Introduction to Bayesian MCMC Techniques: A Practical Guide for the Lay-Actuary

Spencer Miller & Kenny Smart

2025-06-13

Abstract

To be completed later.

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 find themselves using heuristics. As outlined by Amos Tversky and Daniel Kahneman in their ground-breaking paper, there are several factors at play when making predictions:

- *Insensitivity to sample size*: An actuary may predict some future value based on limited data.
- *Insensitivity to prior probability of outcomes*: Example?
- *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.

Basics of Bayesian MCMC

Need an introductory sentence.

- A **Markov process** is a random process for which the future depends only on the present state; it has no memory of how the present state was reached.
- **Monte Carlo** simulation uses random sampling and statistical modeling to estimate mathematical functions and mimic the operations of complex systems.
- **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.
- This process is done using Stan , an open-source software, via common coding languages (e.g., python and R). In our example, we leverage the RStan package, though there are numerous other packages that can be used.
- Need to discuss the key parameters (thin, chains, etc.)

A Simple Walkthrough

Kenny to provide example. Ideally this is a step-by-step with graphics so that's super clear what is going on. We can gloss over the more complex stuff. Let's use a weighted dice as an example where you test frequency of rolling a 1 (e.g., the prior would be 1/6 but the die might be weighted such that a 6 comes up twice as often).

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:

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

data.table::setDT(x = eqlist)

dplyr::glimpse(eqlist)
```

```
## Rows: 14,014
## Columns: 23
## $ time        <chr> "2024-05-26T20:47:09.259Z", "2024-05-25T22:23:16.981Z"~
## $ latitude    <dbl> -19.3769, -17.1088, 52.1586, 14.4518, -15.1394, -3.313~
## $ longitude   <dbl> -174.8692, 167.8842, -170.8669, -92.3630, 167.9978, 13~
## $ depth       <dbl> 112.235, 29.182, 28.000, 75.426, 12.450, 13.099, 63.33~
## $ mag         <dbl> 6.6, 6.3, 6.0, 6.4, 6.1, 6.1, 6.1, 6.5, 6.1, 6.1, 6.3,~
## $ magType     <chr> "mww", "mww", "mww", "mww", "mww", "mww", "mww", "mww"~
## $ nst         <int> 102, 169, 209, 148, 76, 102, 231, 217, 120, 119, 386, ~
## $ gap         <dbl> 21, 28, 59, 40, 33, 28, 19, 24, 22, 35, 29, 42, 23, 31~
## $ dmin        <dbl> 4.679, 4.830, 0.857, 0.788, 6.438, 1.342, 0.637, 2.427~
## $ rms         <dbl> 0.86, 0.77, 1.24, 1.30, 1.30, 0.60, 0.66, 0.60, 0.67, ~
## $ net         <chr> "us", "us", "us", "us", "us", "us", "us", "us", "us", ~
## $ id          <chr> "us6000n158", "us6000n102", "us6000mzme", "us6000my20"~
## $ updated     <chr> "2024-05-27T20:55:30.551Z", "2024-05-26T22:34:09.479Z"~
```

```
## $ place          <chr> "70 km NW of Fangale'ounga, Tonga", "83 km NNW of Port~
## $ type           <chr> "earthquake", "earthquake", "earthquake", "earthquake"~
## $ horizontalError <dbl> 9.27, 7.83, 6.01, 6.98, 5.62, 5.76, 6.02, 9.43, 3.41, ~
## $ depthError      <dbl> 5.482, 4.688, 1.874, 4.585, 3.564, 3.331, 2.947, 3.133~
## $ magError        <dbl> 0.058, 0.039, 0.038, 0.031, 0.045, 0.039, 0.044, 0.035~
## $ magNst          <int> 29, 64, 68, 103, 48, 63, 50, 77, 23, 28, 69, 18, 58, 1~
## $ status          <chr> "reviewed", "reviewed", "reviewed", "reviewed", "revie~
## $ locationSource  <chr> "us", "us", "us", "us", "us", "us", "us", "us", "us", ~
## $ magSource       <chr> "us", "us", "us", "us", "us", "us", "us", "us", "us", ~
## $ day             <date> 2024-05-26, 2024-05-25, 2024-05-19, 2024-05-12, 2024-~
```

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

Number of Worldwide 6.0 Magniture Earthquakes by Year
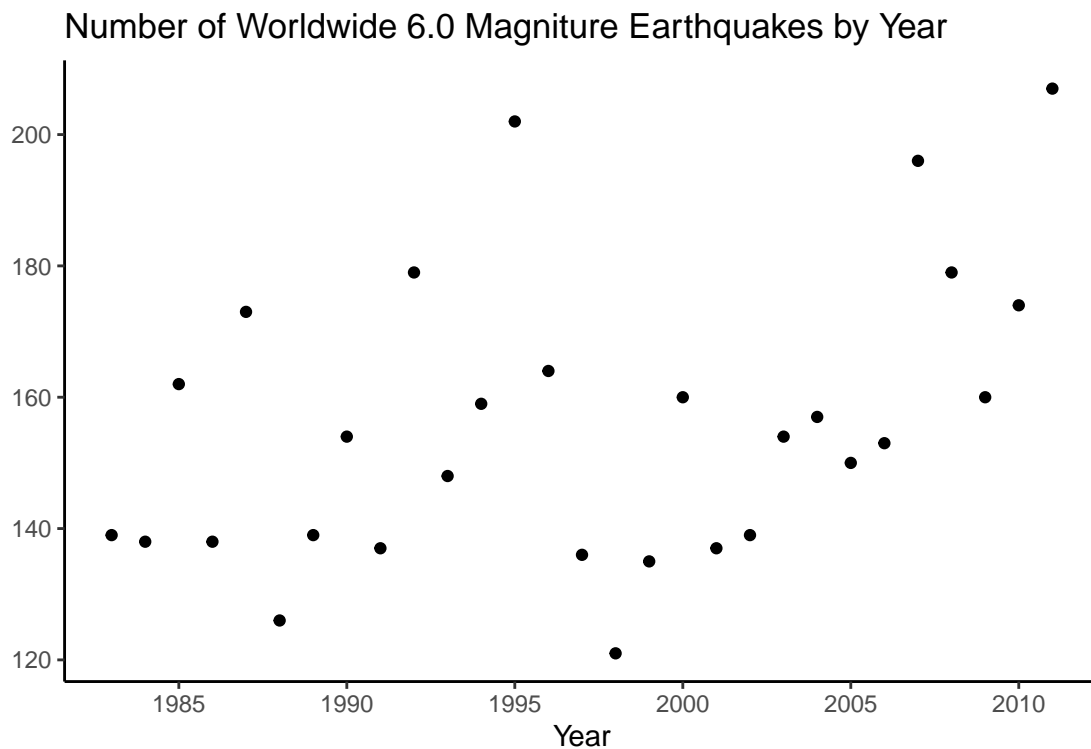
At first glance, you might make the following observations about the data:

- 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).
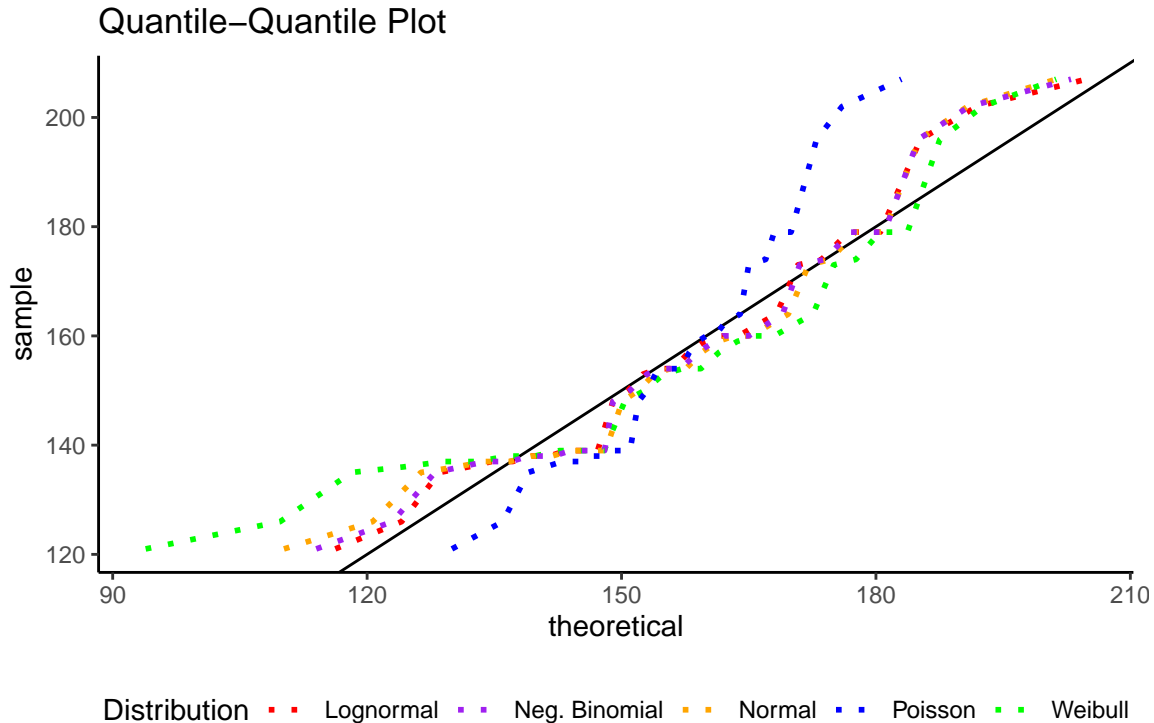
Quantile–Quantile Plot

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* ~ Normal(5.199, 0.26)
- *sigma* ~ 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* ~ Normal(5.199, 0.052)
- *sigma* ~ Exponential(6.704)

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

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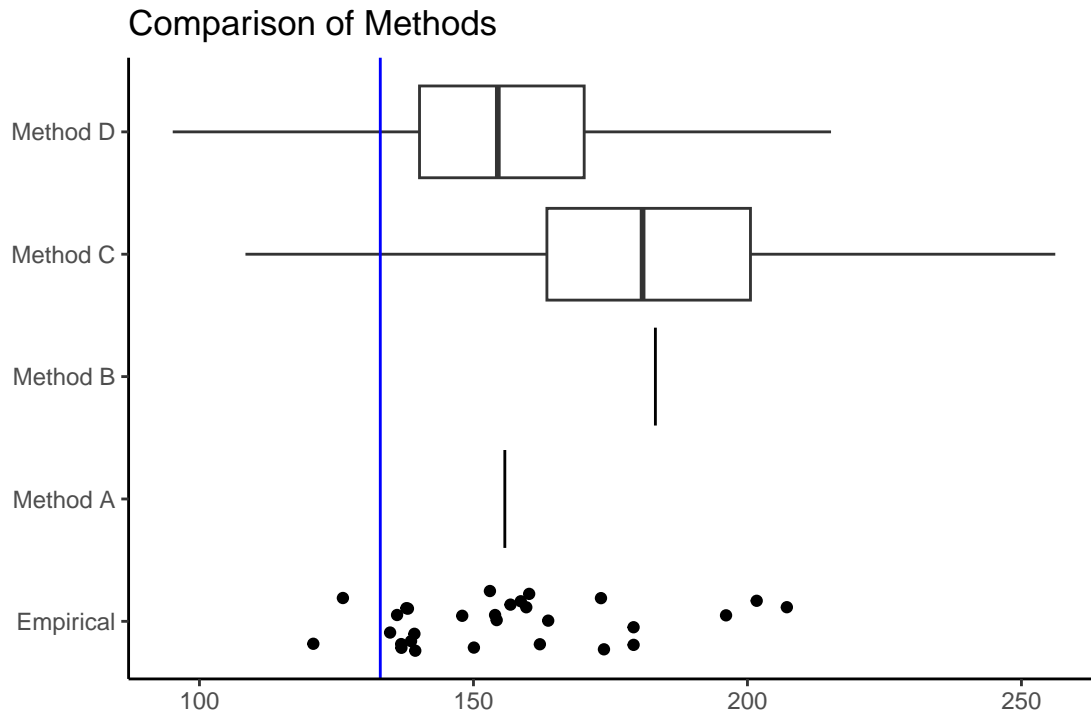
It appears that all four methods severely overestimated the true counts. However, if we look Method C versus Method D, the true count is approximately equal to the 1.7% confidence level of Method C but the 14.4% confidence level of Method D.

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 alternative 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 prior to the rise (1983-2007).

Number of Worldwide 6.0 Magnitude Earthquakes by Year
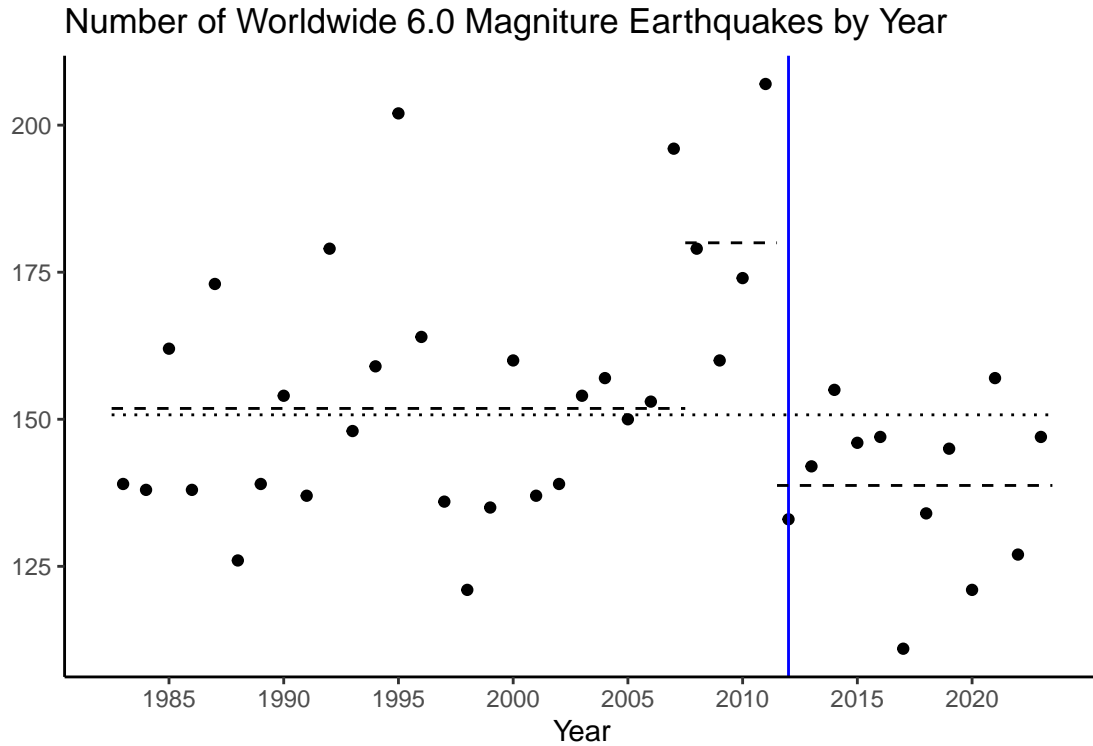
Literature Review

Bayesian MCMC is currently being used by actuaries in various applications:

- Reserving (Meyers 2015)
- Renewal Functions (Aminzadeh & Deng 2022)
- Risk Margins (Meyers 2019)
- Trends (Schmid 2013)

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

Acknowledgements

Thank everyone that peer review

Biographies of the Authors

Spencer Miller

To be completed later

Kenny Smart

To be completed later

Citations

References

A. Tversky, and D. Kahneman, Judgment under uncertainty: Heuristics and biases, Science 185, 1124–1131 (1974).

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

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. 2019. "A Cost-of-Capital Risk Margin Formula for Nonlife Insurance Liabilities." Variance 12 (2).

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

Software

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.