

A Practical Guide to the Bayesian Hamiltonian Monte Carlo Method

Spencer Miller, FCAS, MAAA & Kenny Smart, FCAS, MAAA

2025-07-01

Abstract

To be completed later.

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

Introduction

Some text

Target Audience

- An actuary who has finished exams a few years ago
- An actuary who constantly 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 MCMC models layer in process risk but 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:^[1]

- *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*
Using the prior prediction.

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 r by generating a random value u from the distribution $U \sim \text{Uniform}[0,1]$, where $r = \lceil 6u \rceil$.
- **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 (unknown to us) with a distribution equal to $\text{Prob}(r = i) = i/21$ and that we have observed the following rolls: $y = \{3, 4, 5, 6, 4, 3, 5, 4, 2\}$.

Using a Bayesian approach, we have three components:

- *Likelihood function*: This is the parametric form of our data, $p(y|\theta)$, given a set of parameters, θ .
- *Prior distribution*: This is our belief of the distribution of our parameters prior to seeing the data, $p(\theta)$. In this case, we might assume that the die is fair: $p(\theta) = 1/6$.
- *Posterior distribution*: The result of updating our prior belief about the parameters given the data, $p(\theta|y) \propto p(\theta) * p(y|\theta)$.

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 a simple example, we'll use the procedure to estimate the annual number of worldwide earthquakes greater than 6.0 magnitude. The data includes 14,014 earthquakes occurring since 1900. Let's assume it's 2011 and we are tasked with estimating the number of 6.0 magnitude earthquakes in 2012.

First, let's get the data we'll be using in our example:

```
data("eqlist", package = 'CASdatasets')
```

```
data.table::setDT(x = eqlist)
```

```
print(eqlist)
```

```
##               time latitude longitude   depth   mag magType   nst
##               <char>    <num>    <num>    <num> <num> <char> <int>
##      1: 2024-05-26T20:47:09.259Z -19.3769 -174.8692 112.235 6.60      mww   102
##      2: 2024-05-25T22:23:16.981Z -17.1088  167.8842  29.182 6.30      mww   169
##      3: 2024-05-19T09:35:24.992Z  52.1586 -170.8669  28.000 6.00      mww   209
##      4: 2024-05-12T11:39:14.137Z  14.4518  -92.3630  75.426 6.40      mww   148
##      5: 2024-05-08T08:17:15.442Z -15.1394  167.9978  12.450 6.10      mww    76
##      ---
## 14010: 1901-12-31T09:02:30.000Z  51.4500 -171.0200      NA 7.10      ms    NA
## 14011: 1901-12-30T22:34:00.000Z  52.0000 -160.0000      NA 7.00      ms    NA
## 14012: 1901-11-14T04:39:00.000Z  38.5000 -112.4000      NA 6.60      mw    NA
## 14013: 1901-03-03T07:45:00.000Z  36.0000 -120.5000      NA 6.40      ms    NA
## 14014: 1900-10-09T12:24:00.000Z  58.3900 -152.4200      NA 7.86      mw    NA
##      gap dmin   rms   net      id
##      <num> <num> <num> <char>    <char>
##      1:   21 4.679 0.86      us    us6000n158
##      2:   28 4.830 0.77      us    us6000n102
##      3:   59 0.857 1.24      us    us6000mzme
##      4:   40 0.788 1.30      us    us6000my20
##      5:   33 6.438 1.30      us    us6000mx5c
##      ---
## 14010:      NA      NA      NA    ushis    ushis399
## 14011:      NA      NA      NA    ushis    ushis397
## 14012:      NA      NA      NA      uu    uu19011114043900000
## 14013:      NA      NA      NA    ushis    ushis393
## 14014:      NA      NA      NA official official19001009122400000
##      updated
##      <char>
##      1: 2024-05-27T20:55:30.551Z      70 km NW of Fangale'ounga, Tonga
##      2: 2024-05-26T22:34:09.479Z      83 km NNW of Port-Vila, Vanuatu
##      3: 2024-05-28T09:45:33.298Z      161 km WSW of Nikolski, Alaska
##      4: 2024-05-14T07:31:58.254Z 17 km WSW of Brisas Barra de Suchiate, Mexico
##      5: 2024-05-10T03:06:36.063Z      99 km ENE of Luganville, Vanuatu
##      ---
## 14010: 2024-03-11T00:15:29.211Z      Fox Islands, Aleutian Islands, Alaska
## 14011: 2024-03-11T00:20:09.462Z      South of Alaska
## 14012: 2024-03-12T08:42:50.712Z      Near Marysville, Utah
## 14013: 2023-12-23T01:31:35.183Z      Near Parkfield, California
## 14014: 2024-02-06T00:34:23.351Z      Kodiak Island region, Alaska
##      type horizontalError depthError magError magNst   status
```

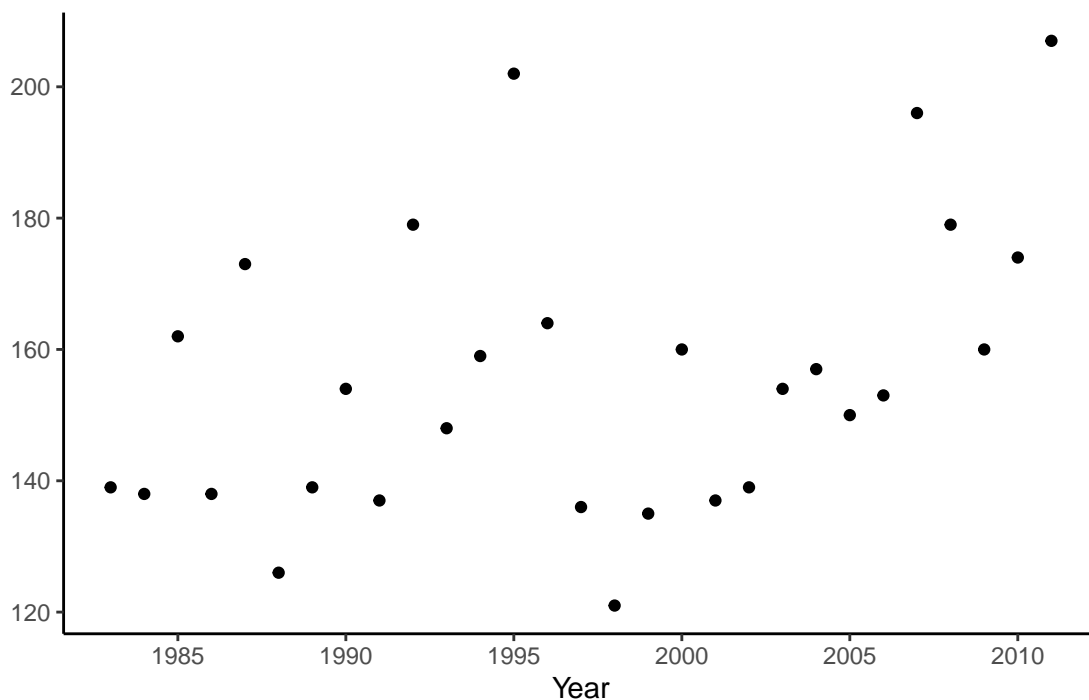
```

##          <char>          <num>      <num>      <num> <int>  <char>
##    1: earthquake      9.27      5.482      0.058    29 reviewed
##    2: earthquake      7.83      4.688      0.039    64 reviewed
##    3: earthquake      6.01      1.874      0.038    68 reviewed
##    4: earthquake      6.98      4.585      0.031   103 reviewed
##    5: earthquake      5.62      3.564      0.045    48 reviewed
##    ---
## 14010: earthquake      NA        NA        NA     NA reviewed
## 14011: earthquake      NA        NA        NA     NA reviewed
## 14012: earthquake      NA        NA        NA     NA reviewed
## 14013: earthquake      NA        NA        NA     NA reviewed
## 14014: earthquake    211.00      NA        NA     NA reviewed
##      locationSource magSource      day
##          <char>      <char>      <Date>
##    1:          us      us 2024-05-26
##    2:          us      us 2024-05-25
##    3:          us      us 2024-05-19
##    4:          us      us 2024-05-12
##    5:          us      us 2024-05-08
##    ---
## 14010:      ushis      abe 1901-12-31
## 14011:      ushis      abe 1901-12-30
## 14012:          uu      uu 1901-11-14
## 14013:      ushis      ell 1901-03-03
## 14014:      official      pt 1900-10-09

```

Next, we'll get the count data by year between 1983 and 2011.

Number of Worldwide 6.0 Magnitude Earthquakes by Year



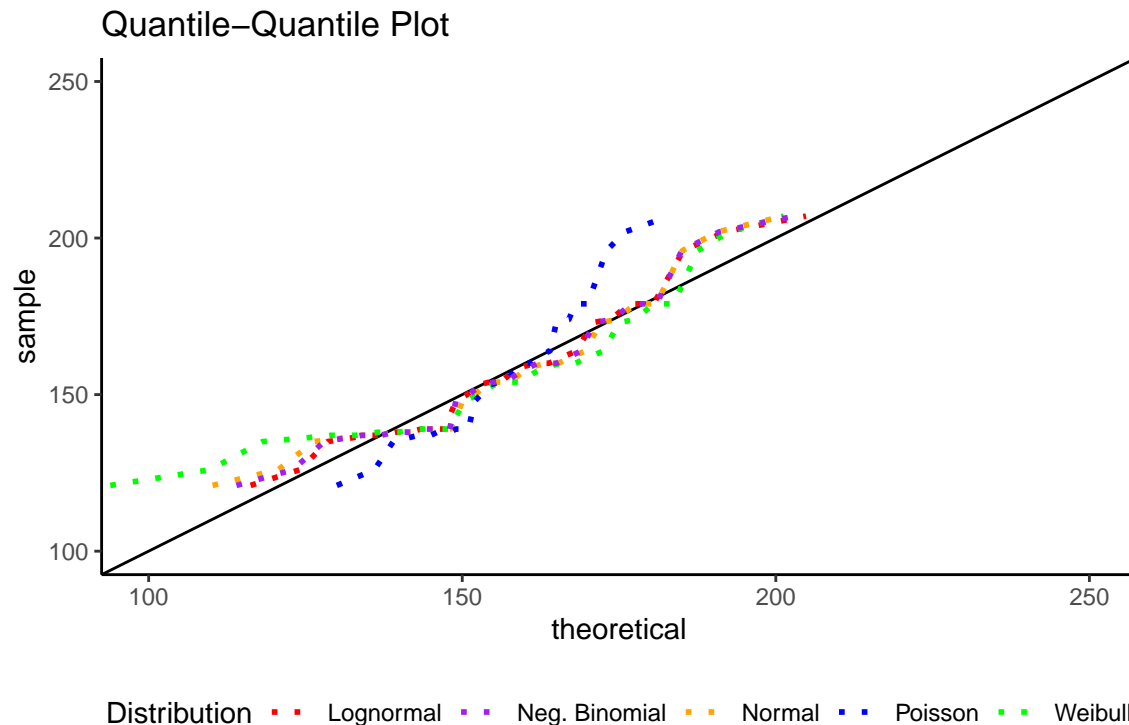
At first glance, you might make the following observations about the data:

- Since around 1997, there appears to be an upward trend.
- Something else.
- Something else.

In the simplest case, an actuary may decide that a straight average of historical counts is the best approximation of future counts. In this case, the average number per year is 156 (“Method A”).

However, you might decide that given the recent upward trend, an all-year average might be low. Instead, you opt to use a five-year average of 183 (“Method B”).

Your next logical step might be to add some process risk around your estimate by fitting a distribution to the data. Using the maximum-likelihood-estimates of a few common distributions, you settle on using a Lognormal distribution parameterized using a mean of 183 and coefficient of variation of 15% (based on all year coefficient of variation).



After running 10,000 simulations with your selected distribution, you’re left with a central estimate of 183 (“Method C”). 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.

So how do we handle these possible biases? That’s where Bayesian MCMC comes in. Our first model will assume a wide prior. We’ll start with our distribution from Method C:

- $\mu \sim \text{Normal}(5.199, 0.26)$
- $\sigma \sim \text{Exponential}(0.149)$

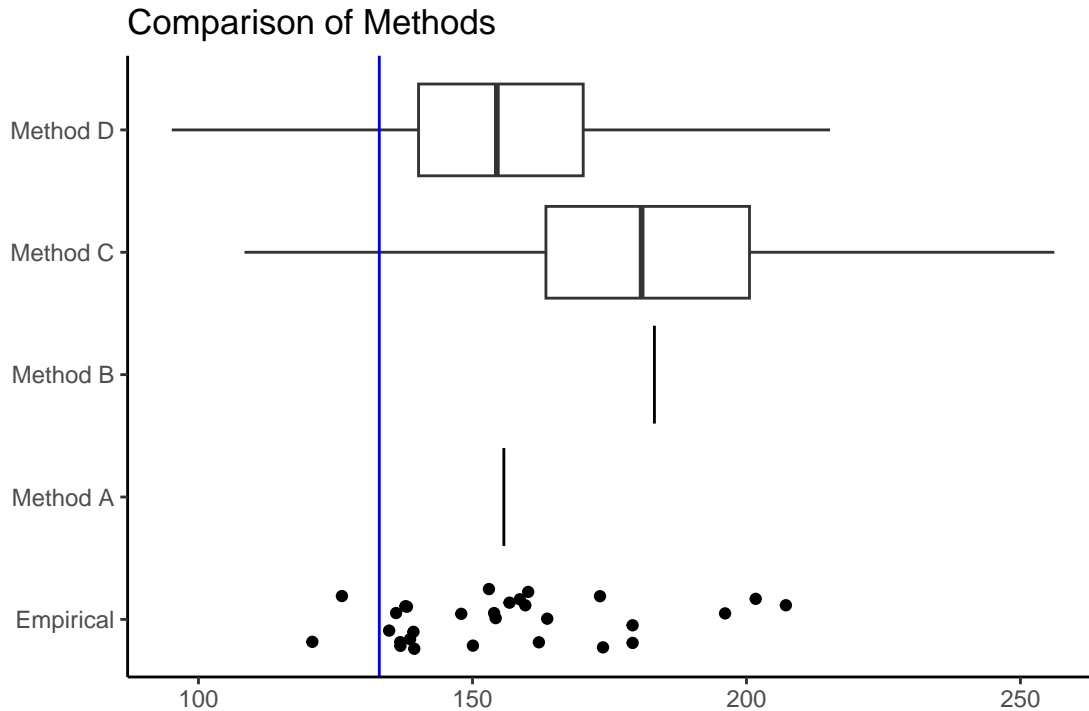
After running 10,000 simulations with your selected distribution, you’re left with a central estimate of 156 (“Method D”). Since you used a wide prior, your result will be fairly close to Method A even though you

had a much higher prior. That is because the data is overwhelmingly lower, thus a higher estimate is less likely.

Next, let's test "tighter" our prior. After all, we're pretty confident of the recent trend. We'll assume that our Lognormal distribution has the following revised prior distributions:

- $\mu \sim \text{Normal}(5.199, 0.052)$
- $\sigma \sim \text{Exponential}(6.704)$

Let's compare each result to the true number of fires that occurred in 2012 (133).

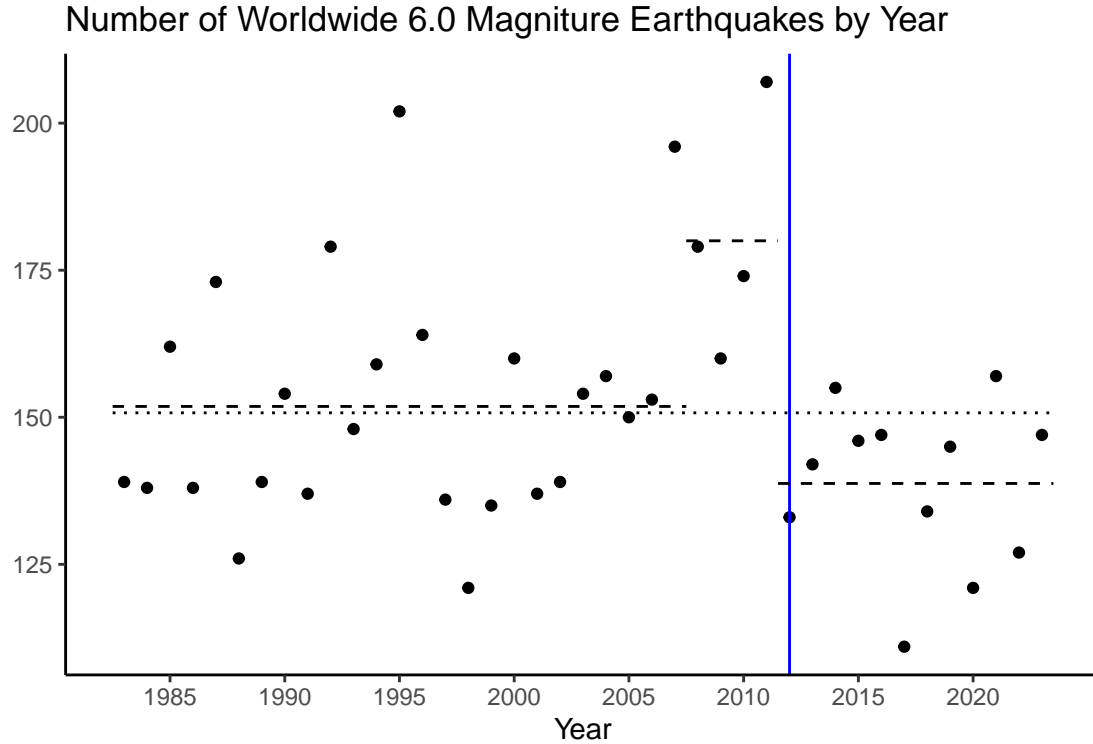


It appears that all four methods severely overestimated the true counts. However, if we look at Method C versus Method D, the true count is approximately equal to the 1.7% confidence level of Method C, but the 14.4% confidence level of Method D.

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 2008-2011 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 (1983-2007).¹

earthquakes: From the USGC



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:

- Reserving²
- Renewal Functions³
- Risk Margins⁴
- Trends⁵

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

- [1] A. Tversky, and D. Kahneman, Judgment under uncertainty: Heuristics and biases, Science 185, 1124–1131 (1974).
- [2] Meyers, G. 2015. “Stochastic Loss Reserving Using Bayesian MCMC Models.” CAS Monograph #1.
- [3] 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).
- [4] Meyers, G. 2019. “A Cost-of-Capital Risk Margin Formula for Nonlife Insurance Liabilities.” Variance 12 (2).
- [5] Schmid, F. 2013. “Bayesian Trend Selection.” Casualty Actuarial Society E-Forum, Spring 2013.

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>.