Day 13 in 2023s

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^*

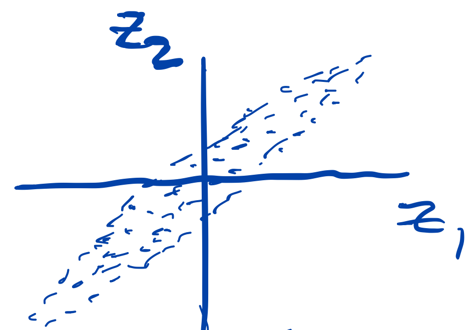
R.V. $\mathbf{z} = [z_1, z_2, \cdots z_D]$ $z_d \in \mathbb{R}$

D entries, each one real value

Simple example:

$D = 2$

$$p^* \triangleq \mathcal{N}\left(\underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{\text{mean}}, \underbrace{\begin{bmatrix} 1 & 0.9 \\ 0.9 & 1 \end{bmatrix}}_{\text{covar}}\right)$$



each point is one "sample"

How to draw JOINT samples?

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

1) Transform simpler r.v.

$u_1^{(s)} \ldots u_D^{(s)} \overset{iid}{\sim} \text{Unif}$

$z_1^{(s)} \ldots z_D^{(s)} \leftarrow T(u)$

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 $2, 3, \ldots S$:

$$z_1', \ldots z_D' \sim Q(\cdot \mid z_1^{(t-1)} \ldots z_D^{(t-1)})$$

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

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

Option 2: Separate proposal for each dimension

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

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

$$z_d' \sim Q_d(\cdot \mid \underbrace{z_1^{(t)} \ldots z_{d-1}^{(t)}}_{\text{new values for dims} < d} \; z_d^{(t-1)} \; \underbrace{z_{d+1}^{(t-1)} \ldots z_D^{(t-1)}}_{\text{old values for dims} > d})$$

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

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

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	$\tilde{p}$ evaluatable EASY	EASY	std dev σ
(3) Gibbs MCMC	FAIR-to-GOOD	can sample from conditionals $p^*(z_d \mid z_1 \ldots z_{d-1}, z_{d+1} \ldots z_D)$ 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

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

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

$$z^{(t)} \xrightarrow[\tau_{Gibbs}]{} z^{(t+1)}$$

Transition τ_{Gibbs} has D steps:

Condition on

1) $z_1^{(t+1)} \sim p^*(z_1 \mid z_2^{(t)}, \cdots z_D^{(t)})$ — all dims except 1

2) $z_2^{(t+1)} \sim p^*(z_2 \mid z_1^{(t+1)}, z_3^{(t)} \cdots z_D^{(t)})$ — all dims except 2

$\vdots$

D) $z_D^{(t+1)} \sim p^*(z_3 \mid z_1^{(t+1)} \cdots z_{D-1}^{(t+1)})$ — all dims except D

If we repeatedly sample using τ_{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: $p^*(z_{1:3} \mid z_{3:D})$

Gibbs example 1: 2D Gaussian

Suppose we can easy 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

$$p^*(z_1 | z_2) = \mathcal{N}\Big(z_1 \,\Big|\, \underbrace{(0.8)\tfrac{1}{2}(z_2 - \overset{0}{\mu_2})}_{= +0.4 z_2},\ \underbrace{1 - \tfrac{1}{2} 0.8^2}_{= 1 - 0.32 = 0.68}\Big)$$

Bishop Eq. 2.81 2.82

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

Thus, Gibbs Sampler would be:

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

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

for t in $2, \dots S$:

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

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

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

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 $\mathcal{T}_{Gibbs}$, but just one.

Consider only the first dim's conditional transition

$$z_1' \sim Q_1(z_1 \mid z_{1:2}^{(t)}) = p^*(z_1 \mid z_2^{(t)})$$

old: $z_1^{(t)}$ $z_2^{(t)}$

proposal: z_1' sampled, z_2' COPIED

then, accept threshold is:

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

for any M-H proposal

$$= \frac{p^*(z_1', z_2')}{p^*(z_1^{(t)}, z_2^{(t)})} \frac{p^*(z_1^{(t)} \mid z_2')}{p^*(z_1' \mid z_2^{(t)})}$$

$$= \frac{p^*(\cancel{z_2'}) \, \cancel{p^*(z_1' \mid z_2')}}{p^*(\cancel{z_2^{(t)}}) \, \cancel{p^*(z_1^{(t)} \mid z_2^{(t)})}} \frac{\cancel{p^*(z_1^{(t)} \mid z_2')}}{\cancel{p^*(z_1' \mid z_2^{(t)})}}$$

recall $z_2' = z_2^{(t)} = c$

(proposal only samples $d=1$, other dim is COPIED)

$$= 1$$

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

Proof of correctness 2:

Show τ_{Gibbs} has stationary distrib. p^*

Strategy: Show that τ_{Gibbs} satisfies DETAILED BALANCE wrt p^*. This implies p^* is the stationary!

Let's fix $D=2$. Focus on transition only for dim $d=1$

DETAILED BALANCE condition says:

for all z_1^a, z_2^a, z_1^b: $p^*(z_1^a, z_2^a)\tau(z_1^b | z_2^a) = p^*(z_1^b, z_2^a)\tau(z_1^a | z_2^a)$

Proof step 1: product rule expand p^* joint

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

Step 2: substitute definition $\tau(x|y) = p^*(x|y)$

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

which is true for all values by definition

thus, DB condition is true for τ_{Gibbs}, and p^* is the Markov chain's stationary distribution