European actuarial academy

an initiative of the 'Deutsche Aktuarvereinigung', the Dutch 'Actuarieel Genootschap', the 'Schweizerische Aktuarvereinigung' and the 'Aktuarvereinigung Österreichs'

Session 7

Generalized Linear Models for IBNR-reserves

Agenda

- Why a multivariate model?
- The linear model
- What are Generalized Linear Models (GLMs)?
- GLMs for reserving

Generalized Linear Models

+ Standard framework for estimation, testing, prediction
+ Broad class of probability functions
+ Additive, multiplicative model structure (+ other possibilities)
+ Software: S-Plus, GLIM, SAS (Genmod)

– Caveat: over-specification because of too many parameters
– Not always specifically aimed at IBNR-applications

Why a multivariate model? (1)

Example:

Weight	Private use	Business use	Total
light	9000 x 200.00	300 x 230.00	9300 x 200.97
middle	6000 x 220.00	700 x 253.00	6700 x 223.45
heavy	3000 x 240.00	1000 x 276.00	4000 x 249.00
Total	18000 x 213.33	2000 x 261.05	20000 x 218.11

On the basis of marginal totals:

- What is the loading for Business vs. Private use?
- What is the loading for middle and heavy vs. light weight?

- What is the resulting loading for a heavy business care vs. a light private car?

Why a multivariate model? (2)

Loadings resulting from marginal totals:

Weight	Private use	Business use	Total
light	9000 x 200.00	300 x 230.00	9300 x 200.97
middle	6000 x 220.00	700 x 253.00	6700 x 223.45
heavy	3000 x 240.00	1000 x 276.00	4000 x 249.00
Total	18000 x 213.33	2000 x 261.05	20000 x 218.11

1.1119

1.2390

1.2237

- Loading heavy business car: 1.2237 x 1.2390 = 1.5161 ⇒ 51.6%
- But the actual loading only is: 1.15 x 1.20 = 276/200 = 1.38 ⇒ **38% !**

Linear Model (1)

$$y_i = \alpha + \beta x_{1i} + \gamma x_{2i} + \ldots + \varepsilon_i$$

- Dependent variables y_i are linked linearly to explanatory variables x.
- Disturbance terms ε are Normally distributed.
- Estimation of parameters with Ordinary Least Squares (= maximum likelihood in this case)

Linear Model (2)

In matrix notation:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

with

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \qquad \boldsymbol{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \qquad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Linear Model (3)

Matrix $\boldsymbol{X}$ is called the design matrix.

For example:

$$\boldsymbol{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 2 \\ 1 & 1 & 0 & 0 & 3 \\ 1 & 1 & 0 & 1 & 4 \\ 1 & 0 & 1 & 0 & 5 \\ 1 & 0 & 1 & 1 & 6 \end{pmatrix} \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix}$$

with for instance: $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$

$$y_4 = \beta_1 + \beta_2 + \beta_4 + 4\beta_5 + \varepsilon_4$$

Generalized Linear Models (1)

Systematic component

$\eta_i = \boldsymbol{x}_i' \beta$ ($\boldsymbol{x}_i'$ is i-th row of design matrix)

Stochastic component

Error-distribution from exponential family, encompassing

- Normal
- Poisson
- Binomial
- Gamma
- Inverse Gaussian

Link function

Linear link is dropped: $E[y_i] = f(\eta_i)$, e.g. $\alpha \times \exp(\beta x_{1i}) \times \exp(\gamma x_{2i})$

Generalized Linear Models (2)

Exponential family of distributions:

$$f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi/w} + c(y,\phi/w)\right)$$

- $b(\cdot)$ and $c(\cdot,\cdot)$ are predetermined functions
- w_i are known weights (e.g. number of cars)
- θ_i are 'canonical parameters'
- ϕ is dispersion/scale parameter

Generalized Linear Models (3)

Normal(μ,σ^2) distribution:

- $\theta = \mu$
- $\phi = \sigma^2$
- $b(\theta) = \theta^2/2$
- $c(y,\phi) = -(y^2/\phi + \ln(2\pi\phi))/2$

Poisson(λ) distribution:

- $\theta = \ln(\lambda)$
- $\phi = 1$
- $b(\theta) = \exp(\theta)$
- $c(y,\phi) = -\ln(y!)$

Generalized Linear Models (4)

Back to exponential family: $f(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi/w} + c(y,\phi/w)\right)$

- $E[Y] = b'(\theta)$
- $Var[Y] = b''(\theta)\ \phi/w$

Link function:

relates E[Y] to linear systematic component η (= $\boldsymbol{x}_i'\beta$)

$\eta = g(E[Y]) = g(b'(\theta))$

- linear: $g(z) = z$
- log: $g(z) = \ln(z)$
- reciprocal: $g(z) = 1/z$

Generalized Linear Models (5)

'Canonical' or 'natural' link-function: $\eta = \theta$

$\Rightarrow$ desirable statistical properties, but no must

- Normal distribution: $b'(\theta) = \theta \Rightarrow g(b'(\theta)) = g(\theta)$

  $\eta = g(b'(\theta)) = \theta \Rightarrow g(\theta) = \theta$ (linear link function)

- Poisson distribution: $b'(\theta) = \exp(\theta) \Rightarrow g(b'(\theta)) = g(\exp(\theta))$

  $\eta = g(b'(\theta)) = \theta \quad \Rightarrow g(\exp(\theta)) = \theta$

  $\Rightarrow g(z) = \ln(z)$ ('log' link function)

Generalized Linear Models (6)

Log-likelihood function:

$$\ell = \sum_i \frac{y_i\theta_i - b(\theta_i)}{\phi/w_i} + c(y_i, \phi/w_i)$$

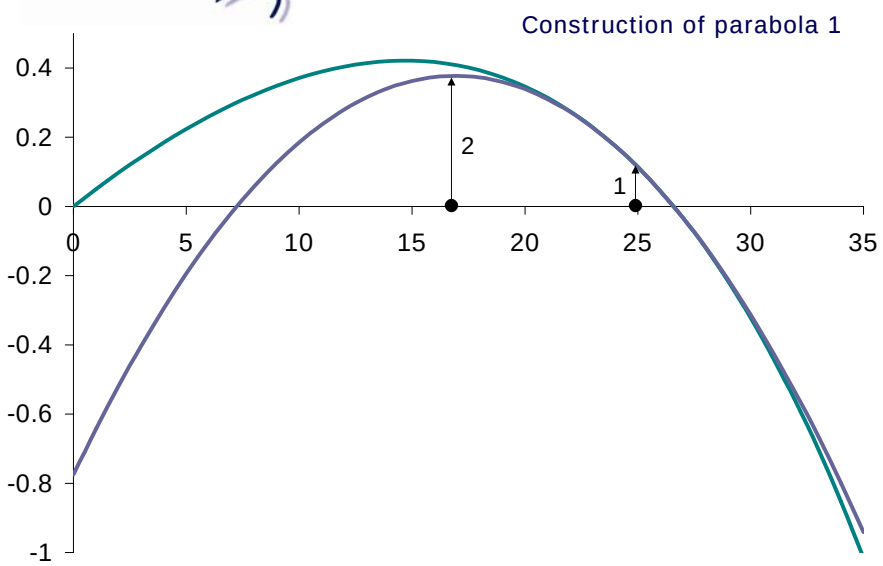
Parameters in $\boldsymbol{\beta}$ (and hence the θ_i) follow by maximization.

General idea numerical maximization likelihood (1)

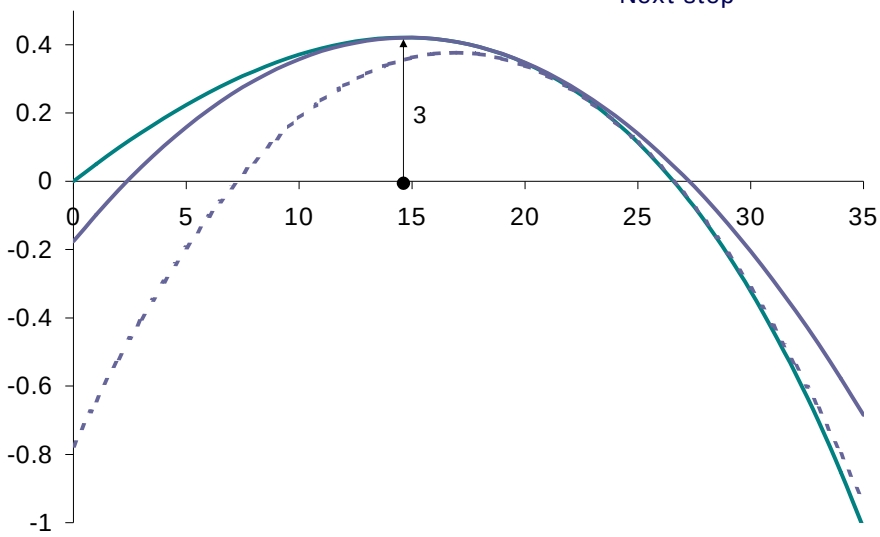
2-dimensional case: max/minimize $f(x)$ with respect to x

- Take starting point x_0
- Construct parabola with same function value, (numerical) first and second derivative
- Calculate maximum/minimum of parabola (point x_1)
- Construct corresponding parabola again
- ....
- Until two successive points x_j and x_{j+1} are close enough.

General idea numerical maximization likelihood (2)

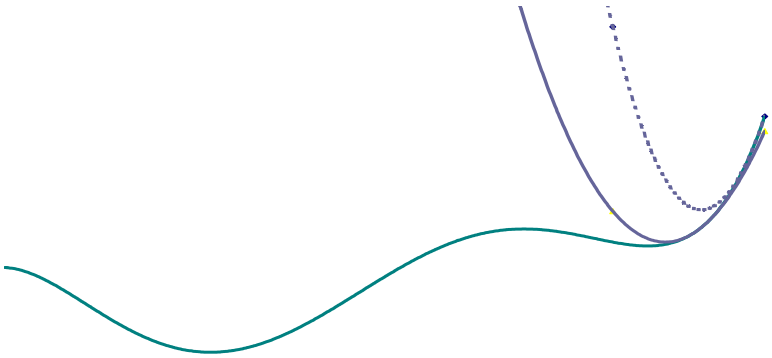


General idea numerical maximization likelihood (3)



General idea numerical maximization likelihood (4)

- For ill-conditioned numerical optimization, no convergence:



- GLMs with canonical link-functions: usually no problems.

Analysis of deviance (1)

Take two models of which one encompasses the other.

Measure of discrepancy between models:

twice difference between maximized log-likelihoods

$$2(\tilde{\ell}-\hat{\ell}) = 2\sum_i w_i\{y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\}/\phi = D/\phi$$

D is called the <u>deviance</u> and $SD=D/\phi$ the <u>scaled deviance</u>

SD is the likelihood-ratio statistic to test the null-hypothesis that the "smaller" model is true.

Under the null-hypothesis it is $\chi^2(q)$ distributed, with *q* the difference of number of parameters.

Analysis of deviance (2)

Forms of deviances

Normal $$D = \sum (y-\hat{\mu})^2$$

Poisson $$D = 2\sum y\ \ln(y/\hat{\mu}) - (y-\hat{\mu})$$

Inverse Gaussian $$D = \sum (y-\hat{\mu})^2/(\hat{\mu}^2 y)$$

Numerical example (1)

Example: numbers of reported claims

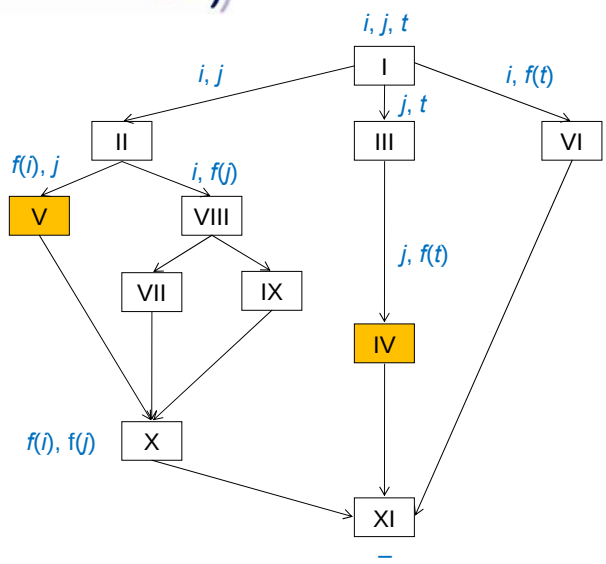
Loss year	Development year 1	2	3	4	5	6	7	8
2000	101	153	52	17	14	3	4	1
2001	99	121	76	32	10	3	1	
2002	110	182	80	20	21	2		
2003	160	197	82	38	19			
2004	161	254	85	46				
2005	185	201	86					
2006	178	261						
2007	168							

Numerical example (2)

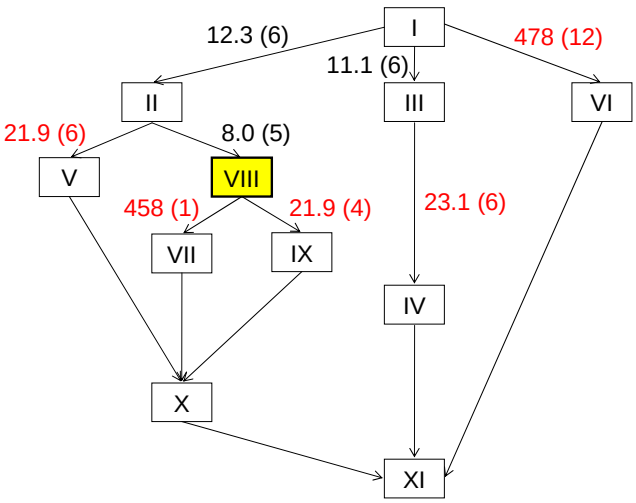
Basis-model: N_{ij} ~ Poisson(exp[$\mu+\alpha_i+\beta_j+\gamma_{i+j-1}$])

Model	Parameters	# observations minus # parameters	Scaled Deviance	"R squared"
I	$\mu, \alpha_i, \beta_j, \gamma_t$	15	25.7	99%
II	μ, α_i, β_j	21	38.0	99%
III	μ, β_j, γ_t	21	36.8	99%
IV	$\mu, \beta_j, t\gamma$	27	59.9	98%
V	$\mu, i\alpha, \beta_j$	27	59.9	98%
VI	$\mu, \alpha_i, t\gamma$	27	504.0	81%
VII	$\mu, \alpha_i, j\beta$	27	504.0	81%
VIII	$\mu, \alpha_i, \beta_1, j\beta$	26	46.0	98%
IX	$\mu, i\alpha, \beta_1, j\beta$	32	67.9	97%
X	$\mu, i\alpha, j\beta$	33	582.0	78%
XI	μ	35	2656.0	0%

Numerical example (3) – Nested structure



Numerical example (4) – Analysis of Deviance



Critical values
95% confidence

$\chi^2(1)$: 3.84

$\chi^2(4)$: 9.49

$\chi^2(5)$: 11.1

$\chi^2(6)$: 12.6

Optimal model (VIII)

$$\widehat{N}_{ij} = 102.3 \times \begin{bmatrix} i=1: & 1.00 \\ i=2: & 0.99 \\ i=3: & 1.21 \\ i=4: & 1.47 \\ i=5: & 1.67 \\ i=6: & 1.56 \\ i=7: & 1.81 \\ i=8: & 1.64 \end{bmatrix} \times 3.2^{\{j \neq 1\}} \times 0.42^{j-1}$$

Predicted number of IBNR claims

	Development year												#
Loss year	1	2	3	4	5	6	7	8	9	10	11	12	**IBNR**
2000	102	137	58	24	10	4	2	1	0	0	0	0	1
2001	101	136	57	24	10	4	2	1	0	0	0	0	1
2002	124	166	70	29	12	5	2	1	0	0	0	0	4
2003	150	202	85	36	15	6	3	1	0	0	0	0	11
2004	171	230	96	41	17	7	3	1	1	0	0	0	29
2005	160	214	90	38	16	7	3	1	0	0	0	0	65
2006	185	249	105	44	18	8	3	1	1	0	0	0	180
2007	168	225	95	40	17	7	3	1	1	0	0	0	389
													680