Bayesian Inference using Markov Chain Monte Carlo (MCMC), Part 2

William Wong

November 16, 2021

Outline

- Review of MCMC
- PyMC3 examples
  - Linear regression
  - Coal mining disasters
  - Time series modeling
  - Mixed media modeling

Review

- What is Bayesian inference?
  Prior assumptions $\rightarrow$ Look at evidence from data $\rightarrow$ update prior beliefs $\rightarrow$ posterior distribution.
- The law of total probability $\Pr(A) = \Pr(A|B_1)\Pr(B_1) + \Pr(A|B_2)\Pr(B_2) + \dots$
- θ: model parameters, X: observations

$$\begin{aligned}
\Pr(\theta|X) &= \frac{\Pr(X|\theta)\Pr(\theta)}{\Pr(X)} \\
\Pr(\theta) &= \text{prior; can make our own choice!} \\
\Pr(X|\theta) &= \text{likelihood; fixed by model} \\
\Pr(\theta|X) &= \text{posterior} \\
\Pr(X) &= \text{evidence; fixed by data}
\end{aligned}$$

- The math gets messy when we rewrite $\Pr(X)$ using the law of total probability.

$$\begin{aligned}
\Pr(\theta|X) &= \frac{\Pr(X|\theta)\Pr(\theta)}{\Pr(X)} \\
&= \frac{\Pr(X|\theta)\Pr(\theta)}{\int \Pr(X|\theta)\Pr(\theta)d\theta}
\end{aligned}$$

- When we update our belief, it gets very expensive because we have to compute an integral.

Question — how is Markov Chain used?
Question — how is Monte Carlo used?

- Stata video, part 1 (00:30 to 4:30).
    - coin toss. $\theta = \Pr(\text{head})$
    - posterior $\propto$ prior $\bullet$ likelihood. Generally it is not a multiplication but note that we are using a conjugate prior!
    - $\Pr(\theta|y) \propto \Pr(\theta)\Pr(y|\theta) = \text{Beta}(1,1) \bullet \text{Binomial}(4, 10, \theta)$ // 4 heads out of 10 tosses.
- Stata video, part 2 (00:00 to 5:01).
    1. Markov chain
    2. Monte Carlo
    3. Metropolis-Hastings algorithm

PyMC3

- PyMC3 supports Bayesian inference on user-defined probabilistic models.
- To go over a Jupyter notebook.
- Example 1 — Linear regression where the response function is normally distributed, where the mean is a linear function of two predictor variables, X_1 and X_2.

$$
\begin{aligned}
Y &\sim \mathcal{N}(\mu, \sigma^2) \\
\mu &= \alpha + \beta_1 X_1 + \beta_2 X_2
\end{aligned}
$$

Priors

$$
\begin{aligned}
\alpha &\sim \mathcal{N}(0, 100) \\
\beta_i &\sim \mathcal{N}(0, 100) \\
\sigma &\sim |\mathcal{N}(0, 1)|
\end{aligned}
$$

PyMC3 (con't)

- Example 2 — Coal mining disasters
  In our model, **the disaster rate** r_t **changes at some time** s.

$$D_t \sim \text{Poisson}(r_t),\, r_t = \begin{cases} e, & \text{if } t \leq s \\ l, & \text{if } t > s \end{cases}$$
$$s \sim \text{Uniform}(t_l, t_h)$$
$$e \sim \exp(1)$$
$$l \sim \exp(1)$$

  The parameters are defined as follows:
  - D_t: The number of disasters in year t
  - r_t: The rate parameter of the Poisson distribution of disasters in year t.
  - s: The year in which the rate parameter changes (the switchpoint).
  - e: The rate parameter before the switchpoint s.
  - l: The rate parameter after the switchpoint s.
  - t_l, t_h: The lower and upper boundaries of year t.

Stationarity

- The stocastic process Y_t is weakly stationary if
  - $\mathbb{E}[Y_t]$ is independent of t.
  - $\mathrm{Var}(Y_t)$ is independent of t.
  - Covariance$[Y_t]$ is independent of t.
- **Question — what is a stochastic process? How is it related to random variables?**
- **Question — what are some examples of stationary and non-stationary time series?**

The Autocorrelation Function

- The autocorrelation of Y_t is $R_{YY}(t_1, t_2) \equiv \mathbb{E}[Y_{t_1} Y_{t_2}]$.
- If Y_t is stationary, $R_{YY}(\tau) = \mathbb{E}[Y_t Y_{t+\tau}]$, which is independent of t.
- $R_{YY}(\tau) = R_{YY}(-\tau)$.
- $|R_{YY}(\tau)| \leq R_{YY}(0)$.
- The Fourier transform of $R_{YY}(\tau)$ is the power spectrum of Y_t.
- **Question — what is the algorithmic complexity to compute $R_{YY}(\tau) = \mathbb{E}[Y_t Y_{t+\tau}]$?**

PyMC3 Time Series Modeling (con't)

The AR(1) Model

- $Y_t = \theta Y_{t-1} + \epsilon$.
- $\epsilon \sim \mathcal{N}(0, \sigma^2)$.
- Y_t is weakly stationary if $|\theta| < 1$.
- $\mathbb{E}[Y_t] = 0$.
- $\mathrm{Var}(Y_t) = \frac{\sigma^2}{1-\theta^2}$.
- $\mathrm{corr}(Y_t, Y_{t-h}) = R_{YY}(h) = \theta^h$.

The Jupyter notebook.

- The HelloFresh engineering blog

  In the model, the number of customers acquired in a week y_t is defined to be

$$y_t = \sum_{m=1}^{M} \beta_m f(x_{mt}) + \sum_{c=1}^{C} \beta_c z_{ct} + \epsilon_t$$
$$\beta_m \sim |\mathcal{N}(0,1)|$$
$$\beta_c \sim \mathcal{N}(0,1)$$

  The parameters are defined as follows:

  - x_{mt}: The amount of money spent on a marketing channel m in week t.
  - $f()$: A saturation and decay function.
  - β_m: The effect of channel m on customer acquisition.
  - z_{ct}: The value of control variable c in week t.
  - β_c: The effect of control variable c.

  After building a model, find out how to spend our marketing budget on the various channels by solving a constrained optimization problem.