Paper review: Elliptical Slice Sampling

Da Long 2022-2-23

Outline

- Brief review of existing MCMC methods: Metropolis Hastings, Gibbs, slice sampling
- Elliptical slice sampling model
- Experiments and results, compared with line slice, Neal M-H, and control point M-H sampling methods

Markov Chain Monte Carlo

- MCMC works by constructing and simulating a Markov Chain whose equilibrium distribution is the distribution of interest
- Burn-in stage and detailed balance (sufficient condition, not necessary)

$$\pi_{i+1}(x^*) = \int \pi_i(x)p(x \to x^*)dx$$

$$\pi_i * p(i, j) = \pi_j * p(j, i)$$

Metropolis-Hastings sampling

- $$\pi(x_i)p(x_i \to x_*) * \frac{\pi(x_*)p(x_* \to x_i)}{\pi(x_i)p(x_i \to x_*)} = \pi(x_*)p(x_* \to x_i)$$

- Algorithm:
  - Initialize x_0
  - Repeat until sufficient number of samples are generated
    - $u \sim U(0,1)$
    - $x_* \sim q(x_*|x_i)$
    - $if\ u < min[1, \frac{\pi(x_*)p(x_* \to x_i)}{\pi(x_i)p(x_i \to x_*)}]$
      - $x_{i+1} = x_*$

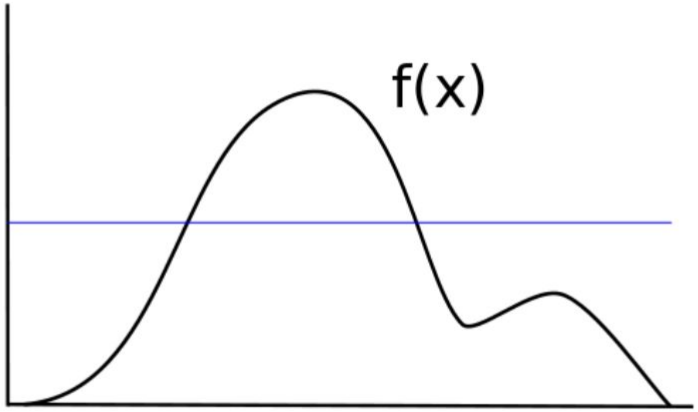
    $else \quad x_{i+1} = x_i$

Gibbs sampling

- Gibbs sampling is a algorithm for a specified multivariate probability distribution.
- A special case of Metropolis-Hastings sampling

Slice sampling

- Sampling uniformly from the region under the graph of its density function
-



Elliptical Slice Sampling

- Starting point: Neal M-H method on posterior distribution over latent variables that is proportional to the product of a multivariate Gaussian prior $p^*(f) = \frac{1}{Z} N(f|0, \Sigma) L(f)$
  - Neal M-H: $f' = \sqrt{1 - \epsilon^2} f + \epsilon v, v \sim N(0, \Sigma), with\, p(accept) = min(1, \frac{L(f')}{f(L)})$
  - Drawback:
    - Step-size ϵ needs to be chosen appropriately to mix efficiently. Usually different step-size parameters are needed as the model parameters are updated.
    - How to automatically search over the step-size parameter?

Elliptical Slice Sampling

- Neal M-H method: $f' = \sqrt{1-\epsilon^2} f + \epsilon v, v \sim N(0, \Sigma), with\, p(accept) = min(1, \frac{L(f')}{f(L)})$

$\rightarrow$ Elliptical slice sampling: $f' = v \sin\theta + f \cos\theta$

- The acceptance rate:

$$\frac{\pi(j)Q(j,i)}{\pi(i)Q(i,j)} = \frac{L(f')N(f'|0,\Sigma)N(f|0,\Sigma)}{L(f)N(f|0,\Sigma)N(f'|0,\Sigma)} = \frac{L(f')}{L(f)}$$

- For a fixed auxiliary random draw v, f' is a full ellipse passing through the current state f and the auxiliary draw v
- They are not equal, but elliptical slice sampling gives a richer choice of updates for a given v

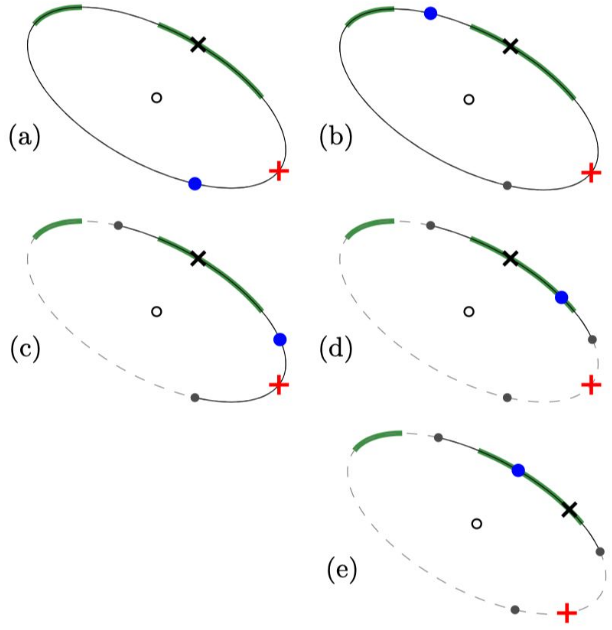
Elliptical Slice Sampling

- Algorithm:

Input: current state $\mathbf{f}$, a routine that samples from $\mathcal{N}(0, \Sigma)$, log-likelihood function $\log L$.

Output: a new state $\mathbf{f}'$. When $\mathbf{f}$ is drawn from $p^\star(\mathbf{f}) \propto \mathcal{N}(\mathbf{f}; 0, \Sigma)\, L(\mathbf{f})$, the marginal distribution of $\mathbf{f}'$ is also $p^\star$.

1. Choose ellipse: $\boldsymbol{\nu} \sim \mathcal{N}(0, \Sigma)$
2. Log-likelihood threshold:
$$u \sim \text{Uniform}[0, 1]$$
$$\log y \leftarrow \log L(\mathbf{f}) + \log u$$
3. Draw an initial proposal, also defining a bracket:
$$\theta \sim \text{Uniform}[0, 2\pi]$$
$$[\theta_{\min}, \theta_{\max}] \leftarrow [\theta - 2\pi,\ \theta]$$
4. $\mathbf{f}' \leftarrow \mathbf{f} \cos\theta + \boldsymbol{\nu} \sin\theta$
5. **if** $\log L(\mathbf{f}') > \log y$ **then:**
6. Accept: **return** $\mathbf{f}'$
7. **else:**
   Shrink the bracket and try a new point:
8. **if** $\theta < 0$ **then:** $\theta_{\min} \leftarrow \theta$ **else:** $\theta_{\max} \leftarrow \theta$
9. $\theta \sim \text{Uniform}[\theta_{\min}, \theta_{\max}]$
10. **GoTo** 4.



Elliptical Slice Sampling

- Proof:

  Goal: prove detailed balance: $\pi_i * p(i,j) = \pi_j * p(j,i)$

  Define:
  $$\boldsymbol{\nu}_k = \boldsymbol{\nu}\cos\theta_k - \mathbf{f}\sin\theta_k$$
  $$\mathbf{f}_k = \boldsymbol{\nu}\sin\theta_k + \mathbf{f}\cos\theta_k, \quad k = 1..K.$$

  The original pair: (f,v), the final state pair: (f',v'), when k=K.

  Needing to prove:

$$\pi_i * p(i,j) : p^*(f)p(y|f)p(v)p(\{\theta_k\}|f,v,y) = \frac{1}{Z}N(f|0,\Sigma)N(v|0,\Sigma)p(\{\theta_k\}|f,v,y)$$

?=

$$\pi_j * p(j,i) : p^*(f')p(y|f')p(v')p(\{\theta_k'\}|f',v',y) = \frac{1}{Z}N(f'|0,\Sigma)N(v'|0,\Sigma)p(\{\theta_k'\}|f',v',y)$$

Elliptical Slice Sampling

- Claim: Invariant joint prior:

$$N(v_k|0,\Sigma)N(f_k|0,\Sigma) = N(v|0,\Sigma)N(f|0,\Sigma)\ for\ all\ k.$$

$$\boldsymbol{\nu}_k = \boldsymbol{\nu}\cos\theta_k - \mathbf{f}\sin\theta_k$$
$$\mathbf{f}_k = \boldsymbol{\nu}\sin\theta_k + \mathbf{f}\cos\theta_k, \quad k = 1..K.$$

Use the Jacobian transformation to prove:

$$f_{v_k f_k}(v_k, f_k)|J|dv_k df_k = f_{vf}(v,f)dvdf$$

Here $|J| = |\begin{pmatrix} \partial v_k/\partial v & \partial v_k/\partial f \\ \partial f_k/\partial v & \partial f_k/\partial f \end{pmatrix}| = |\begin{pmatrix} cos\theta_k & -sin\theta_k \\ sin\theta_k & cos\theta_k \end{pmatrix}| = 1$

Therefore, the joint prior distribution is invariant.

Elliptical Slice Sampling

- Claim: $p(\{\theta_k\}|f, v, y) = p(\{\theta_k'\}|f', v', y)$

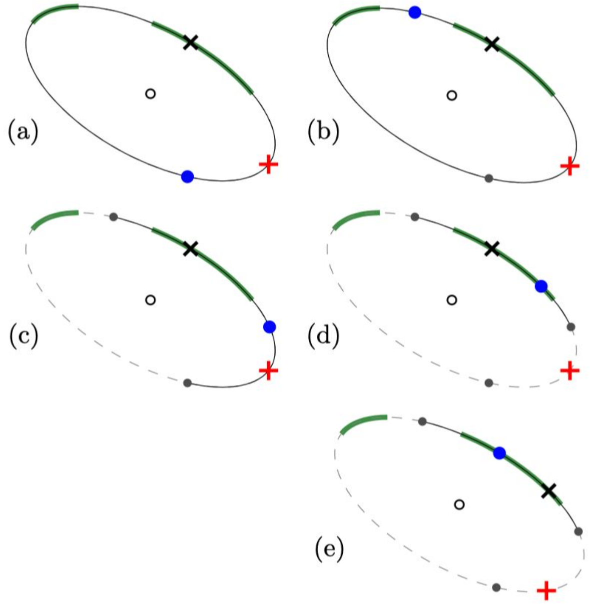
They are all uniform distributed.

First choice made: $$p(\theta_1) = p(\theta_1') = \frac{1}{2\pi}$$

Second choice made: $$p(\theta_2) = p(\theta_2') = \frac{1}{2\pi}$$

…

$$p(\theta_K) = p(\theta_K')$$



Experiments and results

- Three Gaussian process based model tasks as follows, compared with line slice, Neal M-H, and control point M-H sampling methods
  - Gaussian regression, varying the input dimensions from one to ten, data size N=200
  - Gaussian process classification, binary classification, USPS classification problem
  - Log Gaussian Cox process, a Cox process model of the dates of mining disasters

Experiments and results

- Gaussian Process Regression

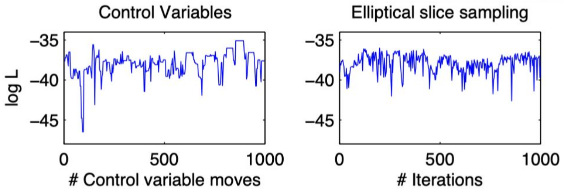


Figure 4: Traces of log-likelihoods for the 1-dimensional GP regression experiment. Both lines are made with 333 points plotted after each sweep through $M = 3$ control variables and after every 3 iterations of elliptical slice sampling.

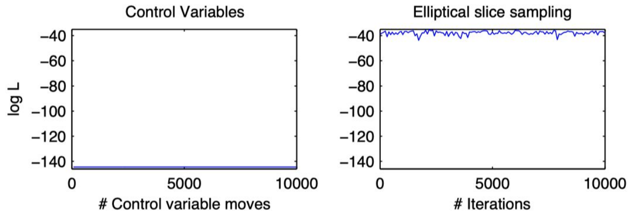


Figure 5: As in Figure 4 but for 10-dimensional regression and plotting every $M = 78$ iterations. (Control variables didn't move on this run.)

Experiments and results

On the USPS classification problem, control variables ran exceedingly slowly and no meaningful results were obtained.

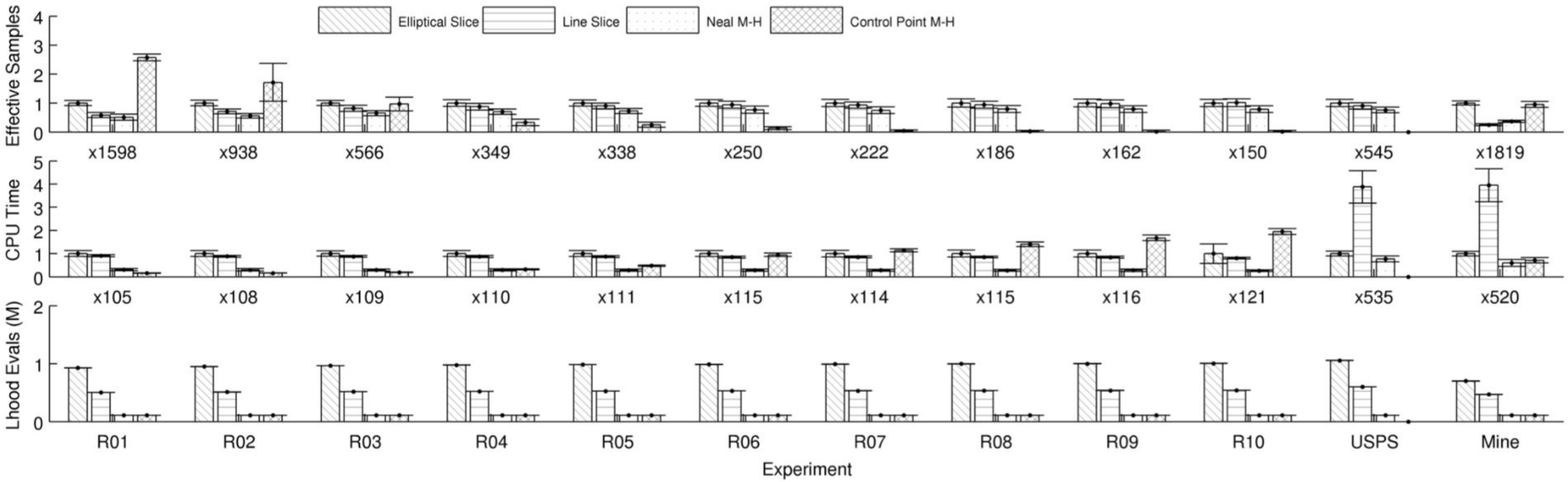


Figure 6: Number of effective samples from 10^5 iterations after 10^4 burn in, with time and likelihood evaluations required. The means and standard deviations for 100 runs are shown (divide the "error bars" by 10 to get standard errors on the mean, which are small). Each iteration involves one $\mathcal{O}(N^2)$ operation (e.g. one $\boldsymbol{\nu}$ draw or updating one control variable). Each group of bars in the top two rows has been rescaled for readability: the numbers beneath each group show the number of effective samples or CPU time in seconds for elliptical slice sampling, which always has bars of height 1.

Thank you!