Comparing Machine Learning and Conventional Statistical Techniques In Claims Models

Douglas A. Noe, Ph.D.
Josh Brady, Ph.D., FCAS, MAAA
Luyang Fu, Ph.D., FCAS

The authors' affiliation with The Cincinnati Insurance Company is provided for identification purposes only and is not intended to convey or imply The Cincinnati Insurance Company's concurrence with or support for the positions, opinions or viewpoints expressed.

Agenda

- Claims Triage Models
- Beyond GLM: Regression Models
- Beyond GLM: Machine Learning models
- Case Study
- Conclusions

Claims Triage Model

Objectives:

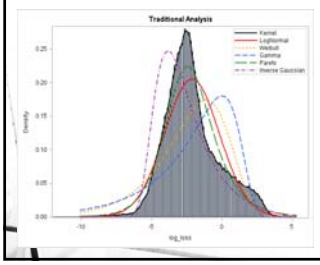
- Identify claims that have the potentials to blow up
- Allocate claims resources efficiently
- Manage claims to mitigate the loss

Challenges for model development:

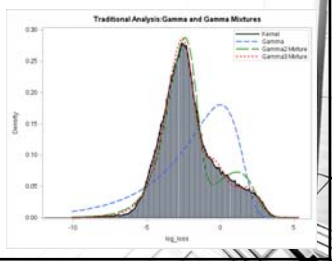
- Time stamps of variables
- Empirical loss distributions are different from exponential family
- Many possible interactions
- Unobserved heterogeneity
- Model validation and benefit quantification

WC Loss Distribution

Common-used distributions do not fit WC loss data well



Need mixtures to fit WC loss data better



Claims Triage Models: Beyond GLM

- Regression Models
  - Finite mixture models, double GLM, quantile regressions
  - Generalized mixed model (random and fixed effects)
  - Elastic nets, LASSO, Ridge
- Machine Learning Models
  - Support Vector Machine
  - Neural Networks
  - Random Forests
  - Gradient Boosted Trees (GBM)

Claims Triage Models: Beyond GLM

Case study to appear at the presentation.

Anti-Trust Notice

- The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.
- Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.
- It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Questions?

Mixed Effects Model Resources

General Sources

- Predictive Modeling Application in Actuarial Science Vol. 1 (Chapters 7, 8, 9, and 16): http://instruction.bus.wisc.edu/jfrees/jfreesbooks/PredictiveModelingVol1/index.htm
- http://www.ats.ucla.edu/stat/mult_pkg/glmm.htm

Mixed Models in R

- https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf
- http://lme4.r-forge.r-project.org/lMMwR/lrgprt.pdf
- http://rpsychologist.com/r-guide-longitudinal-lme-lmer

Elastic Net Resources

General Resources

- An Introduction to Statistical Learning (Ch. 6): http://www-bcf.usc.edu/~gareth/ISL/
- The Elements of Statistical Learning (Ch. 3): http://statweb.stanford.edu/~tibs/ElemStatLearn/

Fitting Elastic Nets in R

- http://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html

Neural Network Resources

General Resources:

- The Elements of Statistical Learning (Ch. 3): http://statweb.stanford.edu/~tibs/ElemStatLearn/
- http://mi.eng.cam.ac.uk/~mjfg/local/i10/i10_hand4.pdf
- http://neuralnetworksanddeeplearning.com/chap1.html

Neural Networks in R

- https://journal.r-project.org/archive/2010-1/RJournal_2010-1_Guenther+Fritsch.pdf
- http://datascienceplus.com/fitting-neural-network-in-r/
- http://bicorner.com/2015/05/13/neural-networks-using-r/

Decision Tree Resources

- Gradient Boosting:
  - Greedy Function Approximation: A Gradient Boosting Machine: https://statweb.stanford.edu/~jhf/ftp/trebst.pdf
  - https://cran.r-project.org/web/packages/gbm/gbm.pdf
- XGBoost (extreme gradient boosting)
  - https://xgboost.readthedocs.io/en/latest/
  - https://cran.r-project.org/web/packages/xgboost/xgboost.pdf
- Random Forest
  - Random Forests: https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf
  - https://cran.r-project.org/web/packages/randomForest/randomForest.pdf