Bayesian Techniques

From Theory to Practice

Manolis Bardis, FCAS, MAAA, PhD, MBA
Enbo Jiang, ACAS

September 17, 2019

Willis Towers Watson

Theoretical Knowledge

Manolis Bardis, FCAS, MAAA, PhD, MBA

Statistics 101 – Bayes Theorem

- Bayes theorem indicates how prior subjective belief changes based on evidence

$$P(A/B) = \frac{P(B/A)\,P(A)}{P(B)}, \text{ where}$$

- P(A) is the prior belief,

- P(A/B) is the posterior belief accounting for B

- P(B/A)/P(B) represents the support B provides to A

Willis Towers Watson

Statistics 101 – Likelihood function

- Usually we think in terms of probabilities, i.e., the probability of an outcome X given a parameter Θ; $P(X|\Theta)$

- The Likelihood instead is a function of Θ given an outcome, i.e. $L(\Theta|X)$

- With an observed outcome X the maximum likelihood principle chooses the parameter Θ that maximizes the $P(X|\Theta)$

- The GLM model can produce a maximum likelihood function

Statistics 101 – Posterior Probability

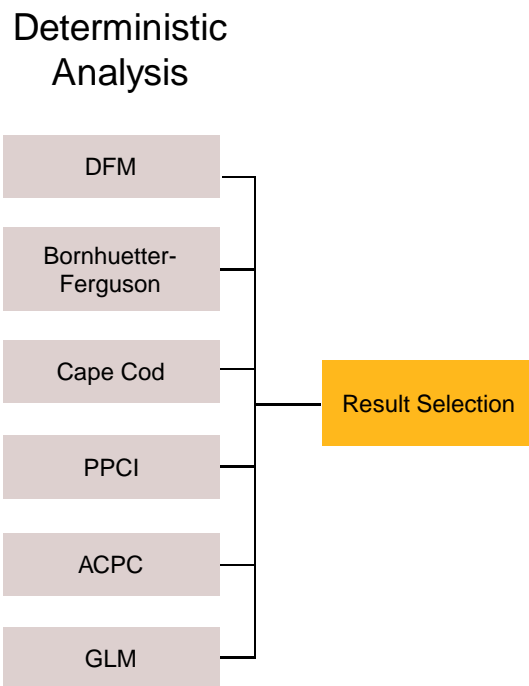Given prior belief p(θ) and observation x with likelihood P(x/θ) the posterior is

$$P(\theta/x) \quad = \quad \frac{P(x/\theta)P(\theta)}{P(x)}, \quad i.e.$$

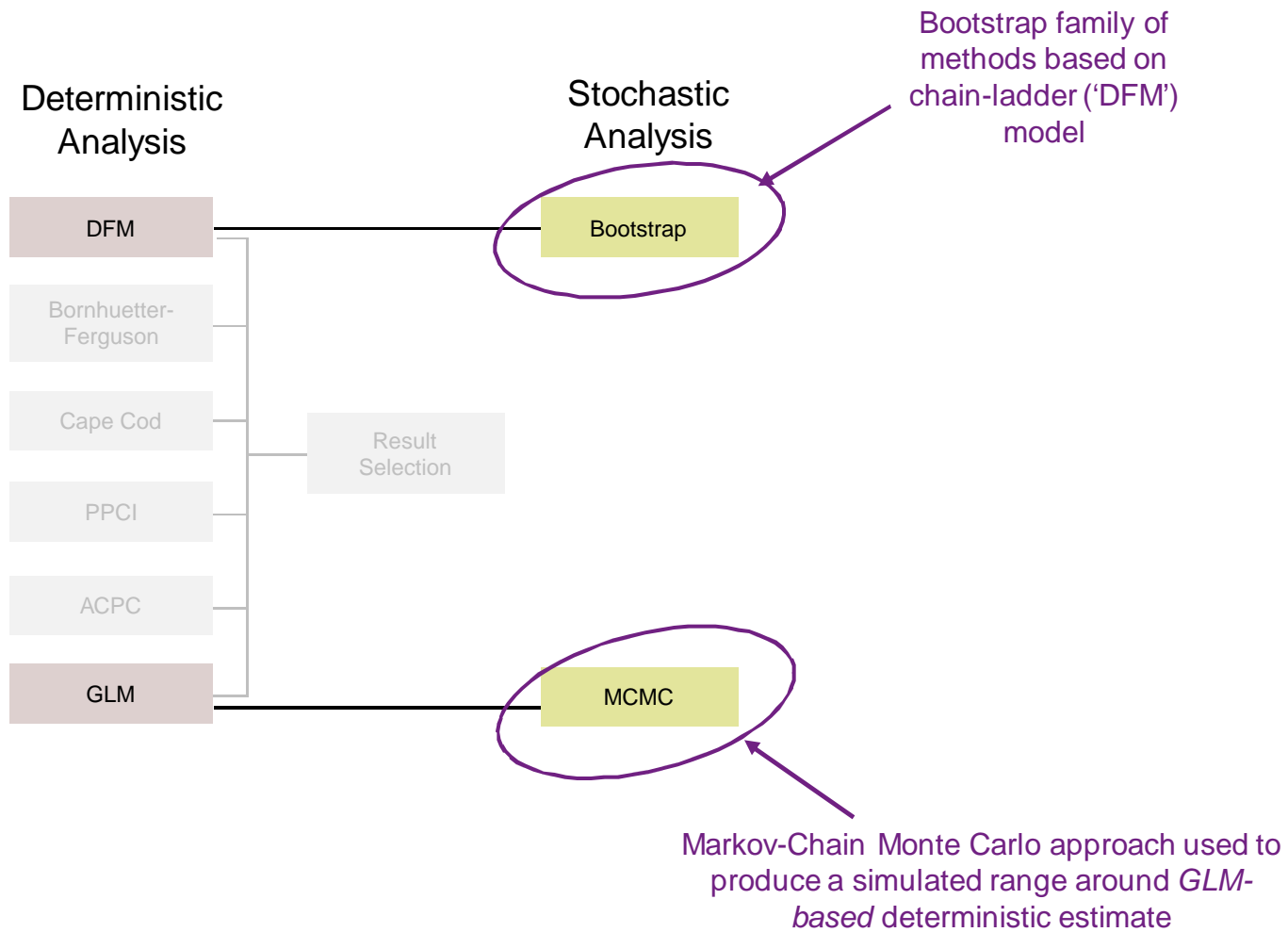Posterior probability $\propto$ Likelihood x Prior Probability

Statistics 101 – MCMC stochastic methods

- MCMC methods are a class of algorithms for sampling from a probability distribution
  - This distribution is usually difficult to approximate with analytical functions
- MCMC constructs a random process that undergoes transition from one state to another, called Markov Chain
  - This process is *memoryless*, i.e. the next state is based only on the current state but not the sequence of the preceding states
  - The quality of the convergence to an *equilibrium* distribution improves with the number of steps employed in the process
  - The first few draws are usually thrown away (called *burn-in*) to ensure target is independent of starting point and improve convergence
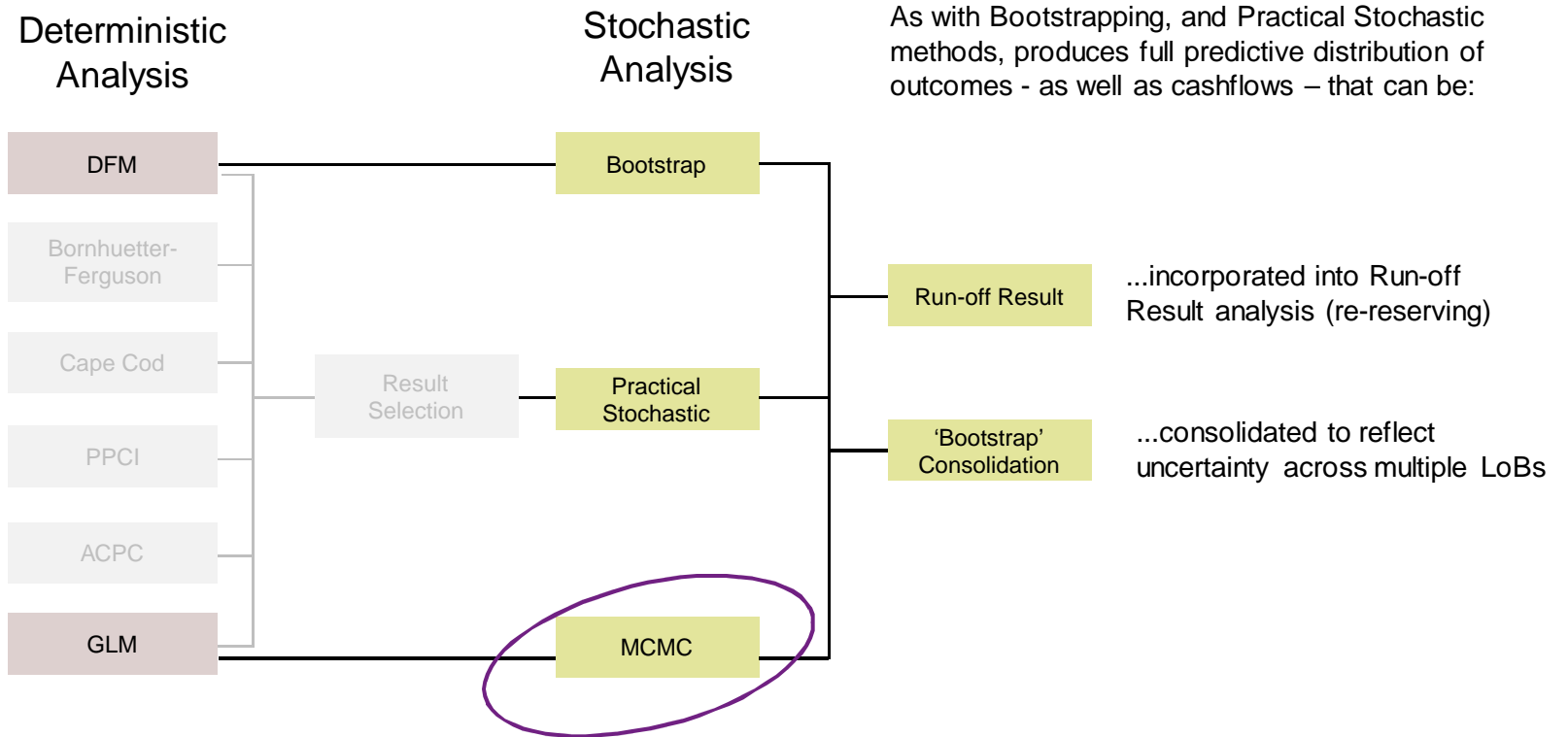
Statistics 101 – MCMC compared to Bootstrapping

Deterministic
Analysis

DFM	
Bornhuetter-Ferguson	
Cape Cod	Result Selection
PPCI	
ACPC	
GLM	

Statistics 101 – MCMC compared to Bootstrapping (cont'd)

Bootstrap family of methods based on chain-ladder ('DFM') model

Deterministic Analysis

Stochastic Analysis

- DFM
- Bornhuetter-Ferguson
- Cape Cod
- PPCI
- ACPC
- GLM

Result Selection

Bootstrap

MCMC

Markov-Chain Monte Carlo approach used to produce a simulated range around *GLM-based* deterministic estimate

Statistics 101 – MCMC compared to Bootstrapping (cont'd)

Deterministic Analysis

DFM	
Bornhuetter-Ferguson	
Cape Cod	
PPCI	
ACPC	
GLM	

Result Selection

Stochastic Analysis

Bootstrap

Practical Stochastic

MCMC

Run-off Result

'Bootstrap' Consolidation

As with Bootstrapping, and Practical Stochastic methods, produces full predictive distribution of outcomes - as well as cashflows – that can be:

...incorporated into Run-off Result analysis (re-reserving)

...consolidated to reflect uncertainty across multiple LoBs

Statistics 101 – MCMC compared to Bootstrapping (cont'd)

Similarities	Differences
Similarities	**Differences**

▪ Both are simulation methods
▪ They produce a full predictive distribution of outcomes including
   ▪ Parameter risk
   ▪ Process risk
	▪ Bootstrapping
	   ▪ Samples with replacement the residuals from an actual versus expected comparison of historical development
	   ▪ Simulations are independent from one another
	   ▪ There is no convergence in the simulations
	▪ MCMC
	   ▪ Samples the parameters of the resulting GLM likelihood function
	   ▪ Simulations are built through a Markov chain
	   ▪ The simulations converge into an *equilibrium* state

Stochastic Models Beyond the Bootstrap: Deploying a Stochastic Modeling Framework

IV. Bayesian Modeling Steps

Bayesian modeling steps

Step 1: Specify probability distribution for the data given some unknown parameters (data distribution)

Step 2: Specify prior probability distribution for the parameters of the data distribution (prior distribution)

Step 3: Derive the likelihood function of the parameters, given the data (likelihood function)

Step 4: Combine prior distribution and likelihood function to derive posterior joint distribution of parameters(posterior distribution)

Step 5: Obtain parameters for posterior distribution

Step 6: Combine data distribution and posterior distributions to obtain forecast of predictive distribution

Bayesian modeling steps (cont'd)

Step 1 & 2: Example is the ODP model.
Prior distributions could assume some distributional shape (i.e. Lognormal, Gamma etc.)
Informative priors with small variance that could affect shape of posterior

Step 4: Posterior is based on Bayesian theory and is proportional to prior and the likelihood function

Step 5: It is easy to obtain the parameters of posterior when the shape of distribution is known. Otherwise special statistical algorithms, like Gibbs MCMC, are needed

Step 6: Like step 5 the complexity of the forecasting depends on whether predictive distribution is recognizable. Generic sampling algorithms such as Adaptive Rejection Sampling (ARS) might be needed

Stochastic Models Beyond the Bootstrap: Deploying a Stochastic Modeling Framework

V. Bayesian Modeling within the Reserving Context

Willis Towers Watson I.I'I'I.I

Bayesian modeling within the reserving context

- Reserving example

Origin	Development Period				
Period	1	2	3	...	n
1	C_{11}	C_{12}	C_{13}	...	C_{1n}
2	C_{21}	C_{22}	C_{23}	...	C_{2n}
3	C_{31}	C_{32}	C_{33}	...	C_{3n}
...	...	...	...	...	...
n	C_{n1}	C_{n2}	C_{n3}	...	C_{nn}

- $C = \{C_{ij}: i+j < n+1\}$ is the upper-left triangle of observed payments, and the reserving problem attempts to estimate the unobserved values in the lower-right triangle

Bayesian modeling within the reserving context (cont'd)

- **Step 1&2**: Assumes C_{ij} follows a probability density distribution of $f(C_{ij} / \theta)$, where θ denotes parameters describing a particular claims generating process and $\pi(\theta)$ is the prior distribution function
- **Step 3:** The likelihood function $L(\theta/\underline{c})$ for the parameters given observe data is:

$$L(\theta / \underline{c}) = \prod_{i+j \leq n+1} f(C_{ij} / \theta)$$

Bayesian modeling within the reserving context (cont'd)

- **Step 4**: Given the data distribution and the prior distribution, the posterior distribution $f(\theta/\underline{c})$ is proportional to the product of the likelihood and the prior:

$$f(\theta / \underline{c}) \propto L(\theta / \underline{c}) \, \pi(\theta)$$

- **Step 5**: Parameters θ are obtained from the posterior distribution and are used in **Step 6**

Bayesian modeling within the reserving context (cont'd)

- **Step 6**: The known data C_{ij} for $i + j \leq n + 1$ is used to predict unobserved values in the lower right triangle C_{ij} for $i + j > n + 1$ by means of the predictive distribution:

$$f(C_{ij} / \underline{c}) = \int f(C_{ij} / \theta) \, f(\theta / \underline{c}) d\,\theta$$

- Predictive distribution can either be obtained in a closed form analytically or through a generic sampling algorithm instead

Stochastic Models Beyond the Bootstrap: Deploying a Stochastic Modeling Framework

VI. Simple Example – No Simulations Needed

Willis Towers Watson

Simple example – no simulations needed

- **Step 1&2**: Assume the loss generating process follows a Poisson distribution with parameter and the parameter follows a Gamma distribution with some known parameters a and b

- $c_{ij} / \theta \sim \rho(\theta)$

- $\theta / a,b \sim Gamma(a,b)$

Willis Towers Watson

Simple example – no simulations needed (cont'd)

- **Step 3**: The likelihood function is given by

- **Step 4**: The posterior distribution is proportional to the product of the likelihood and the prior:

$$f(\theta/\underline{c},a,b) = \prod_{i=1}^{n} \theta^{x_i} e^{-\theta} / x_i! \; b^a/\Gamma(a) \; \theta^{a-1} e^{-b\theta} =$$

$$\sim Gamma\left(a + \sum_{i=1}^{n} x_i, b + n\right), \text{ i.e. posterior follows a}$$
Gamma distribution

Simple example – no simulations needed (cont'd)

- **Step 6**: The product of the posterior and the data distribution, i.e., the product of a Gamma and a Poisson distribution results in a negative binomial distribution:

$$NB(a + \sum_{i=1}^{n} x_i , \frac{1}{1+b_1})$$

- No need for complicating sampling here !

Stochastic Models Beyond the Bootstrap: Deploying a Stochastic Modeling Framework

VII. Examples of Popular Sampling Techniques

Willis Towers Watson **I.I'I'I.I**

Examples of popular sampling techniques – Gibbs sampler

- Gibbs sampler avoids sampling from a complicated bivariate distribution f(x,y) by making random draws instead from univariate conditional distributions (f(x/y) and f(y/x))

- For two parameters and n iterations it produces an

  n x 2 table where xo is the initial value – Next steps:

$$y_1 \tilde{} f(y/x = x_0)$$

$$\dots\dots\dots\dots\dots\dots$$

$$x_i \tilde{} f(x/y = y_{i-1})$$

$$y_i \tilde{} f(y/x = x_i)$$

- Eventually $(x_i, y_i) \rightarrow (x,y) \tilde{} f(x,y)$ for sufficient large number of iterations (so called burn-in sample)

- After burn-in, it is common to define a spacing between accepted points, maybe every m draws, to ensure independence of random draws

Examples of popular sampling techniques – Adaptive Rejection Sampling

Log space

$g_{En}(x)$

f(xi)

f(x)

x_i
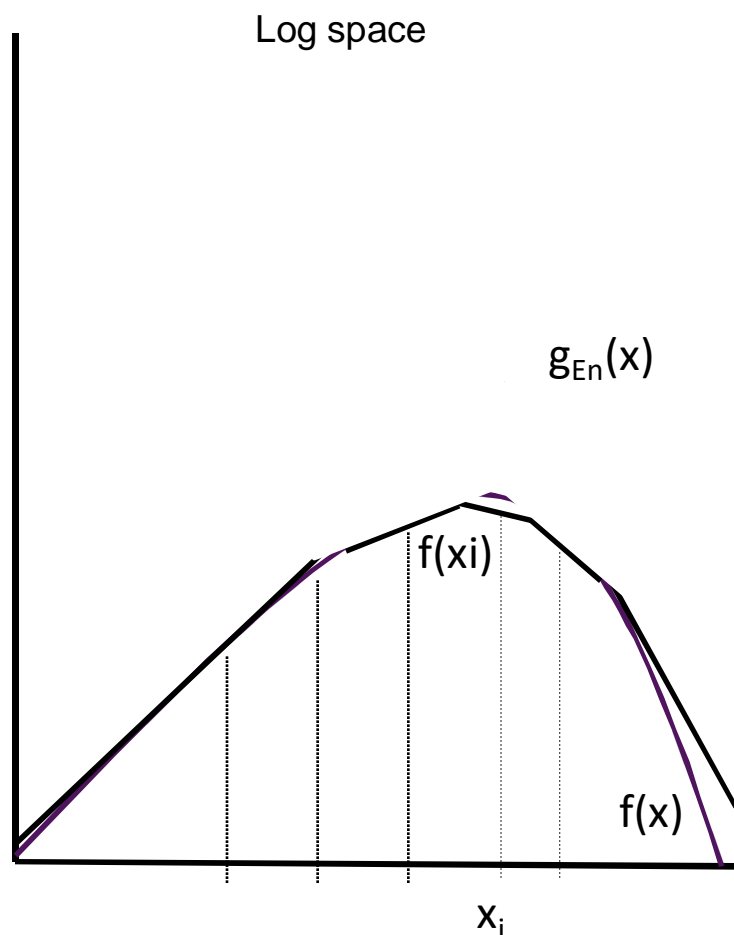
- ARS works with log concave densities f(x)

- An envelope function $g_{En}(x)$ as an upper bound of the log density function is employed

- A random draw x_i from the x-axis is then sampled

- When the resulting $g_{En}(x_i)$ is close to $f(x_i)$ the envelope function remains unchanged

- When the resulting $g_{En}(x_i)$ is much larger to $f(x_i)$ the envelope function changes to incorporate a line that is tangent to $f(x_i)$

Examples of popular sampling techniques – Metropolis-Hastings Algorithm

1. Create GLM with Error distribution $f(x|\mu)$ (typically Poisson)

Data   Ultimate

Parameters

Selected
Parameters
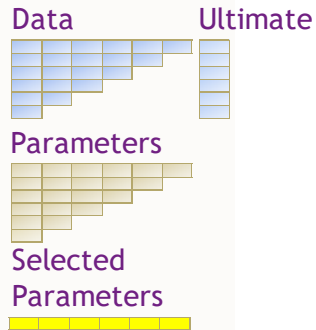
2. GLM produces Parameter estimates with uncertainty

Parameters, Error

3. Set Initial Markov Chain μ_0 equal to parameter estimates from GLM

4. Sample a Candidate Parameter μ^*

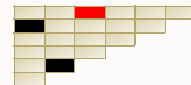p ~ Multivariate Normal

6. Draw U from Uniform(0,1) Distribution

U ~ Uniform(0,1)

5. Calculate Markov Transition Probability R

Based on a ratio of Likelihood Functions, where the fit of μ^* is compared to the fit of μ_{t-1}

7. Accept or Reject the Candidate from step 4

set $\mu_t = \mu^*$ if U<R
Otherwise set $\mu_t = \mu_{t-1}$

8. Calculate Reserves based on the Markov Chain ending value μ_t

Data

Selected μ_t
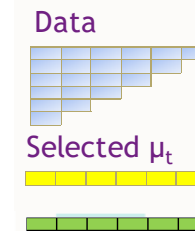
← Add process variance

Simulated Ultimate

9. Repeat steps 4-8 10,000 times for burn-in period

10. Discard the burn-in steps and Repeat steps 4-8 10,000 times for final result

Practical Implementations

Enbo Jiang, ACAS

Willis Towers Watson

Status Quo of Reserving Techniques

***The Actuary and IBNR* (1972) by Bornhuetter & Fergusson:**

"If one were to survey today's IBNR computational techniques and then reread Mr. Tarbell's 1934 paper, one might conclude that it must still be 1934."

At 2019 CLRS, Bardis & Jiang:

"If one were to survey today's IBNR computational techniques and then reread BF's 1972 paper, one might conclude that it must still be 1972."

Moral of the joke:

- Reserving actuaries tend to rely on what works, which is understandable given the gravity of our job (company solvency, regulatory scrutiny, etc.)
- This inertia results in the slow adoption of new and at times superior techniques such as Bayesian MCMC

Roadblocks to a Wider Adoption

Roadblocks:

1. Inertia (per previous slide)
2. For the love of triangles
3. The fact that Bayesian models are generally less understood (hence seen as a black box)
4. More computationally demanding compared to traditional methods

Responses:

1. Bayesian models do not have to be the main method from the get-go but rather can simply be a supplement to the current process until stakeholders are comfortable (much like pilot testing of an enterprise risk model)
2. Dave Clark has long been calling actuaries to abandon triangles
3. This demo will focus on tackling Roadblock #2 by: (a) providing a straight-forward example of the MCMC techniques (in Excel) and (b) reviewing a Bayesian model in Meyers' CAS Monograph Number 1 (in R)
4. Ever-increasing computing power that actuaries have access to

MCMC Example in Excel

A demo

Purpose:

- Excel is still the king of reserving software
- Excel formulae help illustrate "black-box" concepts step-by-step, and are great for learning a new method for the first time
  - Steve Mildenhall on Neural Network at CAS 2019 Spring Meeting: (Paraphrased by Jiang) With a simple example, first do it in Excel and then VBA and finally fly with existing packages in R or Python
- Likewise, first familiarizing oneself with Bayesian MCMC techniques in Excel has tremendous educational value

Instructive and Easy Read:

https://towardsdatascience.com/mcmc-intuition-for-everyone-5ae79fff22b1

Willis Towers Watson **I.I'I'I.I**

A Bayesian Model by Meyers in R

A demo

Purpose:

- Meyers' models are some of the most cutting-edge implementations of Bayesian MCMC techniques
- This demo walks the audience through one of the proposed models
- An modified version of the original codes published by Meyers will be utilized and shared. Edits include:
  - Thorough comments to facilitate learning
  - Corrections of errors identified during the preparation of this talk

Link to Paper and Original Code:

https://www.casact.org/pubs/monographs/index.cfm?fa=meyers-monograph01

Pros and Cons

Advantages

- Automatically produces a full distribution of outcomes. In other words, variability is estimated at the same time as the central estimate, all in one integrated model
- Flexible in model forms (Gamma, ODF, etc.) and allows expert opinions via the specification of prior distributions
- 21st-century solution

Disadvantages

- Generally less understood for a reason: Bayesian models and MCMC techniques are more theoretically sophisticated and computationally heavy
- Harder to secure management buy-in due to lower interpretability (compared link ratios, for example)
- Selection of prior distributions is highly judgmental

"Markov Chain Monte Carlo (MCMC) methods are a clever solution for integrating over highly dimensional spaces and have rapidly gained popularity by the standards of 'technical hype,'" [David Clark] says. "At the same time, much of the related technology is still computationally intensive; it requires compromises in how variables are defined, and the models may not always converge. It may not be wise to throw MCMC at a problem when simpler methods are available." -- Five Laws of Hype for Actuaries

Thank you!

Willis Towers Watson ||.||'||.||

Manolis Bardis, FCAS, MAAA, PhD, MBA
Senior Director

800 Boylston St.
Boston, MA 02199

T   +1 617 638 3807
E   manolis.bardis@willistowerswatson.com

Willis Towers Watson ||.||'||.||

Enbo Jiang, ACAS
Analyst

1500 Market St.
Philadelphia, PA 19102

T   +1 215 246 6748
E   enbo.jiang@willistowerswatson.com

About Willis Towers Watson
Willis Towers Watson (NASDAQ: WLTW) is a leading global advisory, broking and solutions company that helps clients around the world turn risk into a path for growth. With roots dating to 1828, Willis Towers Watson has 40,000 employees serving more than 140 countries. We design and deliver solutions that manage risk, optimize benefits, cultivate talent, and expand the power of capital to protect and strengthen institutions and individuals. Our unique perspective allows us to see the critical intersections between talent, assets and ideas – the dynamic formula that drives business performance. Together, we unlock potential. Learn more at willistowerswatson.com.