CAS Spring Meeting - Session C3
Bayesian Analysis with
Monte-Carlo Markov Chain Methods

Glenn Meyers
FCAS, MAAA, Ph.D.
Vice President – Research
ISO Innovative Analytics
May 25, 2010

DATA • ANALYTICS • DECISION SUPPORT

ISO Innovative Analytics

ISO

Perspectives

- **Actuarial Perspective**
  - Credibility – Approximation to Bayesian Analysis
- **Statistical Perspective**
  - Pure Bayesian Analysis
  - Empirical Bayesian Analysis
- **Statisticians and Actuaries came together in 80's – Empirical Bayesian Credibility**
  - Buhlmann and Straub – "Credibility for Loss Ratios"
  - Efron and Morris – "Stein's Paradox in Statistics"

Bayesian Analysis

$$p(\mu | x) = \frac{f(x | \mu) \cdot g(\mu)}{\int_0^{\infty} f(x | \mu) \cdot g(\mu) d\mu}$$

- *x* – observation
- *f* – conditional distribution of x given μ
- *g* – prior distribution of μ
- *p* – posterior distribution
- Integral difficult to evaluate in multi-parameter situations.

Markov Chain

- A discrete random process which can be in various states, and which changes randomly in discrete steps.
- Transition probabilities to next state depends only on current state.
- Under “conditions” (irreducible and aperiodic) the time spent in a give state converges to an equilibrium distribution.

Markov Chain Monte Carlo Methods in Bayesian Analysis

- **Transition probabilities determined by *f* and *g***
  - Gibbs sampler
  - Metropolis Hastings algorithm
- **Equilibrium distribution is the posterior!!!**
- **MCMC Methods became popular in the 90's**
  - Start with a guess at μ, and simulate a Markov-Chain
  - Ignore first (thousand or so) states – "burning period"
- **WINBUGS/COTOR Challenge**

Gibbs Sampler on a Lognormal
Example from February 2008 Actuarial Review

- Simulate μ from the prior distribution.
- Calculate the likelihood of the data with μ and previous σ.
- Select uniform(0,1) random U.
- Accept μ if $\frac{\text{Likelihood}}{\text{Maximum Likelihood}} > U$
- If otherwise, start over.
- Next iteration, switch role of μ and σ.

Figure 1

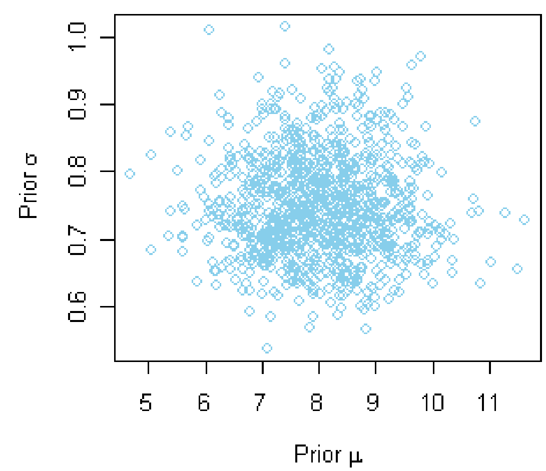
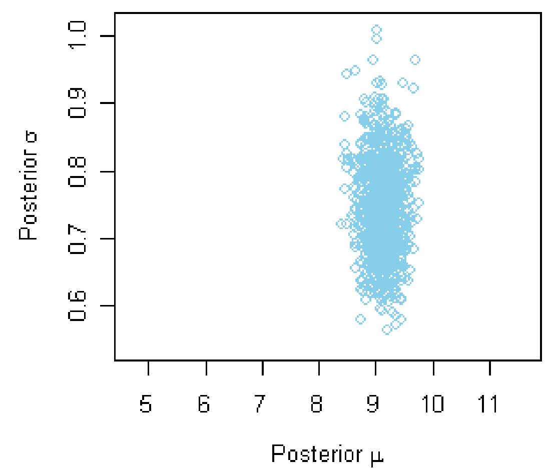


Figure 2



Posterior Distribution of μ and σ is Only of Temporary Interest!

- Most often we are interested in functions of μ and σ.
- For example:

Mean

$$e^{\mu+\sigma^2/2}$$

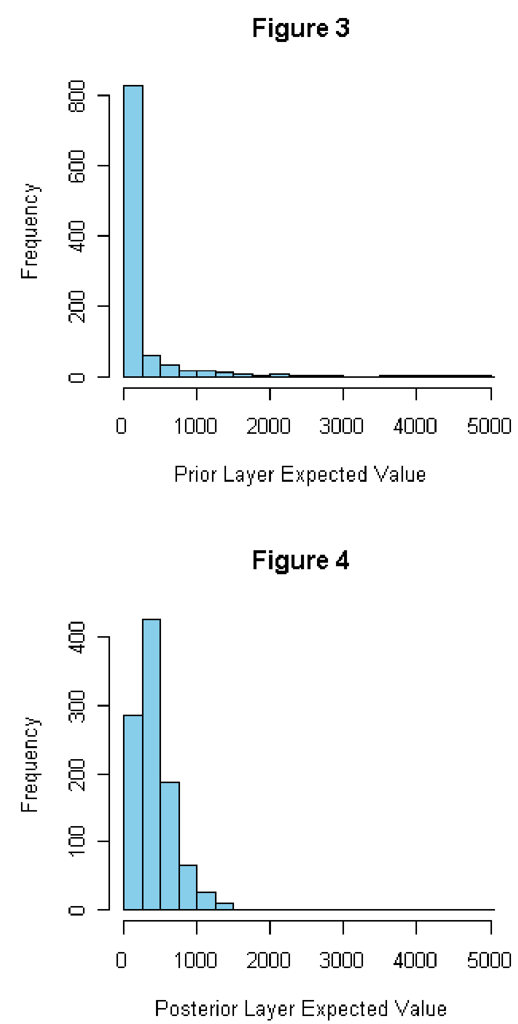
Limited Expected Value

$$e^{\mu+\sigma^2/2}\cdot\Phi\left(\frac{\log(L)-\mu-\sigma^2}{\sigma}\right)+L\cdot\left(1-\Phi\left(\frac{\log(L)-\mu}{\sigma}\right)\right)$$

Layer Expected Value
25,000 to 30,000

- Some posterior parameters generated by Gibbs sampler

μ	σ	LEV
9.194	0.723	392
9.206	0.708	383
8.817	0.707	119
8.944	0.644	120
9.461	0.785	836
9.150	0.651	252
9.043	0.739	280
9.240	0.773	514
9.392	0.863	845
9.018	0.781	311



The Metropolis-Hastings Algorithm

1. Select a random candidate value, μ^* from a proposal density function $p\left(\mu^* \mid \mu_{t-1}\right) = \Gamma\left(\mu^* \mid \mu_{t-1} / \alpha_p, \alpha_p\right)$

2. Compute the ratio $R = \dfrac{f\left(x \mid \mu^*\right) \cdot g\left(\mu^*\right) \cdot p\left(\mu^* \mid \mu_{t-1}\right)}{f\left(x \mid \mu_{t-1}\right) \cdot g\left(\mu_{t-1}\right) \cdot p\left(\mu_{t-1} \mid \mu^*\right)}$

   – *f* comes from the modeled distribution
   – *g* is the prior distribution
   – *f*·*g* is the posterior distribution

3. Select a random *U* from a uniform distribution on (0,1).
4. If $U < R$ then set $\mu_t = \mu^*$. Otherwise set $\mu_t = \mu_{t-1}$.

Introducing the proposal density function can keep big jumps from getting into the random walk

Simple Example

- $Y \sim$ Tweedie($\phi = 1$, $p = 1.5$, μ unknown)
- 25 observed losses

y	0	1	2	3	5	8	10	12	16
Freq	8	6	2	2	2	1	1	1	2

- Prior distribution $g(\mu) = \Gamma(\mu|\alpha = 1, \theta = 5)$
  - (prior mean=5)
- Proposal density function (mean = μ_{t-1})

$$p\left(\mu^* \mid \mu_{t-1}\right) = \Gamma\left(\mu^* \mid \mu_{t-1} / \alpha, \alpha\right)$$

Single Variable Example of Tuning the Metropolis-Hastings Algorithm

Figure 2 - $\alpha_p = 2{,}500$

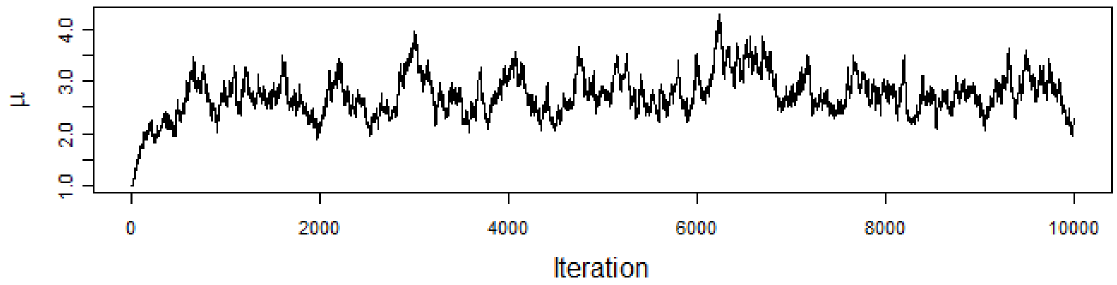


Figure 3 - $\alpha_p = 25$

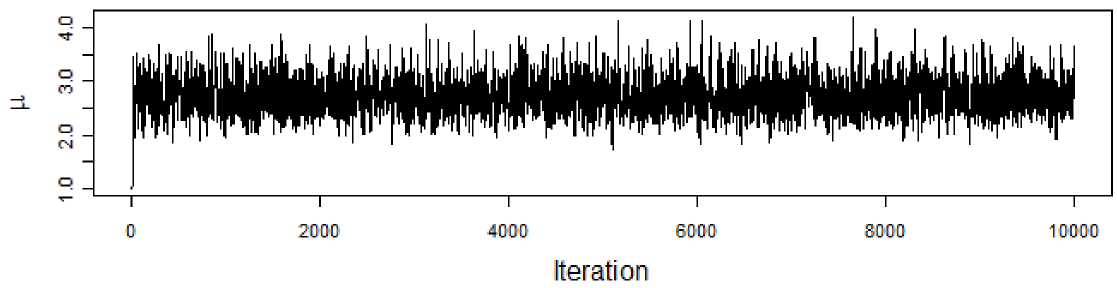
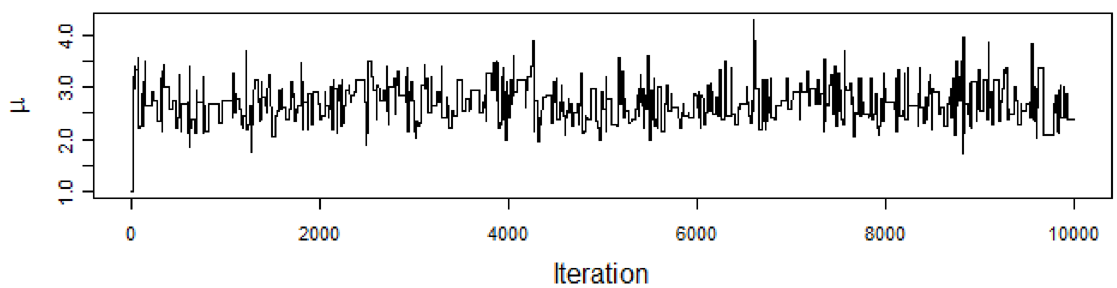
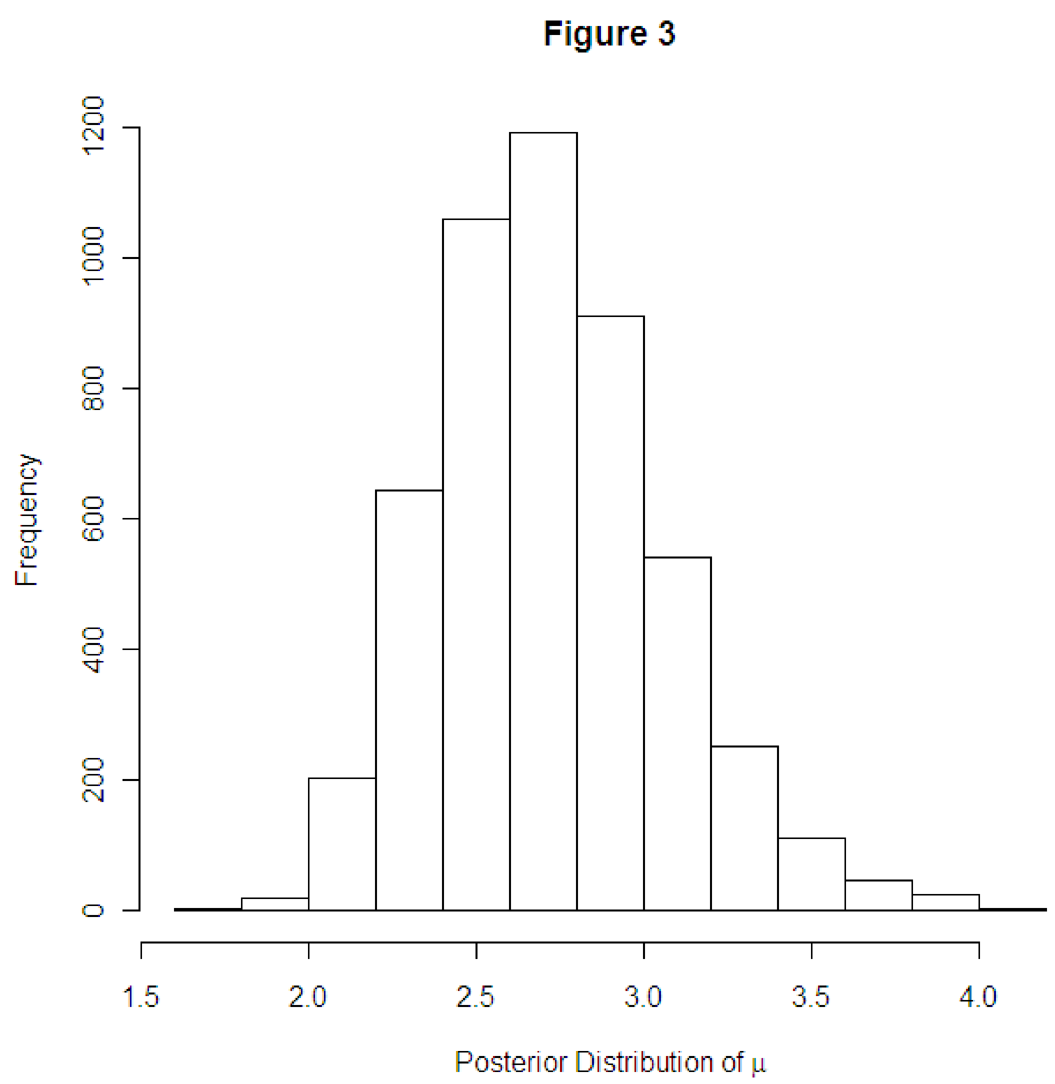


Figure 4 - $\alpha_p = 0.25$



Posterior Distribution of μ

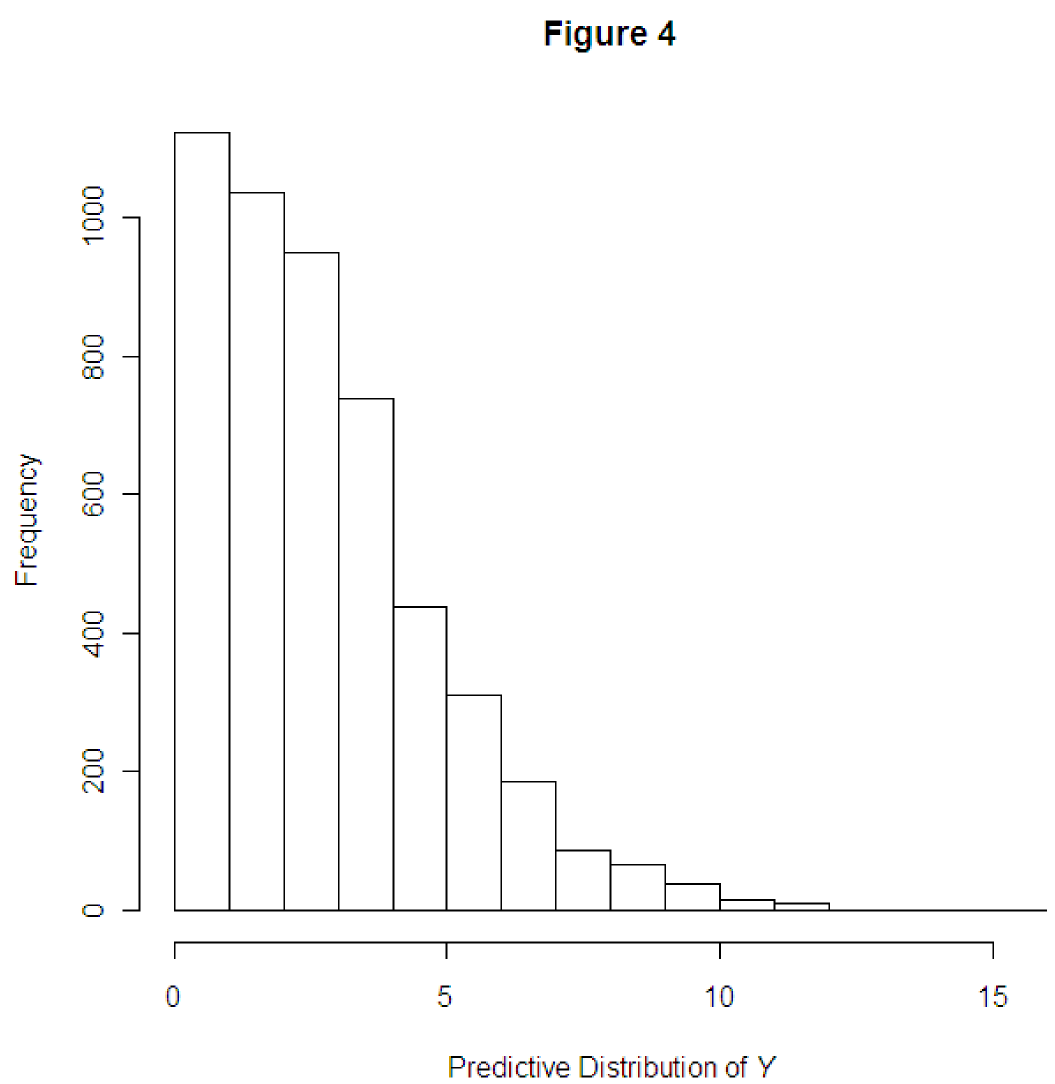


Often
Posterior Distribution is Not the Desired Output

- For each Tweedie(ϕ = 1, p = 1.5, μ_t)
  - Simulate outcome Y_t

- Distribution of Y is called the predictive distribution.

Predictive Distribution of *Y*



Recall Data

- $Y \sim$ Tweedie($\phi = 1$, $p = 1.5$, μ unknown)
- 25 observed losses

y	0	1	2	3	5	8	10	12	16
Freq	8	6	2	2	2	1	1	1	2

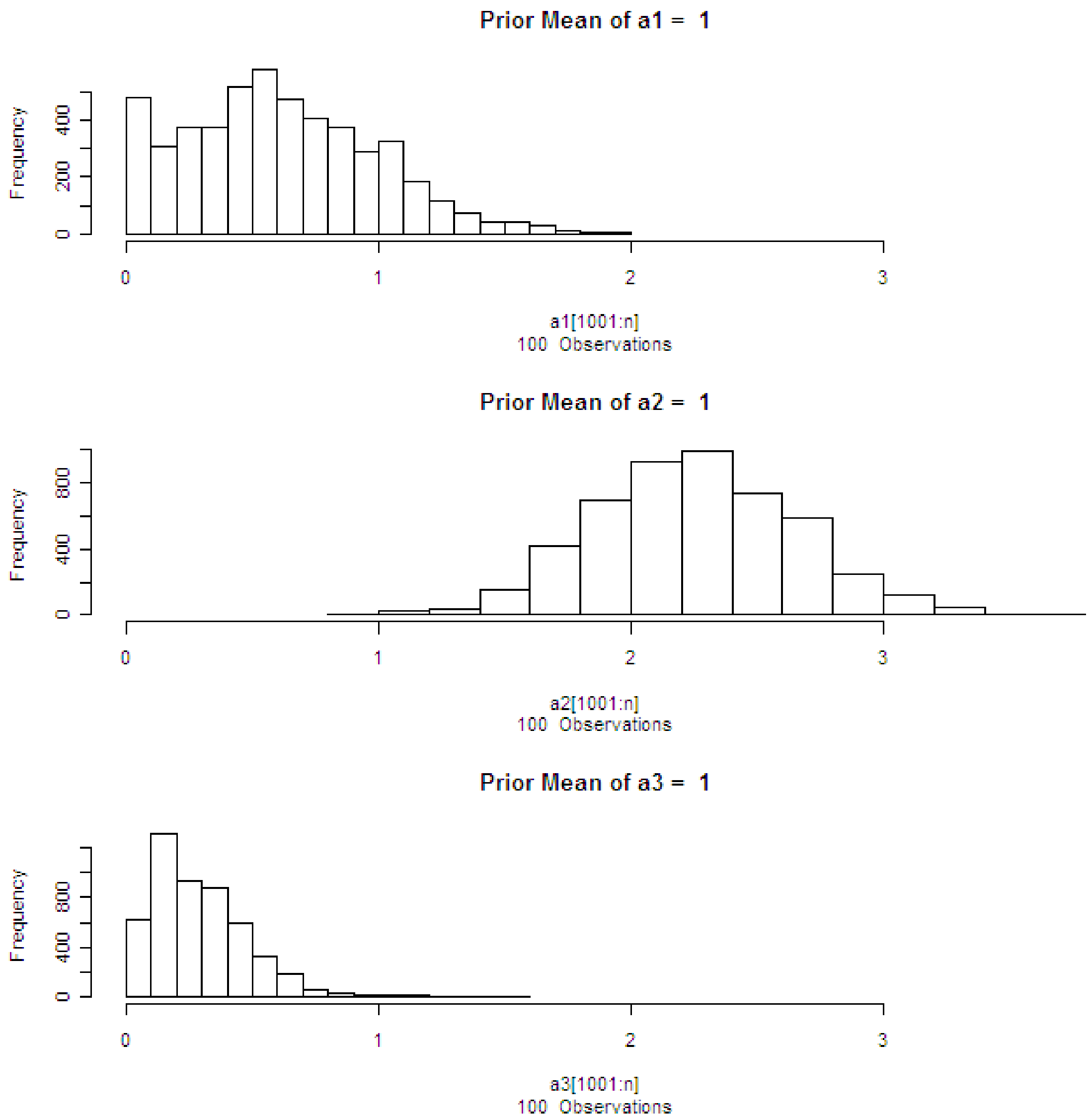
- Prior distribution $g(\mu) = \Gamma(\mu|\alpha = 1, \theta = 5)$

Bayesian Regression Example

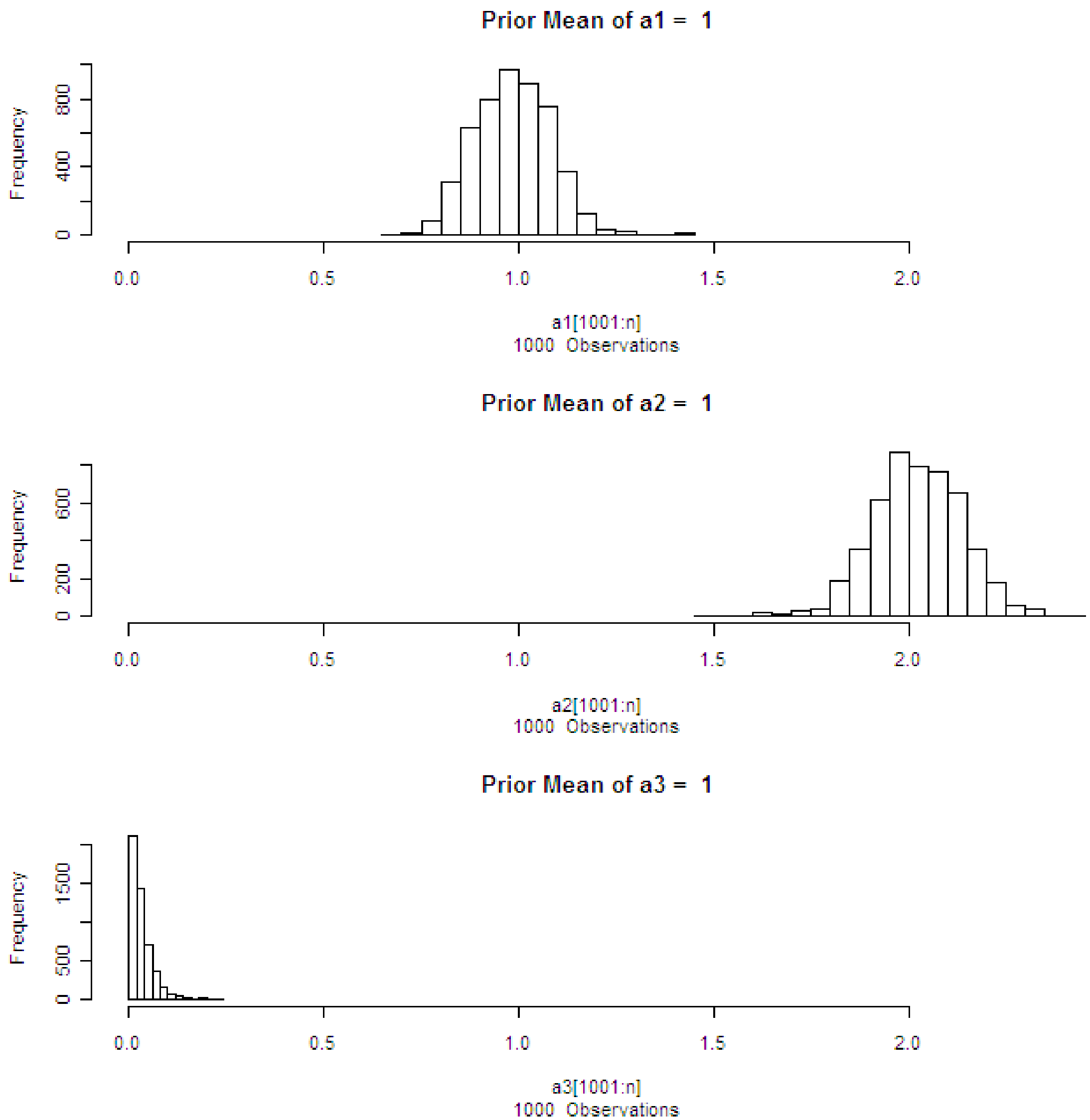
- $Y \sim$ Tweedie($\phi = 1$, $p = 1.5$, μ unknown)
- "Observed" losses simulated from Tweedie
- $\mu = x_1 + 2 \cdot x_2$ - "True" relationship (simulated)
- Model $\mu = a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3$
- Prior distribution $g(a_i) = \Gamma(a_i | \alpha = 1, \theta = 1)$
  - For $i = 1, 2, 3$ (prior mean = 1)
- Proposal density function (mean = $a_{i,t-1}$)

$$p\left(a_i^* \mid a_{i,t-1}\right) = \Gamma\left(a_i^* \mid \alpha = 25, \theta = a_{i,t-1} / 25\right)$$

Posterior with 100 Observations



Posterior with 1,000 Observations



A Non-Linear Regression Example

- Claim Count $N_i \sim$ Poisson ($\lambda = a_0 + a_1 \cdot d_i$)
- Claim Severity $Z_{ij} \sim \Gamma(\alpha,\theta)$

$$X_i = \sum_{j=1}^{N_i} Z_{ij}$$

- a_0, a_1, α and θ are unknown parameters
- Fit model with observed loss X_i and covariate d_i.

A Non-Linear Regression Example

- X ~ Tweedie (μ, p, ϕ)

- With
  - $\mu = \lambda \cdot \alpha \cdot \theta$ (with $\lambda = a_0 + a_1 \cdot d_i$)
  - $p = \dfrac{\alpha + 2}{\alpha + 1}$
  - $\phi = \dfrac{\mu^{1-p} \cdot \lambda}{2 - p}$

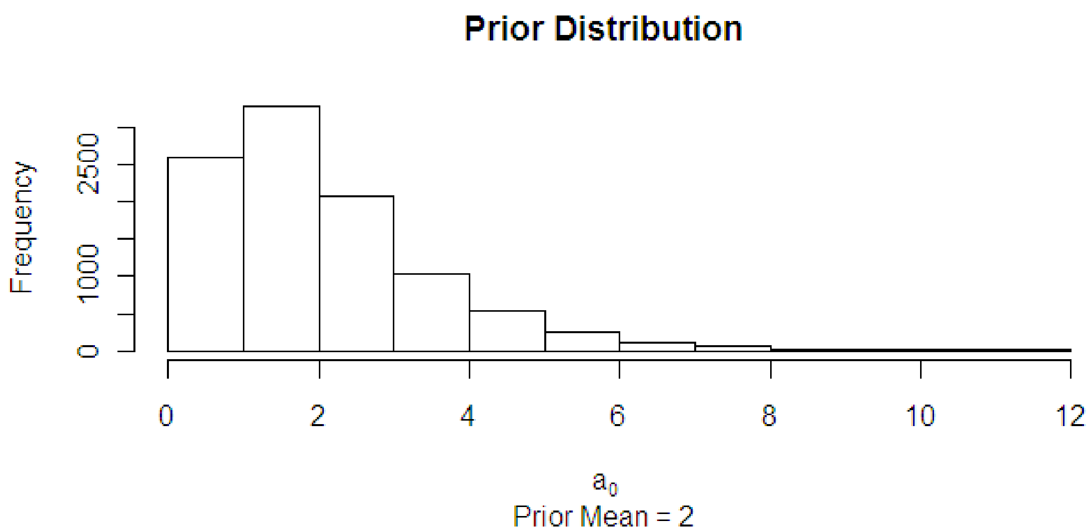
- So given any a_0, a_1, α and θ we can calculate the likelihood of the data for the Tweedie distribution

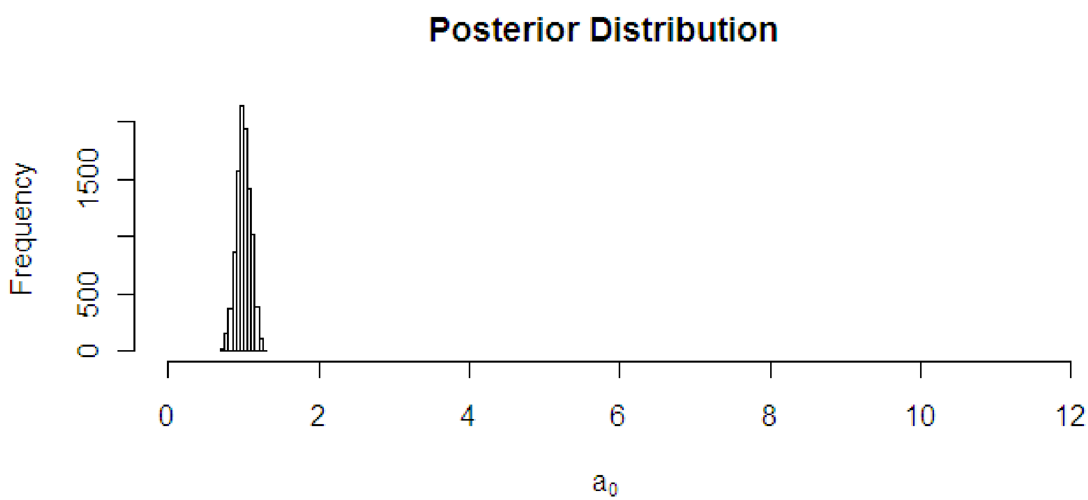
Test Model with Simulated Data

- Simulated Data
  - $N_i \sim \text{Poisson}(\lambda = a_0 + a_1 \cdot d_i) = \text{Poisson}(\lambda = 1 + 3d_i)$
  - $Z_{ij} \sim \Gamma(\alpha = 3, \theta = 5)$
- Prior Distributions
  - $\text{Prior}(a_0) = \text{Prior}(a_1) = \Gamma(\alpha = 2, \theta = 1)$ (Prior Mean = 2)
  - $\text{Prior}(\alpha) = \text{Prior}(\theta) = \Gamma(\alpha = 4, \theta = 1)$ (Prior Mean = 4)
- Proposal Density Function

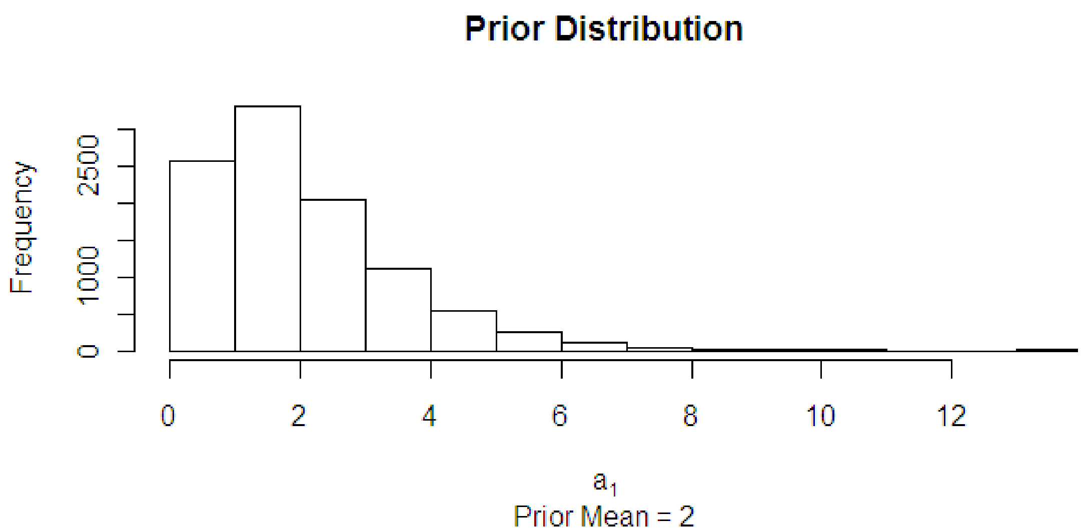
$$p\left(parm_i^* \mid parm_{i,t-1}\right) = \Gamma\left(parm_i^* \mid \alpha = 500, \theta = parm_{i,t-1} / 500\right)$$

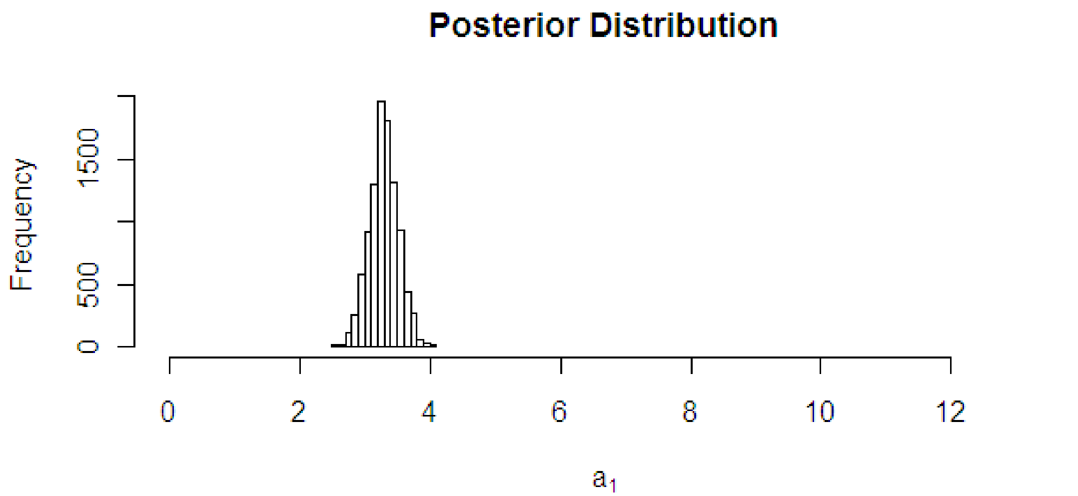
A Non-Linear Regression Model
1,000 Observations



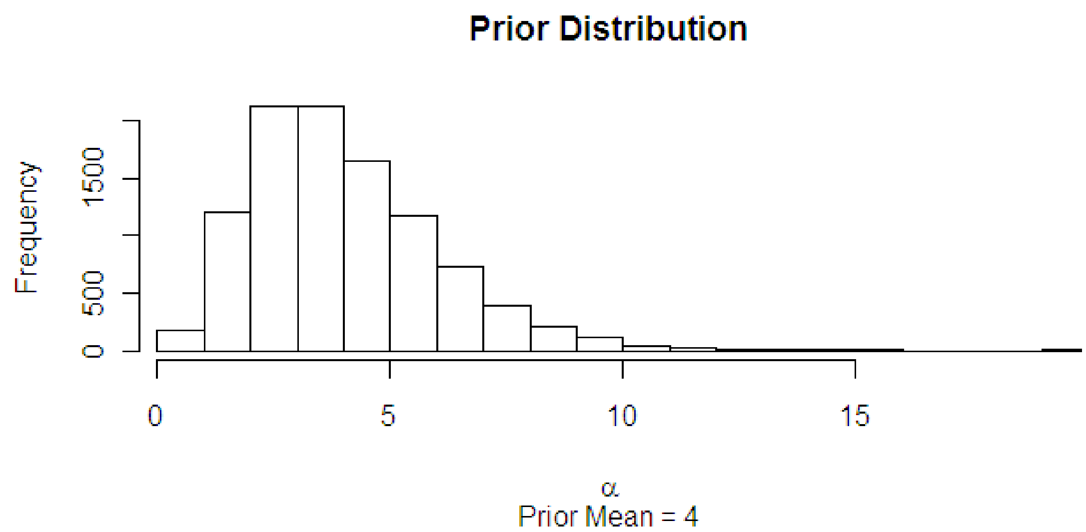


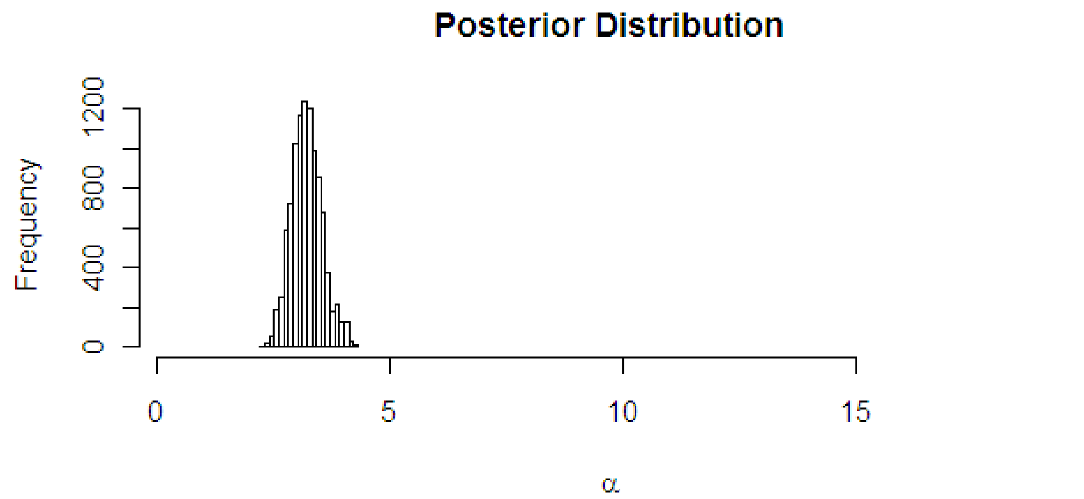
A Non-Linear Regression Model
1,000 Observations



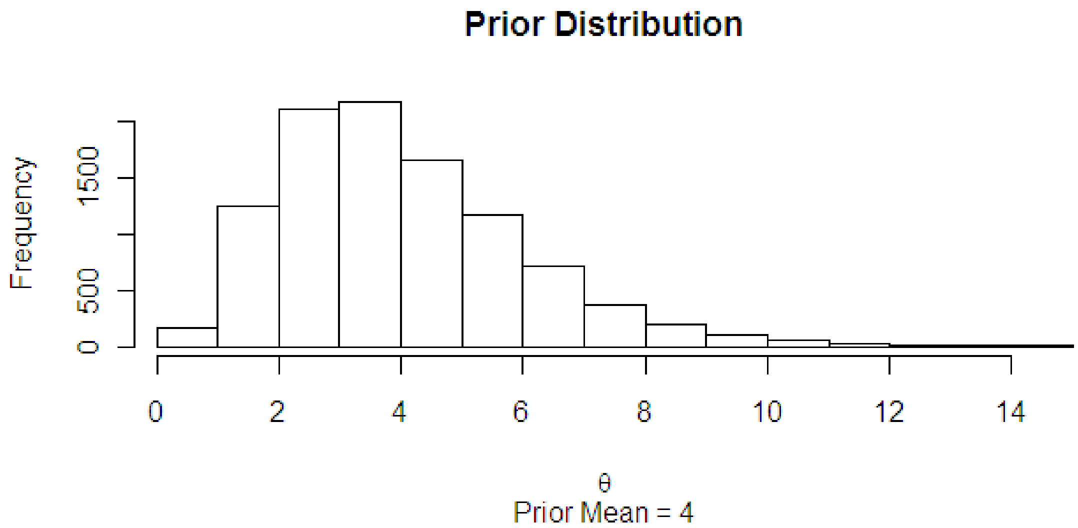


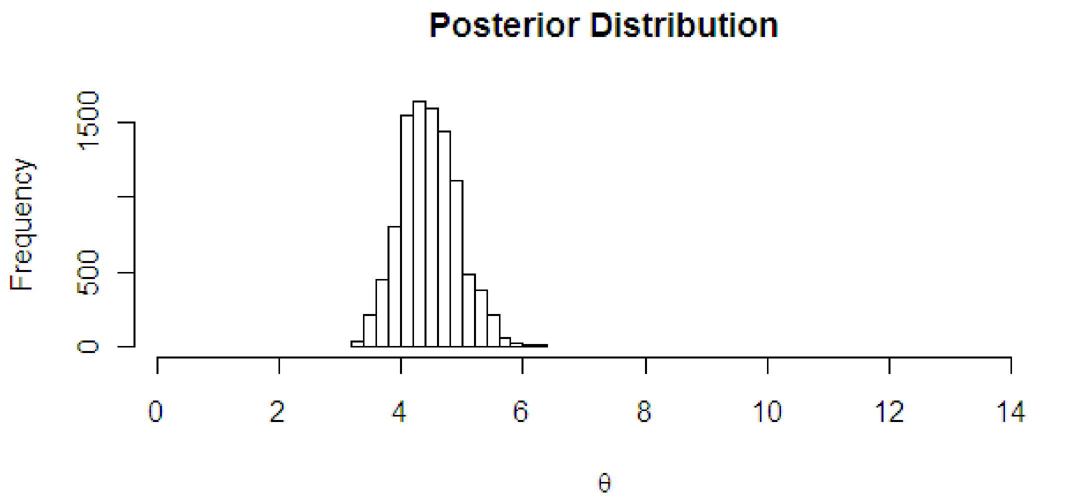
A Non-Linear Regression Model
1,000 Observations





A Non-Linear Regression Model
1,000 Observations



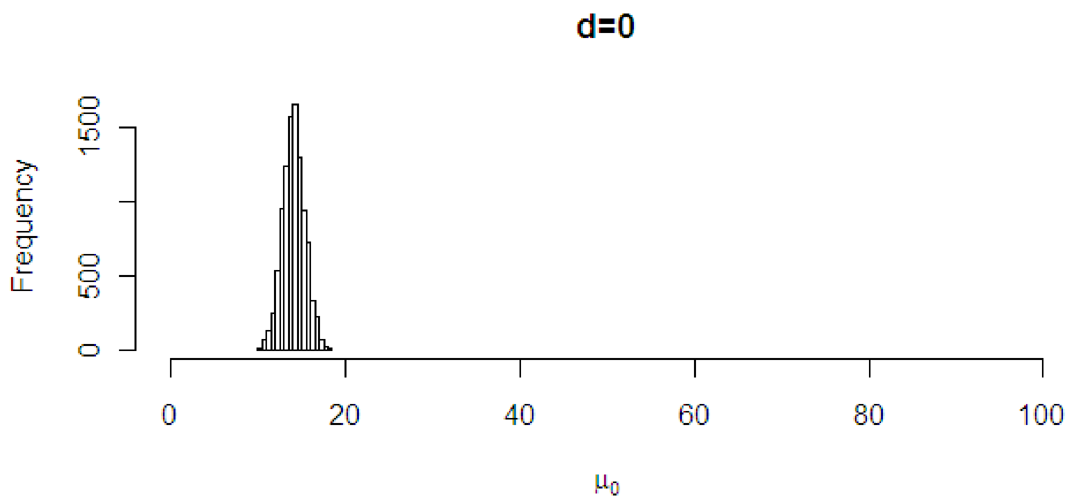


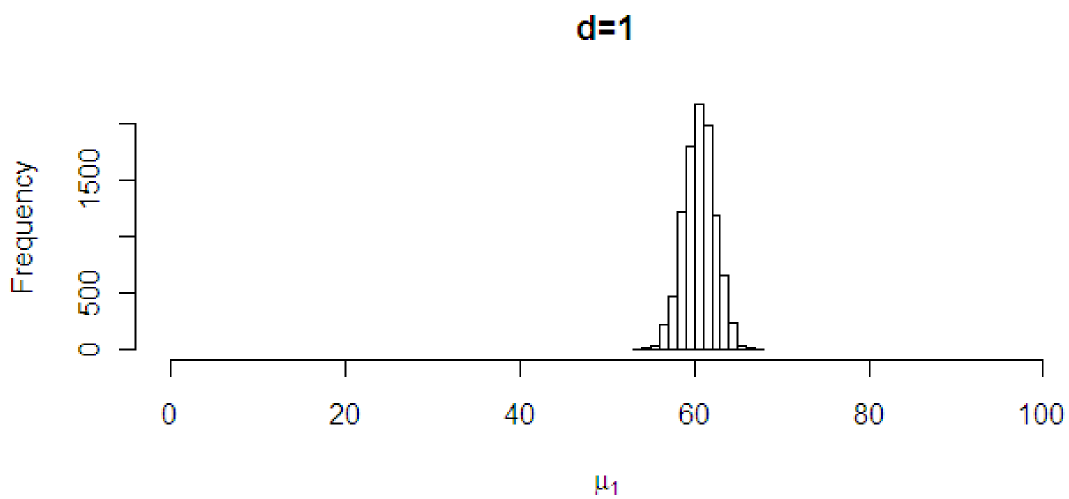
Statistic of Interest – Variability of $\mu|d$

- Create a histogram of μ for a given value of *d*.

- Produce histogram for d = 0 and d = 1

A Non-Linear Regression Model
Range of Estimates Given 1,000 Observations

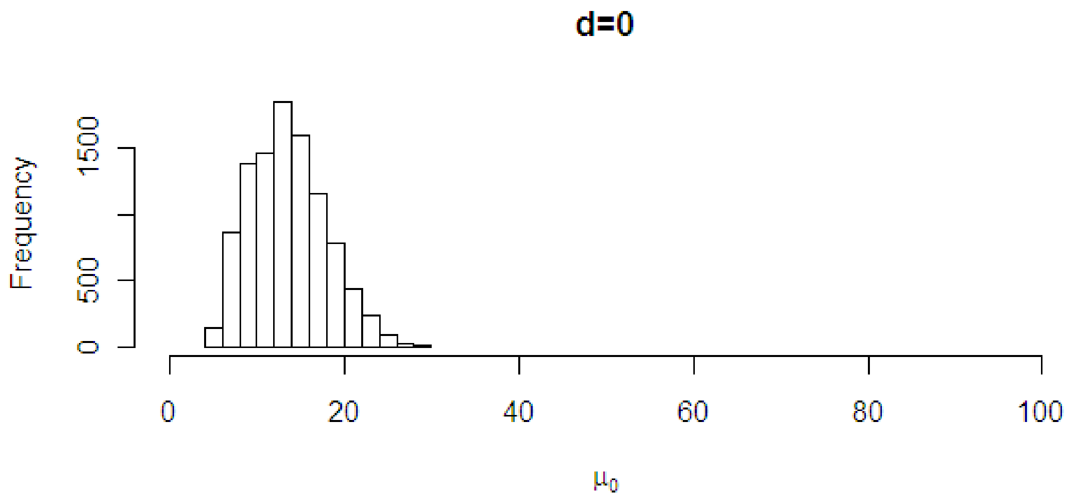


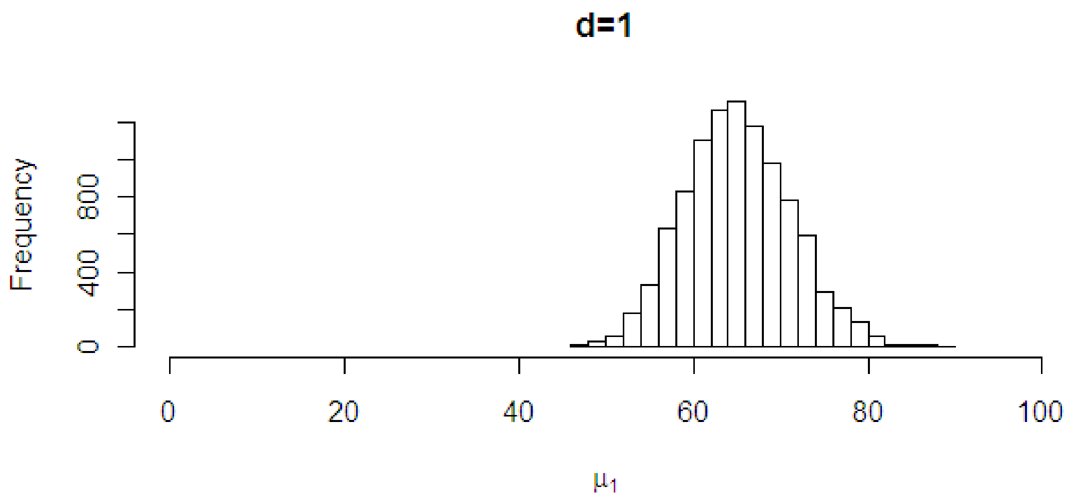


Claim Count N_i ~
Poisson ($\lambda = a_0 + a_1 \cdot d_i$)

Claim Severity Z_{ij} ~
$\Gamma(\alpha,\theta)$

A Non-Linear Regression Model
Range of Estimates Given 100 Observations





A Real Application – Loss Reserving

S&P Report, November 2003

Insurance Actuaries – A Crisis in Credibility

“Actuaries are signing off on reserves that turn out to be wildly inaccurate.”

Method Illustrated on Data

Incremental Paid Losses

AY	Premium	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7	Lag 8	Lag 9	Lag 10
1	29,701	5,234	5,172	3,708	1,783	923	537	175	145	8	0
2	27,526	5,234	5,683	4,392	2,134	1,377	673	155	81	47	-
3	30,750	5,702	5,865	7,966	2,472	NA	143	152	73	-	-
4	35,814	6,349	4,611	3,959	2,522	1,924	622	206	-	-	-
5	42,277	8,377	6,890	4,055	3,795	1,292	1,422	-	-	-	-
6	50,088	9,291	13,836	12,441	4,086	2,293	-	-	-	-	-
7	56,921	12,029	12,462	8,369	7,034	-	-	-	-	-	-
8	61,406	13,119	12,618	9,117	-	-	-	-	-	-	-
9	67,983	15,860	14,893	-	-	-	-	-	-	-	-
10	73,359	16,498	-	-	-	-	-	-	-	-	-

54 observations – 45 unknown cells

Loss Model

- Expected Loss

$$\mu_{AY,Lag} = Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag} \cdot t^{AY+Lag-1}$$

- Variance of Loss

$$Var\left[X_{AY,Lag}\right] = \mu_{AY,Lag} \cdot \tau_{Lag} \cdot \left(1 + 1/\alpha\right) + c \cdot \mu_{AY,Lag}^{2}$$

$$\tau_{Lag} = Sev \cdot \left(1 - \left(1 - \frac{Lag}{10}\right)^{3}\right) \text{ for } Lag = 1,2 \ldots,10.$$

- $\{ELR_{AY}\}$,$\{Dev_{Lag}\}$, *t*, *c*, and *Sev* are unknown parameters,

Tweedie Model of Losses in Each (*AY,Lag)* Cell

$$\mu_{AY,Lag} = Premium_{AY} \cdot ELR_{AY} \cdot Dev_{Lag} \cdot t^{AY+Lag-1}, p = \frac{\alpha+2}{\alpha+1}$$

$$\phi_{AY,Lag} \cdot \mu^{p}_{AY,Lag} = \mu_{AY,Lag} \cdot \tau_{Lag} \cdot (1+1/\alpha) + c \cdot \mu^{2}_{AY,Lag}$$

- Pick a parameter set $\{ELR_{AY}\}$, $\{Dev_{Lag}\}$, t, c, Sev
- Translate parameters into Tweedie parameters
  - $\mu_{AY,Lag}$, p and $\phi_{AY,Lag}$
- Calculation likelihood, prior and proposal density for Metropolis Hastings Algorithm

Perspective on Loss Models

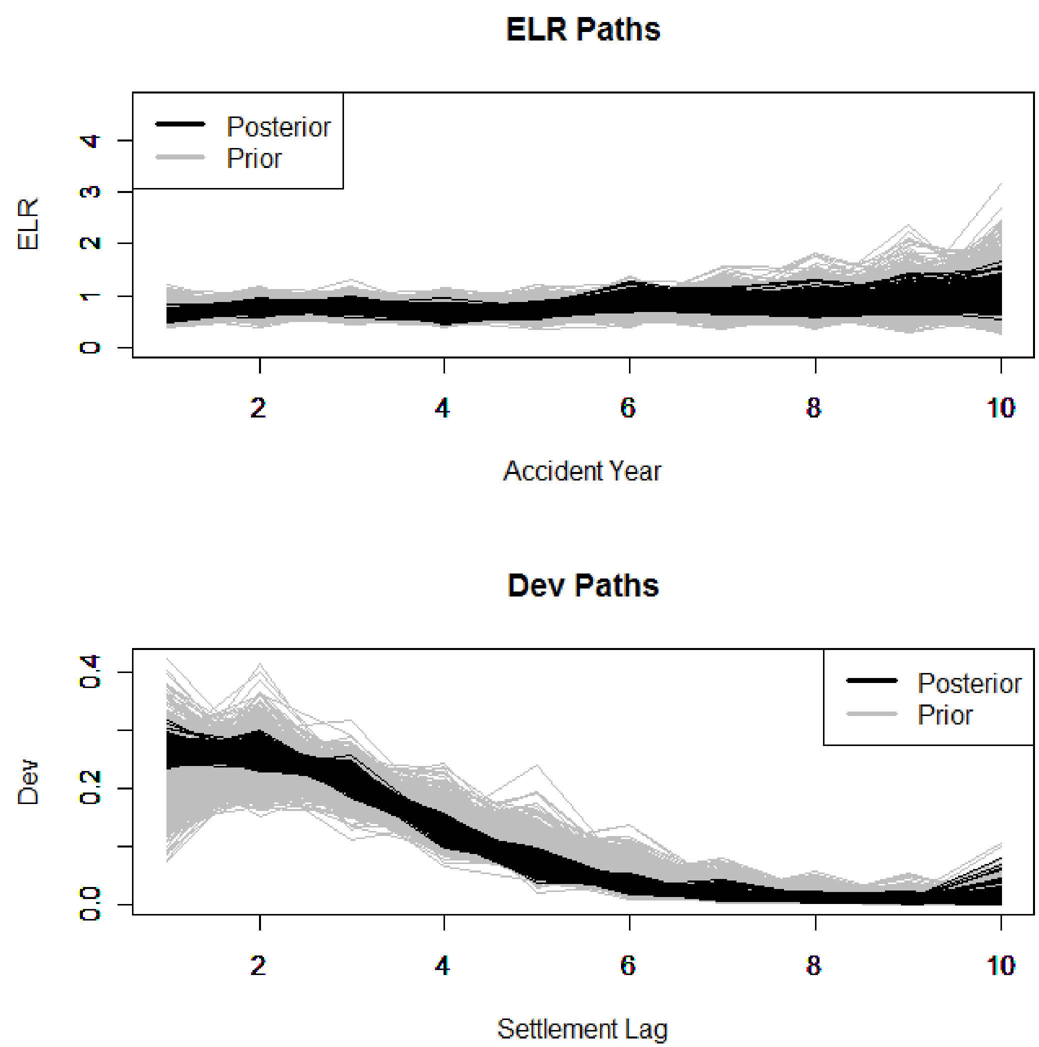
- 55 data points with 23 parameters
- Efforts to formulate models with fewer parameters have been problematic.
- Don’t fight many parameters, figure out how to deal with it.
- Actuaries generally use models with several parameters, and temper their results with “judgment.”
  - Experience gained by looking at data from “similar” lines of insurance and/or from other insurers.
- This calls (screams?) for a Bayesian approach.

Sample from Metropolis-Hastings Algorithm Applied to $\{Dev_{Lag}\}$ and $\{ELR_{AY}\}$ parameters

ELR_1	ELR_2	ELR_3	ELR_4	ELR_5	ELR_6	ELR_7	ELR_8	ELR_9	ELR_{10}
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.91503	0.66796	0.62778	0.58480	0.56635	0.67332	0.56119	0.68528	0.69505	0.70776
0.86193	0.63186	0.67501	0.57013	0.60554	0.64775	0.61769	0.74869	0.68954	0.68855
0.85805	0.62464	0.68672	0.55612	0.58922	0.63364	0.65857	0.70962	0.67289	0.64800

Dev_1	Dev_2	Dev_3	Dev_4	Dev_5	Dev_6	Dev_7	Dev_8	Dev_9	Dev_{10}
0.16546	0.25163	0.22465	0.16499	0.10414	0.05589	0.02427	0.00762	0.00131	0.00005
0.16546	0.25163	0.22465	0.16499	0.10414	0.05589	0.02427	0.00762	0.00131	0.00005
0.16321	0.24844	0.22338	0.16574	0.10598	0.05781	0.02564	0.00827	0.00148	0.00006
0.16613	0.24962	0.22293	0.16463	0.10487	0.05701	0.02520	0.00811	0.00144	0.00006
0.16613	0.24962	0.22293	0.16463	0.10487	0.05701	0.02520	0.00811	0.00144	0.00006
0.16613	0.24962	0.22293	0.16463	0.10487	0.05701	0.02520	0.00811	0.00144	0.00006
0.16613	0.24962	0.22293	0.16463	0.10487	0.05701	0.02520	0.00811	0.00144	0.00006
0.16613	0.24962	0.22293	0.16463	0.10487	0.05701	0.02520	0.00811	0.00144	0.00006
0.15732	0.24804	0.22578	0.16815	0.10736	0.05822	0.02555	0.00810	0.00141	0.00006
0.15732	0.24804	0.22578	0.16815	0.10736	0.05822	0.02555	0.00810	0.00141	0.00006

Graphical Representation of Metropolis-Hastings Sample

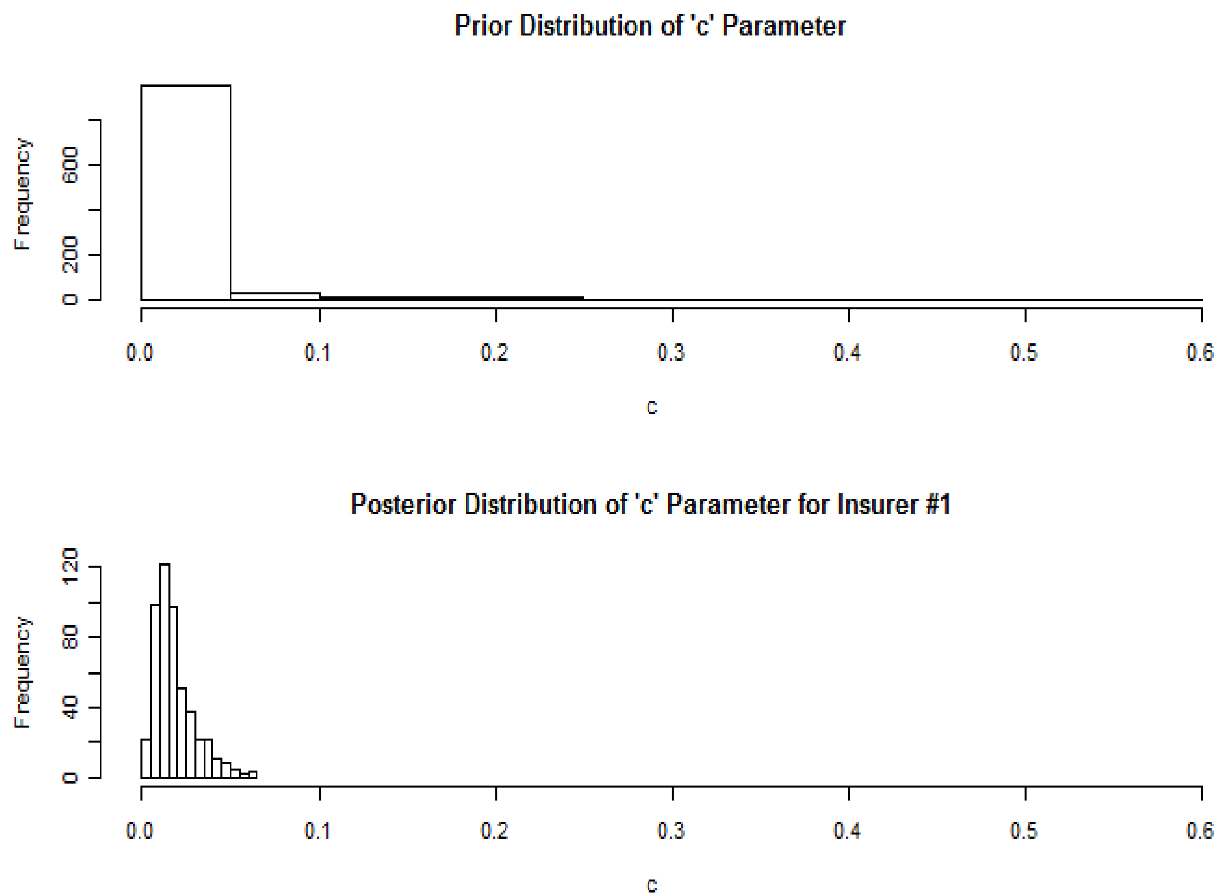


Note that the posteriors are tighter, showing how the data narrows the range of results.

"Information Reduces Uncertainty"

Claude Shannon

Graphical Representation of Metropolis-Hastings Sample

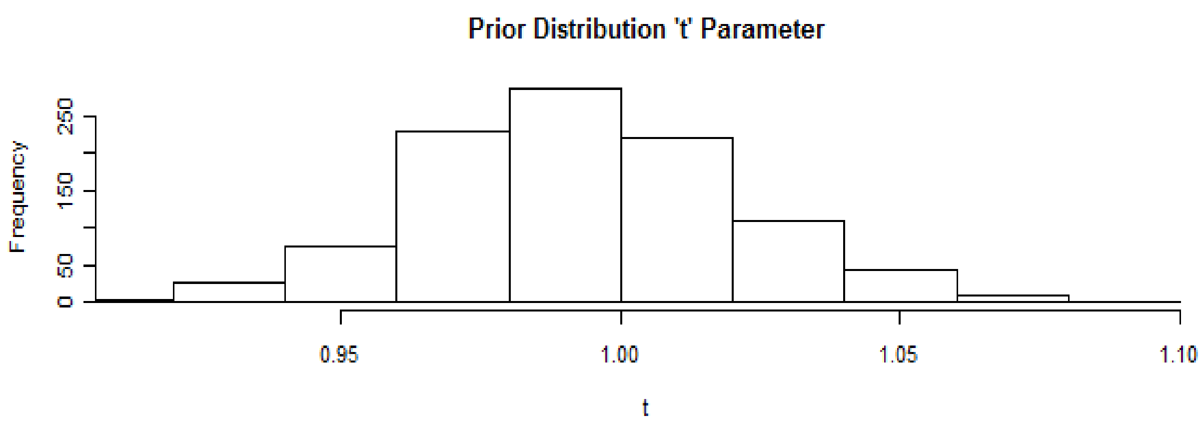


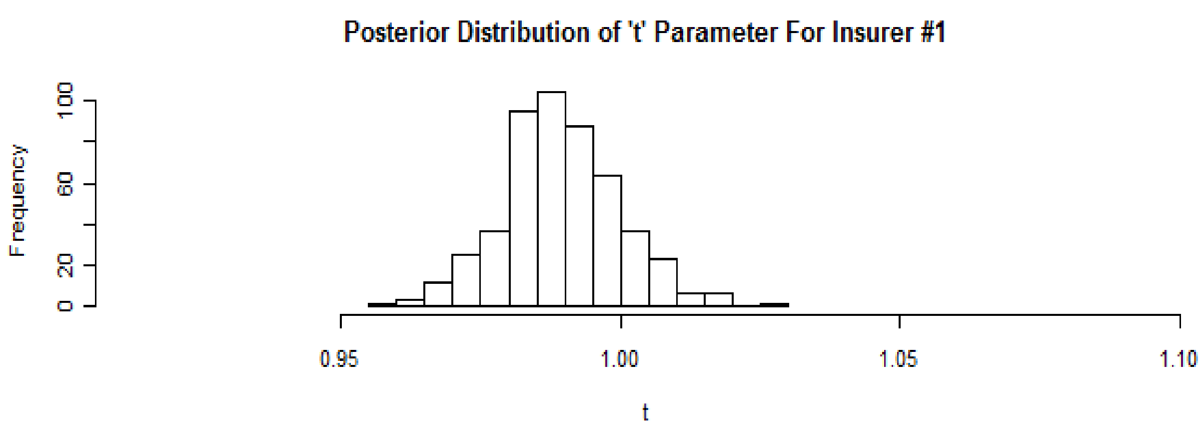
Note that the posteriors are tighter, showing how the data narrows the range of results.

"Information Reduces Uncertainty"

Claude Shannon

Graphical Representation of Metropolis-Hastings Sample



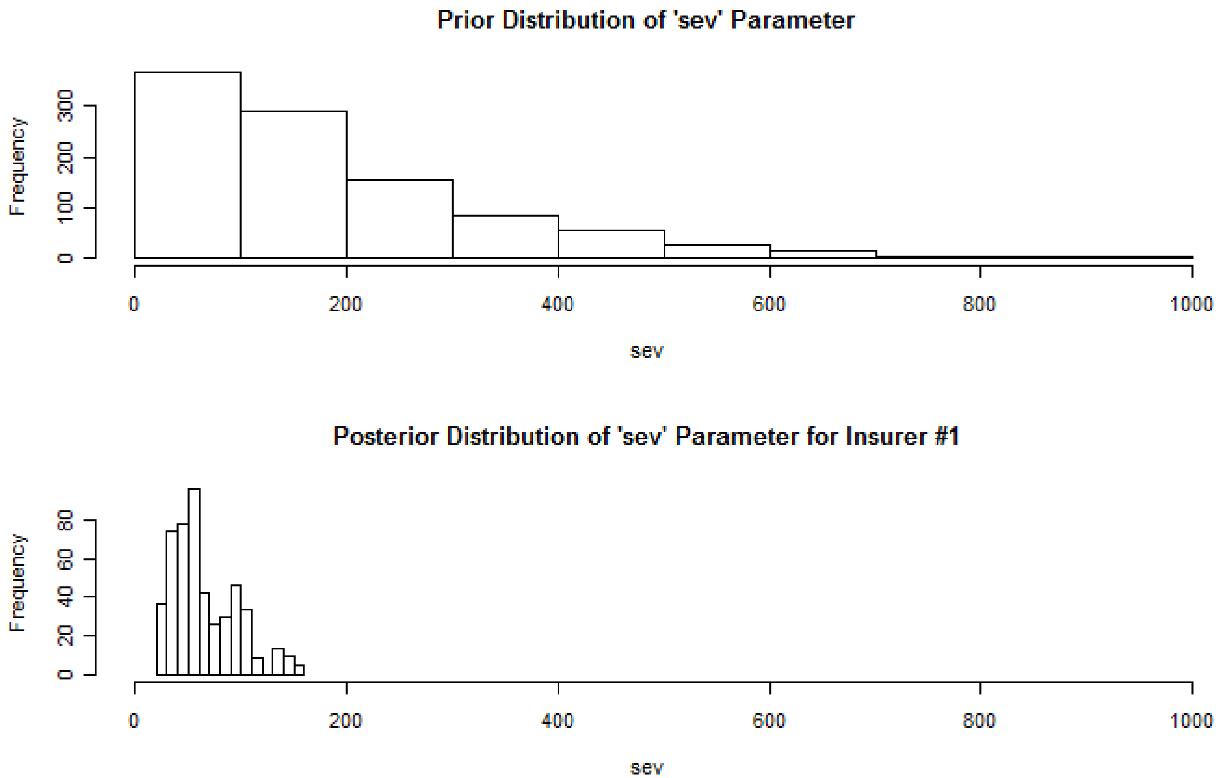


Note that the posteriors are tighter, showing how the data narrows the range of results.

"Information Reduces Uncertainty"

Claude Shannon

Graphical Representation of Metropolis-Hastings Sample



Note that the posteriors are tighter, showing how the data narrows the range of results.

"Information Reduces Uncertainty"

Claude Shannon

Statistics of Intrest

Incremental Paid Losses

AY	*Premium*	Lag_1	Lag_2	Lag_3	Lag_4	Lag_5	Lag_6	Lag_7	Lag_8	Lag_9	Lag_{10}
1	29,701	5,234	5,172	3,708	1,783	923	537	175	145	8	0
2	27,526	5,234	5,683	4,392	2,134	1,377	673	155	81	47	$X_{2,10}$
3	30,750	5,702	5,865	7,966	2,472	NA	143	152	73	$X_{3,9}$	$X_{3,10}$
4	35,814	6,349	4,611	3,959	2,522	1,924	622	206	$X_{4,8}$	$X_{4,9}$	$X_{4,10}$
5	42,277	8,377	6,890	4,055	3,795	1,292	1,422	$X_{5,7}$	$X_{5,8}$	$X_{5,9}$	$X_{5,10}$
6	50,088	9,291	13,836	12,441	4,086	2,293	$X_{6,6}$	$X_{6,7}$	$X_{6,8}$	$X_{6,9}$	$X_{6,10}$
7	56,921	12,029	12,462	8,369	7,034	$X_{7,5}$	$X_{7,6}$	$X_{7,7}$	$X_{7,8}$	$X_{7,9}$	$X_{7,10}$
8	61,406	13,119	12,618	9,117	$X_{8,4}$	$X_{8,5}$	$X_{8,6}$	$X_{8,7}$	$X_{8,8}$	$X_{8,9}$	$X_{8,10}$
9	67,983	15,860	14,893	$X_{9,3}$	$X_{9,4}$	$X_{9,5}$	$X_{9,6}$	$X_{9,7}$	$X_{9,8}$	$X_{9,9}$	$X_{9,10}$
10	73,359	16,498	$X_{10,2}$	$X_{10,3}$	$X_{10,4}$	$X_{10,5}$	$X_{10,6}$	$X_{10,7}$	$X_{10,8}$	$X_{10,9}$	$X_{10,10}$

"Range of Reasonable Estimates"
Distribution of

$$Estimate \sim \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} \mu_{AY,Lag}$$

Predictive Distribution
of Reserve Outcomes

$$Outcome \sim \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} X_{AY,Lag}$$

Calculating Ranges

- For a given parameter set, P_n, sampled from the Markov chain:
- Calculate mean $\mu_{n,AY,Lag}$ and $\phi_{n,AY,Lag}$.
- Simulate $X_{n,AY,Lag}$ from a Tweedie($\mu_{n,AY,Lag}$,p,$\phi_{n,AY,Lag}$) distribution.

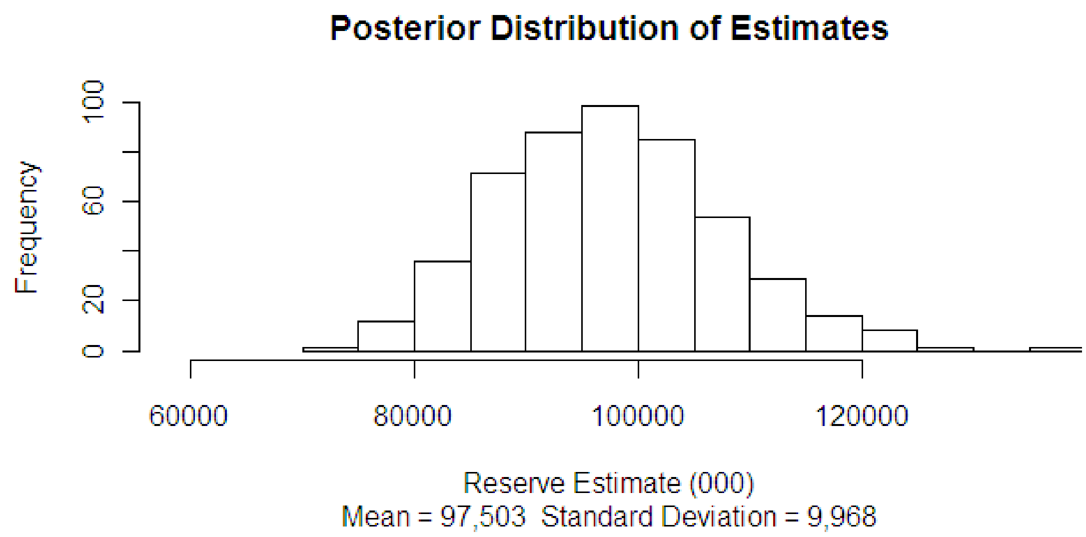
Range of "Reasonable" Estimates Distribution of

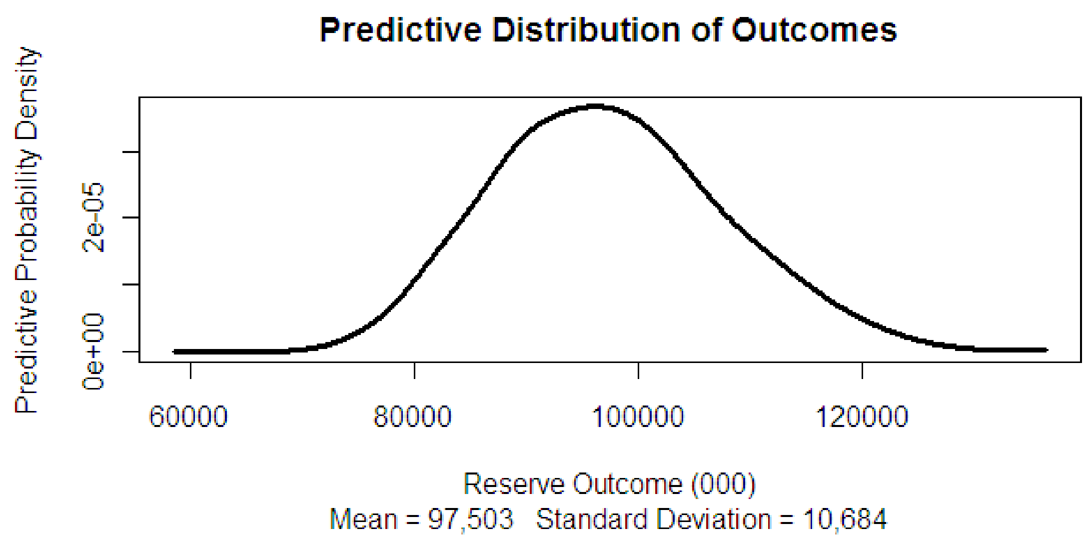
$$Estimate_n \sim \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} \mu_{n,AY,Lag}$$

Predictive Distribution of Reserve Outcomes

$$Outcome_n \sim \sum_{AY=2}^{10} \sum_{Lag=12-AY}^{10} X_{n,AY,Lag}$$

Range of Estimates and Outcomes





Bayesian Sound Bite #1
By George Box – Sung to a Familiar Show Tune

- There's no theorem like Bayes' theorem, it's like no theorem we know.
- Everything about it is appealing
- Everything about it is a wow
- Let out all that a priori feeling, you've been *concealing* till now.

- Bayesian analysis forces one to specify the prior distribution. It is more transparent.

Bayesian Sound Bite #2
Relayed indirectly to me through Stuart Klugman

- To frequentist statisticians, models are real, and data are random.
- To Bayesian statisticians, data are real, and models are random.

References

- **"Quantifying Tail Risk with the Gibbs Sampler"**
  - Brainstorms Column - Actuarial Review – February 2008
  - http://www.casact.org/newsletter/index.cfm?fa=Index&newsletter_id=1
- **"Bayesian Analysis with the Metropolis Hastings Algorithm"**
  - Brainstorms Column - Actuarial Review – November 2009
  - http://www.casact.org/newsletter/index.cfm?fa=Index&newsletter_id=1
- **"Stochastic Loss Reserving with the Collective Risk Model"**
  - Variance – February 2010
  - http://www.variancejournal.org/issues/?fa=article&abstrID=6606
- **Albert, J., *Bayesian Computation with R, NewYork: Springer,* 2007.**
- **Lynch, S. M., *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists, New York: Springer,* 2007**