Baysian Inference using Markov Chain Monte Carlo for Siesmic Data set

Tejas Badakere Gopalakrishna[a]

[a] *University of California San Diego*

Abstract

In this project, a Metropolis hasting MCMC is performed to sample the parameters from an unknown distribution of a static model used in modeling the relative plate motion in central California. Parameters such as fault slip rate, locking depth, and the fault location are found using the Bayesian inference technique. 200 Observational data, such as cross-fault location and the measured strike-slip displacement, are used to carry on the Bayesian inference via MCMC. The posterior distribution of the parameters of the model is found assuming a uniform prior. Prediction of the slip-distance velocity is made by randomly sampling the parameters from the posterior distribution and evaluating at these points. 95% credible and prediction interval for prediction is also plotted. Narrow intervals, even in the absence of data, indicate that the model is accurate in its prediction capability.

1. Siesmic Model

$$v(x; D, x_0, v_0) = \frac{v_0}{\pi} \arctan\left(\frac{x - x_0}{D}\right) \tag{1}$$

where

- v is the strike-slip displacement §,
- x is the cross fault coordinate,
- D is the locking depth,
- x_0 is the location of the fault
- v_0 is the fault slip rate.

The parameters of the model are sampled using bayesian inference, with the help of observational data. We assume that the error in the observation is normally distributed, and hence we write the observational displacement in terms of the model as

$$y_i = v_i(x_i; D, x_0, v_0) + \varepsilon_i \tag{2}$$
$$\varepsilon_i \sim \mathcal{N}(0, 1) \tag{3}$$

2. Derivation of Posterior Distribution

We know that the strike-slip displacement observation is related to the model as depicted by equation 3. From this, we can calculate the likelihood function calculated by

$$\pi(y_1, \ldots, y_n \mid x_0, v_0, D) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left(-\frac{1}{2\sigma^2}\left(y_i - \frac{v_0}{\pi}\arctan\left(\frac{x_i - x_0}{D}\right)\right)^2\right) \tag{4}$$

Therefore, the posterior is given by

$$P(x_0, v_0, D \mid y_1, \ldots, y_n) = \frac{\pi(y_1, \ldots, y_n \mid x_0, v_0, D) \cdot P(x_0, v_0, D)}{\int_{x_0, v_0, D} \pi(y_1, \ldots, y_n \mid x_0, v_0, D) \cdot P(x_0, v_0, D)\, dx_0\, dv_0\, dD} \tag{5}$$

This is the distribution through which we sample the parameters x_0, v_0, D, instead of the frequentist approach that has fixed values for these parameters based on the observational data. This ensures the predicted slip-displacement is much more robust when the data is limited.

Assumptions made

- observational error is normally distributed across the data points
- Error is independent of the parameters x_0, v_0, D chosen;
- The prior is uniform and the parameters x_0, v_0, D are independent of each other.

3. Sampling from the posterior via Metropolis-Hastings Markov Chain Monte Carlo

Equation 5 is pretty much difficult to evaluate due to a high-dimensional integral in the denominator. Therefore, instead of trying to assess the distribution, it is enough to sample the parameters from the distribution, and hence, an efficient sampling technique likethe MCMC algorithm will be used to sample the parameters from the seemingly complicated distribution.

3.1. Metropolis hastings

Metropolis-Hastings is an MCMC algorithm used to sample from the posterior distribution. At each step, it proposes a new sample using a proposal distribution $q(x' \mid x)$, and accepts it with probability α.

Algorithm 1: Metropolis-Hastings Algorithm

1: **Input:** Target density $\pi(X)$ (up to normalization), proposal distribution $q(X' \mid X)$, number of samples T
2: **Initialize:** Choose initial state X_0
3: **for** $t = 1$ to T **do**
4: Sample $X' \sim q(X' \mid X_{t-1})$ {Propose a new candidate}
5: Compute acceptance ratio:

$$\alpha = \min\left(1, \frac{\pi(X') \cdot q(X_{t-1} \mid X')}{\pi(X_{t-1}) \cdot q(X' \mid X_{t-1})}\right)$$

6: Sample $u \sim \text{Uniform}(0,1)$
7: **if** $u < \alpha$ **then**
8: Accept: $X_t = X'$
9: **else**
10: Reject: $X_t = X_{t-1}$
11: **end if**
12: **end for**
13: **Return:** Samples $\{X_1, X_2, \ldots, X_T\}$ =0

3.2. *Random Walk Metropolis*

If the assumption for the proposal distribution is made to be Gaussian normal, then this class of Metropolis-Hastings is called Random Walk Metropolis. In this project, a Random Walk Metropolis is used to sample from the posterior distribution. Therefore, the algorithm 1 simplifies to the algorithm shown in 2

Algorithm 2: Random Walk Metropolis Algorithm

1: **Input:** Target density $\pi(X)$ (unnormalized), number of samples T
2: **Initialize:** Choose initial state X_0
3: **for** $t = 1$ to T **do**
4: Sample $X' \sim \mathcal{N}(X_{t-1}, \Sigma)$ {Propose a new candidate}
5: Compute acceptance ratio: $\alpha = \min\left(1, \frac{\pi(X')}{\pi(X_{t-1})}\right)$
6: Sample $u \sim \mathcal{U}(0,1)$
7: **if** $u < \alpha$ **then**
8: Accept: $x_t = X'$
9: **else**
10: Reject: $X_t = X_{t-1}$
11: **end if**
12: **end for**
13: **Return:** Samples $\{X_1, X_2, \ldots, X_T\}$ =0

Few Observations from the Random Walk Metropolis Algorithm

- The proposal distribution is a Gaussian normal, by construction it is a symmetric distribution, i.e, $q(X_{t-1} \mid X') = q(x' \mid X_{t-1})$. This explains the simplification of acceptance ratio in step 5 of 2
- The variance of the proposal distribution can be found by maximizing the likelihood function given in 4 with the variance. However, this is not essential,and a simple diagonal co-variance assumption can be made
- The initial guess X_0 is also made by maximizing the likelihood with respect to the parameters of the model. This is essential as it can reduce the amount of burnout period.
- A note to be made is that X's in the algorithm refers to the multi-dimensional random variable corresponding to the three parameters x_0, v_0, D
- Target density $\pi(X)$ refers to 5

4. Results

In this section, detailed results of the outcome of the RWC is explained. The initial state and co-variance matrix are assigned based on the likelihood function. A total of 10,000 iterations are being carried, out of which the first few, around 1000, are being rejected. Figure 1 shows the samples of the distribution as a function of iterations. The high amplitude and frequency of the variations of samples indicate that the algorithm has converged well, and the RWC algorithm has successfully converged to the posterior distribution 5. From the figure, we can also see that after 1000 iterations, the algorithm has started converging, and hence, we accept the samples only after the 1000th iteration.

The Random Walk Metropolis assumes normal distribution as a proposal distribution, then means $q(X' \mid X_{t-1})$ is assumed to be normal distribution ($\mathcal{N}(X_{t-1}, \Sigma)$), whose mean is the current sample and the variance is heuristically calculated based on optimal acceptance ratio of 0.234 ((1)). This variance of the proposal distribution is given below.

$$\Sigma = \begin{bmatrix} 0.20 & -0.10 & 0.10 \\ -0.10 & 0.56 & -0.10 \\ 0.10 & -0.10 & 0.80 \end{bmatrix}$$

Figure 2a shows the corner plot, which plots the marginal distribution of one variable at a time. From the figure, we can observe that the variance between D and v_0 is non-diagonal and hence the underlying distributions of the parameters are not independent (at least two of the variables are dependent) Once the distributions of the parameters of the model are found, the accuracy of the fit needs to be checked. This is done by plotting the credible and prediction intervals of the model. The former only takes into account the uncertainty in the model where as the latter takes into account the uncertainty in measurements as well.

4.1. *Credible and Prediction Interval*

The observational displacement is given by the equation 3. Taking the expectation of the equation 3 yields

$$\mathbb{E}[y] = \mathbb{E}[v(x; D, x_0, v_0)] + \mathbb{E}[\varepsilon] \quad (6)$$
$$\mathbb{E}[y] = \mathbb{E}[v(x; D, x_0, v_0)] \quad (7)$$
$$\forall \quad \mathbb{E}[\varepsilon] = 0 \quad (8)$$

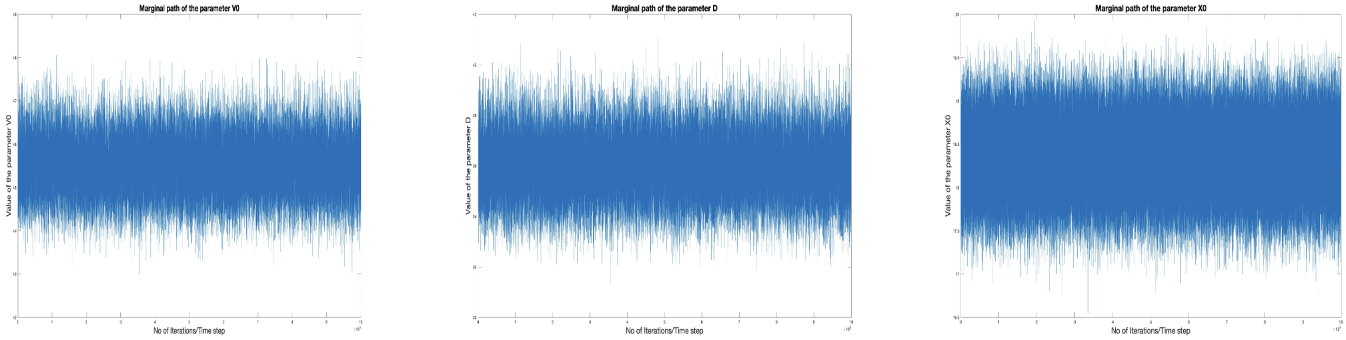

Figure 1: Samples of three Random variables v_0, D, X_0 as the function of iterations. The convergence to the target can be seen right from the initial iterations due to the good choice of initial conditions.

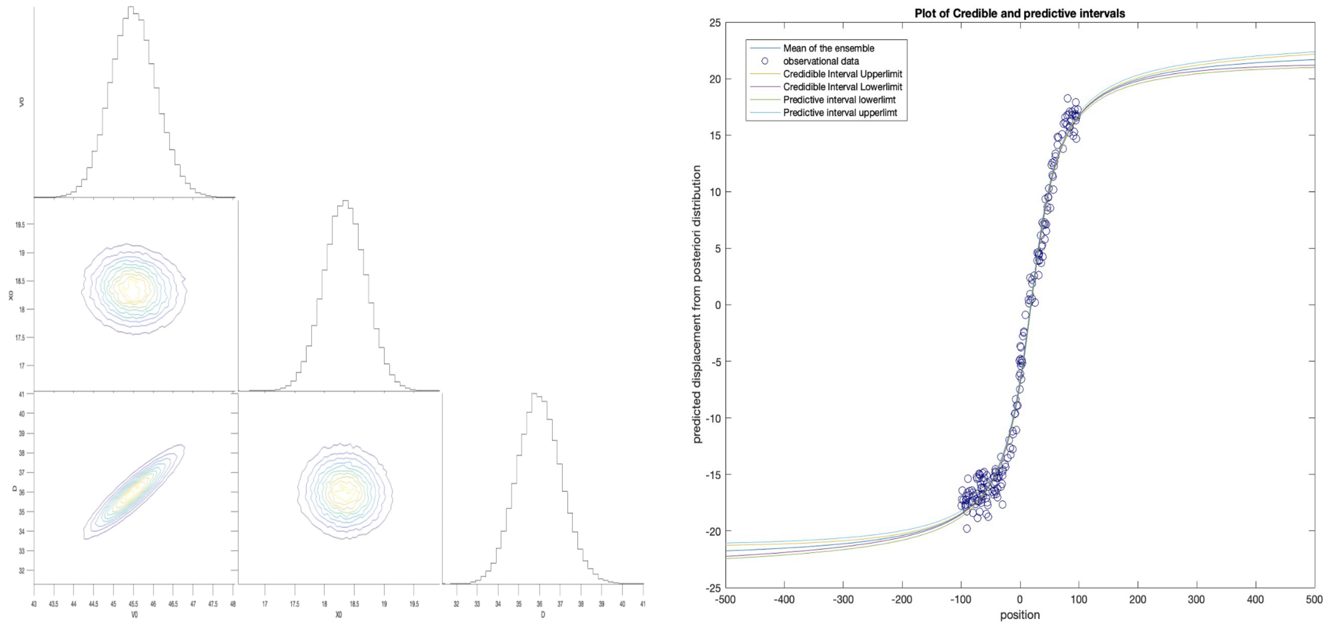

Figure 2: (a) Corner plot of three Random variables v_0, D, X_0 depicting contour plots taking two variables at a time, and marginal distribution for each of the variables. (b) Plot of 95 % prediction and credible intervals indicating confidence in the model's prediction.

We are interested in the 95 % credible and the prediction interval. These are calculated respectively as follows.

$$\begin{aligned}\textbf{95\% Credible interval} &= \mathbb{E}[v(x;D,x_0,v_0)] \\ &\quad \pm 2\cdot \text{std}[v(x;D,x_0,v_0)] \\ \textbf{95\% Prediction interval} &= \mathbb{E}[v(x;D,x_0,v_0)] \\ &\quad \pm 2\cdot \text{std}[v(x;D,x_0,v_0)] + 2\cdot \text{std}[\varepsilon]\end{aligned} \tag{9}$$

Now the only task is to compute this equation for every x. Since we do not have the distribution of the parameters, but the samples from it, the only way to evaluate the expectation is through Monte Carlo sampling. The standard deviation can also be calculated by monte carlo sampling by rewriting the equation for the variance. We know that variance of the variable can be calculated as follows.

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{10}$$

$$\text{Var}\big(v(x;D,x_0,v_0)\big) = \mathbb{E}\left[v(x;D,x_0,v_0)^2\right] - \left(\mathbb{E}\left[v(x;D,x_0,v_0)\right]\right)^2 \tag{11}$$

Equation 8 and 11 can be calculated through Monte Carlo sampling.

Figure 2b shows the plot of model expectation along with the 95% prediction and credible interval. It can be observed that the both intervals are smaller when the data is present, and wider along the edges, where the data is absent. Note that the data observational data is present only for the spatial coordinate between -100 to 100. Hence, in this region the 95 % credible and prediction interval is narrower than the region where the data is absent. Nonetheless, even at the extreme ends, the prediction and credible interval is narrow, which ensures that the model is confident about its perdiction in the region where the data is absent.

References

[1] Andrew Gelman, Walter R. Gilks, and Gareth O. Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.