

Metropolis-Hastings algorithm

- 1 Initialise $x^{(0)}$
- 2 For $t = 1, \dots, n + N$:
 - a Sample $y^{(t)} \sim q(y^{(t)} | x^{(t-1)})$
 - b Compute the acceptance probability
$$\alpha^{(t)} = \min \left\{ 1, \frac{\tilde{p}(y^{(t)})}{q(y^{(t)} | x^{(t-1)})} / \frac{\tilde{p}(x^{(t-1)})}{q(x^{(t-1)} | y^{(t)})} \right\}$$
 - c Acceptance-rejection step: sample $u^{(t)} \sim \mathcal{U}_{[0;1]}$
$$x^{(t)} = \begin{cases} y^{(t)} & \text{if } u^{(t)} \leq \alpha^{(t)} \\ x^{(t-1)} & \text{else} \end{cases}$$

$$\alpha^{(t)} = \min \left\{ 1, \frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})} \frac{q(x^{(t-1)} | y^{(t)})}{q(y^{(t)} | x^{(t-1)})} \right\}$$

⇒ computable even if \tilde{p} is known only up to a constant !
(like the posterior)

Metropolis-Hastings: particular cases

Sometimes $\alpha^{(t)}$ computation simplifies:

- **independent Metropolis-Hastings:** $q(y^{(t)}|x^{(t-1)}) = q(y^{(t)})$
- **random walk Metropolis-Hastings:** $q(y^{(t)}|x^{(t-1)}) = g(y^{(t)} - x^{(t-1)})$
If g is symmetric ($g(-x) = g(x)$), then:

$$\frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})} \frac{q(y^{(t)}|x^{(t-1)})}{q(x^{(t-1)}|y^{(t)})} = \frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})} \frac{\cancel{g(y^{(t)} - x^{(t-1)})}}{\cancel{g(x^{(t-1)} - y^{(t)})}} = \frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})}$$

Pro and cons of Metropolis-Hastings

- 😊 very simple & very general
- 😄 allow sampling from uni- or multi-dimensional distributions
- 😞 choice of the proposal is crucial, but hard
⇒ huge impact on algorithm performances
- 😞 quickly becomes inefficient dimension is too high

NB: a high rejection rate often implies important computation timings

Simulated annealing

Change $\alpha^{(t)}$ computation during the algorithm:

- 1 $\alpha^{(t)}$ must first be large to explore all of the state space
- 2 then $\alpha^{(t)}$ must become smaller when the algorithm converges

Simulated annealing

Change $\alpha^{(t)}$ computation during the algorithm:

- 1 $\alpha^{(t)}$ must first be large to explore all of the state space
- 2 then $\alpha^{(t)}$ must become smaller when the algorithm converges

1 Initialise $x^{(0)}$

2 For $t = 1, \dots, n + N$:

a Sample $y^{(t)} \sim q(y^{(t)} | x^{(t-1)})$

b Compute the acceptance probability

$$\alpha^{(t)} = \min \left\{ 1, \left(\frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})} \frac{q(x^{(t-1)} | y^{(t)})}{q(y^{(t)} | x^{(t-1)})} \right)^{\frac{1}{T(t)}} \right\}$$

c Acceptance-rejection step: sample $u^{(t)} \sim \mathcal{U}_{[0;1]}$

$$x^{(t)} := \begin{cases} y^{(t)} & \text{if } u^{(t)} \leq \alpha^{(t)} \\ x^{(t-1)} & \text{else} \end{cases}$$

Simulated annealing

Change $\alpha^{(t)}$ computation during the algorithm:

- 1 $\alpha^{(t)}$ must first be large to explore all of the state space
- 2 then $\alpha^{(t)}$ must become smaller when the algorithm converges

1 Initialise $x^{(0)}$

2 For $t = 1, \dots, n + N$:

a Sample $y^{(t)} \sim q(y^{(t)} | x^{(t-1)})$

b Compute the acceptance probability

$$\alpha^{(t)} = \min \left\{ 1, \left(\frac{\tilde{p}(y^{(t)})}{\tilde{p}(x^{(t-1)})} \frac{q(x^{(t-1)} | y^{(t)})}{q(y^{(t)} | x^{(t-1)})} \right)^{\frac{1}{T(t)}} \right\}$$

c Acceptance-rejection step: sample $u^{(t)} \sim \mathcal{U}_{[0;1]}$

$$x^{(t)} := \begin{cases} y^{(t)} & \text{if } u^{(t)} \leq \alpha^{(t)} \\ x^{(t-1)} & \text{else} \end{cases}$$

Ex: $T(t) = T_0 \left(\frac{T_f}{T_0} \right)^{\frac{t}{n}} \Rightarrow$ particularly useful for avoiding local optima

Gibbs sampler

When the dimension $\nearrow \Rightarrow$ very hard to propose probable values

Gibbs samplers: re-actualisation coordinate by coordinate, while conditioning on the most recent values (no acceptance-rejection)

- 1 Initialise $x^{(0)} = (x_1^{(0)}, \dots, x_d^{(0)})$
- 2 For $t = 1, \dots, n + N$:
 - a Sample $x_1^{(t)} \sim p(x_1 | x_2^{(t-1)}, \dots, x_d^{(t-1)})$
 - b Sample $x_2^{(t)} \sim p(x_2 | x_1^{(t)}, x_3^{(t-1)}, \dots, x_d^{(t-1)})$
 - c ...
 - d Sample $x_i^{(t)} \sim p(x_i | x_1^{(t)}, \dots, x_{i-1}^{(t)}, x_{i+1}^{(t-1)}, \dots, x_d^{(t-1)})$
 - e ...
 - f Sample $x_d^{(t)} \sim p(x_d | x_1^{(t)}, \dots, x_{d-1}^{(t)})$

NB: if the conditional distribution is unknown for some coordinates, an acceptance-rejection step can be included for this coordinate only (*Metropolis within gibbs*)

Your turn !



Practical: exercise 4