A Deeper Look Inside Generalized Linear Models

University of Minnesota

February 3rd, 2012

Nathan Hubbell, FCAS

Discover · Develop · Succeed

Agenda

- Property & Casualty (P&C Insurance) – in one slide
- The Actuarial Profession
- Travelers
  - Broad Overview
  - Actuarial & Analytics Career Opportunities
- GLMs
  - Background
  - Insurance Applications

What is insurance and what good does it serve?

Insurance restores individuals to the financial state they were in prior to a loss
(e.g. car accident; tree fell on house)

For this benefit, customers pay a premium to the insurance company

If a customer doesn't have a loss, then their premium:

A. Helps the insurance company cover the loss another customer did have
B. Keeps the insurance company functioning so it can continue providing this service

If the customer does have a loss, it's the other insureds that are helping them!

TRAVELERS

The Actuarial Profession

Some statistics on the actuarial profession:

- Approximately 20,000 actuaries in the U.S.
- Projected 21% Increase in Employment from 2008-2018
- Over half of All Actuaries are Employed by Insurance Carriers

TRAVELERS

Actuarial Salaries

Actuary Salary Survey - actuary jobs - D.W. Simpson & Company, Inc. , actuarial, actuaries, Emp - Microsoft Internet Explor...

http://www.dwsimpson.com/salary.html

McAfee

Favorites | Actuary Salary Survey - actuary jobs - D.W. Simpson ...

Property & Casualty October 2011	0-0.5 yrs (excl. sign-on)	0.5-2.5 yrs	2.5-4.5 yrs	4.5-6.5 yrs	6.5-9.5 yrs	9.5-14.5 yrs*	14.5-19.5 yrs*	19.5+ yrs*
1 exam	46-65	53-67	55-72					
2 exams	53-69	56-74	59-78	62-82				
3 exams	54-74	59-78	65-86	68-88	72-100			
4 exams	57-75	61-86	68-92	73-99	76-112	90-139		
5 exams		62-90	72-106	82-114	84-126	98-150		
6 exams		73-95	82-113	88-126	97-140	104-169		
ACAS		81-110	89-122	90-133	100-161	118-188	135-267	149-306+
8 exams			94-140	109-163	112-178	135-225		
FCAS			101-154	125-182	133-224	148-340	175-442	177-483+
CEO/CFO/COO/CIO/CRO/President - 10th-90th percentile TBD data points - from 2011 SEC Def 14A filings							N/A	

Source: DW Simpson October 2011 Salary Survey; includes base pay and bonus.

Two Actuarial Societies

- Casualty Actuarial Society (CAS) and the Society of Actuaries (SOA)
- Exams 1 through 4
  - Lower Level Exams
  - Similar for the two societies
- Exams 5 through 9
  - Upper Level Exams
  - Different for the two societies
- Exams are offered 1 – 6 times a year
- Validation by Educational Experience (VEEs)

Lower Level Exams

- Exam 1/P – Probability
- Exam 2/FM – Financial Mathematics
- Exam 3F – Financial Economics
- Exam 3L – Life Contingences and Statistics
- Exam 4/C – Construction and Evaluation of Actuarial Models

- Exams are multiple choice, range from 2 to 3 hours in length, and require about 300 hours of studying
- Very challenging exams – you are competing against many other exam-takers

TRAVELERS

Actuarial Exams

- Computer-Based Testing Available for first exam.
- Offered at Prometric Testing Locations in Woodbury and Edina.
- Register online at http://www.prometric.com/SOA

About Travelers

- Offers property and casualty solutions to individuals and companies of all sizes
- Second-largest commercial insurer in the U.S.
- Second-largest personal insurer through the independent agency channel
- No. 98 on the Fortune 500 list of largest U.S. companies
- Representatives in every U.S. state, Canada, Ireland and the United Kingdom
- A member of the Dow Jones Industrial Average – the **only** insurance company on the list

TRAVELERS

Analytics at Travelers – Who are we?

Across Travelers, we form a large (300+) and diverse community of Ph.D., Masters and Bachelors holders in the following disciplines:

- mathematics
- statistics
- physics
- actuarial science
- computer science
- business

… and more!

TRAVELERS

Actuaries at Travelers – What do we do?

We ask and answer questions requiring sophisticated analyses

- How much will it cost to insure a customer?
- What amount of loss reserves need to be set aside to pay future claims?
- How likely is it that this customer will purchase our product?
- How many claims adjusters will we need in two years?
- What new statistical methods will help move our business into the future?

Why are these questions hard?

Example: The cost to insure an auto customer

It's impossible to predict if someone is going to

A. Get into an accident
B. The type of accident (hit a telephone pole, hit another vehicle, bodily injury)
C. How bad the accident will be

But if we have enough customers, we can start to group them…

in group A) we expect 1 / 10 to get into an accident, costing on average $1000

in group B) etc.

TRAVELERS

Why it matters…

The more finely we group, the more accurate the price. In that case:

1. If our competitors charge more, the customer will choose us and we will grow profitably
2. If our competitors charge less, the customer will choose them and they'll grow *un*profitably

Either way, we win! (or if we price inaccurately or too coarsely, we lose!)
The trick is finding the right groups, and getting the right price for them…

Now, how might one go about doing this?

Generalized Linear Models

Predictive Modeling

- Using Generalized Linear Models (GLM's) and other statistical methods to predict exposure to loss at detailed level.

– Recently, property-casualty insurance companies have embraced predictive modeling as a strategic tool for competing in the marketplace.
  - Originally introduced as a method of increasing precision for personal auto insurance pricing
  - Extended to homeowners and commercial lines
  - Today, it is applied in areas such as marketing, underwriting, pricing, and claims management

How are rates differentiated?

Automobile	Homeowners
– Age	– Age of home
– Gender	– # occupants
– Marital status	– Primary / Secondary
– # vehicles	– Prior claim experience
– # drivers	– Construction
– Home policy	– Protection
– Driving record	– Roof Type
– Years Licensed	– Location (CAT?)
– Limits	– Amount of Insurance
– Prior Insurance	– Auto Policy
– Student/Nonstudent	– Responsibility of owner
– Location (Garage/driven)	
– Annual Mileage	

Possible Solution:
Multiple Linear Regression / Ordinary Least Squares

- $E[Y] = a_0 + a_1X_1 + \ldots + a_nX_n$

- ***Two Key Assumptions:***
  - *Y* is Normally distributed random variable.
  - Variance of *Y* is constant (homoscedastic).

TRAVELERS

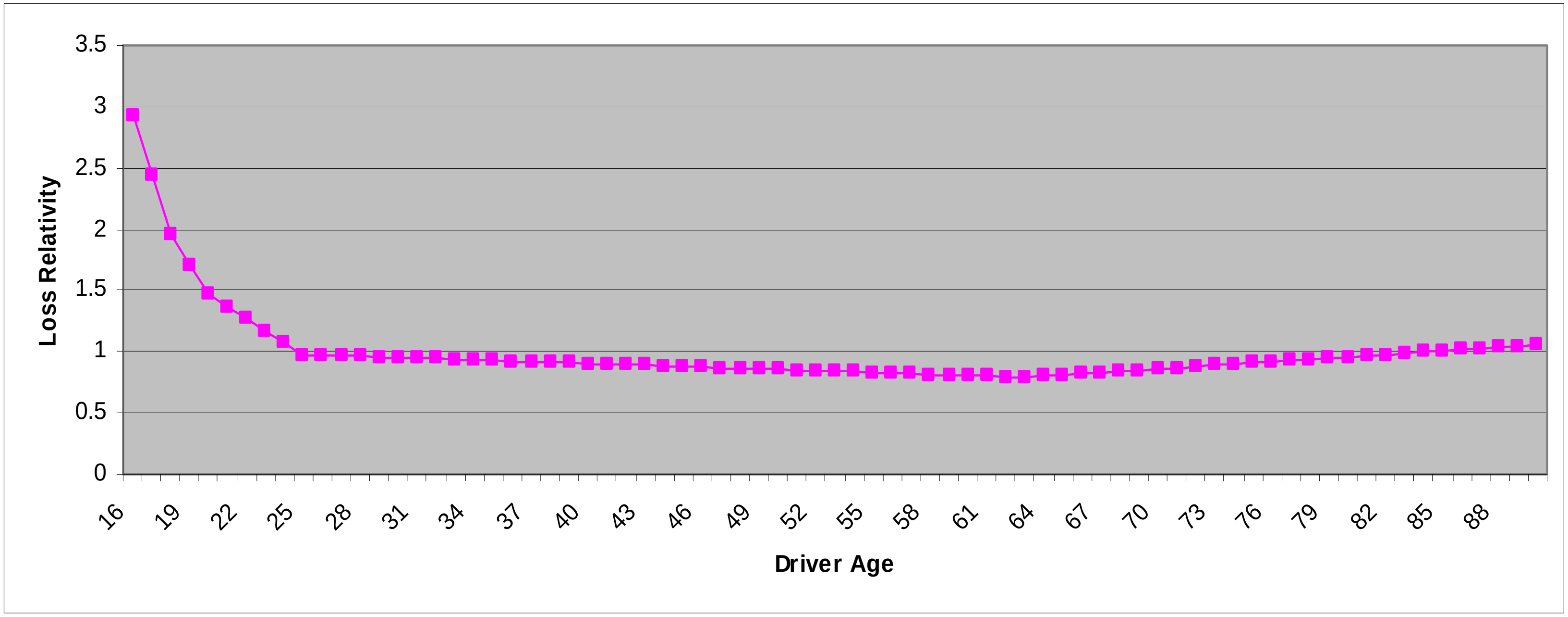
Problems With OLS Regression – Part I

- ***Y* is NOT normally distributed.**
  - Number of claims is discrete (and most policies have zero claims)
  - Claim sizes are skewed to the right
  - Probability of an event is in [0,1]

Problems With OLS Regression – Part II

- **Variance of *Y* is NOT constant.**
  - Varies by expected loss.
    - High frequency losses have less variance.
    - High severity losses have more variance.
  - Varies by exposure.

Problems With OLS Regression – Part III



(numbers are illustrative only)

TRAVELERS

Introducing:
Generalized Linear Models (GLMs)

- $E[Y] = g^{-1}(a_0 + a_1 X_1 + \ldots + a_n X_n)$

- **Fewer restrictions:**
  - Non-linear relationships.
    - g(x) = x → Additive model
    - g(x) = exp(x) → Multiplicative model
    - g(x) = 1 / (1+exp(x)) → Logistic model

Generalized Linear Models (GLMs)

- $E[Y] = g^{-1}(a_0 + a_1X_1 + \ldots + a_nX_n)$

- **Fewer restrictions:**
  - *Y* can be from any exponential family of distributions.
    - Poisson (number of claims)
    - Binomial (probability of renewing)
    - Gamma (loss severity)

Generalized Linear Models (GLMs)

- **$E[Y] = g^{-1}(a_0 + a_1X_1 + \ldots + a_nX_n)$**

- **Fewer restrictions:**
  - Variance depends on the expected mean.
    - Normal: Variance is constant. (σ^2)
    - Poisson: Variance equals mean. ($\mu = E[Y]$)
    - Gamma: Variance equals mean squared ($\mu^2 = E[Y]^2$)

TRAVELERS

Exponential Family ABCs

$$f(y_i;\theta,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}$$

	$a(\phi)$	$b(\theta)$	$c(y,\phi)$
Normal	ϕ/ω	$\theta^2/2$	$-\frac{1}{2}(\omega y^2/\phi + \ln(2\pi\phi/\omega))$
Poisson	ϕ/ω	e^θ	$-\ln y!$
Gamma	ϕ/ω	$-\ln(-\theta)$	$\omega/\phi \ln(\omega y/\phi) - \ln(y) - \ln(\Gamma(\omega/\phi))$
Binomial (m trials)	ϕ/ω	$m.\ln(1+e^\theta)$	$\ln\binom{m}{y}$
Inverse Gaussian	ϕ/ω	$-\sqrt{-2\theta}$	$-\frac{1}{2}\{\ln(2\pi\phi y^3/\omega) + \omega/(\phi y)\}$

Source: "A Practitioner's Guide to Generalized Linear Models"

TRAVELERS

Exponential Family – Mean and Variance

	Notation	ϕ	$\mu(\theta)$	$V(\mu)$
Normal	$N(\mu,\sigma^2)$	σ^2	θ	1
Poisson	$P(\mu)$	1	e^θ	μ
Gamma	$G(\mu,\nu)$	ν^{-1}	$-1/\theta$	μ^2
Binomial	$B(m,\pi)/m$	$1/m$	$e^\theta/(1+e^\theta)$	$\mu(1-\mu)$
Inverse Gaussian	$IG(\mu,\sigma^2/\omega)$	σ^2	$(-2\theta)^{-1/2}$	μ^3

Tweedie $V(\mu) = \mu^p \; ; 1 < p < 2$

Source: “A Practitioner's Guide to Generalized Linear Models”

TRAVELERS

Uncle Tweedie

$$f_Y(y;\theta,\lambda,\alpha) = \sum_{n=1}^{\infty} \frac{\left\{(\lambda\omega)^{1-\alpha}\kappa_\alpha(-1/y)\right\}^n}{\Gamma(-n\alpha)n!\,y} . \exp\{\lambda\omega[\theta_0 y - \kappa_\alpha(\theta_0)]\} \quad \text{for } y > 0$$

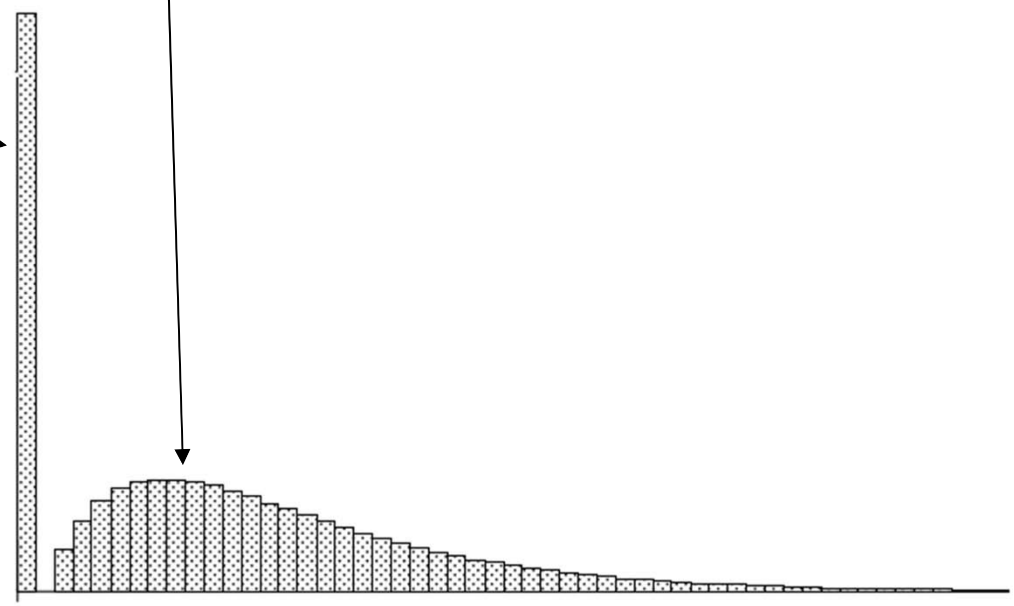
and

$$p(Y=0) = \exp\{-\lambda\omega\kappa_\alpha(\theta_0)\}$$

where

$$\kappa_\alpha(\theta) = \frac{\alpha-1}{\alpha} . \left(\frac{\theta}{\alpha-1}\right)^\alpha$$

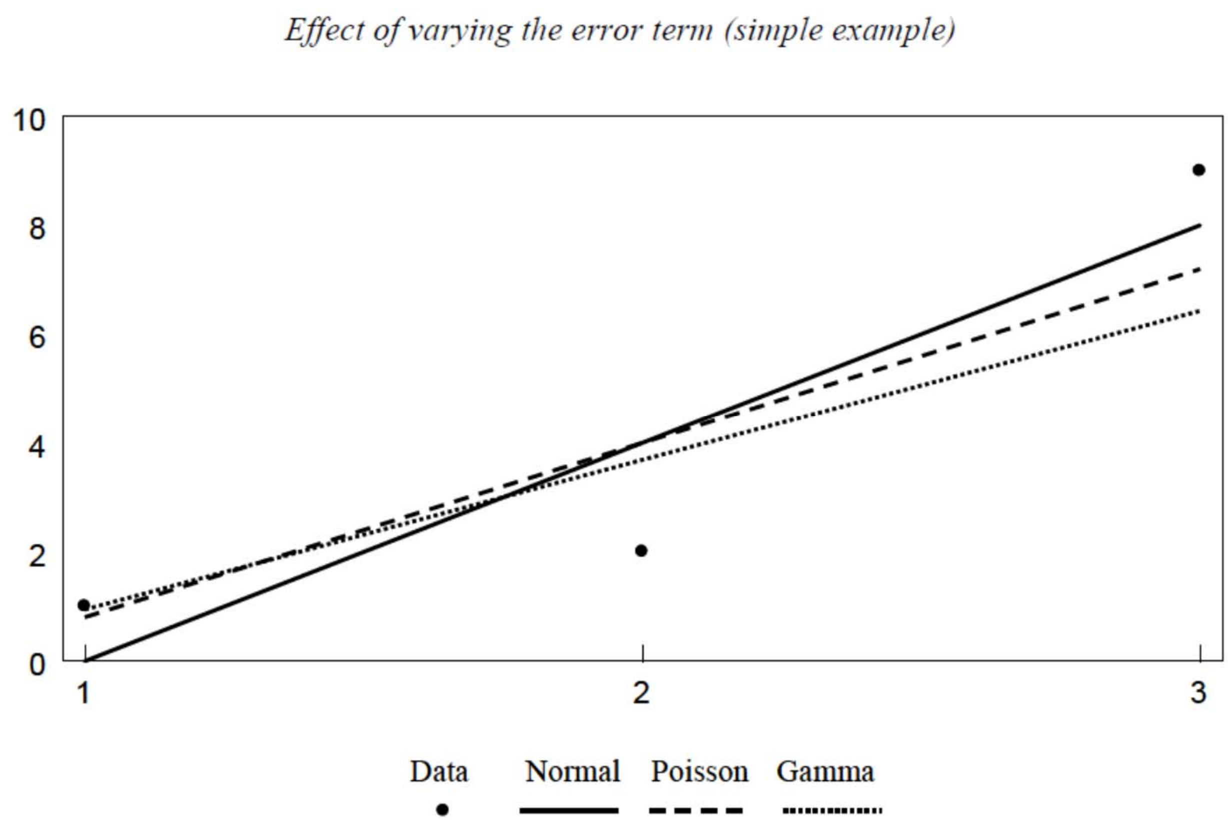
$$\theta_0 = \theta.\lambda^{1/(1-\alpha)}$$



Source: “A Practitioner's Guide to Generalized Linear Models”

TRAVELERS

Effect of Exponential Family Choice on Fit



Source: “A Practitioner's Guide to Generalized Linear Models”

Link Functions: How to Translate Linear Predictors

	$g(x)$	$g^{-1}(x)$
Identity	x	x
Log	$\ln(x)$	e^x
Logit	$\ln(x/(1-x))$	$e^x/(1+e^x)$
Reciprocal	$1/x$	$1/x$

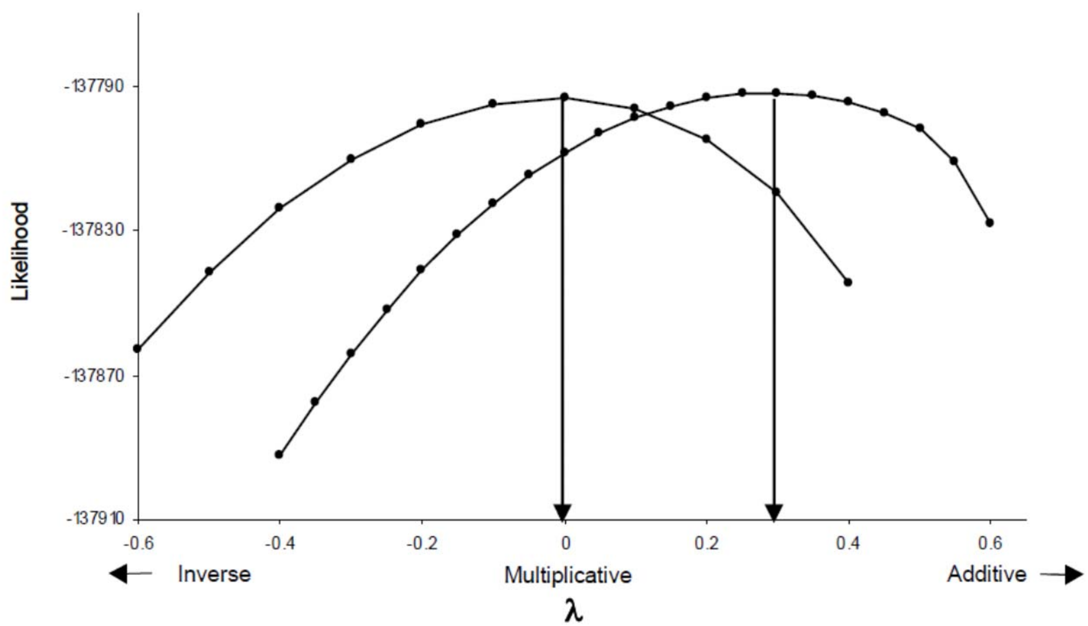
Source: “A Practitioner's Guide to Generalized Linear Models”

TRAVELERS

What Link to Choose?

$$g(x) = \begin{cases} \dfrac{(x^{\lambda} - 1)}{\lambda}, & \lambda \neq 0 \\ \ln(x), & \lambda = 0 \end{cases}$$

Box Cox transformation results on frequency



Source: “A Practitioner's Guide to Generalized Linear Models”

Maximum Likelihood Looks Pretty Ugly

$$f(y_i;\theta,\phi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}$$

$$l = \sum_i \left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}$$

$$0 = \frac{\partial l}{\partial \beta_j} = \sum_i \frac{\partial}{\partial \theta_i}\left(\frac{Y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right) \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_i}$$

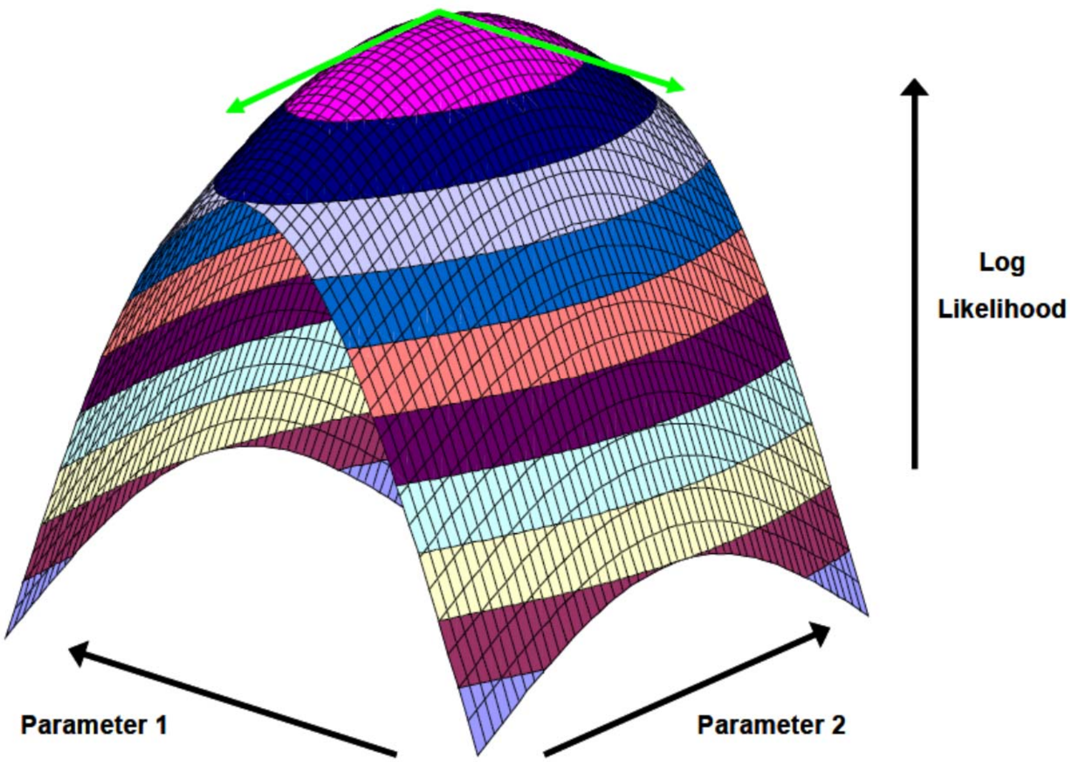
$$\frac{\partial l}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)}{a(\phi)} \cdot \frac{1}{b''(\theta_i)} \cdot \frac{1}{g'(\mu_i)} \cdot x_{ij}\ , \qquad j = 1,\ldots,p$$

$$= \sum_i \frac{\omega_i\,(y_i - \mu_i)\,x_{ij}}{\phi\,V(\mu_i)g'(\mu_i)}\,, \qquad j = 1,\ldots,p$$

Source: “A Practitioner's Guide to Generalized Linear Models”

TRAVELERS

Maximum Likelihood Never Looked So Pretty



SAS:

PROC GENMOD

PROG LOGISTIC

R:

glm {base}

Source: "A Practitioner's Guide to Generalized Linear Models"

TRAVELERS

Common Response Variables in Insurance

$\underline{Y}$	Claim frequencies	Claim numbers or counts	Average claim amounts	Probability (eg of renewing)
Link function $g(x)$	ln(x)	ln(x)	ln(x)	ln(x/(1-x))
Error	Poisson	Poisson	Gamma	Binomial
Scale parameter ϕ	1	1	Estimated	1
Variance function $V(x)$	x	x	x^2	x(1-x)*
Prior weights $\underline{\omega}$	Exposure	1	# of claims	1
Offset $\underline{\xi}$	0	ln(exposure)	0	0

Source: “A Practitioner's Guide to Generalized Linear Models”

TRAVELERS

References

- Generalized Linear Models in Insurance
  - “A Practitioner's Guide to Generalized Linear Models”
  - Casualty Actuarial Society Discussion Paper Program Casualty Actuarial Society - Arlington, Virginia (2004), p. 1-116
  - http://www.casact.org/pubs/dpp/dpp04/04dpp1.pdf
- Actuarial Exams and Career Information
  - http://www.beanactuary.org/
  - http://www.casact.org/
  - http://www.soa.org/
- Travelers Careers
  - http://www.travelers.com/careers
  - Research Internship and Full Time
  - Actuarial & Analytics Leadership Development Program – Internships and Full Time