Hawaiian Shirt: No!

Told Byron: yes.

Date: Thursday, February 14, 2019

Metropolis-Hastings:

- You can simplify any proposal distribution $J_t(\theta^{(t)} \mid \theta^{(t-1)})$ that you'd like.
- Why does this matter? When is this better than Metropolis? When a symmetric proposal does not seem to converge quickly to the desired target density.

Example: The true target is say Poisson(0,1). Then $N(\theta^{t-1}, \sigma^2)$ may not converge as fast as say a Poisson(θ^{t-1}).

Also, in this case any values sampled from the $N(\theta^{t-1}, \sigma^2)$ must be discretized to match the descrete nature of our target density.

Example: Suppose $J_t(\theta \mid \theta^{t-1})$ is the density of a Poisson(θ^{t-1}) random variable.

Sample $\theta^* \sim$ Poisson(θ^{t-1})

$$J_t(\theta^* \mid \theta^{t-1}) = \frac{(\theta^{t-1})^{\theta^*} e^{-\theta^{t-1}}}{\theta^*!}$$

observed sample

Parameter

$$J_t(\theta^{t-1} \mid \theta^*) = \frac{(\theta^*)^{\theta^{t-1}} e^{-\theta^*}}{\theta^{t-1}!}$$

Slide 17/19 ... Fix! Second part is correct.

Chapter 3

Key ideas:
1) The landscape of $\theta | y$
2) "Picking needles in the haystack"
3) MCMC Diagnostics
4) Application to Bayesian Clustering

1) The landscape of $\theta | y$

- Whenever we have multiple dimensions for θ, say N, we can talk about probabilities of N-dimensional coordinates as a curve/shape with a heatmap representing the probability and the boundary representing limits on the values.

Visualization of posteriors

- Our priors represent our origin space and shape
- Our data stretches and pulls the shape but does not change the limits set by the prior.

Example: $\left.\begin{array}{l} Y_1 \sim Po(\lambda_1) \\ Y_2 \sim Po(\lambda_2) \end{array}\right\}$ Priors for λ_1, λ_2 → New visual for posterior of $\lambda_1 | Y_1 ; \lambda_2 | Y_2$

- The more data, the more stretching and pulling of our prior shape.
- This reinforces the idea that the posterior is an updated belief of our parameters once data has been incorporated.

2) Finding Needles in the Haystack

- The aim of MCMC is to identify (1) and sample (2) likely values from our posterior contour plot.

  1) Identify → Estimate the N dimensional shape of the posterior
  2) Sample → Sample coordinates according to heatmap contour.

- Without smart strategies we'd never be able to do this! (we use MCMC, which guarantees that our eventually distribution from which we sample is the true contour of the posterior).

- Moving from one sample to the next via our proposal $J_t(\theta^t \mid \theta^{t-1})$ is essentially moving from a sampled point in the contour to a more probable point (if $J_t(\cdot)$ is reasonable).

- We need lots of samples because we want to explore a contour of the landscape.

- For example, we could easily get stuck in the most probable coordinate.

3) Diagnostics of MCMC

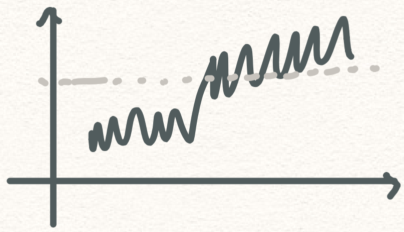
- Keep in mind that we're aiming to sample from a stationary distribution for each parameter.

- To figure out whether or not we have reached a stationary distribution, we look at trace plots of each sampled parameters.

What are we looking for?

- Trace plots just plots the sampled value of each parameter across samples. You have one plot per parameter.

1. A plot that varies around a stationary mean.



2. We do not want flat lines! This suggests that you no longer accept new values from your proposal. If this happens, modify your proposal typically to have lower variance.

3. [Heuristic]: It is recommended that your overall acceptance rate is around 0.25 (somewhere between 0.1 and 0.4).

Metrics you can use to determine stationarity

- Autocorrelation of the chain
  ↳ We want only a lag 1 dependence

- The Geweke statistic
  ↳ Formal hypothesis test statistic for a stationary mean.