A Practical Guide to the Bayesian Hamiltonian Monte Carlo Method

Spencer Miller, fcas, maaa & Kenny Smart, fcas, maaa

2025-09-18

This paper seeks to provide a practical guide to leveraging Bayseian methods, as well as a discussion on behavioral economics as a motivating force.

**Keywords** — Bayesian, Monte Carlo, Markov Chain, MCMC, Hamiltonian, Stan, Forecasting, Predictions, Behavioral Economics

# Introduction

A lot has been written in the actuarial literature about Bayesian methods over the last few decades; few of which as still part of current exam syllabi. While these papers provide a thorough explanation of the process, the math can seem overly complex even to an experienced actuary. The purpose of this paper is twofold. First, we want to provide actuaries with the *basic* understanding of what goes on behind the scenes of a Bayesian model without getting bogged down in the math. Second, we want to champion the continued inclusion of these topics in exam syllabi in the future.

## Target Audience

* An actuary who has finished exams a few years ago
* An actuary who occassionly finds their estimates to be out of line with results
* An actuary interested in learning about Stan but finds the math too onerous or intimidating
* An actuary with limited data

## Current Theories on Prediction

* Point estimates do not provide for an adequate measure of process / parameter / model risk
* Simple simulation models layer in process risk but can still fail to account for parameter risk / model risk
* Using a copula for parameters ignores the likelihood of occurring given the data

## Yes, Even Actuaries are Human

When making predictions, people often rely on heuristics. While these are typically useful from an evolutionary standpoint, they can “… lead to systematic and predictable errors”. Below is a non-exhaustive list of possible factors at play when making actuarial predictions (Tversky & Kahnemann):

* *Insensitivity to sample size*  
  Example?
* *Insensitivity to the prior probability of outcomes*  
  When making a judgment of whether item X belongs to either class A or class B, people will assign it to the class in which the description of item X matches the stereotype of the class, despite the relative likelihood of each class. For example, given the description of a person and a list of occupations, you would likely let the description of the individual inform your estimate of their occupation, while ignoring how people work in a given occupation.
* *Misconception of chance*  
  Say something about process risk or that an actuary may infer some sort of pattern where one does not exist.
* *Insensitivity to predictive accuracy*  
  Taking a risk manager’s assessment of a new claims handling process at face value.
* *The illusion of validity*  
  Using an exposure base as a proxy for loss?
* *Misconceptions of Regression*  
  Letting one bad AY influence the expectation of the next AY.
* *Anchoring*  
  You’re in charge of providing a loss forecast for the upcoming accident year and have been doing so for a few years. When making your selection, you may keep the projection from last year in mind.

While it’s impossible to avoid all possible biases or errors when making predictions, there are methods available that can help mitigate the risk.

# Basics of Bayesian MCMC

Before we get into the Hamiltonian Monte Carlo method, we provide a brief overview of Bayesian MCMC.

* A **Markov process** is a random process for which the future depends only on the present state. For example, the outcome of rolling a die does not depend on the previous roll.
* **Monte Carlo** methods use random sampling to approximate an outcome. Continuing the die roll example, you could replicate the outcome of a single roll by generating a random value from  
  the distribution *U* ~ Uniform[0,1], where .
* **Bayesian statistics** is an approach to data analysis based on Bayes’ theorem, where available knowledge about parameters in a statistical model is updated with the information in observed data. In our example, let’s assume that the die is not fair (unbeknownst to us) with a distribution equal to and that we have observed the following rolls: = {3, 4, 5, 6, 4, 3, 5, 4, 2}.
* Using a Bayesian approach, we have three components:
  + *Likelihood function*: This is the parametric form of our data, , given a set of parameters, .
  + *Prior distribution*: This is our belief of the distribution of our parameters prior to seeing the data, . In this case, we might assume that the die is fair: .
  + *Posterior distribution*: The result of updating our prior belief about the parameters given the data, .

This process can be done using **Stan**, an open-source software, via common coding languages (e.g., python and R). In this paper, we leverage the RStan package, though there are numerous other packages that can be used.

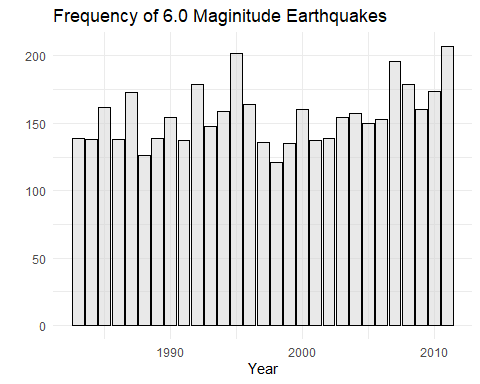
* Need to discuss the key parameters (thin, chains, etc.).
* Discuss sampling method: Gibbs vs Metropolis Hastings vs Hamiltonian Monte Carlo

Applying this process to our loaded die example: *Kenny to complete*

# An Example in Action

For our example, let’s assume you are tasked with estimating the expected frequency of earthquakes over 6.0 magnitude for the upcoming year (2012). You are given a dataset of prior earthquakes since 1983.

First, let’s get the data we’ll be using in our example. Below is the history for the last .



In the simplest case you may decide given the recent trend, that the five-year average is the best approximation of future frequency. In this case, your selected frequency is 183 (“Point Estimate”).

Your next logical step might be to add some process risk around your estimate by fitting a distribution to the data. Based on an examination of the data, you decide to model frequency using the Poisson distribution with lambda = 183.

After running 10,000 simulations with your selected distribution, you’re left with a central estimate of 183. Not surprisingly, this is essentially the same as your point estimate. However, you now have a predictive distribution.

Another possible step would be to layer in parameter risk since we might not be convinced of our selected lambda.

After running another 10,000 simulations, you’re left with a central estimate of 183. Even though we let lambda vary, the result is still quite close to the point estimate.

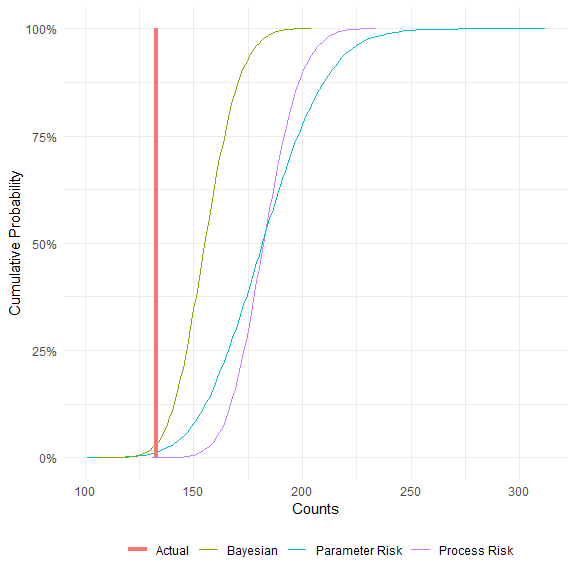
But as we discussed earlier, there are a few problems with just relying on our data:

* The sample size is on the low end (n = 29).
* Regression to the mean.
* It might look like there is a “new normal” but you could be seeing a trend where there isn’t one.

How do we handle these possible biases? That’s where Bayesian MCMC comes in. We’ll continue with the distribution from our parameter risk model.

After running 10,000 simulations with your selected distribution, you’re left with a central estimate of 156. Since you used a wide prior, your result will be closer to the historical average (156) than the point estimate (183).

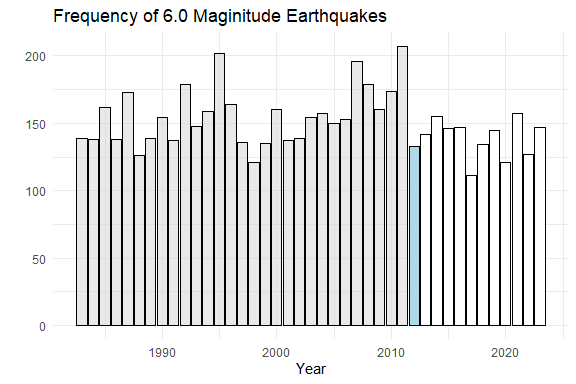
Let’s compare each result to the true number of earthquakes that occurred in 2012.



It appears that all four methods severely underestimated the true average losses. However, if we look at the Bayesian versus the other three, the true count is a more likely outcome. Additionally, there are essentially no “extreme” values (225+) being simulated.

With the benefit of hindsight, let’s see why we overestimated the true counts in all four methods. As it turns out, we fell prey to (at least) two key biases:

* Humans just aren’t that great at identifying random processes. For example, a series of coin tosses with all heads is seen as “less random” than a series with alternating heads and tails, even though both results are equally likely.
* Regression to the mean. As seen in the chart, the high average for 2007-2011 was followed by a low average in 2012-2023. This resulted in the new all-year average (1983-2023) being nearly identical to the all-year average before the rise (1983-2006).



# Supplemental Information

## Literature Review

While our paper focused on the usage of Bayesian MCMC methods for forecasting purposes, the methods can and have been used in other applications. We provide a non-exhaustive list below of papers that utilize these methods to some extent:

* Trends (Schmid)
* Reserving (Meyers 2015)
* Renewal Functions (Aminzadeh & Deng)
* Risk Margins (Meyers 2019)
* Incorporating expert opinion into traditional reserving methods (Verrall)

We also believe there are several other possible applications, ranging from the selection of development factors to estimating reserve variability. We leave it to the reader to explore these.

## Acknowledgements

We would like to extend our gratitude to our colleagues for their thoughtful reviews (Rajesh Sahasrabuddhe, Molly Colleary, Alex Taggart, and Chris Schneider).

## Biographies of the Authors

Spencer is a Senior Manager with Oliver Wyman Actuarial Consulting, Inc., located in Philadelphia. He holds a Bachelor of Science from Lebanon Valley College.

Kenny is a Senior Manager with Oliver Wyman Actuarial Consulting, Inc, located in Chicago. He holds a Bachelor of Science degree in Actuarial Mathematics and Statistics from the University of Pittsburgh.

## Citations

A. Tversky, and D. Kahneman. 1974. “Judgment Under Uncertainty: Heuristics and Biases.” Science 185, 1124–1131.

Aminzadeh, M.S., and Min Deng. 2022. “Bayesian Estimation of Renewal Function Based on Pareto-Distributed Inter-Arrival Times via an MCMC Algorithm.” Variance 15 (2).

Meyers, G. 2015. “Stochastic Loss Reserving Using Bayesian MCMC Models.” CAS Monograph #1.

Meyers, G. 2019. “A Cost-of-Capital Risk Margin Formula for Nonlife Insurance Liabilities.” Variance 12 (2).

Sahasrabuddhe, R. 2021. “The Single Parameter Pareto Revisited.” Casualty Actuarial Society E-Forum, Spring 2021.

Schmid, F. 2013. “Bayesian Trend Selection.” Casualty Actuarial Society E-Forum, Spring 2013.

Verrall, R. J. 2007. “Obtaining Predictive Distributions for Reserves Which Incorporate Expert Opinion.” Variance 1 (1).

## R Packages

Barrett, Tyson, Matt Dowle, Arun Srinivasan, Jan Gorecki, Michael Chirico, Toby Hocking, Benjamin Schwendinger, and Ivan Krylov. 2025. “Data.table: Extension of ’Data.frame‘.” <https://CRAN.Rproject.org/package=data.table>.

Dutang, Christophe, and Arthur Charpentier. 2025. “CASdatasets: Insurance Datasets.” <https://doi.org/10.57745/P0KHAG>.

Stan Development Team. 2025. “{RStan}: The {r} Interface to {Stan}.” <https://mc-stan.org/>.

Wickham, Hadley. 2016. “Ggplot2: Elegant Graphics for Data Analysis.” <https://ggplot2.tidyverse.org>.