Markov Chain Monte Carlo

Felipe José Bravo Márquez

June 30, 2021

Markov Chain Monte Carlo

- This class introduces estimation of posterior probability distributions using a stochastic process known as **Markov chain Monte Carlo** (MCMC).
- Here we'll produce samples from the joint posterior without maximizing anything.
- We will be able to sample directly from the posterior without assuming a Gaussian, or any other, shape.
- The cost of this power is that it may take much longer for our estimation to complete.
- But the benefit is escaping multivariate normality assumption of the Laplace approximation.
- More advanced models such as the generalized linear and multilevel models tend produce non-Gaussian posterior distributions.
- In most cases they cannot be estimated at all with the techniques of earlier classes.
- This class is based on Chapter 9 of [McElreath, 2020] and Chapter 7 of [Kruschke, 2014].

Markov Chain Monte Carlo

- The essence of MCMC is to produce samples from the posterior $f(\theta|d)$ by only accessing a function that is proportial to it.
- This proportial function is the product of the likelihood and the prior $f(d|\theta) * f(\theta)$, which is always available in a Bayesian model.
- So, merely by evaluating $f(d|\theta) * f(\theta)$, without normalizing it by $f(d)$, MCMC allows us to generate random representative values from the posterior distribution.
- This property is wonderful because the method obviates direct computation of the evidence $f(d)$, which, as you'll recall, is one of the most difficult aspects of Bayesian inference.
- It has only been with the development of MCMC algorithms an software that Bayesian inference is applicable to complex data analysis.
- And it has only been with the production of fast and cheap computer hardware that Bayesian inference is accessible to a wide audience.
- The question then becomes this: How does MCMC work? For an answer, let's ask a politician.

A politician stumbles upon the Metropolis algorithm

- Suppose an elected politician lives on a long chain of islands.
- He is constantly traveling from island to island, wanting to stay in the public eye.
- At the end of a day he has to decide whether to:
  1. stay on the current island
  2. move to the adjacent island to the west
  3. move to the adjacent island to east
- His goal is to visit all the islands **proportionally** to their **relative population**.
- But, he doesn't know the total population of all the islands.
- He only knows the population of the current island where he is located.
- He can also ask about the population of an adjacent island to which he plans to move.

A politician stumbles upon the Metropolis algorithm

- The politician has a simple heuristic for travelling accross the islands.
- First, he flips a (fair) coin to decide whether to propose the adjacent island to the left or the adjacent island to the right.
- If the proposed island has a larger population than the current island ($P_{proposed} > P_{current}$), then he goes to the proposed island.
- If the proposed island has a smaller population than the current island ($P_{proposed} < P_{current}$), then he goes to the proposed island only probabilistically
- The probabiltiy of moving in that case is $p_{move} = P_{current}/P_{proposed}$.
- This is done by generating a uniform random number between 0 and 1, and moving when the number is lower than p_{move}.
- In the long run, the probability that the politician is on any one of the islands exactly matches the relative population of the island!

A politician stumbles upon the Metropolis algorithm

- Let's analyze this heuristic in more detail.
- Suppose there are 10 islands in total.
- Each island is neighbored by two others, and the entire archipelago forms a ring.
- The islands are of different sizes, and so had different sized populations living on them.
- The second island is twice as populous as the first, the third three times as populous as the first.
- And so on, up to the largest island, which is 10 times as populous as the smallest.

King Markov was a benevolent autocrat of an island kingdom, a circular archipelago, with 10 islands.

Conclusions

- Blabla

References I

Kruschke, J. (2014).
Doing bayesian data analysis: A tutorial with r, jags, and stan.

McElreath, R. (2020).
Statistical rethinking: A Bayesian course with examples in R and Stan.
CRC press.