

SPR Day 21

Markov Chain Monte Carlo

{ Metropolis Hastings Algorithm

Reading: Bishop PRML See 11.2.1

Markov chains

See 11.2.2

Metropolis
Hastings

Topics: 1) MCMC with Propose-Accept transitions

2) Random Walk proposals: intuition

3) Detailed Balance, proof that RW works!

4) Metropolis + Metropolis-Hastings algorithms

Goal: Sample from target distribution
with PDF $p^*(z)$ over real $z \in \Omega \subseteq \mathbb{R}$

Challenge: Do NOT know how to evaluate $p^*(z)$,
only known up to multiplicative constant

$\tilde{p}(z)$ is evaluateable

$p^*(z) = c \tilde{p}(z)$ is NOT

because $c = \frac{1}{\int \tilde{p}(z) dz}$ is intractable
(cannot do the integral)

Key idea: ratio of PDFs of two possible z values
can be evaluated exactly

$$\frac{p^*(z')}{p^*(z^z)} = \frac{\cancel{c} \tilde{p}(z')}{\cancel{c} \tilde{p}(z^z)} = \frac{\tilde{p}(z')}{\tilde{p}(z^z)}$$

easy to compute!

Want to design Markov proposal that
some how uses rations

Markov Transition Design: Propose then $\begin{matrix} \text{accept} \\ \text{reject} \end{matrix}$

Transition distribution γ takes current "state" value z_t , and produces new "state" value z_{t+1}

Idea: Propose new value $z' \sim Q(z'|z_t)$ then decide to accept w. prob $A(z', z_t)$

$$z_{t+1} = \begin{cases} z' & \text{if } u < A(z', z_t) \\ z_t & \text{otherwise} \end{cases}$$

where $z' \sim Q(z'|z_t)$
 $u \sim \text{Unif}(0, 1)$

Example

Start: $z_1 = 0.3$

- γ 
- 1) Sample $z' \sim Q(\cdot | z_1)$. $z' = -1.8$
 - 2) Sample $u \sim \text{Unif}(0, 1)$. $u = 0.3$
 - 3) Eval $A(z', z_1) = 0.8$. $u < A$ so ACCEPT
- $z_2 = -1.8$

- γ 
- 1) Sample $z' \sim Q(\cdot | z_2)$. $z' = 2.7$
 - 2) Sample $u \sim \text{Unif}(0, 1)$. $u = 0.91$
 - 3) Eval $A(z', z_2) = 0.01$. $u > A$ so REJECT
- $z_3 = -1.8$
(and keep going to get z_1, z_2, \dots, z_S)

Remember, A is an accept probability threshold

We know $0 \leq A \leq +\infty$

$A = 0.0$ means no chance of accept
ALWAYS REJECT

$A \in (0, 0.5)$ means can accept,
but likely reject

$A \in (0.5, 1]$ means can reject
but likely accept

$A \geq 1$ means ALWAYS ACCEPT

So sometimes we write

$$P(\text{accept} | z; z_t) = \min(A(z; z_t), 1)$$

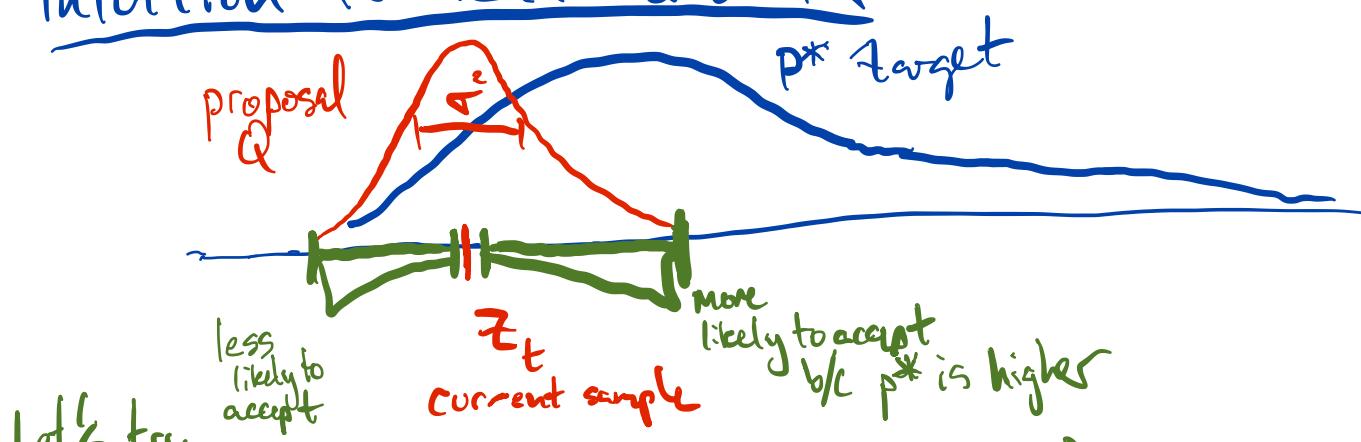
any value ≥ 1
is just set to 1

Proposal Idea: Random Walk

Random Walk $Q(z' | z_t) = \text{Normal}(z' | \bar{z}_t, \frac{\text{var}}{\tau^2})$ You pick $\tau^2 > 0$ hyperparam.

Why: easy to sample
likely to propose "nearby"

Intuition for Random Walk



let's try

$$A(z', z_t) = \frac{\tilde{p}(z')}{\tilde{p}(z_t)} = \frac{p^*(z')}{p^*(z_t)}$$

remember ratios of \tilde{p}
can be computed exactly

Given current z_t , and proposed value z' ,

$p^*(z')$ — than $p^*(z_t)$

ratio $\frac{p^*(z')}{p^*(z_t)}$

proposal will be

Much smaller

$\ll 0.5$

likely rejected

a bit smaller

in $(0.5, 1)$

likely accepted

Same or larger

≥ 1

certainly accepted

Intuitively, random walk proposal Q
 plus p^* ratio A makes sense,
 but how to show this will create
 Markov chain w/ stationary distr. p^* ?

Recall, p^* is stationary for transition $\tilde{\tau}$ if:

$$\text{for all } z_{t+1} \in \Sigma: p^*(z_{t+1}) = \int_{z_t} p^*(z_t) \tilde{\tau}(z_{t+1} | z_t) dz_t$$

Can we show that when $\tilde{\tau}$ uses rand walk Q
 and accept proba A , that this is satisfied?

First, define: $\tilde{\tau}(z_{t+1} | z_t) = \min\left(1, \frac{p^*(z_{t+1})}{p^*(z_t)}\right) Q(z_{t+1} | z_t)$
 PDF: when $z_{t+1} \neq z_t$

Next, show $p^*(z_a) \tilde{\tau}(z_b | z_a) = p^*(z_b) \tilde{\tau}(z_a | z_b)$
 for all $z_a \neq z_b$

$$\min\left(\frac{p^*(z_a)}{p^*(z_b)}, \frac{p^*(z_b)}{p^*(z_a)}\right) Q(z_b | z_a) = \min\left(\frac{p^*(z_b)}{p^*(z_a)}, \frac{p^*(z_a)}{p^*(z_b)}\right) Q(z_a | z_b)$$

and we know $Q(z_b | z_a) = Q(z_a | z_b)$ because $N(z_a | z_b, \sigma^2) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2} (z_a - z_b)^2}$
 Norm PDF $= N(z_b | z_a, \sigma^2)$ Symmetric

Remember, for any $a > 0, b > 0$

$$a \cdot \min\left(1, \frac{b}{a}\right)$$

$$= a \cdot 1 \text{ if } b > a = \min(a, b)$$

$$= a \cdot \frac{b}{a} \text{ if } b \leq a$$

Now, for

$$\begin{aligned} z_t &= z_{t+1} \\ &= \tilde{z} \end{aligned}$$

$$\cancel{P^*(z_t) \tilde{\gamma}(z_{t+1} | z_t)} = \cancel{P^*(z_{t+1}) \tilde{\gamma}(z_t | z_{t+1})}$$

$$\cancel{z_t = z_{t+1}}$$

$$\tilde{\gamma}(\tilde{z} | \tilde{z}) = \tilde{\gamma}(\tilde{z} | \tilde{z})$$
true by symmetry

So we have shown for all $z_a \in \Omega, z_b \in \Omega$, that random walk transition $\tilde{\gamma}$ satisfies:

$$P^*(z_a) \tilde{\gamma}(z_b | z_a) = P^*(z_b) \tilde{\gamma}(z_a | z_b)$$

We call this the DETAILED BALANCE condition

$z_a \sim P^*$
 $z_b \sim \tilde{\gamma}(\cdot | z_a)$

has same
joint
probab
as

$z_b \sim P^*$
 $z_a \sim \tilde{\gamma}(\cdot | z_b)$

but, how does this help show P^* is stationary of $\tilde{\gamma}$?

Proving stationary if γ satisfies DETAILED BALANCE

well,

$$\begin{aligned} p^*(z_{t+1}) &= \int p^*(z_t) \gamma(z_{t+1}/z_t) dz_t \\ &= \int p^*(z_{t+1}) \gamma(z_t/z_{t+1}) dz_t && \text{by detailed balance} \\ &= p^*(z_{t+1}) \int \gamma(z_t/z_{t+1}) dz_t && \text{bringing out term const wrt } z_t \\ &= p^*(z_{t+1}) && \text{because } \gamma \text{ is a PDF and must integrate to 1} \end{aligned}$$

Thus, p^* is stationary distribution, because it meets required definition.

METROPOLIS MCMC Algorithm

Initialize $z_1 \in \Omega$

for t in $1, 2, \dots, S-1$

$$1) z' \sim Q(\cdot | z_t)$$

$$2) u \sim \text{Unit}([0, 1])$$

$$3) z_{t+1} = \begin{cases} z' & \text{if } u \leq \\ & \frac{\tilde{P}(z')}{\tilde{P}(z_t)} \\ z_t & \text{otherwise} \end{cases}$$

return $[z_1, z_2, \dots, z_S]$

Assumes that Q valid PDF over Ω
 easy to sample
 easy to evaluate PDF
must be symmetric

$$Q(z_a | z_b) = Q(z_b | z_a)$$

for all $z_a \in \Omega, z_b \in \Omega$

$$\frac{\tilde{P}(z')}{\tilde{P}(z_t)} = A(z', z_t)$$

if S large enough,
 can consider
 z_{S-B}, \dots, z_S as samples of P^*

METROPOLIS-HASTINGS

Initialize $z_1 \in \Omega$

for t in $1, 2, \dots, S-1$

$$1) z' \sim Q(\cdot | z_t)$$

$$2) u \sim \text{Unit}([0, 1])$$

$$3) z_{t+1} = \begin{cases} z' & \text{if } u < \\ & \frac{\tilde{P}(z')}{\tilde{P}(z_t)} \frac{Q(z_t | z')}{Q(z' | z_t)} \\ z_t & \text{otherwise} \end{cases}$$

return $[z_1, z_2, \dots, z_S]$

Assumes only that
 Q is valid PDF over Ω
 easy to sample
 & easy to evaluate PDF

$$\frac{\tilde{P}(z')}{\tilde{P}(z_t)} \frac{Q(z_t | z')}{Q(z' | z_t)}$$

can slow this A also
 satisfies DETAILED BALANCE

Sanity check: What if we use $Q = p^*$ in Met-Hastings

$$\text{Accept ratio } A = \frac{\tilde{p}(z')}{\tilde{p}(z_t)} \frac{p^*(z_t)}{p^*(z')} = \frac{\tilde{p}(z') \tilde{p}(z_t) c}{\tilde{p}(z_t) \tilde{p}(z') c} = 1$$

So we'd always accept! Makes sense.