Model Selection and Averaging of Health Costs in Episode Treatment Groups

Shujuan(Jane) Huang
University of Connecticut

Joint work with Brian M. Hartman (UConn) and
Vytaras Brazauskas (UW-Milwaukee)

Research supported by CIGNA and the Hickman Scholarship, Travel supported by the Graduate School at the University of Connecticut

49th Actuarial Research Conference, Santa Barbara, July 15, 2014

Outline

- Introduction
- Research Goal
- Review on Potential Models and Metrics
  - AIC/BIC Weights
  - Bayesian Inference and Parallel Model Selection
  - Random Forest
- Speed and Accuracy Comparison on all ETGs among Four Metrics
- Take Home Message

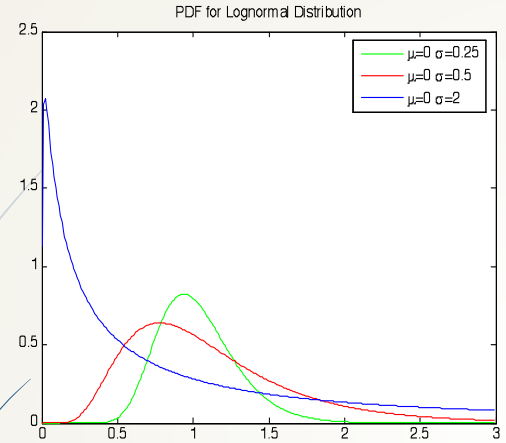
Introduction

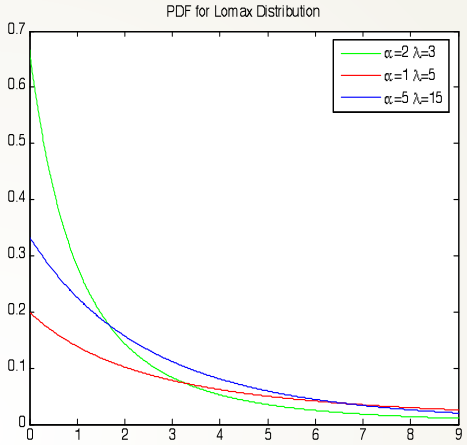
- Risk Assessment and Pricing for Health Insurance
- Episode Treatment Groups (ETG)
- Business Goal and Applications

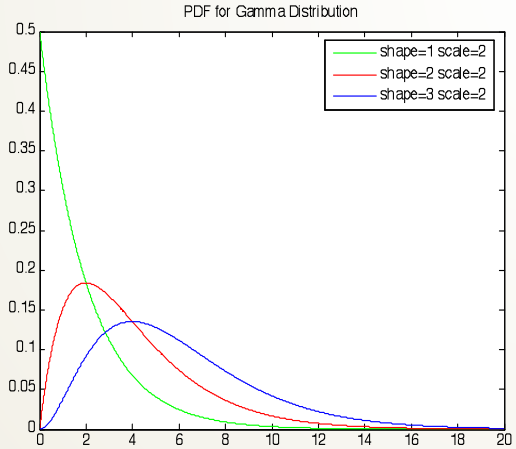
Research Goal

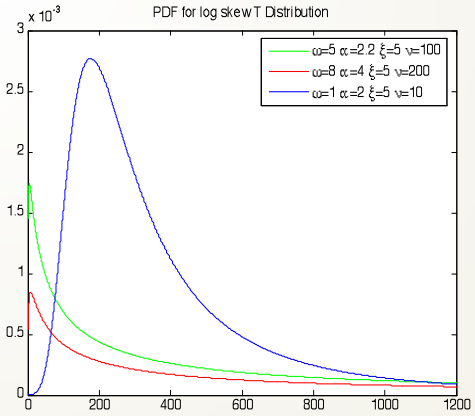
- **Model Selection vs. Model Averaging**
- **Metrics Comparison**
  - AIC and BIC weights (Most widely used)
  - Random Forest Feature Classification (Fastest)
  - Bayesian Model Averaging (Most theoretical accurate)
- **Potential Models**
  - Lognormal, Gamma, Log skew t, Lomax

Review of Potential Models









Data

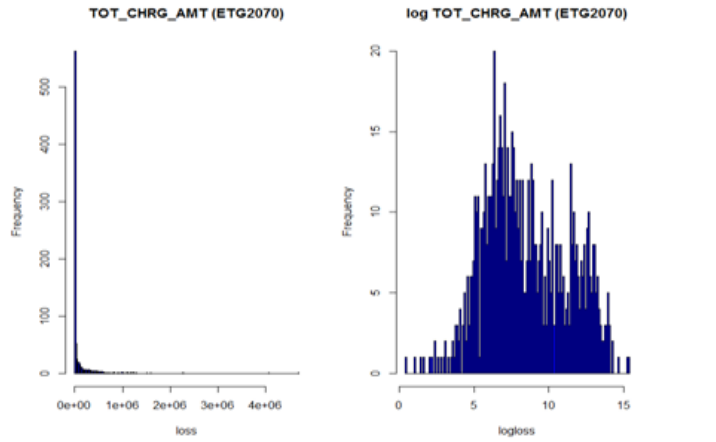
- 33 million samples from a major health insurer
- 8.98 million claimants
- Total Cost per claimant per year on each ETG
- Total ETGs: 347, Non-Routine ETGs: 320
- Only non-zero costs are recorded in the database

Episode Treatment Groups (ETG)

Sample ETGs Dictionary

ETG_NUM	ETG_DESC
1301	AIDS
1302	HIV sero-positive w/o AIDS
1304	Septicemia
1306	Other infectious diseases
1308	Immunodeficiencies
1399	Infectious diseases signs & symptoms
1620	Lipidoses (Gauchers Disease, Fabry Disease, Mucolipidosis I-III)
1621	Hyper-functioning thyroid gland
1622	Hypo-functioning thyroid gland
1623	Non-toxic goiter
1624	Malignant neoplasm of thyroid gland
1625	Non-malignant neoplasm of thyroid gland
1626	Other diseases of thyroid gland
1630	Diabetes

Histogram for Total Charged Amount on individual and annual basis for Sample ETG 2070



Introduction to Metrics (Part 1): AIC/BIC

- **Akaike information criterion (AIC)/Bayesian information criterion(BIC)**
- **AIC/BIC weights**
- **Results on sample ETG:**

AIC				BIC			
lognormal	gamma	log skew T	lomax	lognormal	gamma	log skew T	lomax
5361.11	5468.30	5365.07	5365.11	**5368.24**	5475.43	5372.20	5379.37

AIC weights				BIC weights			
lognormal	gamma	log skew T	lomax	lognormal	gamma	log skew T	lomax
0.785	0.000	**0.109**	**0.106**	**0.876**	0.000	**0.121**	**0.003**

Introduction to Four Metrics (Part 2): Bayesian Parallel Model Selection

Gives the probabilities of each model being the best given the data and all models under consideration, enabling model averaging and providing deeper insights into the relationships between the models.

	Lognormal	Folded T	Gamma	Log skew T
Hemophilia (ETG2070)	100%	0%	0%	0%
Chronic Renal Failure(ETG5554)	0%	0%	0%	100%
Hyper-functioning adrenal gland (ETG1635)	0%	0.0015%	0%	99.9985%
Personality Disorder (ETG2394)	68.74%	0%	0%	31.26%

Introduction to Four Metrics (Part 3): Random Forest Feature Classification

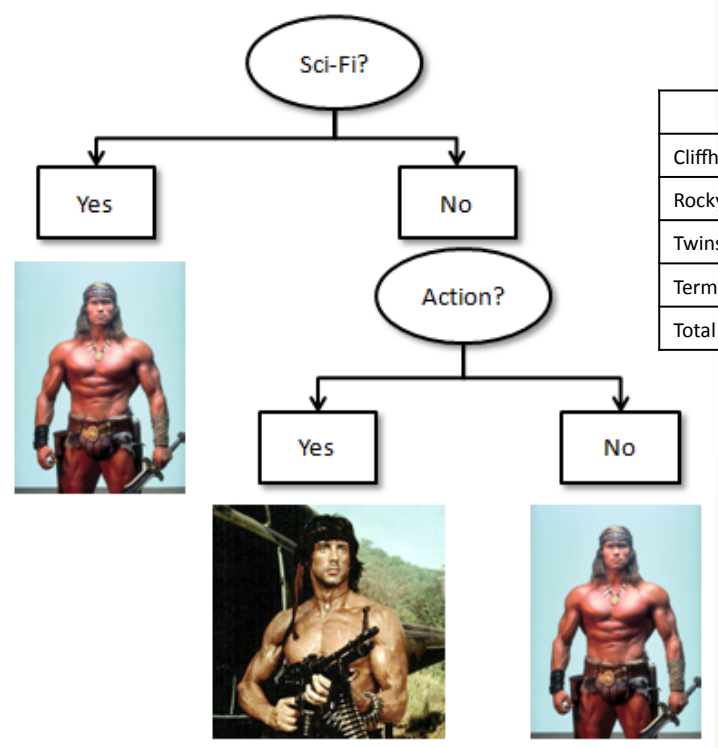
Model selection using database characteristics:
Schwartz and Bradlow(2013)

Random forests: An ensemble learning method for classification (and regression) that operate by constructing a multitude of decision trees.

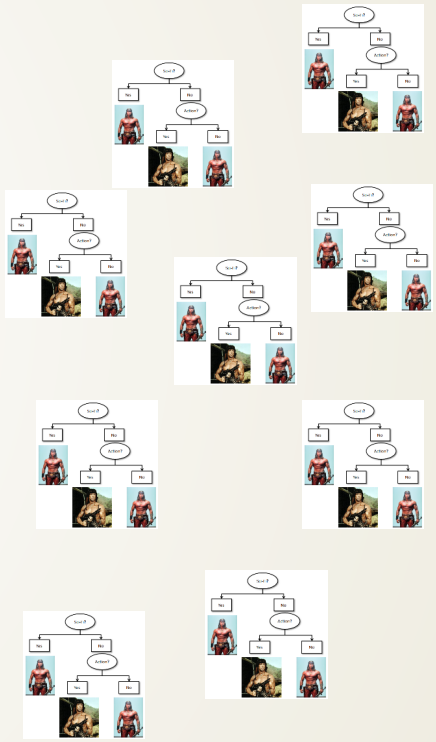
Features:

- Efficiency
- Thousands of input variables without variable deletion
- Missing data
- Prediction

Decision Tree Classification & Random Forest



	Action	Sci-Fi	Actor
Cliffhanger	Yes	No	Stallone
Rocky	Yes	No	Stallone
Twins	No	No	Schwarzenegger
Terminator	Yes	Yes	Schwarzenegger
Total Recall	Yes	Yes	Schwarzenegger



Procedures for Random Forest Model Selection

- Step 1: Domain Specific Feature Extraction

  (mean, variance, skewness, kurtosis, q1, q3, median…)

- Step 2: Train the Random Forest for Prediction

- Step 3: Random Forest Model Selection

ETG_NUM	lognormal	gamma	log skew T	lomax	winner
1301	0.11275	0.24475	0.537	0.1055	log skew T
1302	0.2415	0.211	0.415	0.1325	log skew T
1304	0.9665	0.0105	0.0165	0.0065	lognormal
1306	0.0045	0.00125	0.93225	0.062	log skew T
1308	0.89925	0.001	0.00675	0.093	lognormal

Classification Results on Training and Testing Data

Classification results on training data		
	OOB estimate of error rate	List of distributions
4 distributions	0.25%	lognormal, gamma, log skew T, Lomax
3 distributions	0.00%	lognormal, gamma, Lomax

Classification results on testing data			
	Distributions	Misclassification Rate	Details
Test 1	Four distributions: lognormal, gamma, log skew T, Lomax	23.8%	500 simulated datasets for each distribution (2000 in all)
Test 2	Three distributions: lognormal, gamma, Lomax	1.2%	500 simulated datasets for each distribution (1500 in all)

Speed Comparison on all ETGs among Four Metrics

Models	Time
Random Forest	About 2 minutes
AIC and BIC	About 4 hours
Bayesian	About 4 weeks

Comments: 8 GB Thinkpad with a 2.50 GHz Intel Quad-Core processor

Accuracy Comparison on Simulated Datasets among Three Metrics

	Random Forest				Bayesian				AIC weights			
dataset	lognormal	gamma	log skew T	LOMAX	lognormal	gamma	log skew T	LOMAX	lognormal	gamma	log skew T	LOMAX
1-lognormal	99.70%	0.00%	0.10%	0.20%	100.00%	0.00%	0.00%	0.00%	75.81%	0.00%	24.19%	0.00%
2-gamma	11.30%	62.75%	15.00%	10.95%	1.90%	93.90%	3.14%	1.06%	0.00%	94.42%	5.58%	0.00%
3-log skew T	0.08%	0.03%	67.58%	32.33%	0.00%	0.00%	100.00%	0.00%	0.00%	0.00%	93.91%	6.09%
4-lomax	0.03%	0.00%	43.98%	56.00%	0.23%	0.00%	38.54%	61.23%	0.00%	0.00%	27.81%	72.19%

Model Selection Results Comparison between RF and AIC for all 320 ETGs

	Lognormal	Gamma	Lomax	log Skew T
AIC	24	0	4	292
RF	131	0	27	162

Take Home Message

- Model Selection, Model Averaging, Feature Classification
- Suggested model
  - Depends on the data
  - When data is extremely heavy tailed, Gamma may not be a good choice
- Suggested metrics:
  - Small datasets: Bayesian Model Averaging, AIC/BIC
  - Large datasets: Random Forests

Thank you! Questions?

Contact Information:
Shujuan.Huang@uconn.edu

Prior Distributions Setting

Distributions	Priors	Thinned Samples Per Chain	Burning Sample Per Chain
lognormal	$\tau\sim$ Gamma(4,4.5) $\mu\sim$ Norm(6,5)	30,000	20,000
gamma	$\tau\sim$ Gamma(2,3) $\nu\sim$ Exp(ω) $\omega\sim$Uniform(0.01,10)	50,000	35,000
Log skew T	$\theta\sim$Norm(0,5) $\Omega\sim$ Inverse Gamma(6,1) $\alpha\sim$Norm(50,4) $\xi\sim$ Norm(θ, 7) $\nu\sim$dexp(0.25)	300,000	260,000
Lomax	$\lambda\sim$Gamma(2,3) $\alpha\sim$ Exp(ω) $\omega\sim$unif(0.01,10)	300,000	20,000