Statistical analysis of dependencies within insurance portfolios

Viktor Lukocius

University of Twente
Outline

- Introduction (What and how have been done)
- Development of the models
- The last model
- Data for the last model implementation
- Example of the dependence effect
Introduction

- Rare events and the independence assumption
- Alternative modelling of the aggregated sum $S$
- Dependence effect analysis on the basis of the Stop-Loss contract and Value at Risk (approximation is needed)
- Parameters estimation
- Estimation impact
Development of the models

**Model 1:** \[ S = \sum_{i=1}^{N} C_i, \]

where \( N \sim P(\lambda) \)
Development of the models

**Model 1:** \( S = \sum_{i=1}^{N} C_i, \)

where \( N \sim P(\lambda) \)

**Model 2:** \( S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{g} C_{ij}, \)

where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{g}) \)
Development of the models

**Model 1:**  
\[ S = \sum_{i=1}^{N} C_i, \]  
where \( N \sim P(\lambda) \)

**Model 2:**  
\[ S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{g} C_{ij}, \]  
where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{g}) \)

**Model 3:**  
\[ S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{G} C_{ij}, \]  
where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{\mu_G}), G_i \sim P(\mu_G) \)
Development of the models

**Model 1:**  
\[ S = \sum_{i=1}^{N} C_i, \]
where \( N \sim P(\lambda) \)

**Model 2:**  
\[ S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{g} C_{ij}, \]
where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{g}) \)

**Model 3:**  
\[ S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{G_i} C_{ij}, \]
where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{\mu_G}), G_i \sim P(\mu_G) \)

**Model 4:**  
\[ S = \sum_{i=1}^{N} C_i + \sum_{i=1}^{H} \sum_{j=1}^{G_i} C_{ij}, \]
where \( N \sim P(\lambda(1 - \epsilon)), H \sim P(\epsilon \frac{\lambda}{\mu_G}), G_i \sim P(L_i), \mu_G = \mu_L \)
The last model

- Main difference from the previous models

- Why do we need it?
  - Data heterogeneity because of high aggregation
  - Indirect illustration by the simulation of the flu epidemic inside the company

- Suggested assumptions for the mixing distribution
  - $L \sim Gamma$ (Gamma distribution)
  - $L \sim IG$ (Inverse Gaussian distribution)
Outline

• Introduction (What and how we are doing)

• Development of the models

• The last model

• Data for the Model 4 implementation

• Dependence effect example
Data for the Model 4 implementation

- **Model parameters**
  - $\lambda$, $\epsilon$
  - $\mu_C$, $\gamma_C ( = \sigma_C / \mu_C )$
  - $\mu_G$, $\gamma_L ( = \sigma_L / \mu_G )$

- **Parameters to be estimated**
  - $\lambda$, $\mu_C$, $\gamma_C$
    - Expected number of claims and claim size parameters
    - No information about the dependence is needed
    - Easy part
  - $\epsilon$, $\mu_G$, $\gamma_L$
    - Percentage of the "special" part and group size parameters
    - Information about the dependence structure is needed
"Perfect" data set example

<table>
<thead>
<tr>
<th>Claim amount</th>
<th>Group code</th>
</tr>
</thead>
<tbody>
<tr>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>92</td>
<td>1</td>
</tr>
<tr>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>88</td>
<td>2</td>
</tr>
<tr>
<td>42</td>
<td>2</td>
</tr>
<tr>
<td>68</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>85</td>
<td>0</td>
</tr>
<tr>
<td>92</td>
<td>43</td>
</tr>
<tr>
<td>12</td>
<td>43</td>
</tr>
</tbody>
</table>
Example of creating the "Group code"

<table>
<thead>
<tr>
<th>Nr</th>
<th>Date</th>
<th>Name</th>
<th>Claim amount</th>
<th>Group code</th>
<th>Incident</th>
<th>Place</th>
<th>Claim type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>d1</td>
<td>n1</td>
<td>c1</td>
<td>0</td>
<td>i7</td>
<td>p1</td>
<td>t4</td>
</tr>
<tr>
<td>2</td>
<td>d2</td>
<td>n2</td>
<td>c2</td>
<td>0</td>
<td>i2</td>
<td>p2</td>
<td>t2</td>
</tr>
<tr>
<td>3</td>
<td>d3</td>
<td>n3</td>
<td>c3</td>
<td>d3.i3.p3</td>
<td>i3</td>
<td>p3</td>
<td>t1</td>
</tr>
<tr>
<td>4</td>
<td>d3</td>
<td>n4</td>
<td>c4</td>
<td>d3.i3.p3</td>
<td>i3</td>
<td>p3</td>
<td>t2</td>
</tr>
<tr>
<td>5</td>
<td>d3</td>
<td>n5</td>
<td>c5</td>
<td>d3.i3.p3</td>
<td>i3</td>
<td>p3</td>
<td>t1</td>
</tr>
<tr>
<td>6</td>
<td>d4</td>
<td>n6</td>
<td>c6</td>
<td>0</td>
<td>i4</td>
<td>p4</td>
<td>t8</td>
</tr>
<tr>
<td>7</td>
<td>d4</td>
<td>n7</td>
<td>c7</td>
<td>0</td>
<td>i1</td>
<td>p1</td>
<td>t3</td>
</tr>
<tr>
<td>8</td>
<td>d6</td>
<td>n8</td>
<td>c8</td>
<td>0</td>
<td>i6</td>
<td>p4</td>
<td>t3</td>
</tr>
<tr>
<td>9</td>
<td>d7</td>
<td>n9</td>
<td>c9</td>
<td>d7.i7.p2</td>
<td>i7</td>
<td>p2</td>
<td>t5</td>
</tr>
<tr>
<td>10</td>
<td>d7</td>
<td>n10</td>
<td>c10</td>
<td>d7.i7.p2</td>
<td>i7</td>
<td>p2</td>
<td>t4</td>
</tr>
<tr>
<td>11</td>
<td>d8</td>
<td>n11</td>
<td>c11</td>
<td>0</td>
<td>i2</td>
<td>p1</td>
<td>t2</td>
</tr>
</tbody>
</table>
How do we distinguish special claims?
**Dependence effect example**

\[ C, L \sim \text{Gamma}, \lambda = 400, \mu_C = 10^5, \gamma_C = 0.05, \epsilon = 0.03, \mu_G = 20, \gamma_L = 1 \]
Questions