

- today:
- Gibbs sampling
  - Variational methods

Gibbs sampling alg.

↳ M.H. with a clever choice of proposal  $q_t(x^t | x)$  [clever:  $dx^t | x = 1$ ]

example of applications

UGM:  $\tilde{p}(x) = \prod_c \psi_c(x_c)$

difficult conditional in a DGM:  $\tilde{p}(x) = p(x_{FC}, \tilde{x}_E) \delta(x_E, \tilde{x}_E) \propto p(x | \tilde{x}_E)$

(UGM)

cyclic Gibbs sampling alg: nodes  $i=1, \dots, n$

start at some  $x^{(0)}$

for  $t=1, \dots$

- pick  $i = (t \bmod n) + 1$
- sample  $x_i^{(t)} \sim p(x_i | \{x_j^{(t-1)}\}_{j \neq i})$   
true conditional on  $x_j$  as proposal
- set  $x_j^{(t)} = x_j^{(t-1)}$  for  $j \neq i$

end

⊗ Gibbs sampling is M.H. with a time varying proposal

suppose  $i$  is picked at time  $t$

for  $x_{-i}$  to stay constant

then proposal is  $q_t(x^t | x^{(t-1)}) = p(x_i | x_{-i}^{(t-1)}) \delta(x_{-i}^t, x_{-i}^{(t-1)})$

M.H. acceptance ratio:

$$\alpha(x^t | x^{(t-1)}) = \frac{q_t(x^{(t-1)} | x^t) p(x^t)}{q_t(x^t | x^{(t-1)}) p(x^{(t-1)})} = \frac{p(x_{-i}^{(t-1)}) p(x_i | x_{-i}^{(t-1)})}{p(x_i | x_{-i}^{(t-1)}) p(x_{-i}^{(t-1)})} = 1$$

$\Rightarrow 1 \rightarrow$  always accepts! //

convergence of G-S:

- let  $A$  be a Markov transition kernel of one full cycle of G-S (i.e.  $n$  steps)

→ homogeneous M.C.

if we suppose that  $p(x) > 0 \forall x$

$$\Leftrightarrow p(x_i | x_{-i}) > 0 \forall x_i, x_{-i}$$

⇒ A is irreducible & aperiodic because  $A_{ii} > 0$

$$A_{ij} > 0 \forall i, j$$

state can get to any state with  $n$  steps

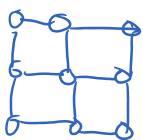
$$\Rightarrow A^t \pi_0 \xrightarrow{t \rightarrow \infty} p$$

⊗ also works for a random scan (pick  $i \sim \text{Unif}(1:n)$  at each step)

example: G.S. for Ising model

Ising model  $x_i \in \{-1, 1\}$

UGM:



$$p(x) = \frac{1}{Z(n)} \exp\left(\sum_i n_i x_i + \sum_{i,j \in E} n_{ij} x_i x_j\right)$$

[minimal exp. family representation]

to do G.S.:

want to compute

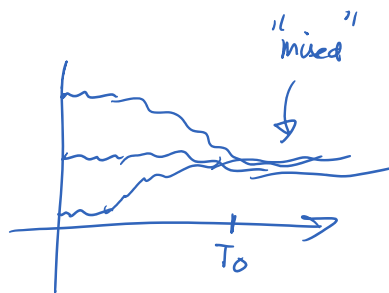
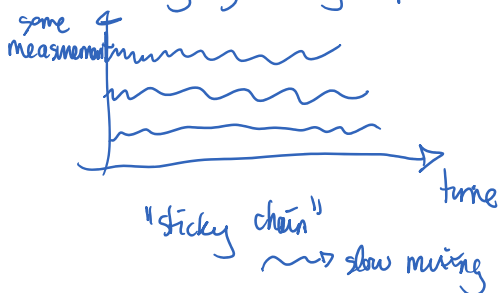
$$p(x_i | x_{-i}) \stackrel{\text{by cond indep}}{=} p(x_i | x_{\text{neighbor}(i)}) \propto \exp\left(n_i x_i + \sum_{j \in N(i)} n_{ij} x_i x_j\right)$$

⇒ renormalize to get conditional

$$p(x_i = 1 | x_{-i}) = \frac{\exp\left(n_i + \sum_{j \in N(i)} n_{ij} x_j^{(t-1)}\right) \exp(\text{rest})}{\left(\exp\left(n_i + \sum_{j \in N(i)} n_{ij} x_j^{(t-1)}\right) + 1\right) \exp(\text{rest})} = \sigma\left(n_i + \sum_{j \in N(i)} n_{ij} x_j^{(t-1)}\right)$$

diagnostic of mixing

monitor mixing by running independent chains



17h39

variational methods

general idea: use <sup>approximate</sup> optimization for approximating a quantity

eg say want to approximate  $\Theta^*$

then, express it as solution to opt. problem

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \mathcal{J}(\theta) \quad ] \text{ OPT}$$

idea: approximate  $\theta^*$  via an approximation to OPT

linear algebra example

say want sol'n to  $Ax = b$  (i.e.  $x^* = A^{-1}b$ )

$$x^* = \underset{x}{\operatorname{argmin}} \|Ax - b\|^2$$

Variational EM (motivation for objective)

recall EM trick:

latent variable  $p(x, z | \theta)$  <sup>unknown</sup>

$$\log p(x | \theta) \geq \mathbb{E}_q \left[ \log \frac{p(x, z | \theta)}{q(z)} \right] \triangleq \mathcal{J}(q, \theta)$$

$$\log p(x | \theta) - \mathcal{J}(q, \theta) = \text{KL}(q(\cdot) \| p(z | x; \theta))$$

$$\text{E step: } \underset{q \in \text{all distributions on } z}{\operatorname{argmax}} \mathcal{J}(q, \theta^{(t)}) \iff \underset{q}{\operatorname{argmin}} \text{KL}(q \| p(z | x; \theta^{(t)}))$$

\* a variational approx for E step:

$$\text{do } q_{\text{approx}}^{(t+1)} = \underset{q \in \mathcal{Q}_{\text{simple}}}{\operatorname{argmin}} \text{KL}(q \| p(z | x; \theta^{(t)}))$$

↑ choice of approx to approximate  $p(z | x; \theta^{(t)})$

approximate M step:

$$\theta^{(t+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathbb{E}_{q_{\text{approx}}^{(t+1)}} \left[ \log p(x, z | \theta) \right] \triangleq \mathcal{J}(q_{\text{approx}}^{(t+1)}, \theta)$$

still a lower bound on  $\log p(x | \theta)$

but base monotonicity guarantee on  $\log p(x | \theta^{(t)})$  as a fun of  $\theta$

more generally, using  $\underset{q \in \mathcal{Q}}{\operatorname{argmin}} \text{KL}(q \| p)$  is a variational approach to approximate  $p$

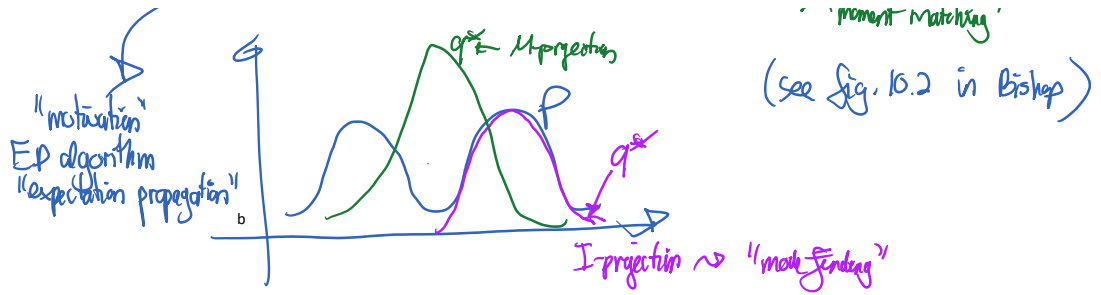
note: I-projection; if  $q$  is simple,  $\mathbb{E}_q[\log q]$  can compute

$$\text{alternative: } \underset{q \in \mathcal{Q}}{\operatorname{argmin}} \text{KL}(p \| q)$$

M-projection

↳ "moment matching"

(see fig. 10.2 in Bishop)



## Mean-field approximation (section 10.1 in Bishop)

let's suppose that  $p(z)$  is in exp. family

$$z_1, \dots, z_p \quad p(z) = \exp(\eta^T T(z) - A(\eta))$$

mean field approximation  $\mathcal{Q}_{MF} = \{q(z) = \prod_i q_i(z_i)\}$   
set of fully factorized distributions

$$\begin{aligned} KL(q \| p) &= \mathbb{E}_q [\log \frac{q}{p}] \\ &= -\eta^T \mathbb{E}_q [T(z)] + A(\eta) + \sum_z q(z) \log q(z) \\ &\quad \left( \sum_i q_i(z_i) \right) \left( \sum_j \log q_j(z_j) \right) \\ &= \sum_i \sum_z q_i(z_i) q_i(z_i) \log q_i(z_i) + \sum_j \sum_z q_j(z_j) \log q_j(z_j) \end{aligned}$$

### coordinate descent on $q_i$ 's

fix  $q_j$  for  $j \neq i$      $\min_{q_i} KL(q_i q_{-i} \| p)$

$$= -\mathbb{E}_{q_i} [\eta^T \mathbb{E}_{q_{-i}} [T(z)]] + \text{const.} + \sum_z q_i(z_i) \log q_i(z_i)$$

$\triangleq f_i(z_i)$

(Or like Max-ENT)

add Lagrange multiplier  $\sum_z q_i(z_i) = 1 \rightarrow \lambda (1 - \sum_z q_i(z_i))$

$$\frac{\partial}{\partial q_i(z_i)} = 0 \Rightarrow -f_i(z_i) + \log q_i(z_i) + 1 - \lambda = 0$$

$$\Rightarrow q_i^*(z_i) \propto \exp(f_i(z_i))$$

⊛ general mean field update when target  $p$  is in exp family

$$q_i^*(z_i) \propto \exp(\eta^T \mathbb{E}_{q_{-i}^{(t)}} T(z))$$

Ising model:

$$T(z) \Rightarrow (z_i)_{i \in V} \quad z_i \in \{0, 1\}$$

$$(z_i z_j)_{i, j \in E}$$

$$j \in \mathcal{N}(i) : \mathbb{E} q_{ji}^{(k)}(z_j) = q_j^{(k)}(z_j=1) \stackrel{\Delta}{=} \mu_j^{(k)}$$

$$\mathbb{E} q_{ji}^{(k)}(z_i z_j) = z_i \mu_j^{(k)}$$

$$n^T \mathbb{E} q_{ji}^{(k)} T(z) = n_i z_i + \sum_{j \in \mathcal{N}(i)} n_j \underbrace{\mathbb{E} q_{ji}^{(k)}[z_j]}_{\mu_j^{(k)}} + \sum_{j \in \mathcal{N}(i)} n_j \underbrace{\mathbb{E} q_{ji}^{(k)}[z_i z_j]}_{z_i \mu_j^{(k)}} + \text{rest (no } z_i)$$

$$\text{result: } q_i^{(k+1)}(z_i) \propto \exp(n_i z_i + z_i \sum_{j \in \mathcal{N}(i)} n_j \mu_j^{(k)})$$

$$\mu_i^{(k+1)} = \sigma(n_i + \sum_{j \in \mathcal{N}(i)} n_j \mu_j^{(k)})$$

MF update for  $q_i(z_i)$   
[with parameter  $\mu_i$ ]

compare with

$$\text{GS, update where } z_i^{(k+1)} = 1 \text{ with prob } \sigma(n_i + \sum_{j \in \mathcal{N}(i)} n_j z_j^{(k)})$$