

Day 14 in 2021s

SPR Day 22

Gibbs Sampling

Reading: Bishop PRML Sec 11.3

Topics:	Goal: Vector Samples + overview of options	page 2-3
	Table of PRO/CONS	page 4
	The Gibbs Sampling MCMC Alg	page 5
	A Simple Example	page 6
	Proof: why Gibbs is a MCMC chain w/ stationary P*	page 7-8

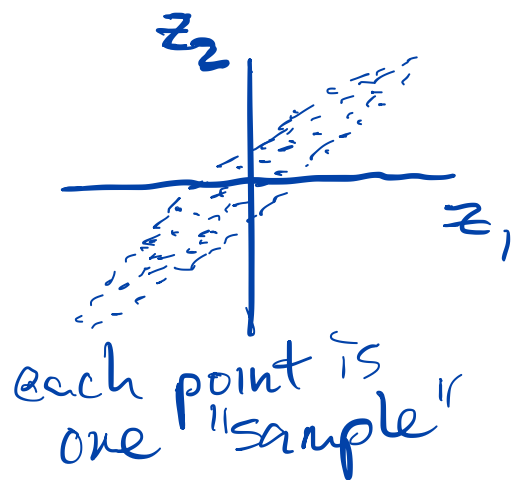
Goal: Sample a vector-valued random variable \mathbf{z} from a target distribution p^*

R.V. $\mathbf{z} = [z_1, z_2, \dots, z_D]$ $z_j \in \mathbb{R}$
 D entries, each one real value

Simple example:

$D=2$

$p^* \stackrel{\text{mean}}{=} \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}\right)$ *Cov*



How to draw ^{DIST} samples?

$[z_1^{(s)} \ z_2^{(s)} \ \dots \ z_D^{(s)}] \sim p^*$

1) Transform simpler r.v.

$u_1^{(s)} \dots u_D^{(s)} \stackrel{\text{iid}}{\sim} \text{Unif}$
 $z_1^{(s)} \dots z_D^{(s)} \leftarrow T(u)$

2) Random walk (Metropolis MCMC)

3) Gibbs sampling

Metropolis-Hastings for vector-valued random variables

Option 1: Propose joint sample, accept or reject entire vector

for t in $2, 3, \dots, S$:

$$z'_1, \dots, z'_D \sim Q(\cdot | z_1^{(t-1)}, \dots, z_D^{(t-1)})$$

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

$$z_1^{(t)}, \dots, z_D^{(t)} = \begin{cases} z'_1, \dots, z'_D & \text{if } u < A(z', z^{(t)}) \\ z_1^{(t-1)}, \dots, z_D^{(t-1)} & \text{o.w.} \end{cases}$$

Option 2:

Separate proposal for each dimension

for t in $2, 3, \dots, S$:

for d in $1, 2, \dots, D$:

$$z'_d \sim Q_d(\cdot | \overbrace{z_1^{(t)}, \dots, z_{d-1}^{(t)}}^{\text{new values for dims } < d}, \overbrace{z_{d+1}^{(t-1)}, \dots, z_D^{(t-1)}}^{\text{old values for dims } > d})$$

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

$$z_d^{(t)} = \begin{cases} z'_d & \text{if } u < A(z_1^{(t)}, \dots, z_{d-1}^{(t)}, z'_d, z_{d+1}^{(t-1)}, \dots, z_D^{(t-1)}) \\ z_d^{(t-1)} & \text{o.w.} \end{cases}$$

Option 1 (joint samples) can make big changes, but could have high rejections

Option 2 may accept more often, but slow to make big changes

Comparison of Sampling Methods

	practical effectiveness sample $z^{(s)}$ is really from p^*	required properties of p^*	ease of derivation for a new model	hyperparams
(1) Transformation Methods	EXCELLENT	can prove $T(u) \sim p^*$ HARD	CHALLENGING to IMPOSSIBLE	none
(2) Random Walk MCMC	TERRIBLE in worst case, OK in best case	\tilde{p} evaluable EASY	EASY	std dev σ
(3) Gibbs MCMC	FAIR - to - GOOD	can sample from conditionals $p^*(z_1 z_2, \dots, z_{d-1}, z_{d+1}, \dots, z_D)$ MODERATE	DOABLE with EFFORT for most directed graphical models	none

Punchline: For most "interesting" models with many variables (large $D \gg 1$) random walk is unlikely to be useful b/c will take way too long to converge even though convergence guaranteed eventually

Using Gibbs sampling is one of our best options for effective sampling

Gibbs sampling: Core Idea

5

Goal: Sample $z_1^{(s)} \dots z_D^{(s)} \sim p^*$

Challenge: Cannot sample using analytical methods

Insight: Samples from joint are hard, but samples from conditional might be easier

So, set up markov chain where

$$z^{(t)} \xrightarrow{\mathcal{Z}_{\text{Gibbs}}} z^{(t+1)}$$

Transition $\mathcal{Z}_{\text{Gibbs}}$ has D steps:

$$1) z_1^{(t+1)} \sim p^*(z_1 | z_2^{(t)}, \dots, z_D^{(t)}) \quad \begin{array}{l} \text{Condition on} \\ \text{all dims} \\ \text{except 1} \end{array}$$

$$2) z_2^{(t+1)} \sim p^*(z_2 | z_1^{(t+1)}, z_3^{(t)}, \dots, z_D^{(t)}) \quad \begin{array}{l} \text{all dims} \\ \text{except 2} \end{array}$$

$$\vdots$$
$$D) z_D^{(t+1)} \sim p^*(z_D | z_1^{(t+1)}, \dots, z_{D-1}^{(t+1)}) \quad \begin{array}{l} \vdots \\ \text{all dims} \\ \text{except} \\ D \end{array}$$

If we repeatedly sample using $\mathcal{Z}_{\text{Gibbs}}$
can show that p^* is stationary distribution

Extension: can sample dimensions in any order
can group or "block" several dims in one conditional
e.g. step 1: $p^*(z_{1:3} | z_{3:D})$

Gibbs example 1: 2D Gaussian 6

Suppose we can easily sample from a 1D Gaussian, but not a Multivariate one. How to sample from:

$$p^*(z_1, z_2) = \mathcal{N}\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \middle| \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.8 \\ 0.8 & 2 \end{bmatrix}\right)$$

Recall formula for conditionals given joint Gaussian

$$p^*(z_1 | z_2) = \mathcal{N}\left(z_1 \middle| \underbrace{(0.8) \frac{1}{2} (z_2 - \mu_2^0)}_{= +0.4 z_2}, \underbrace{1 - \frac{1}{2} 0.8^2}_{= 1 - 0.32 = 0.68}\right) \quad \begin{array}{l} \text{Bishop} \\ \text{Eq.} \\ 2.81 \\ 2.82 \end{array}$$

$$p^*(z_2 | z_1) = \mathcal{N}\left(z_2 \middle| \underbrace{(0.8) \frac{1}{1} (z_1 - \mu_1^0)}_{= +0.8 \cdot z_1}, \underbrace{2 - \frac{1}{1} 0.8^2}_{= 2 - 0.64 = 1.36}\right)$$

Thus, Gibbs Sampler would be:

Initialize $z^{(1)} = [0, 0]$

for t in $2, \dots, S$:

$$z_1^{(t)} \sim \text{SAMPLE_1D_NORM}\left(+0.4 \cdot z_2^{(t-1)}, 0.68\right)$$

$$z_2^{(t)} \sim \text{SAMPLE_1D_NORM}\left(+0.8 \cdot z_1^{(t)}, 1.36\right)$$

return $[z^{(1)}, z^{(2)}, \dots, z^{(S)}]$

Assume `SAMPLE_1D_NORM`
takes two arguments:
- a mean scalar
- a variance scalar

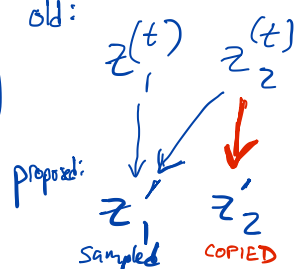
Proof of correctness 1

Consider using the Gibbs transition distrib. as the proposal distrib. Q of Met-Hastings

We'll focus not on all D steps of $\mathcal{Z}_{\text{Gibbs}}$, but just one!

Consider only the first dim's conditional transition

$$z'_1 \sim Q_1(z_1 | z_{1:2}) = p^*(z_1 | z_2)$$



then, accept threshold is:

$$A(z', z^{(t)}) = \frac{p^*(z') Q(z^{(t)} | z')}{p^*(z^{(t)}) Q(z' | z^{(t)})}$$

for any M-H proposal

$$= \frac{p^*(z'_1, z'_2) p^*(z_1^{(t)} | z'_2)}{p^*(z_1^{(t)}, z_2^{(t)}) p^*(z_1 | z_2^{(t)})}$$

$$= \frac{p^*(z'_2) p^*(z_1 | z'_2) p^*(z_1^{(t)} | z'_2)}{p^*(z_2^{(t)}) p^*(z_1 | z_2^{(t)}) p^*(z_1 | z_2^{(t)})}$$

recall $z'_2 = z_2^{(t)} = c$
(proposal only samples z_1 , other dim is COPIED)

$$= 1$$

Punchline: We would always accept a proposal from the Gibbs conditional

Proof of correctness 2:

8

Show $\mathcal{L}_{\text{Gibbs}}$ has stationary distrib. p^*

Strategy: Show that $\mathcal{L}_{\text{Gibbs}}$ satisfies DETAILED BALANCE wrt p^* . This implies p^* is the stationary!

Let's fix $D=2$. Focus on transition only for $\text{dim}=1$
DETAILED BALANCE condition says:

$$\text{for all } z_1^a, z_2^a, z_1^b: p^*(z_1^a, z_2^a) \mathcal{L}(z_1^b | z_2^a) = p^*(z_1^b, z_2^a) \mathcal{L}(z_1^a | z_2^a)$$

Proof step 1:
product rule
expand p^* joint

$$\cancel{p^*(z_2^a)} p^*(z_1^a | z_2^a) \mathcal{L}(z_1^b | z_2^a) = \cancel{p^*(z_2^a)} p^*(z_1^b | z_2^a) \mathcal{L}(z_1^a | z_2^a)$$

Step 2: substitute
definition
 $\mathcal{L}(x|y) = p^*(x|y)$

$$p^*(z_1^a | z_2^a) p^*(z_1^b | z_2^a) = p^*(z_1^b | z_2^a) p^*(z_1^a | z_2^a)$$

which is true for
all values by definition

thus, DB condition is true for $\mathcal{L}_{\text{Gibbs}}$,
and p^* is the Markov chain's stationary distribution