MCMC for Structure Learning

In the context of structure learning, MCMC methods can be used to explore the space of possible causal structures and estimate the posterior probability of different causal graphs given the observed data.

- General MCMC process:

1. Define a prior $p_r(G) = \begin{cases} \text{dirichlet} & \text{if } D \text{ is Discrete} \\ \text{Wilshaw} & \text{if } D \text{ is continuous} \end{cases}$

2. Define a likelihood function: $p_r(D \mid G) = ?$

3. Initialize the Markov Chain with some graph.

4. Iteratively modify the current graph to generate a new graph

5. Estimate the posterior probability of each graph: $p_r(G \mid D) \propto \underbrace{P_r(D \mid G)}_{\text{step 2}} \underbrace{P_r(G)}_{\text{step 1}}$

How do I know the MCMC on graphs converged? Expected value

Effective sample size

conv. diag

chains

BN:

$X \to Y \to Z$

Joint distribution of BN:

$P_r(X, Y, Z) = P_r(X) P_r(Y \mid X) P_r(Z \mid Y)$

Likelihood of the data given the graph.

$P_r(D \mid G) = \prod P_r(X = x_i, Y = y_i, Z = z_i)$

D:

X	Y	Z
x_1	y_1	z_1
x_2	y_2	z_2
$\vdots$	$\vdots$	$\vdots$
x_n	y_n	z_n

BN:

$$X \to Y \to Z$$

DATA D:

$$\begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & & \\ x_n & y_n & z_n \end{bmatrix}$$

the joint distribution of the network is

$$P_r(X, Y, Z) = P_r(X) P_r(Y|X) P_r(Z|Y)$$

the likelihood of the data given the network

$$P_r(D|G) = \prod P_r(X = x_i, Y = y_i, Z = z_i)$$
$$= \prod p_r(x_i)\, p_r(y_i | x_i)\, p_r(z_i | y_i)$$

$$P_r(G|D) = P_r(D|G) P_r G$$

Log likelihood: (step 4)

$$\log P_r(D|G) = \sum \log(p_r(x_i)) + \log(p_r(y_i|x_i)) + \log(p_r(z_i|y_i))$$

Step 3: Initialize the Markov Chain on <u>some graph</u>:

↳ you can start with the PC algo after skeleton phase

Step 4. Metropolis Hastings:

- proposal step
- acceptance step
- Iteration step

Convergence: if the distribution is symmetric add its negative the expected value should be 0.

→ Proposal Step: the proposal step involves generating a new graph from the current distribution $Q(G'|G)$ which is the probability of proposing a move from graph G to graph G'

Let G be the current graph and G' be a proposed new graph. the proposal distribution $Q(G'|G)$ is a conditional probability distribution that defines the probability of proposing each possible new graph G' given the current graph G.

A common choice for $Q(G'|G)$ is a uniform distribution over all graphs that can be obtained by adding, deleting or reversing a single edge.

Under this definition, the proposal distribution is symmetric:

$Q(G'|G) = Q(G|G')$ for all G and G'. this is because each possible move from G to G' is also possible move in the reverse direction from G' to G.

the proposal step of the Metropolis-Hastings algorithm involves drawing a sample from this proposal distribution. this can be done by adding, removing or reversing, and proposing the resulting graph as the new graph.

→ **Acceptance Step**: Consists in accepting or rejecting the proposed new graph.

Let G be the current graph and G' be the proposed new graph G'. the acceptance probability $A(G, G')$ is defined as follows:

$$A(G, G') = \min\left(1, \frac{Pr(G'|D) \times Q(G|G')}{Pr(G|D) \times Q(G'|G)}\right)$$

where

$Pr(G|D)$ and $Pr(G'|D)$ are the posterior probabilities of the current and proposed graphs.

$Q(G|G')$ and $Q(G'|G)$ are the proposal probabilities for moving from G to G' and from G' to G.

the acceptance step of the Metropolis-Hastings involves generating a random number u from a uniform distribution between 0 and 1.

$$\begin{cases} \text{accept} & \text{if } u < A(G, G') \quad G \leftarrow G' \\ \text{reject} & \text{otherwise} \quad G \text{ unchanged} \end{cases}$$

if the proposal distribution is symmetric, then

$$A(G, G') = \min\left(1, \ Pr(G'|D) / Pr(G|D)\right)$$ ratio of the posterior probabilities of the proposed and curr graph

ALGORITHM:

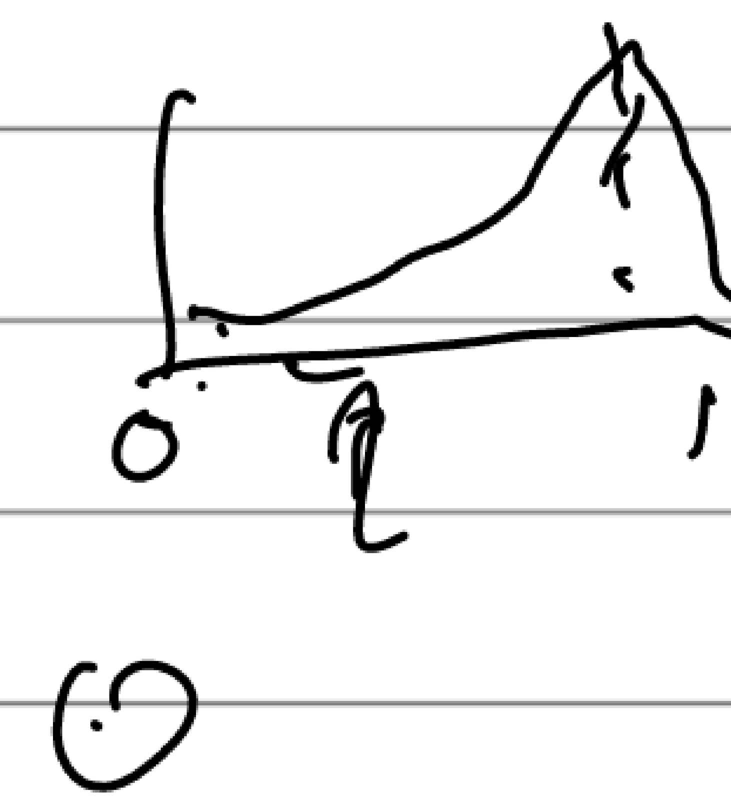
Input: Data $\in \mathbb{R}^{N\times m}$
max iterations

1. Initialize a graph G — we can already generate the graph resulting from the end of skeleton phase in PC-algo

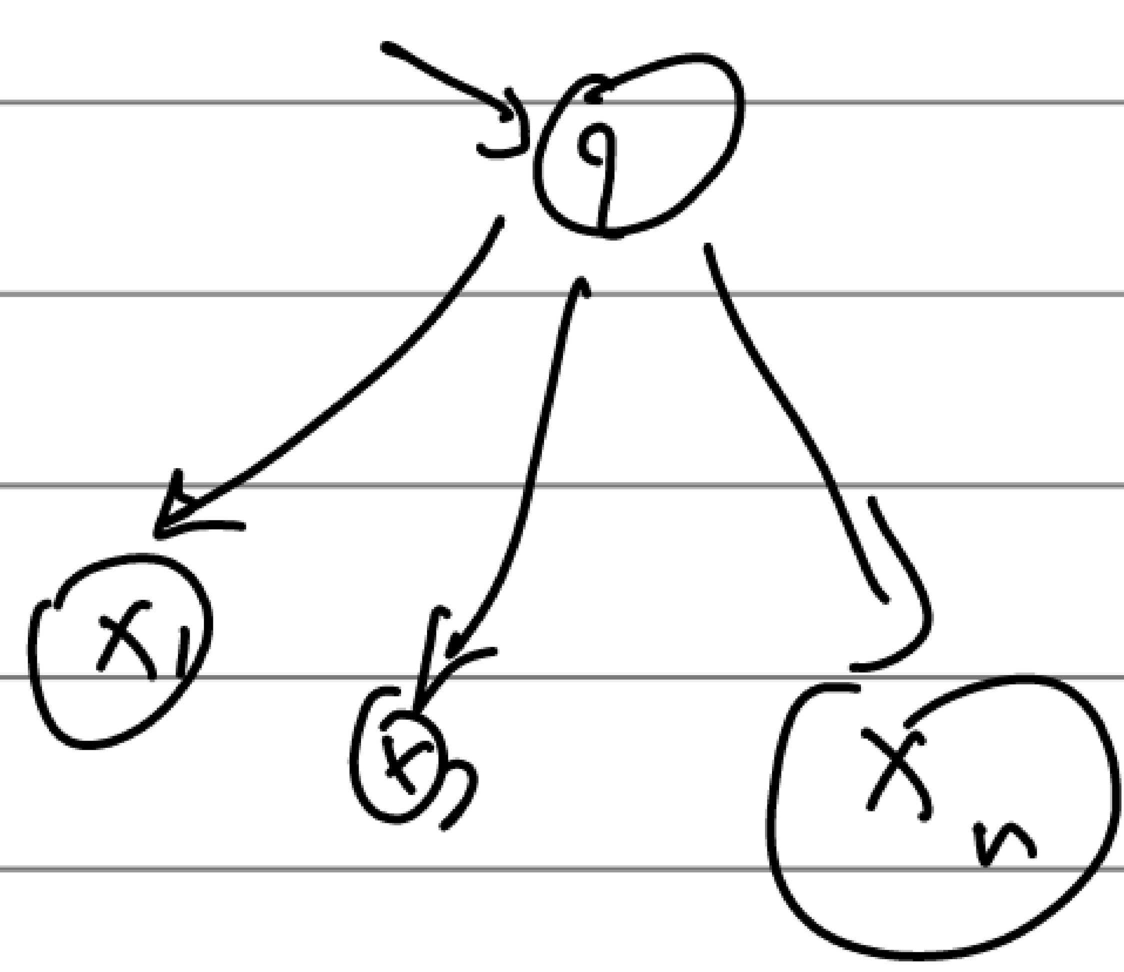
2. While max iterations is not reached do:
   2.1. Propose a new graph G' by making a small random change to G. (e.g. add, remove, reverse a random edge)
   2.2. Compute the acceptance probability $A(G, G')$ as follows:
      2.2.1. Compute the posterior probabilities $P(G|D)$ and $P(G'|D)$
      If the proposal distribution is symmetric:
      $$A(G, G') = \min(1, P_r(G'|D) / P_r(G|D))$$
      else:
      Compute the proposal probabilities $Q(G|G')$ and $Q(G'|G)$
      $$A(G|G') = \min\left(1, \frac{P_r(G'|D) \times Q(G, G')}{P_r(G|D) \times Q(G', G)}\right)$$
   2.3. Generate a random number u from a unifor distr $[0, 1]$
      If $u < A(G, G')$,
         Accept the proposed graph: $G = G'$

return G.

$$p(D|G) = \int p(D|G,\theta) \overset{p(\theta|G)}{p(\theta)} d\theta$$



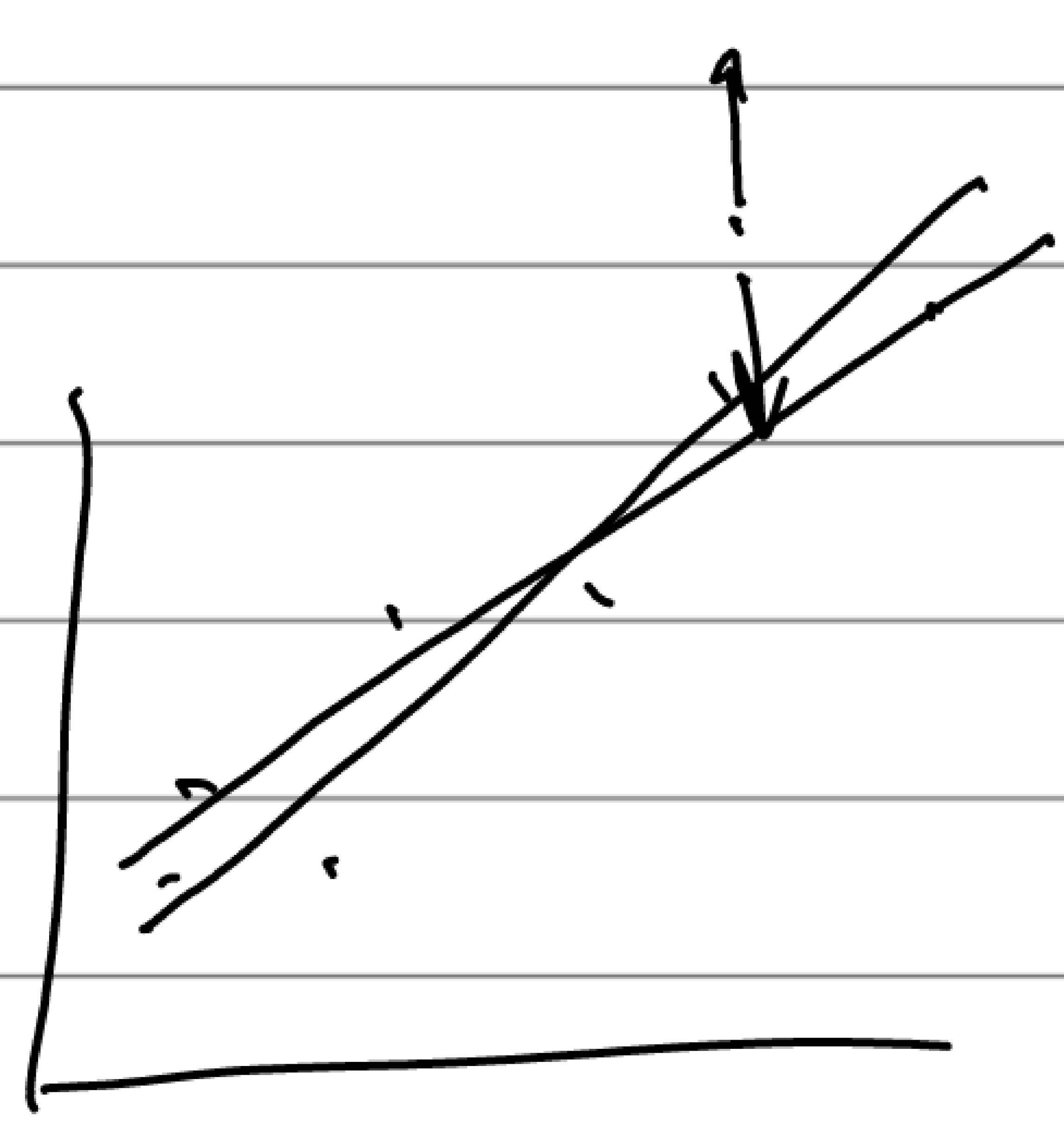
$$= \prod \int$$



$$P(D|G) = \prod_{i=} P(V_i | Pa(V_i))$$

$$= \int \prod P(V_i | Pa(V_i)) \, d\theta$$

$$= \prod \iiint P(V_i | Pa(V_i)) \, d\theta_i$$



$$\min \| mx + b \|^2 \frac{1}{2}$$