2$$

using $F_f(t_1, t_2) = S_f(t_1, t_2) + F_{1,f}(t_1) + F_{2,f}(t_2) - 1$
\[ \tau = 2p - 1 = 4 \int_0^\infty \int_0^\infty S_f(t_1, t_2) f_f(t_1, t_2) dt_1 dt_2 - 1 \]

Use

\[ S_f(t_1, t_2) = \mathcal{L}(H_{1,c}(t_1) + H_{2,c}(t_2)) \]
\[ f_f(t_1, t_2) = (-1)^2 \mathcal{L}^{(2)}(H_{1,c}(t_1) + H_{2,c}(t_2)) h_{1,c}(t_1) h_{2,c}(t_2) \]
\[ x = H_{1,c}(t_1) \quad dx = h_{1,c}(t_1) dt_1 \]
\[ y = H_{2,c}(t_2) \quad dy = h_{2,c}(t_2) dt_2 \]

to obtain

\[ \tau = 4 \int_0^\infty \int_0^\infty \mathcal{L}(x + y) \mathcal{L}^{(2)}(x + y) dx dy - 1 = 4 \int_0^\infty \int_0^s \mathcal{L}(s) \mathcal{L}^{(2)}(s) dv ds - 1 \]
\[ (s = x + y, v = y) \]
\[ = 4 \int_0^\infty s \mathcal{L}(s) \mathcal{L}^{(2)}(s) ds - 1 \]

Kendall's \( \tau \)
- Applied to the gamma frailty we obtain

\[ \mathcal{L}(s) = (1 + \theta s)^{-1/\theta} \]

\[ \mathcal{L}^{(2)}(s) = (1 + \theta)(1 + \theta s)^{-1/\theta - 2} \]

\[ \tau = \frac{\theta}{\theta + 2} \]
The cross ratio function: a local measure of dependence

\[ \zeta(t_1, t_2) = \frac{h_{1,f}(t_1 \mid T_2 = t_2)}{h_{1,f}(t_1 \mid T_2 > t_2)} \]

- See Example 3: time to blood-milk barrier reconstitution
  Positive experience: reconstitution at time \( t_2 \)
  For positively correlated data, we assume that hazard in numerator > hazard in denominator
Cross ratio as odds ratio

Write

\[ \zeta(t_1, t_2) = \frac{f_{1,f}(t_1 \mid T_2 = t_2)/S_{1,f}(t_1 \mid T_2 = t_2)}{f_{1,f}(t_1 \mid T_2 > t_2)/S_{1,f}(t_1 \mid T_2 > t_2)} \]

\[ = \frac{f_f(t_1, t_2) / \frac{\partial S_f(t_1, t_2)}{\partial t_2}}{\frac{\partial S_f(t_1, t_2)}{\partial t_1} / S_f(t_1, t_2)} \]
Consider conditional 2x2 table

<table>
<thead>
<tr>
<th>Reconstitution Status</th>
<th>Treated Udder Quarter</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Untreated</td>
<td>0</td>
<td>$P_{00}$</td>
</tr>
<tr>
<td>Udder Quarter</td>
<td>1</td>
<td>$P_{10}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$P_{+0}$</td>
</tr>
</tbody>
</table>

with $P_{11} = P(t_1 < T_1 \leq t_1 + \delta, t_2 < T_2 \leq t_2 + \delta \mid T_1 > t_1, T_2 > t_2)$

$P_{01} = P(T_1 > t_1 + \delta, t_2 < T_2 \leq t_2 + \delta \mid T_1 > t_1, T_2 > t_2)$

etc...

$OR(t_1, t_2, \delta) = \frac{P_{11}/P_{01}}{P_{10}/P_{00}} \xrightarrow{\delta \rightarrow 0} \zeta(t_1, t_2)$
Cross ratio as local Kendall’s $\tau$

$$\tau = 2p - 1$$

$$\frac{p}{1 - p} = \frac{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) > 0)}{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) < 0)}$$

$$\frac{p(t_1, t_2)}{1 - p(t_1, t_2)} = \frac{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) > 0 \mid T_{i1} \wedge T_{k1} = t_1, T_{i2} \wedge T_{k2} = t_2)}{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) < 0 \mid T_{i1} \wedge T_{k1} = t_1, T_{i2} \wedge T_{k2} = t_2)}$$

$$= \frac{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) > 0, T_{i1} \wedge T_{k1} = t_1, T_{i2} \wedge T_{k2} = t_2)}{P ((T_{i1} - T_{k1}) (T_{i2} - T_{k2}) < 0, T_{i1} \wedge T_{k1} = t_1, T_{i2} \wedge T_{k2} = t_2)}$$

$$\zeta(t_1, t_2) = \frac{p(t_1, t_2)}{1 - p(t_1, t_2)} = -S_f(t_1, t_2) \frac{(L^{-1})^{(2)}(S_f(t_1, t_2))}{(L^{-1})^{(1)}(S_f(t_1, t_2))}$$

The cross ratio function
Other local dependence measures

\[ \psi (t_1, t_2) = \frac{S_f (t_1, t_2)}{S_{1,f}(t_1)S_{2,f}(t_2)} = \frac{P(T_1 > t_1 \mid T_2 > t_2)}{P(T_1 > t_1)} \]

\[ \phi (t_1, t_2) = \frac{f_f (t_1, t_2)}{f_{1,f}(t_1)f_{2,f}(t_2)} \]
An example: the gamma frailty for $\theta = 2/3$

\[ S_f(t_1, t_2) = \left( \frac{1}{S_{1,f}^\theta(t_1)} + \frac{1}{S_{2,f}^\theta(t_2)} - 1 \right)^{-1/\theta} \]

\[ f_f(t_1, t_2) = \frac{(\theta + 1)f_{1,f}(t_1)f_{2,f}(t_2)}{S_{1,f}^{\theta + 1}(t_1)S_{2,f}^{\theta + 1}(t_2)\left(S_{1,f}^{-\theta}(t_1) + S_{2,f}^{-\theta}(t_2) - 1\right)^{2+1/\theta}} \]

\[ \tau = \frac{\theta}{\theta + 2} \quad \text{and} \quad \zeta(t_1, t_2) = \theta + 1 \]

\[ \psi(t_1, t_2) \quad \phi(t_1, t_2) \quad \text{see figures next slide for} \]

\[ \theta = 2/3 \quad (\tau = 0.25) \]
Dependence measures graphically for $\theta = 2/3$ \hspace{1em} (\tau = 0.25)

The ratio of the joint survival function and the product of the population survival functions (a,b) and the ratio of the joint density function and the product of the population density functions (c,d) for the gamma frailty distribution with $\tau=0.25$. 
Other choices for the frailty densities
Overview

- The positive stable frailty
  - The positive stable (frailty) density
  - The marginal likelihood for a positive stable frailty model
  - The positive stable frailty density: functions at the population level
  - An example: udder quarter infection data
  - Updating (no covariates, positive stable frailty)
  - Bivariate data: positive stable Laplace transform

- The lognormal frailty
  - The lognormal (frailty) density
  - Statistical inference for the parametric lognormal frailty model
  - An example: udder quarter infection data
The positive stable frailty density

- The positive stable density

\[ f_U(u) = -\frac{1}{\pi u} \sum_{k=1}^{\infty} \frac{\Gamma(k\theta + 1)}{k!} (-u^{-\theta})^k \sin(\theta k \pi) \]

with \( 0 < \theta < 1 \)

- Laplace transform \( \mathcal{L}(s) = \exp\left(-s^\theta\right) \)

- \( \mathcal{L}(s) \) does not exist for \( s < 0 \) and

\[ \lim_{s \to 0^+} \mathcal{L}^{(1)}(s) = -\theta \lim_{s \to 0^+} \frac{\exp\left(-s^\theta\right)}{s^{1-\theta}} = -\infty \]

- This gives evidence for the fact that the mean of the positive stable density is infinite
The positive stable density: pictures

Fig. 4.20. Positive stable density functions.
Joint survival function

\[ S_{x,f}(t_n) = \int_0^\infty \exp (-uH_{x,c}(t_n)) f_U(u) \, du \]

\[ = \mathbb{E} \left[ \exp (-UH_{x,c}(t_n)) \right] = \mathcal{L}(H_{x,c}(t_n)) \]

\[ S_{x,f}(t_n) = \exp (-H_{x,c}^\theta(t_n)) \]

This quantity is the key for obtaining the marginal likelihood
The marginal likelihood for a positive stable frailty model

- We consider the special case of quadruple data (clusters of size 4). See example 4 on udder quarter infection data.
  - Five different types of contributions, according to \( d_i \), the number of events in cluster \( i \) (\( d_i \in \{0, 1, 2, 3, 4\} \))
  - Order subjects, put uncensored observations first (1,…, l)
  - Contribution of cluster \( i \) to the marginal likelihood is

\[
\begin{align*}
  d_i = 0 & \quad S_{x_i, f}(t_n) = \exp \left( -H_{x_i, c}^\theta(t_n) \right) \\
  d_i > 0 & \quad (-1)^l \frac{\partial^l}{\partial t_1 \ldots \partial t_l} S_{x_i, f}(y_i) \\
  & = (-1)^l \mathcal{L}^{(l)}(H_{x_i, c}(y_i)) \prod_{j=1}^{l} h_0(y_{ij}) \exp \left( x_{ij}^t \beta \right)
\end{align*}
\]
Note that the previous formula can (in terms of $\delta_{ij}$ and $d_i$) be written as $(n_i = 4)$

$$\left( \prod_{j=1}^{4} h_{x_{ij},c(y_{ij})}^{\delta_{ij}} \right) (-1)^{d_i} \mathcal{L}^{(d_i)} \left( \sum_{j=1}^{4} H_{x_{ij},c(y_{ij})} \right)$$
Derivatives of Laplace transforms

\[ \mathcal{L}^{(1)}(s) = -\theta \mathcal{L}(s)s^{\theta-1} \]
\[ \mathcal{L}^{(2)}(s) = \theta^2 \mathcal{L}(s)s^{2(\theta-1)} \left(1 + \theta^{-1}(1 - \theta)s^{-\theta}\right) \]
\[ \mathcal{L}^{(3)}(s) = -\theta^3 \mathcal{L}(s)s^{3(\theta-1)} \left(1 + 3\theta^{-1}(1 - \theta)s^{-\theta}\right. \]
\[ \quad \quad \quad \quad \quad \left. + \theta^{-2}(1 - \theta)(2 - \theta)s^{-2\theta}\right)\]
\[ \mathcal{L}^{(4)}(s) = \theta^4 \mathcal{L}(s)s^{4(\theta-1)} \left(1 + 6\theta^{-1}(1 - \theta)s^{-\theta} + \theta^{-2}(1 - \theta)(11 - 7\theta)s^{-2\theta}\right. \]
\[ \quad \quad \quad \quad \quad \left. + \theta^{-3}(1 - \theta)(2 - \theta)(3 - \theta)s^{-3\theta}\right) \]
Marginal likelihood expression cluster $i$

$$
\sum_{j=1}^{4} \delta_{i,j} \log \left( h_0 (y_{ij}) \exp \left( x_{ij}^t \beta \right) \right) - H_{\infty, c}^{\theta} (y_i) + d_i \log \theta \\
+ d_i (\theta - 1) \log H_{\infty, c}^{\theta} (y_i) + C_{i,d_i}
$$

- $C_{i,0} = C_{i,1} = 0$
- $C_{i,2} = \log \left( 1 + \theta^{-1} (1 - \theta) H_{\infty, c}^{-\theta} (y_i) \right)$
- $C_{i,3} = \log \left( 1 + 3\theta^{-1} (1 - \theta) H_{\infty, c}^{-\theta} (y_i) + \theta^{-2} (2 - \theta) (1 - \theta) H_{\infty, c}^{-2\theta} (y_i) \right)$
- $C_{i,4} = \log \left( 1 + 6\theta^{-1} (1 - \theta) H_{\infty, c}^{-\theta} (y_i) + \theta^{-2} (1 - \theta) (11 - 7\theta) H_{\infty, c}^{-2\theta} (y_i) \\
+ \theta^{-3} (3 - \theta) (2 - \theta) (1 - \theta) H_{\infty, c}^{-3\theta} (y_i) \right)$

The marginal likelihood for a positive stable frailty model
The positive stable frailty density: functions at the population level

- Population survival function (integrate out the frailty term from the conditional survival function).

\[ S_{x,f}(t) = \int_0^\infty \exp(-u H_{x,c}(t)) f_U(u) du = \mathcal{L}(H_{x,c}(t)) \]

\[ S_{x,f}(t) = \exp\left(-H_{x,c}^\theta(t)\right) \]

- Population density function

\[ f_{x,f}(t) = \frac{d(1 - S_{x,f}(t))}{dt} = -\mathcal{L}^{(1)}(H_{x,c}(t)) h_{x,c}(t) \]

\[ f_{x,f}(t) = S_{x,f}(t) \theta H_{x,c}^{\theta-1}(t) h_{x,c}(t) \]
Population hazard function

\[ h_{x,f}(t) = \theta H_{x,c}^\theta(t) h_{x,c}(t) \]

Ratio of population and conditional hazard

\[ \frac{h_{x,f}(t)}{h_{x,c}(t)} = \theta H_{x,c}^\theta(t) \]

\[ \frac{h_{x,f}(t)}{h_{x,c}(t)} = \theta \left(- \log S_{x,f}(t)\right)^{1-1/\theta} \]
Ratio of population and conditional hazard for the positive stable frailty

\[
\frac{h_{x,f}(t)}{h_{x,c}(t)} = \theta \left( -\log S_{x,f}(t) \right)^{1-1/\theta}
\]

**Fig. 4.21.** The ratio of the population and the conditional hazard function for the positive stable frailty distribution for different values of \(\theta\).
- Population hazard ratio for a binary (0-1) covariate

\[ HR_P(t) = \frac{\mathcal{L}^{(1)}(H_0(t) \exp(\beta))}{\mathcal{L}(H_0(t) \exp(\beta))} \frac{\mathcal{L}(H_0(t))}{\mathcal{L}^{(1)}(H_0(t))} \exp(\beta) \]

- Applied to the positive stable frailty we have

\[ HR_P(t) = \frac{\theta H_0^{\theta-1}(t) \exp((\theta - 1)\beta) h_0(t) \exp(\beta)}{\theta H_0^{\theta-1}(t) h_0(t)} = \exp(\theta \beta) \]
Population hazard ratio for the positive stable frailty: \( HR_p(t) = \exp(\theta/\beta) \)

Fig. 4.22. The evolution of the population hazard ratio as a function of the population survival function for the conditional hazard ratio equal to two for the positive stable frailty distribution for different values of \( \theta \).
An example: udder quarter infection data

- See Example 4
- Binary covariate: heifer/multiparous ($\beta$ is the heifer effect)
  - Weibull baseline hazard
  - Positive stable frailty
- Maximising the marginal likelihood we obtain
  - $\hat{\lambda} = 0.177$ (se=0.052)
  - $\hat{\rho} = 2.129$ (se=0.114)
  - $\hat{\beta} = 0.537$ (se=0.351)
  - $\hat{\theta} = 0.529$ (se=0.039)
The conditional hazard ratio is estimated as 
\[ \exp(\hat{\beta}) = 1.71 \]

The population hazard ratio is estimated as 
\[ \exp(\hat{\theta}/\hat{\beta}) = 1.33 \]

Fig. 4.23. The conditional and population hazard functions for the udder quarter infection for (a) primiparous and (b) multiparous cows for the positive stable frailty distribution.
First read in udder quarter infection data

```r
# Read data
udder <- read.table("c:\udderinfect.dat", header = T, skip = 2)
coword <- order(udder$cowid)
cowid <- udder$cowid[coword]
timeto <- 4*(round(udder$timek[coword]*365.25/4))/365.25
stat <- udder$censor[coword]; laktnr <- udder$LAKTNR[coword]
cluster <- as.numeric(levels(as.factor(udder$cowid)))
G <- length(cluster)
mklist <- function(clusternr) {
  list(stat[cowid == clusternr],
       timeto[cowid == clusternr], laktnr[cowid == clusternr])
}
clus <- lapply(cluster, mklist)
```
Function to fit positive stable frailty model


likelihood.posttab<-function(p){ll<-0
for (clusternr in (1:G)){
  stat<-clus[[clusternr]][[1]];time<-clus[[clusternr]][[2]]
  trt<-clus[[clusternr]][[3]];Di<-sum(stat)
  SHij<-sum(exp(p[1])*time^(exp(p[2]))*exp(p[3]*trt))
  part1<-sum(stat*log(exp(p[1])*exp(p[2])*(time^(exp(p[2])-1))*exp(p[3]*trt)))
  part2<-Di*log(exp(p[4]))+Di*(exp(p[4])-1)*log(SHij)-SHij^(exp(p[4]))
  part3<-0;if (Di==2) {part3<-log(1+((1-exp(p[4]))*SHij^(-exp(p[4])))/exp(p[4])))}
  if (Di==3) {part3<-log(1+(3*(1-exp(p[4])))*SHij^(-exp(p[4])))/exp(p[4]))+( (exp(p[4])^(-2))*(2-exp(p[4]))*(1-exp(p[4]))*SHij^(-2*exp(p[4]))))}
  if (Di==4) {part3<-log(1+(6*(1-exp(p[4])))*SHij^(-exp(p[4])))/exp(p[4]))+( (exp(p[4])^(-2))*(1-exp(p[4]))*(11-7*exp(p[4]))*SHij^(-2*exp(p[4])))+( (exp(p[4])^(-3))*(3-exp(p[4])))*SHij^(-2*exp(p[4])))+( (exp(p[4])^(-3))*(3-exp(p[4])))*SHij^(-3*exp(p[4])))}
ll<-ll+(part1+part2+part3)}

An example: udder quarter infection data in R
Results fit positive stable frailty model

\[
\begin{align*}
\lambda &= \exp(t[1]) \\
\rho &= \exp(t[2]) \\
\beta &= t[3] \\
\theta &= \exp(t[4])
\end{align*}
\]

\[
\begin{align*}
\lambda &= 0.1766017 \\
\rho &= 2.129374 \\
\beta &= 0.5374521 \\
\theta &= 0.5285567
\end{align*}
\]

An example: udder quarter infection data in R
Updating (no covariates, positive stable frailty)

- In general we have
  \[ f_U(u \mid T > t) = \frac{\exp(-uH_0(t)) f_U(u)}{\mathcal{L}(H_0(t))} \]

- Applied to the positive stable frailty we have
  \[ f_U(u \mid T > t) = \frac{\exp(-uH_0(t)) f_U(u)}{\exp(-H_0^{\theta}(t))} \]
  \[ = \exp\left(-H_0(t)\left(u - H_0^{\theta-1}(t)\right)\right) f_U(u) \]

- The updated density is not a positive stable density but is still a member of the power variance function family (see further)
Bivariate data: positive stable Laplace transform

\[ S_f(t_1, t_2) = \exp \left\{ - \left[ (-\log S_{1,f}(t_1))^{1/\theta} + (-\log S_{2,f}(t_2))^{1/\theta} \right]^{\theta} \right\} \]

\[ f_f(t_1, t_2) = \exp (-z^\theta) h_{1,f}(t_1) h_{2,f}(t_2) \frac{z^{2(\theta-1)} + (1/\theta - 1)z^{\theta-2}}{(z_1z_2)^{\theta-1}} \]

with \( z = z_1 + z_2 \) \( z_j = [-\log (S_{j,f}(t_j))]^{1/\theta} \)

\( \tau = 1 - \theta \)
\[ \zeta(t_1, t_2) = 1 + \frac{\theta - 1}{\theta \log S_f(t_1, t_2)} \]

**Fig. 4.26.** Cross ratio function for the positive stable frailty distribution with \(\tau=0.25\).
\[ \zeta(t_1, t_2) = 1 + \frac{\theta - 1}{\theta \log S_f(t_1, t_2)} \]

**Fig. 4.27.** The estimate of \( \gamma(r) \) plotted against \( r/s \) for the time to diagnosis data. Adjacent \( r_a \) values are combined in groups 1–10, 11–20, \ldots, 91–100, and the average of the \( r_a \) values within each group, divided by \( s \), is used as value to be plotted on the x-axis. The solid line is the cross ratio function estimated from the positive stable distribution frailty model.
\[ \psi(t_1, t_2) = \frac{S_f(t_1, t_2)}{S_{1,f}(t_1) S_{2,f}(t_2)} \]

**Fig. 4.25.** The ratio of the joint survival function and the product of the population survival functions (a,b) for the positive stable frailty distribution with \( \tau=0.25 \).
\[ \phi(t_1, t_2) = \frac{f_f(t_1, t_2)}{f_{1,f}(t_1) f_{2,f}(t_2)} = \frac{S_f(t_1, t_2)}{S_{1,f}(t_1) S_{2,f}(t_2)} \frac{z^{2(\theta-1)} + (1/\theta - 1)z^{\theta-2}}{(z_1 z_2)^{1/\theta-1}} \]

Fig. 4.25. The ratio of the joint density function and the product of the population density functions (c,d) for the positive stable frailty distribution with \(\tau=0.25\).
The lognormal frailty density

- Introduced by McGilchrist (1993) as
  \[ h_{ij}(t) = h_0(t) \exp \left( x_{ij}^t \beta + w_i \right) \]
  \[ w_i \sim \mathcal{N}(0, \gamma) \]

- Therefore, for the frailty \( u_i = \exp(w_i) \) we have
  \[ f_U(u) = \frac{1}{u \sqrt{2\pi \gamma}} \exp \left( -\frac{(\log u)^2}{2\gamma} \right) \]
  \[ E(U) = \exp(\gamma/2) \quad \text{Var}(U) = \exp(2\gamma) - \exp(\gamma) \]
The lognormal frailty density: pictures

**Fig. 4.36.** Lognormal density functions.
Statistical inference for the parametric lognormal frailty model

- No explicit expression for Laplace transform ... difficult to compare
- Maximisation of the likelihood is based on numerical integration of the normally distributed frailties
An example: udder quarter infection data

- See Example 4
- Binary covariate: heifer/multiparous (\( \beta \) is the heifer effect)
  - Weibull baseline hazard
  - Lognormal frailty density
- Maximising the marginal loglikelihood (using numerical integration: Gaussian quadrature – nlmixed procedure) we obtain
  - \( \hat{\lambda} = 0.317 \) (se=0.094)
  - \( \hat{\rho} = 2.490 \) (se=0.135)
  - \( \hat{\beta} = 0.460 \) (se=0.378)
  - \( \hat{\gamma} = 2.999 \) (se=0.610)
Difficult to compare with e.g. the gamma frailty model since, for the lognormal frailty model, the mean and variance both depend on $\gamma$. We therefore convert the results to the density function of the median event time.

**Fig. 4.37.** Density of median time to infection for the gamma and the lognormal frailty distribution with (a) primiparous and (b) multiparous cows.
data mast;
  infile 'c:\mastitis.dat' delimiter=',' firstobs=2;
  input LAKTNR censor rightk leftk cowid timek;
  if leftk=0 then leftk=0.001;
  if censor=1 then midp=(leftk+rightk)/2;
  if censor=0 then midp=rightk;y=1;

proc nlmixed data=mast qpoints=10 cov;
  ebetaxb = exp(beta1*LAKTNR + b);
  G_t = exp(-lambda*ebetaxb*(midp**g));
  f_t = (lambda*ebetaxb*g*midp**(g-1))*exp(-lambda*ebetaxb*(midp**g));
  if (censor=1) then lik=f_t;
  else if (censor=0) then lik=G_t;
  llik=log(lik);
  model y~general(llik);
  random b~normal(0,theta) subject =cowid;
run;
The semiparametric frailty model
Overview

- The semiparametric frailty model
- The marginal likelihood ... a problem
- The EM-algorithm for the semiparametric frailty model
- The modified EM-algorithm
- An example: the DCIS (EM)
- The penalised partial likelihood approach for the semiparametric frailty model
- An example: the DCIS (PPL: normal, resp. loggamma, random effect)
- The performance of the PPL approach in estimating the heterogeneity parameter
The semiparametric frailty model

- A semiparametric frailty model is a model with unspecified (nonparametric) baseline hazard $h_0(t)$

$$h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t \beta)$$

with $u_i \sim f_U(u)$ and $h_0(t)$ unspecified

- For $f_U(.)$ we will consider the one-parameter gamma and the lognormal frailty
The marginal likelihood ... a problem

- We first discuss the one-parameter gamma frailty
- For parametric frailty models with gamma frailty distribution we maximised the marginal log likelihood obtained from cluster contributions of form

\[
L_{\text{marg},i}(\xi) = \frac{\Gamma(d_i + 1/\theta) \prod_{j=1}^{n_i} (h_0(y_{ij}) \exp(x_{ij}^t \beta))^\delta_{ij}}{\left(\frac{1}{\theta} + \sum_{j=1}^{n_i} H_0(y_{ij}) \exp(x_{ij}^t \beta)\right)^{1/\theta + d_i} \theta^{1/\theta} \Gamma(1/\theta)}
\]

- For the semiparametric frailty model, this is no longer possible since the baseline hazard is unspecified
Semiparametric survival models (Cox models) are typically fitted through partial likelihood maximisation.

We will study two techniques to fit semiparametric frailty models, which combine the partial likelihood maximisation idea (classical approach for the Cox model) with the fact that, since the random frailty terms are unobserved, we have incomplete information (where the EM approach can help):

- The EM-algorithm
- The penalised partial likelihood approach
The EM algorithm for the semiparametric frailty model

- The general idea is as follows
  - The EM-algorithm iterates between an Expectation and Maximisation step
  - In the Expectation step, expected values for the frailties are obtained, conditional on the observed information and the current parameter estimates
  - In the Maximisation step, the expected values for the frailties are considered to be known (fixed offset terms), and partial likelihood is used to obtain new estimates
The full likelihood

Consider the full loglikelihood for observed information \( z \) (the \( y_{ij} \)'s and \( \delta_{ij} \)'s) and unobserved information \( u \)

\[
l_{full}(h_0(.), \theta, \beta) = \log f(z, u \mid h_0(.), \theta, \beta)
\]

\[
= \log f(z \mid h_0(.), \beta, u) + \log f(u \mid \theta)
\]

\[
= l_{full,1}(h_0(.), \beta) + l_{full,2}(\theta)
\]

\[
l_{full,1}(h_0(.), \beta) = \sum_{s}^{s} \sum_{n_i}^{n_i} \left[ \delta_{ij} \log \left( h_0(y_{ij})u_i \exp(x_{i,j}^t\beta) \right) - H_0(y_{ij})u_i \exp(x_{i,j}^t\beta) \right]
\]

\[
l_{full,2}(\theta) = \sum_{i=1}^{s} \log f_U(u_i)
\]
Maximisation step using $l_{\text{part},1}(\beta)$ instead of $l_{\text{full},1}(h_0(.), \beta)$

- Replace $u_i$ and $\log u_i$ in $l_{\text{full},1}(h_0(.), \beta)$ by expected values $E_{(k)}(U_i)$ and $E_{(k)}(\log U_i)$ obtained in iteration $k$

- Profile $l_{\text{full},1}(h_0(.), \beta)$ to a partial likelihood

$$l_{\text{part},1}(\beta) = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{i,j} \left[ F_{(k)}(\log U_i) + x_{i,j}^t \beta \right]$$

$$- \log \left( \sum_{q,w \in R(y_{i,j})} E_{(k)}(U_q) \exp \left( x_{q,w}^t \beta \right) \right)$$

with $R(y_{i,j})$ the risk set corresponding to $y_{i,j}$

- Maximise $l_{\text{part},1}(\beta)$ to obtain estimate $\beta^{(k)}$
Maximisation step for $l_{full,2}(\theta)$

- Replace $u_i$ in

$$l_{full,2}(\theta) = \sum_{i=1}^{s} \log \left( \frac{u_i^{1/\theta - 1} \exp(-u_i/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)} \right)$$

by expected values $E_{(k)}(U_i)$ and maximise wrt $\theta$ to obtain $\theta^{(k)}$

With $y_{(1)} < \ldots < y_{(r)}$ the ordered event times ($r$ is the number of different events) and $N_{(l)}$ the number of events at time $y_{(l)}$, $l = 1, \ldots, r$, obtain the Nelson-Aalen estimator for the baseline hazard with the frailties as fixed offset terms

$$H_0^{(k)}(t) = \sum_{y_{(l)} \leq t} h_0^{(k)} ; \quad h_0^{(k)} = \frac{N_{(l)}}{\sum_{q \in R(y_{(l)})} E_{(k)}(U_q) \exp(x_q^t \beta^{(k)})}$$
The expectation step

We need expressions for \( E_{(k)}(U_i) \) and \( E_{(k)}(\log U_i) \)

- Obtain expectation \( E_{(k+1)}(U_i) = E_{\zeta(k)}(U_i \mid z) \)
  conditional on current parameter estimate

\[ \zeta^{(k)} = \left( h_0^{(k)}(.), \theta^{(k)}, \beta^{(k)} \right) \text{ and } z \]

- The conditional density of \( u_i \) is (Bayes)

\[
\begin{align*}
   f_{U}(u_i \mid z) &= \frac{L_i(h_0(\cdot), \beta \mid u_i) f_{U}(u_i)}{\int_0^\infty L_i(h_0(\cdot), \beta \mid u_i) f_{U}(u_i) \, du_i} \\
   &= \frac{L_i(h_0(\cdot), \beta \mid u_i) f_{U}(u_i)}{L_{\text{marg},i}(h_0(\cdot), \theta, \beta)}
\end{align*}
\]
Working out the previous expression we get

\[ f_U (u_i \mid z) = \]

\[
\frac{u_i^{d_i+1/\theta-1} \exp \left( -u_i \left( \frac{1}{\theta} + H_{x_i,c} (y_i) \right) \right) \left( \frac{1}{\theta} + H_{x_i,c} (y_i) \right)^{d_i+1/\theta}}{\Gamma \left( d_i + \frac{1}{\theta} \right)}
\]

with \( H_{x_i,c} (y_i) = \sum_{j=1}^{n_i} H_{x_{ij},c} (y_{ij}) \)

This corresponds to a gamma distribution with parameters \( (d_i + 1/\theta) \) and \( (1/\theta + H_{x_i,c} (y_i)) \)
Therefore we have

\[ E_{(k+1)}(U_i) = \frac{(d_i + 1/\theta^{(k)})}{1/\theta^{(k)} + H_{x_i, c}^{(k)}(y_i)} \]

Furthermore, \( \log(U_i) \) has a loggamma distribution and thus, with \( \psi \) the digamma function

\[ E_{(k+1)}(\log U_i) = \psi \left( d_i + 1/\theta^{(k)} \right) - \log \left( 1/\theta^{(k)} + H_{x_i, c}^{(k)}(y_i) \right) \]
Asymptotic variances of estimates are obtained as entries of the inverse of the observed information matrix obtained from the marginal loglikelihood

\[
-\frac{\partial^2 l_{marg}}{\partial \theta^2} = \sum_{i=1}^{s} \left[ \sum_{l=0}^{d_i-1} -\frac{1 + 2l\theta}{(l\theta^2 + \theta)^2} + \frac{2\log \theta - 3}{\theta^3} + \frac{2}{\theta^3} \log \left( H_{x_{i,c}}(y_i) + 1/\theta \right) \right. \\
+ \left. \frac{3/\theta + d_i + (4 + 2\theta d_i) H_{x_{i,c}}(y_i)}{(\theta + \theta^2 H_{x_{i,c}}(y_i))^2} \right] \\
\]

\[
-\frac{\partial^2 l_{marg}}{\partial h_{0c} \partial h_{0d}} = \sum_{i=1}^{s} \frac{-(d_i + 1/\theta) \sum_{y_{ij} \geq y(c)} \exp (x_{ij}^t \beta) \sum_{y_{ij} \geq y(d)} \exp (x_{ij}^t \beta)}{(H_{x_{i,c}}(y_i) + 1/\theta)^2} + I(c = d) \frac{N(c)}{(h_{0c})^2}
\]
Fig. 5.1. The EM algorithm for the semiparametric gamma frailty model. $\hat{\beta}_{IW M}$ are the estimates for the regression coefficients in the classical Cox regression model (without frailties). For further explanation, see text.
The modified EM-algorithm

- Convergence rate can be improved by using modified profile likelihood (Nielsen et al., 1992)

- The algorithm consists of two iteration levels
  - Outer loop: maximisation for $\theta$
  - Inner loop: maximisation for all other parameters conditional on chosen $\theta$ value
**Fig. 5.2.** The modified profile likelihood EM algorithm for the semiparametric gamma frailty model. $\hat{\beta}_{IW}M$ are the estimates for the regression coefficients in the classical Cox regression model (without frailties). For further explanation, see text.
An example: the DICS (EM)

- See Example 6
- We fit the semiparametric gamma frailty model
- Ductal Carcinoma in Situ (DCIS) is a type of benign breast cancer
  - Breast conserving therapy with/without radiotherapy
  - Time to event: time to local recurrence
  - Large number of clusters (46) with small cluster size \( N = \sum_{i=1}^{46} n_i = 1010 \)
- Statistical analysis (using the SAS macro of Klein and Moeschberger (1997))

Radiotherapy effect \( \hat{\beta} = -0.63 \) (se=0.17)  
\( \hat{HR} = 0.53 \) [0.38;0.71]

Heterogeneity \( \hat{\theta} = 0.086 \) (se=0.80)

---

**Fig. 5.3.** The survival functions for the treated and untreated group of the ductal carcinoma in situ trial based on the Nelson–Aalen estimator in the semiparametric gamma frailty model in a centre with frailty equal to one.
DCIS and the semiparametric gamma frailty model in SAS (Klein)

%macro gamfrail(mydata,factors,initbeta,convcrit,leftend,stepsize,options,data1, data2,data3);
use &mydata;
...
log(gamma((1/thetacr)+di[,i]))
....

- Macro works for data with clusters with few events, such as DCIS
- Does not work when clusters have many events, as gamma((1/thetacr)+di[,i])) gets very large
- SAS macro requires IML
The PPL approach starts from model

$$h_{ij}(t) = h_0(t) \exp(x_{ij}^t \beta + w_i)$$

with $\exp(w_i) = u_i$ the frailty with variance $\theta$ and $w_i$ the random effect with variance $\gamma$.

The two main ideas are

- Random effects are considered to be just another set of parameters
- For values of $w_i$ far away from the mean (zero) the penalty term has a large negative contribution to the full data (partial) loglikelihood.
The PPL is

\[ l_{ppl}(\gamma, \beta, w) = l_{part}(\beta, w) - l_{pen}(\gamma, w) \]

with

\[ l_{pen}(\gamma, w) = -\sum_{i=1}^{s} \log f_W(w_i) \]

\[ l_{part}(\beta, w) = \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{ij} \left[ \eta_{ij} - \log \left( \sum_{qw \in R(y_{ij})} \exp(\eta_{qw}) \right) \right] \]

with \[ \eta_{ij} = x_{i,j}^t \beta + w_i \]
Maximisation step of the PPL

- An iterative procedure using an inner and an outer loop

**Inner loop:** Newton Raphson procedure to maximise
\[ l_{ppl}(\gamma, \beta, \mathbf{w}) \] for \( \beta \) and \( \mathbf{w} \) (BLUP’s) given provisional value for \( \gamma \).

**Outer loop:** Obtain new value for \( \gamma \) using the current estimates for \( \beta \) and \( \mathbf{w} \) in the iteration process. How this step works depends on the frailty density used in the model. We will consider the outer loop for random effects with a normal distribution and random effects with a loggamma distribution.
The inner loop of the PPL approach

Some notation

- $l(k)$ denotes the outer (inner) loop index
- $\gamma^{(l)}$ is the estimate for $\gamma$ at iteration $l$
- $\beta(l,k)$ and $w(l,k)$ are estimates at iteration $k$ given $\gamma^{(l)}$

At iteration step $k$, the estimate is

$$\begin{bmatrix}
\beta(l,k) \\
w(l,k)
\end{bmatrix}
= \begin{bmatrix}
\beta(l,k-1) \\
w(l,k-1)
\end{bmatrix}
- V \begin{bmatrix}
0 \\
(\gamma^{(l)})^{-1}w(l,k-1)
\end{bmatrix}
+ V \begin{bmatrix}
X & Z
\end{bmatrix} dl_{part}(\beta, w)
\frac{d\eta}{d\eta}
$$

with $\eta = (\eta_{11}, \ldots, \eta_{sn_s})$ and $\eta_{ij} = x_{ij}^t \beta + w_i$
and

\[ V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} \]

the inverse of matrix

\[ A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \]

\[
= \begin{bmatrix} X^t \\ Z^t \end{bmatrix} \left( -d^2 l_{\text{part}}(\beta, w) \right) \begin{bmatrix} X & Z \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & (\gamma^{(l)})^{-1} I_G \end{bmatrix}
\]
The outer loop for the PPL approach: normal random effect

- For normal random effects

\[ l_{pen}(\gamma, w) = \frac{1}{2} \sum_{i=1}^{s} \left( \frac{w_i^2}{\gamma} + \log(2\pi\gamma) \right) \]

- An estimate for \( \gamma \) is given by REML

\[ \gamma^{(l+1)} = \frac{\sum_{i=1}^{s} \left( w_i^{(l,k)} \right)^2}{s - r} \text{ where } r = \text{trace} \left( V_{22} \right) / \gamma^{(l)} \]

- Convergence obtained when

\[ | \gamma^{(l)} - \gamma^{(l-1)} | \text{ is sufficiently small} \]
The variance estimates: normal random effect

\[ \hat{\text{Var}}(\hat{\beta}) = V_{11} \]

\[ \hat{\text{Var}}(\hat{\gamma}) = \frac{2\gamma^2}{s - 2r + \gamma^{-2}\text{trace}(V_{22}^2)} \]
The PPL approach for the semiparametric frailty model

\[
\begin{align*}
    \mathbf{w}^{(1,0)} &= 0 \\
    \gamma^{(1)} &= 1, \quad \gamma^{(0)} = 0 \\
    \mathbf{\beta}^{(1,0)} &= \hat{\mathbf{\beta}}_{IW,M}
\end{align*}
\]

\[l = 1; k = 1\]

\[k = 1, \mathbf{w}^{(l,0)} = \bar{\mathbf{w}}_{\gamma^{(l-1)}}, \quad \mathbf{\beta}^{(l,0)} = \hat{\mathbf{\beta}}_{\gamma^{(l-1)}}\]

Find new values \(\mathbf{\beta}^{(l,k)}\)

\(w^{(l,k)}\) by Newton–Raphson

New value \(\gamma^{(l)}\) by REML estimator based on \(\bar{w}_{\gamma^{(l-1)}}\)

\[k \geq 2\]

\[k = k + 1\]

\[\text{NO}\]

\[\text{YES}\]

\[
|l_{\text{pp}}(\gamma^{(l)}, \mathbf{\beta}^{(l,k)}, \mathbf{w}^{(l,k)}) - l_{\text{pp}}(\gamma^{(l)}, \mathbf{\beta}^{(l,k-1)}, \mathbf{w}^{(l,k-1)})| < \epsilon
\]

\[t = l + 1\]

\[\text{NO}\]

\[\text{YES}\]

\[|\gamma^{(l)} - \gamma^{(l-1)}| < \epsilon\]

\[\text{YES}\]

Set \(\hat{\mathbf{\beta}}_{\gamma^{(l)}} = \mathbf{\beta}^{(l,k)}, \bar{\mathbf{w}}_{\gamma^{(l)}} = \mathbf{w}^{(l,k)}\)

\[\text{STOP}\]

**Fig. 5.4.** The penalised partial likelihood maximisation algorithm for the semiparametric normal random effects model.
The outer loop for the PPL approach: loggamma random effect

- The loggamma distribution is

\[ f_W(w) = \frac{(\exp(w))^{1/\theta} \exp(-\exp(w)/\theta)}{\theta^{1/\theta} \Gamma(1/\theta)} \]

- Therefore, the penalty function is

\[ l_{pen}(\theta, w) = -\frac{1}{\theta} \sum_{i=1}^{s} (w_i - \exp(w_i)) \]
Outer loop: No REML estimator available!

Alternative: similar to modified EM-algorithm approach:
maximise marginal likelihood \( l_{\text{marg}}^{(l)} \) profiled for \( \gamma \)

- For fixed value \( \theta^{(l)} \) obtain estimates \( \hat{\beta}_{\theta^{(l)}} \) \( \hat{u}_{\theta^{(l)}} \) by maximising PPL

- Plug in estimates in Nelson-Aalen estimator to obtain estimates of the (cumulative) baseline hazard

- Use estimates to obtain marginal likelihood
The PPL approach for the semiparametric frailty model

\[ w^{(1,0)} = 0, \theta^{(0)} = 0, \theta^{(1)} = 1, \beta^{(1,0)} = \beta_{IWM} \]

\[ l = 1; k = 1 \]

\[ k = 1, w^{(1,0)} = \hat{w}_{\hat{\theta}(l-1)}, \beta^{(1,0)} = \hat{\beta}_{\hat{\theta}(l-1)} \]

Find new values \( \beta^{(l,k)} \)

\[ w^{(1,k)} \] by Newton-Raphson

\[ k \geq 2 \]

NO \[ k = k + 1 \]

YES \[ |l_{ppl}(\theta^{(l)}, \beta^{(l,k)}, w^{(l,k)}) - l_{ppl}(\theta^{(l)}, \beta^{(l,k-1)}, w^{(l,k-1)})| < \epsilon \]

NO

YES \[ t = t + 1 \]

NO

YES \[ |\theta^{(l)} - \theta^{(l-1)}| < \epsilon \]

STOP

\textbf{Fig. 5.5.} The penalised partial likelihood maximisation algorithm for the semiparametric gamma frailty model.
An example: the DCIS (PPL)

- See Example 6

- We fit the PPL with normal random effect
  - Radiotherapy effect $\hat{\beta} = -0.63$ (se=0.17)
    \[ \hat{HR} = 0.53 [0.38;0.74] \]
  - Heterogeneity $\hat{\gamma} = 0.087$ (se=0.80)

- We fit the PPL with loggamma random effect
  - Radiotherapy effect $\hat{\beta} = -0.63$ (se=0.17)
    \[ \hat{HR} = 0.53 [0.38;0.74] \]
  - Heterogeneity $\hat{\theta} = 0.087$ (se=0.80)
  - The PPL can only be used to estimate $\hat{\beta}$, since the PPL considered as a function of $\theta$ increases (see figure)
- Statistical analysis (using the SAS macro of Klein and Moeschberger (1997))

  Radiotherapy effect \( \hat{\beta} = -0.63 \) (se=0.17)

  Hazard Ratio \( \hat{HR} = 0.53 \) [0.38;0.71]

  Heterogeneity \( \hat{\theta} = 0.086 \) (se=0.80)
Fig. 5.6. Profile penalised partial likelihood for $\theta$: in (a) the penalised partial likelihood (dashed line) and the partial likelihood part (solid line); in (b) the penalty term.
DCIS and the semiparametric gamma frailty model in R

dcis <- read.table("c://dcis.dat", header=T)
dcis.gamfrail <- coxph(Surv(lrectime,lrecst) ~ trt +
  frailty(hospno, dist="gamma", eps=0.00000001), outer.max=100,
data=dcis)
>
summary(dcis.gamfrail)

coef  se(coef) se2 Chisq DF  p
tr -0.628 0.167 0.167 14.1  1.00 0.00018

frailty(hospno, dist = "g"
  exp(coef) exp(-coef) lower .95 upper .95
tr  0.534   1.87  0.385   0.741

Iterations: 10 outer, 29 Newton-Raphson
  Variance of random effect= 0.0865   I-likelihood = -1005
Degrees of freedom for terms= 1.0 7.8
Rsquare= 0.034   (max possible= 0.866 )
Likelihood ratio test= 34.8 on 8.82 df,  p=5.69e-05
Wald test            = 14.1 on 8.82 df,  p=0.112 ....

An example: DCIS in SAS 263
The performance of PPL in estimating heterogeneity

Simulation setting

- Number of clusters: 15 or 30
- Number of subjects/cluster: 20, 40 or 60
- Constant hazard: 0.07 or 0.22
- Accrual/follow-up period: 5/3 years resulting in 70% to 30% censoring
- Variance $\theta$ or $\gamma$: 0, 0.1, 0.2
- HR: 1.3
- 6500 data sets generated
Size for semiparametric gamma frailty model

$\theta = 0$

<table>
<thead>
<tr>
<th>Statistic</th>
<th>20/15</th>
<th>40/15</th>
<th>60/15</th>
<th>20/30</th>
<th>40/30</th>
<th>60/30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hazard rate 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>50%</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>95%</td>
<td>0.086</td>
<td>0.042</td>
<td>0.027</td>
<td>0.061</td>
<td>0.030</td>
<td>0.020</td>
</tr>
<tr>
<td>Mean</td>
<td>0.016</td>
<td>0.008</td>
<td>0.005</td>
<td>0.012</td>
<td>0.006</td>
<td>0.004</td>
</tr>
<tr>
<td>Size</td>
<td>0.012</td>
<td>0.016</td>
<td>0.017</td>
<td>0.014</td>
<td>0.022</td>
<td>0.021</td>
</tr>
<tr>
<td>Hazard rate 0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5%</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>50%</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>95%</td>
<td>0.041</td>
<td>0.020</td>
<td>0.014</td>
<td>0.029</td>
<td>0.015</td>
<td>0.009</td>
</tr>
<tr>
<td>Mean</td>
<td>0.007</td>
<td>0.004</td>
<td>0.003</td>
<td>0.006</td>
<td>0.003</td>
<td>0.002</td>
</tr>
<tr>
<td>Size</td>
<td>0.011</td>
<td>0.013</td>
<td>0.014</td>
<td>0.018</td>
<td>0.022</td>
<td>0.022</td>
</tr>
</tbody>
</table>
Testing for heterogeneity

- Hypothesis test: \( H_0: \theta = 0 \) vs \( H_a: \theta > 0 \)
  \( \theta \)-value in \( H_0 \) is at boundary of parameter space

- Use mixture of \( \chi^2_0 \) and \( \chi^2_1 \)

**Fig. 5.8.** The estimated density function of the likelihood ratio test statistic from the simulated data of two different settings (\( s=30, n=60, h_0=0.22 \) and \( s=15, n=20, h_0=0.07 \)). The other two lines correspond to the density functions for \( \chi^2_0 \) and \( \chi^2_1 \).
<table>
<thead>
<tr>
<th>$\theta/\gamma$</th>
<th>Statistic</th>
<th>20/15</th>
<th>40/15</th>
<th>60/15</th>
<th>20/30</th>
<th>40/30</th>
<th>60/30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma distributed frailties — Hazard rate 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>5%</td>
<td>0.000</td>
<td>0.000</td>
<td>0.015</td>
<td>0.000</td>
<td>0.027</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>0.070</td>
<td>0.078</td>
<td>0.082</td>
<td>0.086</td>
<td>0.089</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.245</td>
<td>0.200</td>
<td>0.188</td>
<td>0.207</td>
<td>0.175</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.088</td>
<td>0.087</td>
<td>0.090</td>
<td>0.093</td>
<td>0.093</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>PowerLR</td>
<td>0.254</td>
<td>0.568</td>
<td>0.802</td>
<td>0.478</td>
<td>0.881</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>PowerSim</td>
<td>0.425</td>
<td>0.751</td>
<td>0.904</td>
<td>0.654</td>
<td>0.941</td>
<td>0.991</td>
</tr>
<tr>
<td>0.2</td>
<td>5%</td>
<td>0.000</td>
<td>0.041</td>
<td>0.056</td>
<td>0.054</td>
<td>0.087</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>50%</td>
<td>0.156</td>
<td>0.169</td>
<td>0.172</td>
<td>0.180</td>
<td>0.186</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>95%</td>
<td>0.424</td>
<td>0.360</td>
<td>0.344</td>
<td>0.352</td>
<td>0.317</td>
<td>0.303</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>0.179</td>
<td>0.181</td>
<td>0.183</td>
<td>0.188</td>
<td>0.192</td>
<td>0.191</td>
</tr>
<tr>
<td></td>
<td>PowerLR</td>
<td>0.556</td>
<td>0.870</td>
<td>0.969</td>
<td>0.879</td>
<td>0.995</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>PowerSim</td>
<td>0.737</td>
<td>0.945</td>
<td>0.986</td>
<td>0.937</td>
<td>0.999</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Gamma distributed frailties — Hazard rate 0.22
| 0.1            | 5%        | 0.002  | 0.023  | 0.030  | 0.029  | 0.045  | 0.050  |
|                | 50%       | 0.080  | 0.084  | 0.086  | 0.090  | 0.093  | 0.093  |
|                | 95%       | 0.203  | 0.180  | 0.171  | 0.176  | 0.157  | 0.151  |
|                | Mean      | 0.089  | 0.091  | 0.092  | 0.095  | 0.095  | 0.096  |
|                | PowerLR   | 0.614  | 0.893  | 0.975  | 0.892  | 0.999  | 1.000  |
|                | PowerSim  | 0.770  | 0.961  | 0.990  | 0.953  | 0.999  | 1.000  |
| 0.2            | 5%        | 0.048  | 0.072  | 0.075  | 0.089  | 0.107  | 0.110  |
|                | 50%       | 0.171  | 0.176  | 0.176  | 0.186  | 0.188  | 0.190  |
|                | 95%       | 0.361  | 0.333  | 0.333  | 0.317  | 0.298  | 0.295  |
|                | Mean      | 0.184  | 0.186  | 0.186  | 0.192  | 0.193  | 0.194  |
|                | PowerLR   | 0.899  | 0.992  | 0.998  | 0.995  | 1.000  | 1.000  |
|                | PowerSim  | 0.959  | 0.998  | 1.000  | 0.998  | 1.000  | 1.000  |

267
## Power For Normal Frailty model

<table>
<thead>
<tr>
<th>$\theta/\gamma$</th>
<th>Statistic</th>
<th>20/15</th>
<th>40/15</th>
<th>60/15</th>
<th>20/30</th>
<th>40/30</th>
<th>60/30</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal distribution — Hazard rate 0.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 5%</td>
<td>0.001</td>
<td>0.008</td>
<td>0.019</td>
<td>0.005</td>
<td>0.032</td>
<td>0.040</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.079</td>
<td>0.084</td>
<td>0.086</td>
<td>0.089</td>
<td>0.092</td>
<td>0.092</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.252</td>
<td>0.200</td>
<td>0.190</td>
<td>0.204</td>
<td>0.172</td>
<td>0.163</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.095</td>
<td>0.092</td>
<td>0.093</td>
<td>0.095</td>
<td>0.095</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>0.2 5%</td>
<td>0.013</td>
<td>0.049</td>
<td>0.060</td>
<td>0.058</td>
<td>0.086</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.164</td>
<td>0.171</td>
<td>0.173</td>
<td>0.178</td>
<td>0.183</td>
<td>0.181</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.406</td>
<td>0.364</td>
<td>0.351</td>
<td>0.338</td>
<td>0.311</td>
<td>0.300</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.183</td>
<td>0.184</td>
<td>0.184</td>
<td>0.185</td>
<td>0.188</td>
<td>0.188</td>
<td></td>
</tr>
<tr>
<td>Normal distribution — Hazard rate 0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1 5%</td>
<td>0.008</td>
<td>0.027</td>
<td>0.032</td>
<td>0.031</td>
<td>0.044</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.088</td>
<td>0.087</td>
<td>0.088</td>
<td>0.092</td>
<td>0.093</td>
<td>0.093</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.211</td>
<td>0.181</td>
<td>0.174</td>
<td>0.176</td>
<td>0.157</td>
<td>0.151</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.096</td>
<td>0.093</td>
<td>0.093</td>
<td>0.096</td>
<td>0.096</td>
<td>0.096</td>
<td></td>
</tr>
<tr>
<td>0.2 5%</td>
<td>0.052</td>
<td>0.068</td>
<td>0.074</td>
<td>0.088</td>
<td>0.102</td>
<td>0.109</td>
<td></td>
</tr>
<tr>
<td>50%</td>
<td>0.171</td>
<td>0.174</td>
<td>0.175</td>
<td>0.184</td>
<td>0.187</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>95%</td>
<td>0.368</td>
<td>0.342</td>
<td>0.329</td>
<td>0.313</td>
<td>0.300</td>
<td>0.292</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.186</td>
<td>0.186</td>
<td>0.185</td>
<td>0.191</td>
<td>0.192</td>
<td>0.192</td>
<td></td>
</tr>
</tbody>
</table>

The performance of PPL in estimating heterogeneity
Bayesian analysis of the semiparametric frailty model
Overview

- Introduction
- Frailty model for grouped data and gamma process prior for the cumulative hazard
- Frailty model for observed event times and gamma process prior for the cumulative hazard
- Examples
Bayesian techniques can be used to fit frailty models.

Important references are:
- Kalbfleisch (1978): Bayesian analysis for the semiparametric Cox model
- Clayton (1991): Bayesian analysis for frailty models
- Ibrahim et al. (2001): Bayesian survival analysis (book)
Metropolis algorithm: sampling from the joint (multivariate) posterior density of the parameters of interest

For semiparametric frailty model: a problematic approach since hazard rate at each event time is considered as a parameter

→ high dimensional sampling problem
  (Metropolis no longer efficient)

Gibbs sampling: sampling from posterior density of each parameter conditional on the other parameters (iterative procedure)
Frailty model for grouped data and gamma process prior for $H(t)$

- Consider the model $h_{ij}(t) = h_0(t)u_i \exp(x_{ij}^t \beta)$ with $u_i$ actual value of $U_i \sim \text{Gamma}(1/\theta, 1/\theta)$

- We need the conditional posterior density of each parameter (conditional on all other parameters). We therefore need
  - the list of all parameters involved
  - the prior densities of the parameters
  - the (conditional) data likelihood

- Grouped data: time axis is partitioned in $z$ disjoint intervals $(L(0), L(1)], \ldots, (L(z-1), L(z)]$

  with $(L(0), L(1)] \equiv (0, L(1)]$
For a particular interval \((L_{(k-1)}, L_{(k)})\) and a particular subject three situations are possible

(i) the subject experienced the event in \((L_{(k-1)}, L_{(k)})\)

(ii) the subject did not experience the event in \((L_{(k-1)}, L_{(k)})\) and is still at risk at \(L_{(k)}\)

(iii) the subject is no longer in the study at time \(L_{(k)}\) and is lost to follow-up in the interval \((L_{(k-1)}, L_{(k)})\)

The increase of the cumulative (unspecified) baseline hazard in \((L_{(k-1)}, L_{(k)})\) is

\[
h_{(k)} = H_0(L_{(k)}) - H_0(L_{(k-1)})
\]

For these increments (considered as parameters) we will assume an independent increments gamma process prior (see further)
The ‘parameters’ of interest are

\[ \beta \] the vector with the regression parameters
\[ u_i, i = 1, \ldots, s \]
\[ h(k), k = 1, \ldots, z \]
\[ \theta \] an hyperparameter

To see how the conditional data likelihood for grouped data is constructed we look at a concrete example

Frailty model for grouped data
Fig. 5.14. Survival data for two clusters with three subjects with: (a) information on the actual times to event/censoring; (b) grouped data information given prespecified intervals. Information on event/censoring times is denoted as ⋄/▲, respectively.
The likelihood contributions of the six subjects are

\[ L(y_{a3}) = 1 \]

\[ L(y_{b2}) = 1 - \exp \left( -u_b \exp(x_{b2}^t \beta) h_{(1)} \right) \]

\[ L(y_{b1}) = \exp \left( -u_b \exp(x_{b1}^t \beta) h_{(1)} \right) \left[ 1 - \exp \left( -u_b \exp(x_{b1}^t \beta) h_{(2)} \right) \right] \]

\[ L(y_{a1}) = \exp \left( -u_a \exp(x_{a1}^t \beta) h_{(1)} \right) \]

\[ L(y_{b3}) = \exp \left( -u_b \exp(x_{b3}^t \beta) \left( h_{(1)} + h_{(2)} \right) \right) \]

\[ L(y_{a2}) = \exp \left( -u_a \exp(x_{a2}^t \beta) \left( h_{(1)} + h_{(2)} \right) \right) \left[ 1 - \exp \left( -u_a \exp(x_{a2}^t \beta) h_{(3)} \right) \right] \]
General expression for the conditional data likelihood is

- in terms of the contributions of the subjects

\[
\prod_{i=1}^{s} \prod_{j=1}^{n_i} \exp \left( -u_i \exp(x_{ij}^t \beta) \sum_{k : L(k) < y_{ij}} h(k) \right) \\
\times \left[ 1 - \exp \left( -u_i \exp(x_{ij}^t \beta) h(m, ij) \right) \right]^{\delta_{ij}}
\]

with \( m, ij = \min\{k : L(k) \geq y_{ij} \} \)

- in terms of the intervals

\[
\prod_{k=1}^{z} \left\{ \exp \left( -h(k) \sum_{q \wedge y_{q \wedge} > L(k)} u_q \exp(x_{q \wedge}^t \beta) \right) \\
\times \prod_{q \wedge : L(k-1) < y_{q \wedge} \leq L(k)} \left[ 1 - \exp \left( -h(k) u_q \exp(x_{q \wedge}^t \beta) \right) \right]^{\delta_{ij}} \right\}
\]
This is different from previous survival expressions since we only specify cumulative hazard increments (and not the baseline hazard itself). The expression resembles the likelihood used for interval-censored data.
Specification of the prior densities (distributions)

- Independent increments gamma process for the cumulative baseline hazard

\[ h_{(k)} \sim \text{Gamma} \left( c \left( H_0^* \left( L_{(k)} \right) - H_0^* \left( L_{(k-1)} \right) \right), c \right) \]

with \( H_0^* \) an increasing continuous function such that \( H_0^*(0) = 0 \) and independence between the increments contained in \( \mathbf{h} = (h_{(1)}, \ldots, h_{(z)})^t \)

For the cumulative baseline hazard we have

\[ H_0 \left( L_{(l)} \right) = \sum_{k=1}^{l} h_{(k)} \sim \text{Gamma} \left( cH_0^* \left( L_{(l)} \right), c \right) \]

with \( \mathbb{E} \left( H_0 \left( L_{(l)} \right) \right) = H_0^* \left( L_{(l)} \right) \) and \( \text{Var} \left( H_0 \left( L_{(l)} \right) \right) = H_0^* \left( L_{(l)} \right) / c \)

Note. Often \( h_0^* \) a time-constant hazard and thus

\[ H_0^* \left( L_{(l)} \right) = h_0^* L_{(l)} ; c \text{ large means variance small and therefore strong prior belief in } h_0^* \]
- \( f(\beta_j) \propto 1 \) \quad (f(\beta) = \prod_{j=1}^{p} f(\beta_j) \propto 1) \)
  improper uniform prior

- \( f(u_i) \sim \text{Gamma}(1/\theta, 1/\theta) \)

- \( \theta \sim \text{Gamma}(\eta, \mu) \) \quad Clayton (1991)

- \( \sqrt{\theta} \sim \text{Uniform}(0, A) \) \quad Gelman (2003)

Frailty model for grouped data
Derivation of posterior densities: the general principle

\[ P(A \mid B \cap C) = \frac{P(B \mid A \cap C)P(A \mid C)}{P(B \mid C)} \]

\[ f(\omega_i \mid y, \omega_{(-i)}) = \frac{f(y \mid \omega)f(\omega_i \mid \omega_{(-i)})}{f(y \mid \omega_{(-i)})} \]

with \( \omega = (h^t, \theta, \beta^t, u^t)^t \) a \((z + p + s + 1) \times 1\)-vector and \( \omega_{(-i)} \) is the \( \omega \) vector with component \( i \) deleted
Since $f(\omega_i \mid \omega_{(-i)}) = f(\omega_i)$ (we assume independence of prior densities) and since $f(y \mid \omega_{(-i)})$ does not depend on $\omega_i$

$$f(\omega_i \mid y, \omega_{(-i)}) \propto f(y \mid \omega) f(\omega_i)$$

unnormalised conditional posterior density

Notes

(i) Normalising factor often difficult to obtain

(ii) If obtained the resulting posterior density does not take a form from which it is easy to sample

(iii) An approximation for the derived likelihood expression will lead to more simple conditional posterior densities from which sampling is easy
Frailty model for event times and gamma process prior for $H(t)$

- To obtain an appropriate data likelihood in case of (censored) event times, we use ideas developed for grouped data with intervals $(0, y_{(1)}], \ldots, (y_{(r-1)}, y_{(r)}]$ with $y_{(1)} < y_{(2)} < \ldots < y_{(r)}$ the ordered event times.

**Fig. 5.14.** Survival data for two clusters with three subjects with: (b) grouped data information given prespecified intervals; (c) grouped data information with event times as interval boundaries. Information on event/censoring times is denoted as ●/▲, respectively.
Approximation for the derived likelihood. For instance, for the event time $y_{b1}$,

$$L(y_{b1}) = \exp \left( -u_b \exp(x_{b1}^t \beta) h_{(1)} \right) - \exp \left[ -u_b \exp(x_{b1}^t \beta)(h_{(1)} + h_{(2)}) \right]$$

$$= \phi(h_{(1)}) - \phi(h_{(1)} + h_{(2)})$$

$$L(y_{b1}) \approx (h_{(1)} - (h_{(1)} + h_{(2)}) \phi'(h_{(1)} + h_{(2)})$$

$$= -h_{(2)} \left[ -u_b \exp(x_{b1}^t \beta) \exp \left( -u_b \exp(x_{b1}^t \beta)(h_{(1)} + h_{(2)}) \right) \right]$$

$$= u_b h_{(2)} \exp(x_{b1}^t \beta) \exp \left( -u_b \exp(x_{b1}^t \beta)(h_{(1)} + h_{(2)}) \right)$$

For censored subjects the contribution is based on the cumulative hazard corresponding to the sum of the cumulative hazard increments of all event times before the actual censoring time (like in grouped data)
This leads to the following likelihood expression

$$
\prod_{i=1}^{s} \prod_{j=1}^{n_i} \exp \left( -u_i \exp(x_{ij}^t \beta) \sum_{k:y(k) \leq y_{ij}} h(k) \right) \left( u_i \exp(x_{ij}^t \beta) h(m,ij) \right)^{\delta_{ij}}
$$

with \( m, ij = \min\{k : y(k) \geq y_{ij} \} \)

or

$$
\prod_{i=1}^{s} \prod_{j=1}^{n_i} \prod_{k:y(k) \leq y_{ij}} (u_i h(k) \exp(x_{ij}^t \beta))^{\delta_{ij}(y(k))} \exp \left( -u_i h(k) \exp(x_{ij}^t \beta) \right)
$$

with \( \delta_{ij}(y(k)) = 1 \) if \( \delta_{ij} = 1 \) and \( y_{ij} = y(k) \)

This expression resembles the likelihood of Poisson distributed data.

Frailty model for observed event times
Prior densities: as before

Posterior densities

- regression parameters

\[
f \left( \beta_a \mid y, h, \theta, \beta_{(-a)}, u \right) \propto f \left( y \mid h, \theta, \beta, u \right)
\]

\[
= \prod_{i=1}^{s} \prod_{j=1}^{n_i} \prod_{k:y(k) \leq y_{ij}} \left( h_{(k)}u_i \exp(\mathbf{x}_{ij}^t \beta) \right)^{\delta_{ij}(y(k))} \exp \left( -h_{(k)}u_i \exp(\mathbf{x}_{ij}^t \beta) \right)
\]

this density is logconcave  adaptive rejection
sampling can be used to generate a sample
cumulative hazard increments

\[ f \left( h_{(k)} \mid y_{(k)}, h_{(-k)}, \theta, \beta, u \right) = \frac{f \left( y_{(k)} \mid h, \theta, \beta, u \right) f \left( h_{(k)} \mid h_{(-k)}, \theta, \beta, u \right)}{f \left( y_{(k)} \mid h_{(-k)}, \theta, \beta, u \right)} \]

\[ = \frac{f \left( y_{(k)} \mid h_{(k)}, \theta, \beta, u \right) f \left( h_{(k)} \right)}{f \left( y_{(k)} \mid \theta, \beta, u \right)} \]

After a tedious derivation we obtain

\[ f \left( h_{(k)} \mid y_{(k)}, \theta, \beta, u \right) = \frac{f \left( y_{(k)} \mid h_{(k)}, \theta, \beta, u \right) f \left( h_{(k)} \right)}{f \left( y_{(k)} \mid \theta, \beta, u \right)} \]

\[ = \left( c + N_{(k)} B_{(k)} \right)^{ch_{(k)} + N_{(k)}} h_{(k)}^{ch_{(k)} + N_{(k)} - 1} \]

\[ \times \exp \left( -h_{(k)} \left( c + N_{(k)} B_{(k)} \right) \right) \left( \Gamma \left( ch_{(k)}^{*} + N_{(k)} \right) \right)^{-1} \]
This is a gamma density

\[ f \left( h_{(k)} \mid y_{(k)}, \theta, \beta, u \right) = \text{Gamma} \left( ch_{(k)}^* + N_{(k)}, c + N_{(k)} B_{(k)} \right) \]

with \( B_{(k)} = \sum_{qw \in R(y_{(k)})} u_q \exp (x^t_{qw} \beta) \)

with \( R(y_{(k)}) \) the risk set at time \( y_{(k)} \)

and \( N_{(k)} \) the number of events at event time \( y_{(k)} \)

sample from a gamma density
- frailties

\[ f(u_i \mid y, \beta, h, \theta) = \text{Gamma}(d_i + 1/\theta, H_{i,c}(y_i) + 1/\theta) \]

→ sample from a gamma density

- the hyperparameter \( \theta \) with \( \text{Gamma}(\eta, \mu) \)

\[ f(\theta \mid u) \propto f(u \mid \theta) f(\theta) \]

\[ \propto \frac{\mu^\eta \theta^{-1-s/\theta} \exp(-\mu \theta - \sum_{i=1}^{s} u_i/\theta + (1/\theta - 1) \sum_{i=1}^{s} \log u_i)}{\Gamma(\eta)} \left( \frac{1}{\Gamma(1/\theta)} \right)^s \]

→ difficult density, slice sampling is needed

Frailty model for observed event times
- the hyperparameter $\theta$ with $\sqrt{\theta}$ uniform on $(0, A)$

$$f(\theta \mid u) \propto \frac{\theta^{-1/2-s/\theta} \exp \left( - \sum_{i=1}^{s} u_i/\theta + (1/\theta - 1) \sum_{i=1}^{s} \log u_i \right)}{2A (\Gamma(1/\theta))^s} I \left( \theta \in (0, A^2) \right)$$

- difficult density, slice sampling is needed

- A similar discussion can be given for the normal frailty model based on Poisson likelihood
Examples

- DCIS example (Example 6)
  We focus on $\beta$ and $\theta$ (the parameters of interest)
  Four chains: burn-in 1000 iterations, followed by 5000 iterations

- Gamma frailties

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.e.</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>-0.628</td>
<td>0.167</td>
<td>-0.962</td>
<td>-0.740</td>
<td>-0.626</td>
<td>-0.514</td>
<td>-0.303</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.216</td>
<td>0.895</td>
<td>0.005</td>
<td>0.054</td>
<td>0.109</td>
<td>0.187</td>
<td>0.451</td>
</tr>
</tbody>
</table>

Table 5.5. Summary statistics of posterior density functions based on the semi-parametric frailty model with gamma distributed frailties for ductal carcinoma in situ data set.
Fig. 5.15. The estimated univariate posterior densities of $\theta$ and $\beta$ based on the semi-parametric frailty model with gamma distributed frailties for the ductal carcinoma in situ data set.
Fig. 5.16. The trace of $\theta$ and $\beta$ based on the semiparametric frailty model with gamma distributed frailties for the ductal carcinoma in situ data set.
- Normal random effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.e.</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>$-0.628$</td>
<td>$0.170$</td>
<td>$-0.960$</td>
<td>$-0.742$</td>
<td>$-0.627$</td>
<td>$-0.513$</td>
<td>$-0.294$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$0.120$</td>
<td>$0.104$</td>
<td>$0.003$</td>
<td>$0.044$</td>
<td>$0.094$</td>
<td>$0.168$</td>
<td>$0.388$</td>
</tr>
</tbody>
</table>

**Table 5.6.** Summary statistics of marginal posterior density functions based on semiparametric frailty model with normal distributed random effects for ductal carcinoma in situ data set.
Posterior densities – normal random effects: pictures

Fig. 5.17. The estimated univariate posterior densities of $\gamma$ and $\beta$ based on the semiparametric frailty model with normal distributed random effects for the ductal carcinoma in situ data set.
Fig. 5.18. The trace of $\gamma$ and $\beta$ based on the semiparametric frailty model with normal distributed random effects for the ductal carcinoma in situ data set.
Conclusion: Problem with convergence
(for $\theta$, resp. $\gamma$)

- Udder infection data (Example 4) with normal random effects

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>s.e.</th>
<th>2.5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>97.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.411</td>
<td>0.362</td>
<td>-0.301</td>
<td>0.165</td>
<td>0.411</td>
<td>0.654</td>
<td>1.111</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>2.676</td>
<td>0.577</td>
<td>1.724</td>
<td>2.266</td>
<td>2.612</td>
<td>3.022</td>
<td>3.969</td>
</tr>
</tbody>
</table>

**Table 5.7.** Summary statistics of posterior density functions based on the semiparametric frailty model with normal distributed random effects for the udder quarter infection data set.
Fig. 5.19. The estimated univariate posterior densities of $\gamma$ and $\beta$ based on the semiparametric frailty model with normal distributed random effects for the udder infection data set.
Fig. 5.20. The trace of $\gamma$ and $\beta$ based on the semiparametric frailty model with normal distributed random effects for the udder infection data set.
Conclusion: No problem with convergence (for $\gamma$)
Multilevel and multifrailty models
Overview

- Multifrailty versus multilevel
  - Only one cluster, two frailties in cluster
    - e.g., prognostic index (PI) analysis, with random center effect and random PI effect
  - More than one clustering level
    - e.g., child mortality in Ethiopia with children clustered in village and village in district

- Modelling techniques
  - Bayesian analysis through Laplacian integration
  - Bayesian analysis through MCMC
  - Frequentist approach through numerical integration
Bayesian analysis through Laplacian integration

Consider the following model with one clustering level:

\[ h_{ij}(t) = h_0(t) \exp \left( w_{0i} + (\beta_1 + w_{1i}) x_{ij1} + x_{ij(-1)}^t \beta_{(-1)} \right) \]

with \( w_{0i} \) the random center effect,
\( x_{ij1} \) covariate information of first covariate, e.g. PI,
\( \beta_1 \) fixed effect of first covariate,
\( w_{1i} \) random first covariate by cluster interaction,
\( x_{ij(-1)}^t = (x_{ij2}, \ldots, x_{ijp}) \) other covariate information,
\( \beta_{(-1)}^t = (\beta_2, \ldots, \beta_p) \) other fixed effects,
\( \beta^t = (\beta_1, \beta_{(-1)}^t) \)
- Aim of Bayesian analysis: obtain posterior distributions for parameters of interest
- Ducrocq and Casella (1996) proposed a method based on Laplacian integration rather than Gibbs sampling
- Laplacian integration is much faster than Gibbs sampling, which makes it more suitable to their type of data, i.e., huge data sets in animal breeding, looking at survival traits
- Emphasis is placed on heritability and thus estimation of variance components
- Posterior distributions are only provided for variance components, fixed effects are only considered for adjustment
The joint posterior density is given by

\[ f(\beta, w, \gamma \mid y) \propto L(\beta, w \mid y) \times f(w \mid \gamma) \times f(\beta) \times f(\gamma) \]

- with \( w^t = (w^t_0, w^t_1) = (w_{01}, \ldots, w_{0s}, w_{11}, \ldots, w_{1s}) \)
- \( \gamma^t = (\gamma_0, \gamma_1) \)
- \( \gamma_0 \) the variance of the random centre effect
- \( \gamma_1 \) the variance of random covariate by cluster interaction

\[ L(\beta, w \mid y) \] likelihood

- furthermore, we have the prior distributions
  - random effects are independent from each other

\[ f(w \mid \gamma) = \prod_{i=1}^{s} \frac{1}{2\pi \sqrt{\gamma_0 \gamma_1}} \exp \left( -\frac{1}{2} \left( \frac{w_{0i}^2}{\gamma_0} + \frac{w_{1i}^2}{\gamma_1} \right) \right) \]
and for the other prior parameters flat priors are assumed

\[
f(\beta) \propto 1 \quad \beta_i \in \mathbb{R} \\
f(\gamma) \propto 1 \quad \gamma_0 \in [0, +\infty), \quad \gamma_1 \in [0, +\infty)
\]

We leave baseline hazard unspecified, and therefore use partial likelihood (Sinha et al., 2003)

\[
L_{part}(\beta, w \mid y) =
\]

\[
\prod_{i=1}^{s} \prod_{j=1}^{n_i} \frac{\exp \left( w_{0i} + (\beta_1 + w_{1i}) x_{ij1} + x_{i,j(-1)}^{t} \beta_{(-1)} \right)}{\sum_{q \in R(y_{ij})} \exp \left( w_{0q} + (\beta_1 + w_{1q}) x_{qw1} + x_{q,w(-1)}^{t} \beta_{(-1)} \right)} \delta_{ij}
\]

Bayesian analysis through Laplacian integration
The logarithm of the joint posterior density is then proportional to

\[
\log f (\beta, w, \gamma \mid y) \propto \sum_{i=1}^{s} \sum_{j=1}^{n_i} \delta_{ij} \left\{ w_{0i} + (\beta_1 + w_{1i}) x_{ij1} + x_{ij(-1)}^t \beta(-1) \right\} \\
- \log \left[ \sum_{q_w \in \mathcal{R}(y_{ij})} \exp \left( w_{0q} + (\beta_1 + w_{1q}) x_{qw1} + x_{qw(-1)}^t \beta(-1) \right) \right]
\]

\[
- s \log (2\pi \sqrt{\gamma_0 \gamma_1}) - \frac{1}{2} \sum_{i=1}^{s} \left( \frac{w_{0i}^2}{\gamma_0} + \frac{w_{1i}^2}{\gamma_1} \right)
\]

The posterior distribution of the parameters of interest \((\gamma_0, \gamma_1)\) can be obtained by integrating out other parameters

\[
f (\gamma \mid y) = \int \ldots \int f (\beta, w, \gamma \mid y) \, d\beta d\omega
\]
No analytical solution available for

\[ f(\gamma \mid y) = \int \ldots \int f(\beta, w, \gamma \mid y) \, d\beta dw \]

Approximate integral for fixed value \( \gamma = \gamma^* \)
by Laplacian integration

For fixed value \( \gamma^* \) we use notation \( f(\beta, w \mid y, \gamma^*) \)

First we obtain the mode of the joint posterior density function at \( \gamma = \gamma^* \)

\[ \hat{\Psi}_{\gamma^*} = \left( \hat{\beta}_{\gamma^*}^t, \hat{w}_{\gamma^*}^t \right)^t = \text{Arg max}_{\Psi} f(\Psi \mid y, \gamma^*) \]

by maximising the logarithm of the joint posterior density wrt \( \beta \) and \( w \)
(limited memory quasi-Newton method)
We first rewrite integral as
\[ \int \ldots \int f(\beta, w | y, \gamma^*) \, d\beta dw = \int \ldots \int \exp \left( \log f(\beta, w | y, \gamma^*) \right) \, d\beta dw \]
and replace \( \log f(\beta, w | y, \gamma^*) \) by the first terms of its Taylor expansion around the mode \( \hat{\Psi}_{\gamma^*} \).

The second term of the Taylor expansion cancels
\[ \left. \frac{\partial \log f(\Psi | y, \gamma^*)}{\partial \Psi} \right|_{\Psi = \hat{\Psi}_{\gamma^*}} = 0 \]

For the third term we need the Hessian matrix
\[ H_{\gamma^*} = -\left. \frac{\partial^2 \log f(\Psi | y, \gamma^*)}{\partial \Psi \partial \Psi^t} \right|_{\Psi = \hat{\Psi}_{\gamma^*}} \]
The integral is then approximately

\[ f(\gamma^* \mid y) \approx \int \ldots \int \exp \left( \log f \left( \hat{\Psi}_{\gamma^*} \mid y, \gamma^* \right) \right) \]

\[ = -\frac{1}{2} \left( \Psi - \hat{\Psi}_{\gamma^*} \right)^t H_{\gamma^*} \left( \Psi - \hat{\Psi}_{\gamma^*} \right) d\Psi \]

As \( \hat{\Psi}_{\gamma^*} \) has asymptotically a multivariate normal distribution with variance covariance matrix \( H_{\gamma^*}^{-1} \), we have

\[ \int \ldots \int (2\pi)^{-\frac{2(s+p)}{2}} \left| H_{\gamma^*}^{-1} \right|^{-1/2} \]

\[ \times \exp \left( -\frac{1}{2} \left( \Psi - \hat{\Psi}_{\gamma^*} \right)^t H_{\gamma^*} \left( \Psi - \hat{\Psi}_{\gamma^*} \right) \right) d\Psi = 1 \]
The integral can therefore be simplified to

\[ f(\gamma^* \mid y) \approx (2\pi)^{(2s+p)/2} |H_{\gamma^*}|^{1/2} f\left(\hat{\Psi}_{\gamma^*} \mid y, \gamma^*\right) \]

or on the logarithmic scale

\[ \log f(\gamma^* \mid y) \approx \text{constant} + \log f\left(\hat{\Psi}_{\gamma^*} \mid y, \gamma^*\right) - \frac{1}{2} \log |H_{\gamma^*}| \]

- Estimates for \( \gamma_0 \) and \( \gamma_1 \) are provided by the mode of this approximated bivariate posterior density.

- To depict the bivariate posterior density and determine other summary statistics by evaluating the integral on a grid of equidistant points

\[ \gamma_{ab} = (\gamma_{0a}, \gamma_{1b}) \quad a = 1, \ldots, n_a \quad b = 1, \ldots, n_b \]
The values can be standardised by computing

\[ p(\gamma_{ab} \mid y) = \frac{f(\gamma_{ab} \mid y)}{\sum_{l=1}^{n_a} \sum_{k=1}^{n_b} f(\gamma_{lk} \mid y) \varepsilon^2} \]

with \( \varepsilon \) the distance between two adjacent points

Univariate posterior densities for each of the two variance components can be obtained as

\[ p(\gamma_{0a} \mid y) \approx \frac{\sum_{k=1}^{n_b} f(\gamma_{ak} \mid y)}{\sum_{l=1}^{n_a} \sum_{k=1}^{n_b} f(\gamma_{lk} \mid y) \varepsilon} \]

Bayesian analysis through Laplacian integration
This discretised version of the posterior densities can also be used to approximate the moments.

- The first moment for $\gamma_0$ is approximated by

$$\mu_{1,\gamma_0} = \sum_{l=1}^{n_a} \gamma_0 \rho(\gamma_0 | y)$$

$$= \sum_{l=1}^{n_a} \sum_{k=1}^{n_b} \gamma_{0l} f(\gamma_{lk} | y)$$

$$= \frac{\sum_{l=1}^{n_a} \sum_{k=1}^{n_b} \gamma_{0l} f(\gamma_{lk} | y)}{\sum_{l=1}^{n_a} \sum_{k=1}^{n_b} f(\gamma_{lk} | y)}$$

- In general, the $c^{th}$ moment is given by

$$\mu_{c,\gamma_0} = \sum_{l=1}^{n_a} \gamma_{0l}^c \rho(\gamma_0 | y)$$

Bayesian analysis through Laplacian integration.
From these non-central moments, we obtain

- The mean
  \[
  E (\gamma_0) = \mu_{1,\gamma_0}
  \]

- The variance
  \[
  Var (\gamma_0) = \mu_{2,\gamma_0} - \mu_{1,\gamma_0}^2
  \]

- The skewness
  \[
  \text{Skewness} (\gamma_0) = \frac{\mu_{3,\gamma_0} - 3\mu_{1,\gamma_0}\mu_{2,\gamma_0} + 2\mu_{1,\gamma_0}^3}{(Var (\gamma_0))^{3/2}}
  \]
Bayesian analysis through Laplacian integration: Example

- Prognostic index heterogeneity for a bladder cancer multicentre trial

- Traditional method to validate prognostic index
  - Split database into a training dataset (typically 60% of the data) and validation dataset (remaining 40%)
  - Develop the prognostic index based on the training dataset and evaluate it based on the validation dataset

- Flaws in the traditional method
  - How to split the data, at random or picking complete centers?
  - This will always work if sample sizes are sufficiently large
  - It ignores the heterogeneity of the PI completely
We fit the following frailty model

\[ h_{ij}(t) = h_0(t) \exp \left( w_{0i} + (\beta_1 + w_{1i}) x_{ij1} + x_{ij1}^t \beta_{1(-1)} \right) \]

with \( \beta_1 \) the overall prognostic index effect
\( w_{0i} \) the random hospital effect
\( w_{1i} \) the random hospital* PI interaction

and consider disease free survival (i.e., time to relapse or death, whatever comes first).

Parameter estimates

\[ \hat{\beta}_1 = 0.737 \text{ (se}=0.0964) \]
\[ \hat{\gamma}_0 = 0.095 \]
\[ \hat{\gamma}_1 = 0.016 \]
Interpreting the parameter estimates

- Good prognosis in center $i$, $x_{ij} = 0$
  \[ h_{ij}(t) = h_0(t) \exp(w_{0i}) \]

- Poor prognosis in center $i$, $x_{ij} = 1$
  \[ h_{ij}(t) = h_0(t) \exp(w_{0i}) \exp(\beta_1 + w_{1i}) \]

- Hazard ratio in center $i$, $HR_i = \exp(\beta + w_{1i})$

- If for cluster $i$, $w_{1i} = 0$, the mean of the distribution
  \[ HR_i = \exp(0.737) = 2.09 \]
  with 95% CI [1.73; 2.52]

- This CI refers to the precision of the estimated conditional hazard ratio
- Bivariate posterior density for $\gamma$

**Fig. 6.1.** The bivariate posterior density of $\gamma_0$ and $\gamma_1$. 
Univariate posterior densities for $\gamma$

**Fig. 6.2.** The univariate posterior densities of $\gamma_0$ and $\gamma_1$. 
Plotting predicted random center and random interactions effects

Fig. 6.3. Predicted frailties $\exp(w_{0i})$ (a) and $\exp(\beta + w_{1i})$ (b) of the different centres.
Interpretation of variance component $\gamma_0$

- Heterogeneity of median event time quite meaningless, less than 50% of the patients had event, therefore rather use heterogeneity of five-year disease free percentage

- Fit same model with Weibull baseline hazard

$$\hat{\lambda} = 0.7182 \quad \hat{\rho} = 0.1548$$

leading to the following estimate

$$\hat{S}_i(t) = \exp \left( -\hat{\lambda} t^{\hat{\rho}} \exp (w_{0i}) \right)$$
The density of the five-year disease-free percentage using \( \hat{S}_i(t) = \exp\left( -\hat{\lambda} t^{\hat{\rho}} \exp(w_{0i}) \right) \) is then given by

\[
P(\hat{S}_i(5) \leq a) \approx P\left( -\hat{\lambda} 5^{\hat{\rho}} \exp(w_{0i}) \leq \log a \right)
\]

\[
= P\left( w_{0i} \leq \log \left( -\frac{\log a}{\hat{\lambda} 5^{\hat{\rho}}} \right) \right)
\]

\[
\approx F_N \left( \frac{1}{\sqrt{\gamma_0}} \log \left( -\frac{\log a}{\hat{\lambda} 5^{\hat{\rho}}} \right) \right)
\]

where \( 0 \leq a \leq 1 \) and \( F_N \) standard normal.

Bayesian analysis through Laplacian integration
The density of the five-year disease-free percentage

Fig. 6.4. Density of five-year disease-free percentage over centres in the good prognosis group.
Interpretation of variance component $\gamma_1$

- The hazard ratio for center $i$ is given by

$$HR_i = \exp(\beta + w_{1i})$$

- Derive the density function for this expression.

- What are the 5th and 95th quantiles for this density assuming the parameter estimates

$$\hat{\beta}_1 = 0.737 \text{ (se}=0.0964)$$

$$\hat{\gamma}_1 = 0.016$$

are the population parameters
We look for

\[ w_{1i} \sim N(0, \gamma_1) \rightarrow HR_i = \exp(\beta + w_{1i}) \sim???
\]

d. General rule

\[
y = g(x) \rightarrow f_Y(y) = f_X(g^{-1}(y)) \mid \frac{dg^{-1}(y)}{dy} \mid
\]

d. Applied to \( y = \exp(\beta + w_{1i}) \quad g^{-1}(y) = \log y - \beta \)

\[
f_{HR}(h) = \frac{1}{h \sqrt{2\pi \gamma_1}} \exp \left( \frac{-1}{2\gamma_1} (\log h - \beta)^2 \right)
\]

= lognormal with parameters \( \beta \) and \( \gamma_1 \)
The density function for $HR_i$

**Fig. 6.5.** Density of prognostic index effect over centres.
For lognormal distribution we have

\[ P(HR < h) = F_N \left( \frac{\log h - \beta}{\sqrt{\gamma_1}} \right) \]

For the \( x\% \) quantile of HR, \( h_{x\%} \), we have

\[ z_x = \frac{\log h_{x\%} - \beta}{\sqrt{\gamma_1}} \rightarrow h_{x\%} = \exp(\beta + z_x \sqrt{\gamma_1}) \]

Therefore, the 5\(^{th}\) and 95\(^{th}\) quantiles are given by

\[ h_{5\%} = \exp(0.737 - 1.645 \sqrt{0.016}) = 1.70 \]
\[ h_{95\%} = \exp(0.737 + 1.645 \sqrt{0.016}) = 2.57 \]
Consider the following model with two clustering levels

\[ h_{ijk}(t) = h_0(t) u_i z_{ij} \exp \left( x_{ijk}^t \beta \right) \]

with \( u_i \) the frailty term for cluster \( i \)

\( z_{ij} \) the frailty term for subcluster \( j \) nested in cluster \( i \)

We further assume that frailty terms are independent and gamma distributed

\[ U_i \sim \text{Gamma}(1/\theta, 1/\theta) \]

\[ Z_{ij} \sim \text{Gamma}(1/\eta, 1/\eta) \]
The conditional likelihood is then given by

\[ L_i (h_0(\cdot), \beta \mid u_i, z_{i1}, \ldots, z_{is_i}) = \prod_{j=1}^{s_i} \prod_{k=1}^{n_{ij}} (h_0(y_{ijk})u_iz_{ij} \exp(x_{ijk}^t\beta))^{\delta_{ijk}} \times \exp(-H_0(y_{ijk})u_iz_{ij} \exp(x_{ijk}^t\beta)) \]

The marginal likelihood for cluster \( i \) is

\[ L_{\text{marg},i} (h_0(\cdot), \theta, \eta, \beta) = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \prod_{j=1}^{s_i} \prod_{k=1}^{n_{ij}} [(h_0(y_{ijk})u_iz_{ij} \exp(x_{ijk}^t\beta))^{\delta_{ijk}} \times \exp(-H_0(y_{ijk})u_iz_{ij} \exp(x_{ijk}^t\beta))] \]

\[ \times \left( \frac{1}{\eta^{1/\eta} \Gamma(1/\eta)} \right)^{s_i} \prod_{j=1}^{s_i} z_{ij}^{1/\eta - 1} \exp(-z_{ij}/\eta) \]

\[ \times \frac{1}{\theta^{1/\theta} \Gamma(1/\theta)} u_i^{1/\theta - 1} \exp(-u_i/\theta) \, dz_{i1} \ldots dz_{is_i} \, du_i \]
Integrating out the frailties $\zeta_{i1}, \ldots, \zeta_{i s_i}$ analytically, we obtain

$$L_{\text{Marg},i}(h_0(\cdot), \theta, \eta, \beta) =$$

$$\prod_{j=1}^{s_i} \prod_{k=1}^{n_{ij}} \left( h_0(y_{ijk}) \exp(x_{ijk}^t \beta) \right)^{\delta_{ijk}} \frac{1}{\theta^{1/\theta} \Gamma(1/\theta)} \prod_{j=1}^{s_i} \frac{\Gamma(1/\eta + d_{ij})}{(1/\eta)^{d_{ij}} \Gamma(1/\eta)}$$

$$\times \int_0^\infty \frac{u_i^{1/\theta - 1 + d_i}}{\prod_{j=1}^{s_i} \left( 1 + \eta u_i \sum_{k=1}^{n_{ij}} H_0(y_{ijk}) \exp(x_{ijk}^t \beta) \right)^{1/\eta + d_{ij}}} \exp \left( -u_i / \theta \right) \, du_i$$

with $d_{ij}$ the number of events in subcluster $j$ nested in cluster $i$.

$d_i$ number of events in cluster $i$
Taking the logarithm we obtain the marginal log-likelihood
\[ l_{\text{marg}}(h_0(\cdot), \theta, \eta, \beta) \]

The marginal loglikelihood still contains the baseline hazard \( h_0(\cdot) \) and cumulative baseline hazard \( H_0(\cdot) \).

Hazard functions are modeled using splines (Rondeau et al., 2006)
- Cubic M-splines for baseline hazard function
- Integrated M-splines for baseline cumulative hazard

Penalty term added to marginal likelihood to control the degree of smoothness of hazard function

\[
l_{p\text{marg}}(h_0(\cdot), \theta, \eta, \beta, \kappa) = l_{\text{marg}}(h_0(\cdot), \theta, \eta, \beta) - \kappa \int_{0}^{\infty} \left( h_0^{(2)}(t) \right)^{2} dt
\]
The smoothing parameter $\kappa$ can be chosen by visual inspection or by maximising a likelihood cross-validation criterion.

The penalised marginal loglikelihood is then maximised using the Marquardt algorithm.

For evaluation of the penalised marginal loglikelihood, the frailties corresponding to the large clusters are integrated out numerically using Gaussian quadrature.

This technique is implemented in the software package ‘frailtypack’, also available in R.
Frequentist approach for multilevel frailty model: Example

- Multilevel child mortality data (Example 12)
  - Gender effect: $\hat{\beta} = -0.137$ (se=0.071)
    HR (female versus male) = 0.87 [0.76; 1.00]
  - Variance of village frailties: 0.0450 (se=0.028)
  - Variance of district frailties: 0.0073 (se=0.00007)