Advanced Claims Reserving Methods

Michal Gerthofer

Aktuársky seminár 18.5.2018

ACTUARIA

Introduction

ACTUARIA

Outline

Advanced Claims Reserving Methods - Motivation

- **Complex reserving problem** in non-life insurance
- Claims reserve is the **biggest balance sheet item** for non-life insurance company
- Utilization of the **micro data** within insurance
- **Regulatory framework**
    - Stochastic modeling
    - Utilization of available data within the company

ACTUARIA

Outline

Advanced Claims Reserving Methods - Objective

- Chain ladder - **distribution free** approach $\rightsquigarrow$ **does not provide distributional properties**
  - Chain ladder Bootstrap - **mimics unknown distribution**
  - **Watch out!** There is also **second stage of the Bootstrap** - simulation of the process error. Distribution must be defined anyway
- **Stochastic models** based on aggregated data able to cope with **dependent variables**
  - Generalized Estimation Equation
  - Generalized Linear Mixed Models
- Development of the **micro-model**
  - Utilization of **all data about the individual claims**
  - Increased **precision of prediction**
  - **Machine learning methods** used
    - Neural networks
    - Hurdle models

Outline

Linear Mixed Models

Let's assume that Y_{it} follows Linear Mixed model

$$Y_{it} = \mathbf{X}_{it}{}^{T}\beta + u_{it}$$

- The breakup of residuals:

$$u_{it} = b_i + \varepsilon_{it}$$

- In a model with random effects, simple correlation structure is introduced:

$$\begin{gathered}
Y_{it} = \mathbf{X}_{it}{}^{T}\beta + b_i + \varepsilon_{it}, \\
b_i \sim iid\left(0, \sigma_b^2 > 0\right), \\
\varepsilon_{it} \sim iid\left(0, \sigma_\varepsilon^2 > 0\right), \\
E\left(\varepsilon_{it} b_j\right) = 0, \quad \forall\ i, j \ and\ t, \qquad E\left(b_i b_j\right) = 0, \quad i \neq j
\end{gathered}$$

Linear Mixed Models

- Introducing the correlation:

$$\begin{aligned} Cov\left(u_{it}, u_{is}\right) &= Cov\left(\varepsilon_{it} + b_i, \varepsilon_{is} + b_i\right) \\ &= Var\left(b_i\right) + Cov\left(\varepsilon_{it}, \varepsilon_{is}\right) = \\ &= \begin{cases} \sigma_b^2 & t \neq s, \ \forall i, \\ \sigma_b^2 + \sigma_\varepsilon^2 & t = s, \ \forall i. \end{cases} \end{aligned}$$

$$Corr\left(u_{it}, u_{js}\right) = \begin{cases} \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\varepsilon^2} & t \neq s, \ i = j, \\ 1 & t = s, \ i = j, \\ 0 & \text{otherwise} \end{cases}$$

- LMM (in general):

$$Y_{it} = \mathbf{X}_{it}{}^T \beta + \mathbf{Z}_{it}{}^T \mathbf{b}_i + \varepsilon_{it}$$

Bayes Approach in Linear Mixed Models

- Prediction of random variable translates into the problem of predicting the conditional mean of $\mathbf{b}_i$, given the vector of response $\mathbf{Y}_i$. Using properties of joint multivariate normal distribution of $\mathbf{b}_i$ and ε_i, it can be written as $E\left(\mathbf{b}_i|\mathbf{Y}_i\right) = \mathbf{G}\mathbf{Z}_i^T\mathbf{\Sigma}_i^{-1}\left(\mathbf{Y}_i - \mathbf{X}_i\widehat{\beta}\right)$, where $Cov\left(\mathbf{Y}_i\right) = \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \mathbf{R}_i$

$$\widehat{\mathbf{Y}}_i = \mathbf{X}_i\widehat{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i, \widehat{\mathbf{Y}}_i = \left(\widehat{\boldsymbol{R}}_i\widehat{\mathbf{\Sigma}}_i^{-1}\right)\mathbf{X}_i$$

$$\widehat{\mathbf{Y}}_i = \left(\widehat{\boldsymbol{R}}_i\widehat{\mathbf{\Sigma}}_i^{-1}\right)\mathbf{X}_i\widehat{\beta} + \left(\mathbf{I}_{n_i} - \widehat{\boldsymbol{R}}_i\widehat{\mathbf{\Sigma}}_i^{-1}\right)\mathbf{Y}_i,$$

- where $\mathbf{R}_i = Cov(\varepsilon_i) = \sigma_\varepsilon^2\mathbf{I}_{n_i}$, $\mathbf{G} = Cov(\mathbf{b}_i)$ and $\mathbf{\Sigma}_i = Cov(\mathbf{Y}_i)$
- If the within-subject variability $\mathbf{R}_i$ is relatively large compared to the between-subject variability $\mathbf{\Sigma}_i$, more weight is given to $\mathbf{X}_i\widehat{\beta}$ than to the i-th observed response

Generalization of Linear Mixed Models

- Given b_i, components of $\mathbf{Y}_i$ are conditionally independent, with density belonging to the **exponential family of distributions**:
$$f(y_{it}|\mathbf{b}_i) = \exp\left\{\frac{Y_{it}\theta_{it} - b(\theta_{it})}{\varphi} + c(Y_{it},\ \varphi)\right\}$$
- Then the **conditional mean** of Y_{it} given $\mathbf{b}_i$ is
$$\mu_{it} \equiv E(Y_{it}|\mathbf{b}_i) = b'(\theta_{it})$$
and the **conditional variance** of Y_{it} given $\mathbf{b}_i$ has the following form:
$$Var(Y_{it}|\mathbf{b}_i) = \varphi b''(\theta_{it}) \equiv \varphi V(\mu_{it})$$
- Furthermore, it is assumed that μ_{it} is related to the **linear predictor**
$$\eta_{it} = \mathbf{X}_{it}{}^T\boldsymbol{\beta} + \mathbf{Z}_{it}^T\boldsymbol{b}_i$$
through the **link function** $g(\mu_{it}) = \eta_{it}$
- Conditional Y_{it} given $\mathbf{b}_i$ satisfies the GLM and the inclusion of $\mathbf{b}_i$ in all η_{it} **brings in correlation** between $Y_{i1}, \ldots, Y_{in_i}$

Outline

Basics of Generalized Estimation Equations

- The main idea behind GEE is to generalize and extend the usual likelihood equations from GLM by including the **covariance matrix** of the vector $\mathbf{Y}$
- The biggest advantage of this model:
  - **No need to specify the whole distribution** of the response
- On the other hand, the following have to be defined:
  - The mean structure
  - The **mean-variance relationship**
  - Specification of the covariance structure
- The first two conditions are similar to the GLM

Definition of Generalized Estimation Equations

- Unbalanced design with independence between individuals $\mathbf{Y}_i,\ i = 1, 2, \ldots,\ N$ is assumed like in GLMM.
- Denote **expected value** of response

$$\mu_{it} \equiv E(Y_{it}),$$

which depends on covariates, $\mathbf{X}_{it}$ as follows

$$g(\mu_{it}) = \eta_{it} = \mathbf{X}_{it}^T \boldsymbol{\beta}$$

- It is also assumed that the variance of each Y_{it} depends on the mean according to

$$Var(Y_{it}) = \varphi V(\mu_{it}),$$

where $V(\cdot)$ is a **known variance function** and $\varphi > 0$ is a **scale or dispersion parameter**, that can be known or may need to be estimated

Impact of Correlation Matrix

- Furthermore, correlation between components of $\mathbf{Y}_i$ is represented by a ***working correlation matrix***

$$\mathbf{C}_i \equiv \mathbf{C}_i(\boldsymbol{\alpha}),$$

  where $\boldsymbol{\alpha}$ is $s \times 1$ vector of unknown parameters
- The name "working" comes from the fact that the **structure of $\mathbf{C}_i$ does not need to be correctly specified and asymptotic properties of estimate still hold**
- The corresponding working covariance matrix for i-th subject can be constructed as the product of standard deviations and working correlation matrix

$$\mathbf{V}_i = \varphi \mathbf{A}_i^{1/2} \mathbf{C}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2},$$

  where $\mathbf{A}_i$ is diagonal matrix with $V(\mu_{it})$ on the diagonal

Outline

ACTUARIA

Claims Reserving using Generalized Linear Mixed Models

- Due to interpretation and useful properties **log link function** was chosen
- Three various **distribution functions**:
  - Gamma
  - Poisson
  - Negative binomial
- **Linear predictor**:

$$log[E(Y_{it}|b_i)] = log(\mu_{it}) = \eta_{it} = \beta_0 + b_i + \beta_t,$$

  - β_0 is intercept
  - b_i is random effect (**Gaussian distribution with zero mean**)
  - β_t captures the impact of change for particular development year
  - b_i is not estimated but **predicted**

ACTUARIA

Claims Reserving using Generalized Estimation Equations

- **Log link** used as well as in GLMM
- **Variance function**:

$$V(\mu_{it}) = \begin{cases} 1, \\ \mu_{it}, \\ \mu_{it}^2 \end{cases}$$

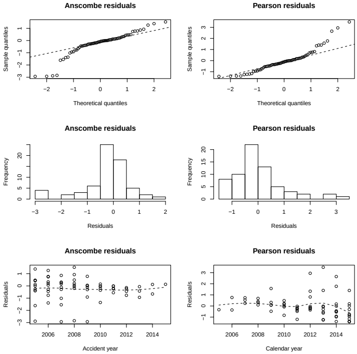
- **Working correlation matrix**: independent, AR(1) and exchangeable
  - Low number of parameters
- **Linear predictor**:

$$log[E(Y_{it})] = log(\mu_{it}) = \eta_{it} = \gamma + \alpha_i + \beta_t,$$

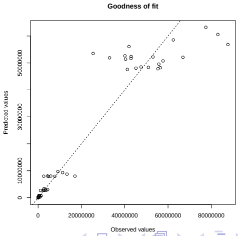
  - $\alpha_1 = \beta_1 = 0$ and α_i represents effect of accident year i, β_t effect of development year t

Residual Diagnostic of Gamma Model

- Based on residual diagnostic **Gamma model** was chosen
- **Predictions** of the models was further **compared with new diagonal** see Table below



Distribution	Gamma	Poison	Negative binomial
RBNS + IBNR (ths.)	84 926	91 485	88 706
Diff. real./est. diag (ths.)	-37 721	-44 245	-41 875
MSEP (bil.)	866	1 466	1 297

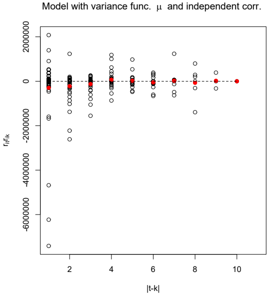


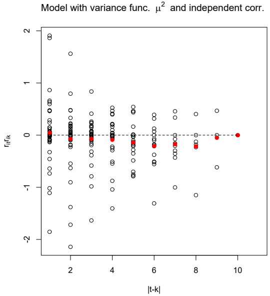
Residual Diagnostic for Generalized Estimation Equations

- Choice of **working correlation structure**
  - If the mean structure is correct, the following should hold for Pearson residuals

$$
\begin{aligned}
r_{i,t} &\approx 0 \\
Var(r_{i,t}) &\approx \varphi \\
r_{i,t}r_{i,k} &\approx \varphi \{\mathbf{C}_i\}_{t,k}, \quad i = 1, 2, \ldots, N, \quad t \neq k \in \{1, 2, \ldots, n_i\}
\end{aligned}
$$

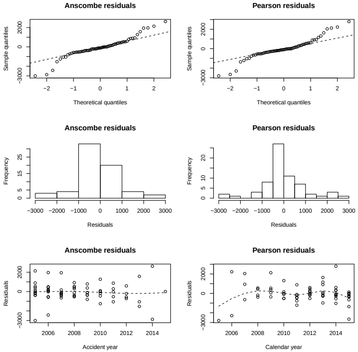
Not possible to clearly identify matrix correlation structure based on Figures below

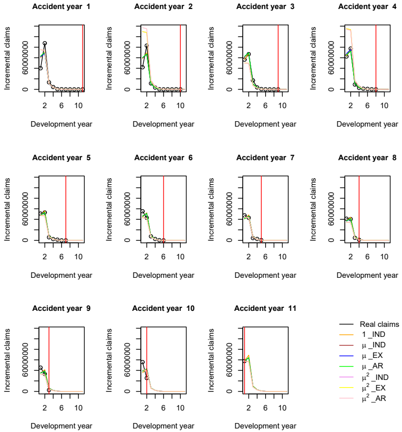




Residual Diagnostic for Generalized Estimation Equations

- Based on residual diagnostic model with **variance function** $\mu_{i,t}$ and **independent correlation structure** is chosen
- Choice of variance function is approved by **QIC**





Comparison of All the Models

- **Predictions** of all the models were further **compared with the new diagonal** see Table below

Model	RBNS + IBNR (ths.)	Diff. diag. (ths.)	MSEP (bil.)
Mack ch. l. (paid triangle)	91 485	-44 245	1 466
GLMM Gamma	84 926	-37 721	866
GLMM Poison	91 485	-44 245	1 466
GLMM Negative binomial	88 706	-41 875	1 297
GEE 1 - IND	97 266	-49 039	1 847
GEE $\mu_{i,t}$ - IND	91 485	-44 245	1 466
GEE $\mu_{i,t}$ - EX	91 762	-44 279	1 466
GEE $\mu_{i,t}$ - AR	91 934	-44 391	1 467
GEE $\mu_{i,t}^2$ - IND	80 919	-35 967	956
GEE $\mu_{i,t}^2$ - EX	80 919	-35 967	956
GEE $\mu_{i,t}^2$ - AR	80 625	-35 793	949

ACTUARIA

Conclusions and Discussion

- Chain ladder is **not always the best** approach
  - Necessary to **check assumptions** via residual diagnostic, back-testing
- **Be careful by using incurred triangles** - internal processes may change by the time
- Most of the mentioned stochastic methods are **not able to deal with zeros** in incremental run-off triangle
  - Might be changed with small values but needs further investigation and testing of sensitivity
- Particular GEE or GLMM models **lead to classical GLM** or have **the same predictions as Chain ladder**
- **Powerful generalization of** well known **GLM models** for dependent variables

Outline

Proposed Approach

- Frequency model
  - Predicted number of claims in a given **accident year** i and **reporting delay** t (in years), $N_{i,t}$
  - Consequently, this number will be split into $N_{i,t,c}$, where $N_{i,t} = \sum_{c=1}^{C} N_{i,t,c}$ and C is maximal assumed length of a claim development in years (time from reporting to closing)
  - $N_{i,t,c}$ **represents number of claims in accident year** i, **development year** t **that are settled after** c **years after reporting**
- Severity model
  - j-th **expected yearly payment** $\mathrm{E}Y_{o,t,c,j} = \mu_{t,c,j}$ of claim o will be modeled with respect to reporting delay t and length of claim settlement c, where $j = 1, 2, \ldots, c$
  - Main objective is the modeling of **development patterns** $\boldsymbol{\mu}_{t,c} = (\mu_{t,c,1}, \mu_{t,c,2}, \ldots, \mu_{t,c,c})^T$

Proposed Approach

- Overall reserve R

$$\mathrm{E}R = \sum_{i=1}^{N}\sum_{t=1}^{N}\sum_{c=1}^{C}\mathrm{E}N_{i,t,c}\sum_{j=1}^{c}\mathrm{E}Y_{o,t,c,j}$$

  - This equation holds if $N_{i,t,c}$ is independent with $Y_{o,t,c,j}$
- Estimation of overall reserve

$$\widehat{R} = \sum_{i=1}^{N}\sum_{t=1}^{N}\sum_{c=1}^{C}\widehat{N}_{i,t,c}\sum_{j=1}^{c}\widehat{\mu}_{t,c,j}$$

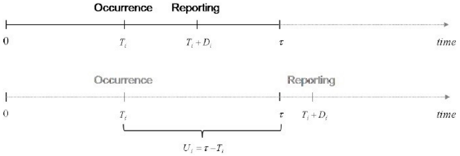
Outline

Dataset

Run-off triangle for incremental number of IBNR claim amounts

Accident year i	Development year j: 1	2	3	4	5	6	7	8	9	10	11	12
2005	**2612**	**617**	**24**	**12**	**0**	**0**	*0*	*0*	*0*	*0*	*0*	*0*
2006	**2405**	**622**	**30**	**8**	**1**	*0*	*0*	*0*	*0*	*0*	*0*	
2007	**2416**	**587**	**41**	**5**	*6*	*0*	*0*	*0*	*0*	*0*		
2008	**2619**	**563**	**42**	*10*	*4*	*1*	*0*	*0*	*0*			
2009	**2081**	**429**	*39*	*8*	*3*	*1*	*0*	*0*				
2010	**1980**	*415*	*61*	*16*	*0*	*0*	*0*					
2011	1745	354	45	13	1	0						
2012	1635	322	27	4	1							
2013	1752	284	17	9								
2014	1705	253	13									
2015	1608	219										
2016	1667											

Maximum Likelihood Estimate



Assumptions

$N_k = \sum_{i=1}^{N} \chi(T_i \in \delta_k)$, $N_k^{(1)} \sim Bi(N_k, p_k)$.
where N_k denotes the number of claims that occurred within the time interval δ_k and p_k is corresponding probability to be reported until the current time τ.

- It is assumed that within this short time interval δ_k, the intensity of the claims process is constant
- In our application δ_k equals to k-th accident week

Maximum Likelihood Estimate

- $N_k^{(1)}$ denotes number of claims reported until the current time τ

Likelihood function

$$l(N_k) = \frac{N_k!}{(N_k - N_k^{(1)})!N_k^{(1)}!} p_k^{N_k^{(1)}} (1 - p_k)^{N_k - N_k^{(1)}}.$$

- Different approach - $N_k^{(1)}$ known, N_k unknown.
- Must estimate p_k, $\forall k$ (distribution of reporting delay)

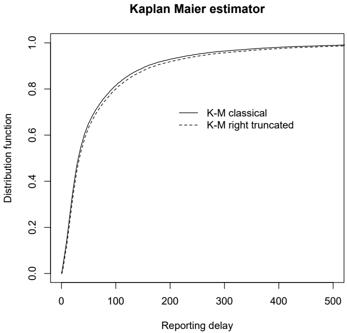
Distribution of Reporting Delay

- Nonparametric estimator
  - Kaplan Maier for truncated data
- Semi-parametric estimator
  - Cox regression for truncated data
- Parametric estimator for truncated data - MLE
  - Weibull
  - Pareto
  - Log-normal

Kaplan Maier for Right Truncated Data

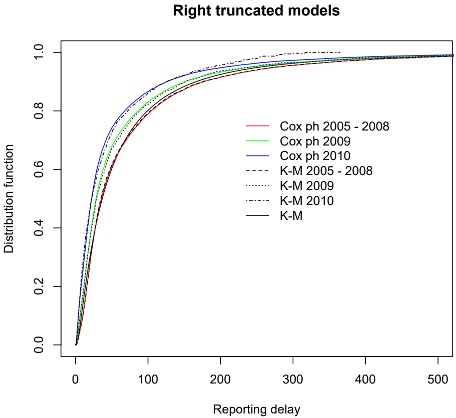
$$\widehat{F}_D^*(d) = \prod_{j:d_{(j)}>d}\left(1-\frac{r_{(j)}}{H_{(j)}}\right),\ r_{(j)} \equiv \sum_{i'=1}^{n'} \chi(d_{i'} = d_{(j)}),\ \ H_{(j)} \equiv \sum_{i'=1}^{n'} \chi(d_{i'} \le d_{(j)} \le u_{i'})$$

where $(u_{i'}, d_{i'})$ are observations of $(U_{i'}, D_{i'})$ and $i' = 1, 2, \ldots, n'$, where n' is the number of observations. Then, $d_{(1)} < d_{(2)} < \ldots < d_{(J)}$ are ordered distinct values of $d_1, d_2, \ldots, d_{n'}$. $F_D^*(d) = \frac{F_D(d)}{F_D(d_{(J)})}$ is just conditional distribution!!!



Cox Regression for Truncated Data

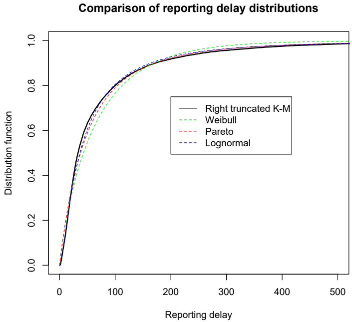
- Assumptions of proportionality holds
- Transformation of variable needed
- All necessary theory can be found in [4]



Parametric Approach - MLE

Estimation

$$\widehat{\boldsymbol{\theta}}_D = \arg\max_{\boldsymbol{\theta}_D} \log \prod_{i'} \frac{f_D(d_{i'}, \boldsymbol{\theta}_D)}{F_D(u_{i'}, \boldsymbol{\theta}_D)},$$



Distribution	Weibull	Pareto	Log-normal
Log-Likelihood	-88523	-87896	-87007

Based on the log-likelihood comparison in the Table above log-normal distribution is preferred

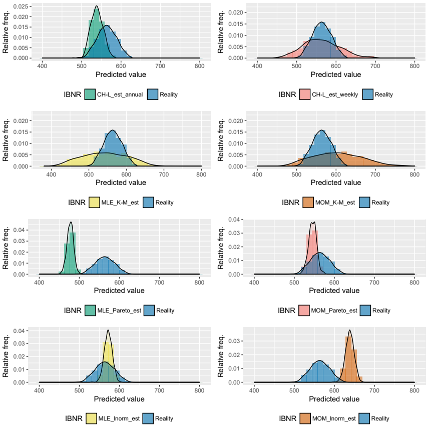
Comparison of Results

- Prediction of IBNR number of claims and comparison with reality

Real	Mack	Pred.	MLE	Pred.	Moment	Pred.
564	Annual	538	K-M classic	447	K-M classic	499
	Weekly	561	K-M truncated	542	K-M truncated	602
			Cox prop. haz.	272	Cox prop. haz.	339
			Weibull	490	Weibull	525
			Pareto	464	Pareto	533
			Log-normal	531	Log-normal	598

- Also **moment method** was used in addition to maximum likelihood
- **Cox regression and classical K-M** are **not very precise**
- It is **not possible to assess prediction of reserve** based on one realization therefore **simulation analysis must be performed**

Simulation Analysis

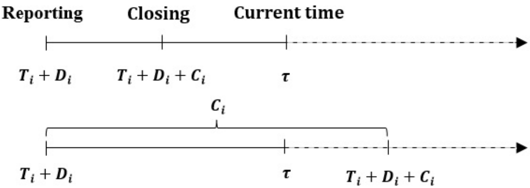


	Mack annual	Mack weekly	MLE K-M	MLE Pareto	MLE L-nor.	Moment K-M	Moment Pareto	Moment L-nor.
Bias	25	1	21	87	-8	-40	18	-75
MSPE	1422	2712	4526	8084	699	6060	953	6199

Outline

ACTUARIA

Survival Analysis for Censored Data

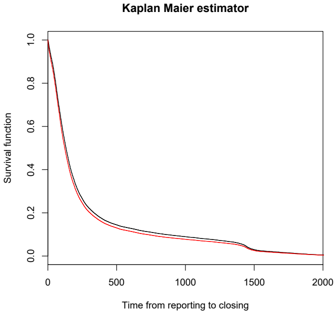


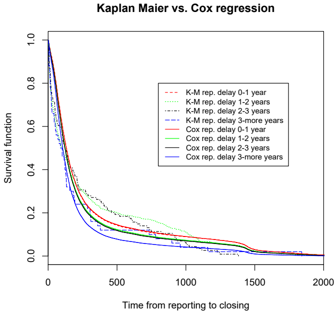
K-M estimation of survival function

$\widehat{S}_C(t) = 1 - \widehat{F}^*_C(t) \sum_{j:c_j > t} \frac{n_j - d_j}{n_j}$,
where k_i is either time from reporting to closing or in case claim hasn't closed yet, censoring time. Further, δ_i is indicator of closed claim and n is the number of observation,
$d_j = \sum_{i=1}^{n} \chi(k_i = c_j, \delta_i = 1)$ represents number of claims at time c_j a $n_j = \sum_{i=1}^{n} \chi(k_i \leq c_j)$ is number of claims in risk at time c_j

Kaplan Maier for Censored Data

- Cox regression not possible to use due to violation of assumptions
- Comparison of Kaplan Maier for censored data and classical Kaplan Maier





Conditional Expected Value

- For claims which are still open the closing time must be predicted. Therefore, conditional expected value $E(T \mid T > t)$ must be predicted in our application.
- Using survival function conditional expected value can be expressed as follows

$$E(T \mid T > t) = t + \int_t^\infty \frac{S(u)}{S(t)} du$$

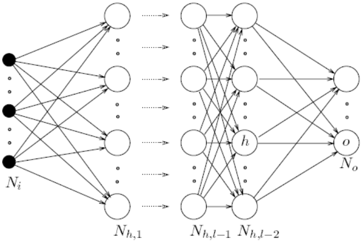
- Estimation of conditional expected value can be expressed using estimated survival function as

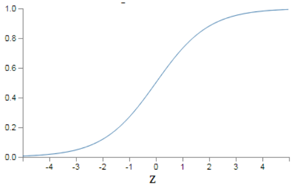
$$E(\widehat{T \mid T} > t) = t + \int_t^\infty \frac{\widehat{S}(u)}{\widehat{S}(t)} du$$

Outline

Structure of Neural Networks

- Hidden layers
  - One
  - Two
- Number of neurons within layers
- Activation function $F\left(\sum_i w_i x_i + b\right)$
  - Sigmoid $F(z) = \frac{1}{1+e^{-z}}$
  - Hyperbolic tangent $F(z) = \tanh(z) = \frac{e^{2z}-1}{e^{2z}+1}$
  - Linear $F(z) = z$
  - ReLu (rectified linear unit) $F(z) = \max(0, z)$





Optimization Algorithm

- **Batch methods** which use the **full training set** to compute the next update to parameters tend to **converge very well**
  - Very few hyper-parameters to tune
  - In practice **computing the cost** and gradient for the entire training set can be very slow and sometimes intractable on a single machine
  - Doesn't give an easy way to incorporate new data in an 'online' setting

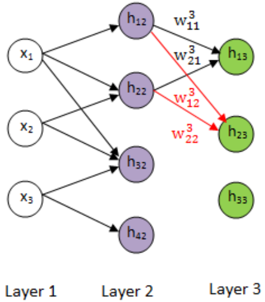
$$\theta = \theta - \gamma \nabla_\theta E\left[L(\theta)\right]$$

  where θ is unknown parameter, γ learning rate and L is loss (cost) function.
- **Stochastic Gradient Descent** (SGD) **addresses all these issues** by following the negative gradient of the objective after seeing only a single or a **few training examples**

$$\theta = \theta - \gamma \nabla_\theta L\left(\theta; x_i, y_i\right)$$

Weights

- θ in SGD denotes all weights w_{jk} and biases b_k for all k and j within neural network
- Upper index w^s_{jk} in figure below denotes hidden layer but it is not necessary



Back Propagation - Generalized Delta Rule

$$L=\sum_{p=1}^{P} L^p, \qquad L^p = \frac{1}{2}\sum_{i=1}^{N_o}(o_i^p - y_i^p)^2, \qquad y_k^p = F(s_k^p) = F\left(\sum_{j=1}^{n} w_{jk}\, y_j^p + b_k\right)$$

where o_i^p is i-th element of p-th observation

- Utilizing **differentiability of the activation function, partial derivation** and **chain rule**

$$\Delta_p w_{j,k} = \gamma y_j^p \delta_k^p, \qquad \delta_k^p = F'(s_k^p)\sum_{o=1}^{N_o} \delta_o^p w_{k,o}$$

- At initial step $\delta_o^p = (o_o^p - y_o^p)F'(s_o^p)$
- Back propagation is **feasible** because the computation of deltas within one hidden layer **needs** only one **value of input from given neuron, weights as the input into following layer** and **values of deltas from previous recursive step**

Setting of Hyper-parameters

- Different results for various setting of algorithm
- Hyper-parameters
  - Validation - training ratio
  - Momentum
  - Batch size
  - Number of epochs
  - Learning rate
    - Fixed learning rate for each epochs
    - Adaptive learning rate method (ADAM, RMSPROP)
- Shortcoming of algorithm
  - Using different random seed you might obtain different results
    - Random initial weights
    - Randomness within algorithm (e.g. batching)

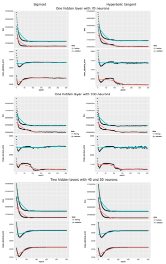
Practical Application - Neural Networks

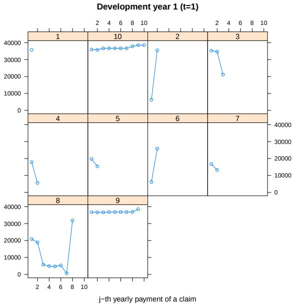
- Our objective
  - j-th **expected yearly payment** $\mathrm{E}Y_{o,t,c,j} = \mu_{t,c,j}$ of claim o will be modeled with respect to reporting delay t and length of claim settlement c, where $j = 1, 2, \ldots, c$
  - Main objective is the modeling of **development patterns** $\boldsymbol{\mu}_{t,c} = (\mu_{t,c,1}, \mu_{t,c,2}, \ldots, \mu_{t,c,c})^T$
- Various structures, activation functions, number of neurons and approaches were used
- One stage approach
- Two stage approach
  - One model for probability of non-zero claim
  - Another model just for positive claims

ACTUARIA

Practical Application - One Stage Approach

- Based on **validation loss function** (mean square error) model with model with **one hidden layer with 70 hyperbolic tangent neurons** output layer **ReLu** neuron

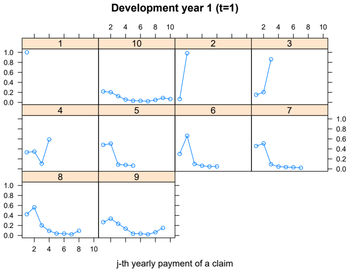


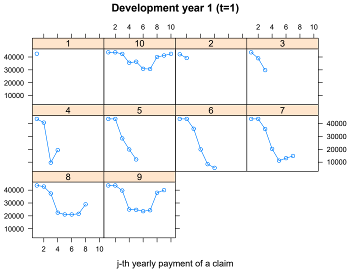


- Figure displays patterns $\mu_{1,1}, \mu_{1,2}, \mu_{1,3}, \mu_{1,4}, \mu_{1,5}, \mu_{1,6}, \mu_{1,7}, \mu_{1,8}, \mu_{1,9}, \mu_{1,10}$
- Chosen approach is **not suitable**

Practical Application - Two Stage Approach

- Stage one - model with **one hidden layer of 20 sigmoid neurons** output neuron **sigmoid**
- Stage two - model with **one hidden layer of 90 sigmoid** output neuron **linear**

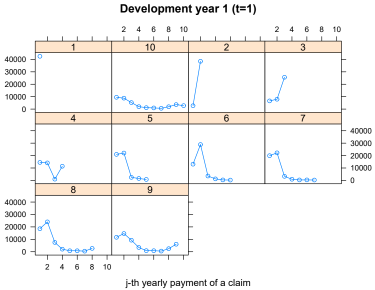


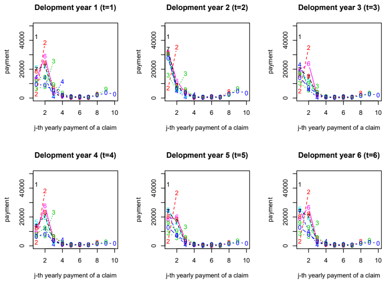


- **Reasonable** results

Practical Application - Two Stage Approach

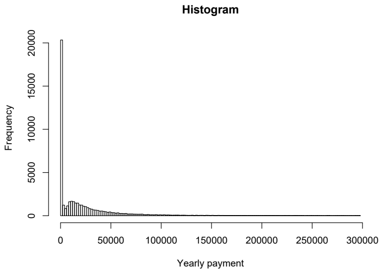
- Joined patterns for all development years/reporting delay





Hurdle Models

- Our data has mass at zero - reason for two stage approach within neural networks



- Hurdle model is proposed for such a situation

$$f_{hurdle}\left(Y_i = y_i\right) = \begin{cases} f_1\left(0\right), & y_i = 0 \\ \left(1 - f_1\left(0\right)\right)\frac{f_2(y_i)}{1-f_2(0)} = \Phi f_2\left(y_i\right) & y_i = 1, 2, \dots. \end{cases}$$

Poisson Hurdle Models

- Basic example of Poisson hurdle model

$$f_{hurdle}\left(Y_i = y_i\right) = \begin{cases} 1 - p_i, & y_i = 0 \\ p_i \frac{e^{\{-\lambda_i\}}\lambda_i^{y_i}}{\left(1 - e^{\{-\lambda_i\}}\right) y_i!} = \Phi f_2\left(y_i\right) & y_i = 1, 2, \ldots, \end{cases}$$

where p_i is modeled by **logistic regression**

$$\begin{aligned} L(\beta, \gamma) &= \left[\prod_{y_i=0} P(y_i = 0)\right]\left[\prod_{y_i>0} P(y_i > 0) f_2(y_i)\right] \\ &= \left[\prod_{y_i=0} \frac{1}{1 + e^{\mathbf{X}_i\beta}}\right]\left[\prod_{y_i>0} \frac{e^{\mathbf{X}_i\beta}}{1 + e^{\mathbf{X}_i\beta}} \frac{e^{-e^{\mathbf{Z}_i\gamma}} e^{\mathbf{Z}_i\gamma y_i}}{\left(1 - e^{-e^{\mathbf{Z}_i\gamma}}\right) y_i!}\right] \\ &= \left[\prod_{y_i>0} e^{\mathbf{X}_i\beta} \prod_{i=1}^{n} \frac{1}{1 + e^{\mathbf{X}_i\beta}}\right]\left[\prod_{y_i>0} \frac{e^{-e^{\mathbf{Z}_i\gamma}} e^{\mathbf{Z}_i\gamma y_i}}{\left(1 - e^{-e^{\mathbf{Z}_i\gamma}}\right) y_i!}.\right] \end{aligned}$$

- Due to the **independence** of the Y_i's, it is **possible to factor the likelihood function** into $L_1(\beta)$ and $L_2(\gamma)$
- The factorization allows us to **maximize** $L_1(\beta)$ and $L_2(\gamma)$ **separately**

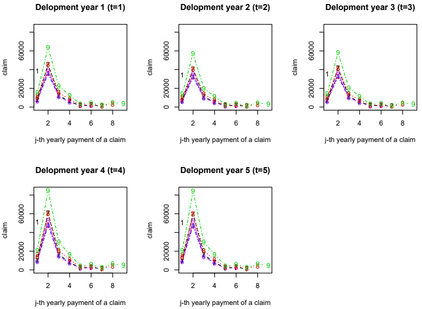
Practical Application - Hurdle Models

- Due to the specifics of non-zero part of our data it can be **viewed as count data**
  - Poisson vs. negative binomial
  - Possible improvement Gamma ...
- Due to the **convergence** issue **merging last two levels** within covariates
- Based on AIC **negative binomial hurdle model** was chosen
- **Simplification of linear predictor** for hurdle component
  - Excluding development year/reporting delay (not statistically significant)

ACTUARIA

Practical Application - Hurdle Models

- Less flexible than patterns from neural networks due to the lower number of parameters - may lead to better prediction (further investigation and back-testing)
- Possible improvement by including interactions - leads to more flexible patterns



The End

Questions?

Thank you for your attention.

Michal Gerthofer

Further reading I

B.M. Bolker and M.E. Brooks and C.J. Clark
Generalized linear mixed models: a practical guide for ecology and evolution.
Trends in ecology & evolution, 24(3):127–135, 2009.

P.J. Diggle and P. Heagerty and K.Y. Liang and S.L. Zeger
Analysis of Longitudinal Data 2nd ed..
Oxford University Press, Reading, Massachusetts, 2002.

E.L. Kaplan.
Nonparametric estimation from incomplete observations.
J. Amer. Stat. Assoc., 53:457–481, 1958.

J.F. Lawless
Statistical Models and Methods for Lifetime Data, 2nd ed..
Wiley, USA, New Jersey, 2003.

ACTUARIA

Further reading II

R.J. Williams and D.E. Rumelhart and G.E. Hinton
Learning representations by back-propagating errors.
Nature, 323(9):533–536, 1986.

J. Mullahy
Specification and testing in some modified count data models.
Journal of Econometrics, 33:341–365, 1986.