www.ICA2014.org

LEARN
INTERACT
GROW

Predictive Modeling in Actuarial Science

Glenn Meyers
April, 3, 2014

What is New About Predictive Modeling?

- Is it just in the name?
  - Actuaries have been doing it for centuries.
- What is new in predictive modeling?
  - Better data.
  - Computers are widely available.
  - Additional Techniques.

The Development of Predictive Modeling in the USA – The Data

- US General Insurers have to report detailed data to regulators through a statistical agent.
  - Policy level
  - Individual claim level
- Why? – An interesting legal history.

Key Supreme Court Decision Paul v. Virginia - 1869

- Insurance is not commerce!
  - Not subject in antitrust laws.
  - Cartels controlled insurance rates.
- As a result, insurance rates came to be regulated by state insurance departments.
- Hence detailed data reporting.

Key Supreme Court Decision
U.S. v. Southeast Underwriters - 1944

- Insurance is commerce
  - Hence subject to antitrust laws.
- State rate regulation of cartels and data reporting was already well established.
- McCarran-Ferguson Act – 1945
  - Maintained state regulation and data reporting
    - Cartels are not permitted and insurers are able to set their own rates.

The American Situation in the 1950's and 1960's

- Insurers have standardized data.
- Insurers are getting computers.
- Insurers begin simple applications such as policy and claim data storage and simple manipulations of that data with tight restrictions on size.

The Start of Modern Predictive Modeling

- "Two Studies in Automobile Insurance Ratemaking"
  - Robert A. Bailey and LeRoy J. Simon (1960)
- "Insurance Rates with Minimum Bias"
  - Robert A. Bailey (1963)
    - Formulated cross classified models in modern statistical terms

Example – Bailey Additive Model

- Estimate relative loss ratios, r_{ij}, for use class i and merit rating class j.
- n_{ij} = earned car years for class ij
- Model – $r_{ij} = \alpha_i + \beta_j$

Estimating α_i and β_j

1. For each *j*, calculate the initial estimate β_j

$$\beta_j = \sum_i n_{ij} \cdot r_{ij}$$

2. For each *i*, use the "balance" criteria and solve for α_i

$$\sum_j n_{ij} \cdot \left(r_{ij} - \alpha_i - \beta_j \right) = 0$$

ICA 2014 CIA
WASHINGTON DC

Iterate on α_i and β_j until converges

- For each i, set $$\alpha_i = \frac{\sum_j n_{ij} \cdot \left(r_{ij} - \beta_j \right)}{\sum_j n_{ij}}$$

- For each j, set $$\beta_j = \frac{\sum_i n_{ij} \cdot \left(r_{ij} - \alpha_j \right)}{\sum_j n_{ij}}$$

A GLM Solution to the Same Problem

- Normal distribution with identity link

- Log likelihood function

$$L = \sum_{i,j} n_{ij} \cdot \left(r_{ij} - \alpha_i - \beta_j \right)^2$$

ICA 2014 CIA
WASHINGTON DC

Maximize the Log-Likelihood

$$\frac{\partial L}{\partial \beta_j} = 2 \cdot \sum_i n_{ij} \cdot \left(r_{ij} - \alpha_i - \beta_j \right) = 0 \quad \text{for all } j$$

$$\frac{\partial L}{\partial \alpha_j} = 2 \cdot \sum_j n_{ij} \cdot \left(r_{ij} - \alpha_i - \beta_j \right) = 0 \quad \text{for all } i$$

- Exactly the Bailey “balance” criteria.

Recognizing the Connection

- Brown (1988) – Zehnwirth(1994) – DeJong
- Mildenhall (1999)
  - For any given GLM, there is a set of weights (w_{ij}) for which

$$\sum_j w_{ij} \cdot \left(r_{ij} - \mu_{ij} \right) = 0 \;\; \forall i \text{ and } \sum_i w_{ij} \cdot \left(r_{ij} - \mu_{ij} \right) = 0 \;\; \forall j$$

ICA 2014 CIA
WASHINGTON DC

Advantages of GLMs

- Continuous independent variables
- Statistical diagnostics
- Etc.

ICA 2014 CIA
WASHINGTON DC

Moving Beyond Traditional Actuarial Problems

- Sales and Marketing
- Productivity Analysis
- Sales and Marketing
- Compensation Analysis

ICA 2014 CIA
WASHINGTON DC

Other Tools Besides GLM

- Generalized Additive Models
- Longitudinal Analysis
- Mixed Models
- Bayesian MCMC Models
- Spatial Analysis
- Unsupervised Models

Purpose of Volume 1

- Introduce tools that the editors think will be helpful for actuaries going forward.
- Cover a wide range of models in current use.
- Provide data and model coding when available and appropriate.