Parrellel Tempering

Master Thesis Defense

Mateusz Łącki

Uniwersytet Warszawski

14 January 2014

Today's Agenda

The Metropolis-Hastings Algorithm

Goal Generate a representatative sample

Mean Searching state-space for probability clusters in 2 steps

1. Choose x_0 in any way
2. Given previous step $X^{[n-1]} = x^{[n-1]}$, draw proposal
$$Y \sim \mathcal{Q}(x^{[n-1]}, \circ)$$
3. Draw
$$U \sim \mathcal{U}(0, 1)$$
4. Accept the proposal ($X_n := Y$) if
$$U \leq \alpha(x^{[n-1]}, Y)$$

where $\alpha(x^{[n-1]}, y) = 1 \wedge \frac{\pi(y)}{\pi(x^{[n-1]})}$

Problems with Metropolis-Hastings

- Choice of α : self-adjointess of corresponding kernel:

$$\mathcal{M}(x,A) = \underbrace{\int_A \alpha(x,y)\mathcal{Q}(x,\mathrm{d}\,y)}_{\mathcal{P}\left(X^{[n+1]}\neq x \text{ and } X^{[n+1]}\in A\right)} + \mathbb{I}_A(x)\underbrace{\int \Big(1-\alpha(x,y)\Big)\mathcal{Q}(x,\mathrm{d}\,y)}_{\mathcal{P}\left(X^{[n+1]}=x\right)}$$

  - Kernel preserves the distribution $\pi(x)\mathrm{d}\,x$

- Problem: $\frac{\pi(y)}{\pi(x^{[n-1]})}$
  - promotes acceptances of points with not so much smaller probability

If $\mathcal{Q}(x,\circ)$ centered to much around x,

then Difficult to leave x's surroundings.

Liang-Wang Example

Metropolis-Hastings algorithm's flaw: LOCALNESS

$$f(x) = \sum_{i=1}^{20} \frac{\omega_i}{\sigma_i\sqrt{2\pi}} \exp\left(- \frac{(x-\mu_i)'(x-\mu_i)}{2\sigma_i^2} \right)$$

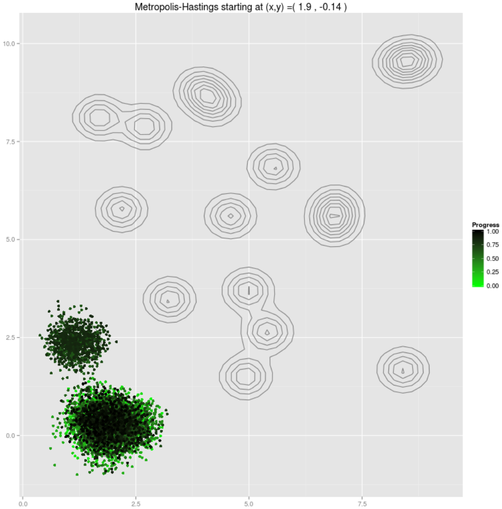
where $\sigma_1 = \cdots = \sigma_{20} = 0.1, \omega_1 = \cdots = \omega_{20} = 0.05$

and the means μ_i are given by

1	2	3	4	5	6	7	8	9	10
2.18	8.67	4.24	8.41	3.93	3.25	1.70	4.59	6.91	6.87
5.76	9.59	8.48	1.68	8.82	3.47	0.50	5.60	5.81	5.40

11	12	13	14	15	16	17	18	19	20
5.41	2.70	4.98	1.14	8.33	4.93	1.83	2.26	5.54	1.69
2.65	7.88	3.70	2.39	9.50	1.50	0.09	0.31	6.86	8.11

Multimodial Disaster



Solution: Parallel Tempering

- Enlarge the state-space!

Chain $X^{[n]} = \left(X_1^{[n]}, \ldots, X_L^{[n]}\right)$

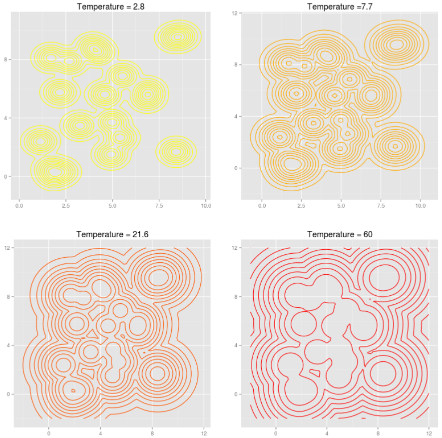
Aim $\pi_\beta \propto \pi^{\beta_1} \times \cdots \times \pi^{\beta_L}$

where $\underbrace{\beta = (\beta_1, \ldots, \beta_L)}_{\text{inverse temperatures}}$

and $\beta_1 \equiv 1$

- Move independently with each chain
- $\alpha_{\beta_l} = \left(\frac{\pi(y)}{\pi(x)}\right)^{\beta_l} > \frac{\pi(y)}{\pi(x)} = \alpha_{\beta_1}$

then Exchange accepted proposals ...



Swaping proposals

- Last States: $x \equiv x^{[n-1]} = \left(x_1^{[n-1]}, \dots, x_L^{[n-1]}\right)$
- Proposal Kernel: $\mathcal{T}(x, A) \equiv \sum_{i<j} p_{ij}(x) \mathbb{I}_A(S_{ij}x)$

where $S_{ij}x = (x_1, \dots, x_{i-1}, x_j, x_{i+1}, \dots, x_{j-1}, x_i, x_{j+1}, \dots, x_L)$

- To assure related kernel self-adjointness

$$\alpha_{\text{swap}}(x, \underbrace{S_{ij}x}_{\text{proposal}}) = \left[\left(\frac{\pi(x_j)}{\pi(x_i)}\right)^{\beta_i - \beta_j} \frac{p_{ij}(S_{ij}x)}{p_{ij}(x)}\right] \wedge 1$$

So different distributions on swaps = STRATEGIES

Also possibly state-dependent: $p_{ij} = p_{ij}(x)$

Our choices for Strategies

Strategy	Proportional to	Strategy	Proportional to
I	$\frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)}$	IV	$\left(\frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)}\right)^{\frac{\lvert\beta_i-\beta_j\rvert}{1+\rho(x_i,x_j)}}$
II	$\frac{\pi(x_j)}{\pi(x_i)} \wedge 1$	V	$\frac{2}{L(L-1)}$
III	$\left(\frac{\pi(x_j)}{\pi(x_i)} \wedge \frac{\pi(x_i)}{\pi(x_j)}\right)^{\lvert\beta_i-\beta_j\rvert}$	VI	$\frac{1}{L-1}\{\lvert i-j\rvert = 1\}$