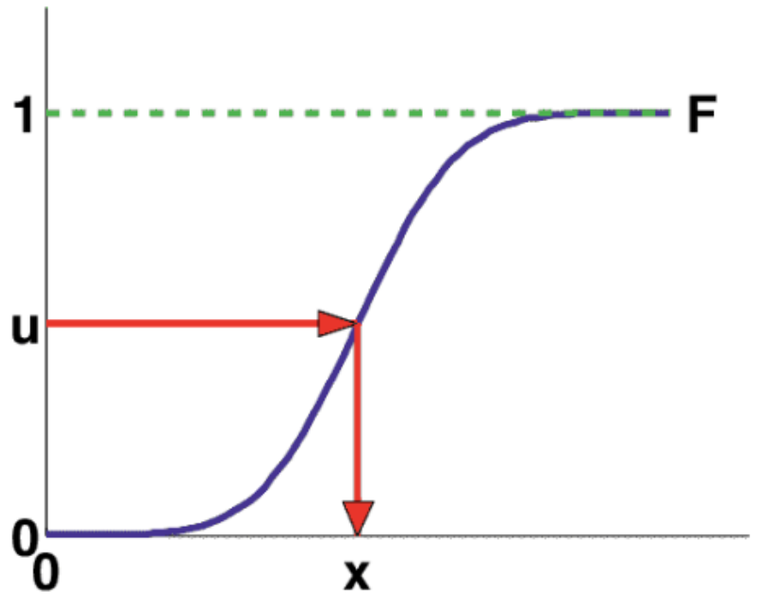
Chapter 23: Sampling and Monte Carlo inference

When you interview a data scientist...

Sampling using the inverse CDF

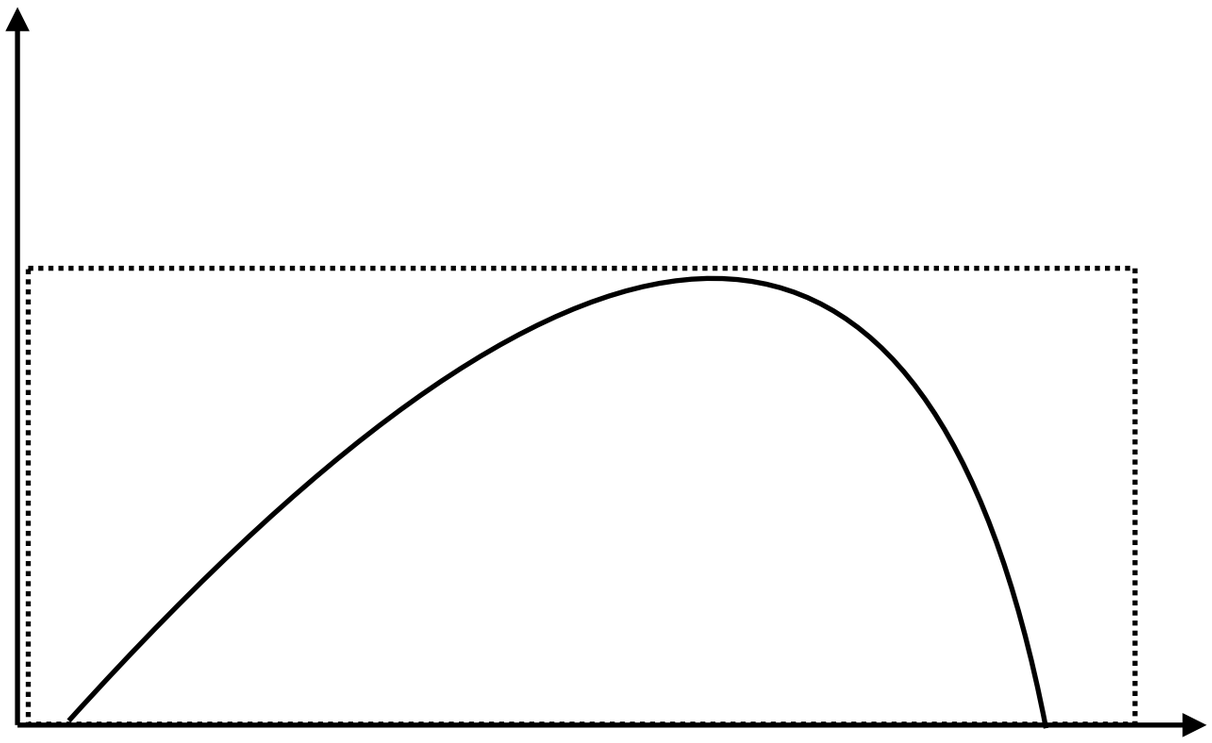


$$\text{Expon}(x|\lambda) \triangleq \lambda e^{-\lambda x}\ \mathbb{I}(x \geq 0)$$

$$F(x) = 1 - e^{-\lambda x}\ \mathbb{I}(x \geq 0)$$

$$F^{-1}(p) = -\frac{\ln(1-p)}{\lambda}$$

Rejection sampling



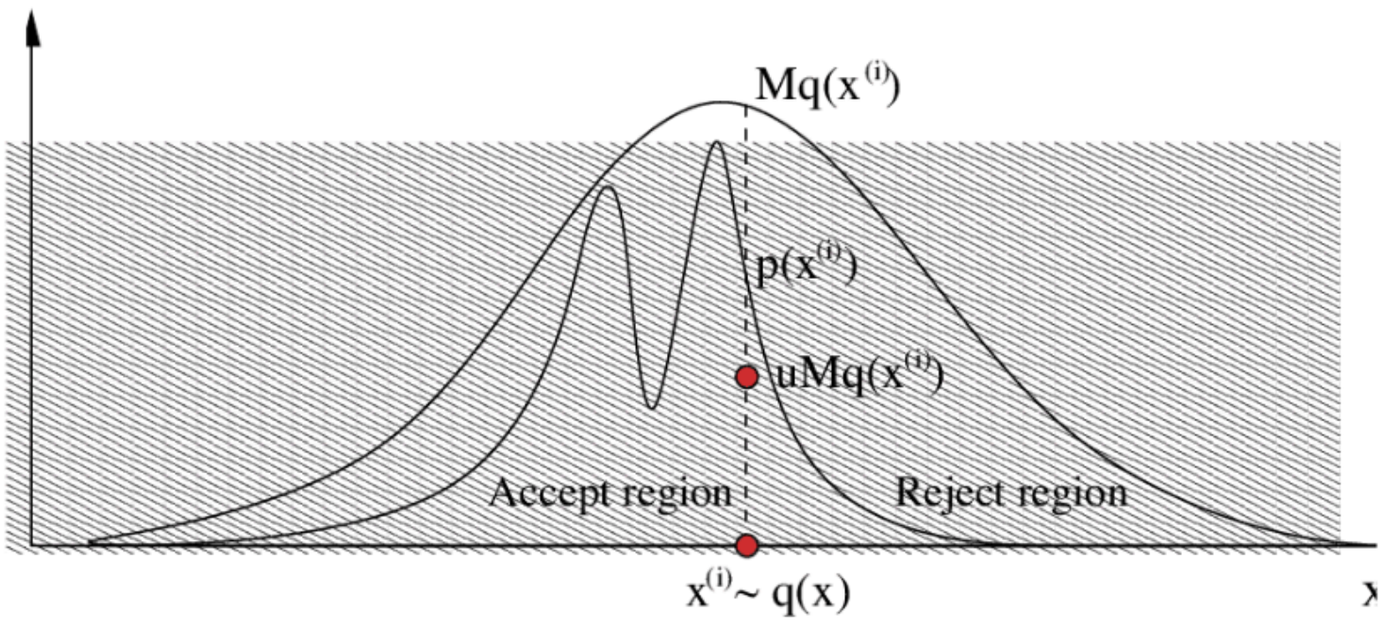
$$M \triangleq \max_x p(x)$$

$$x \sim U(0, D)$$

$$u \sim U(0, 1)$$

$$\text{Accept if } u < \frac{p(x)}{MD}$$

Rejection sampling



$$Mq(x) \geq \tilde{p}(x)$$

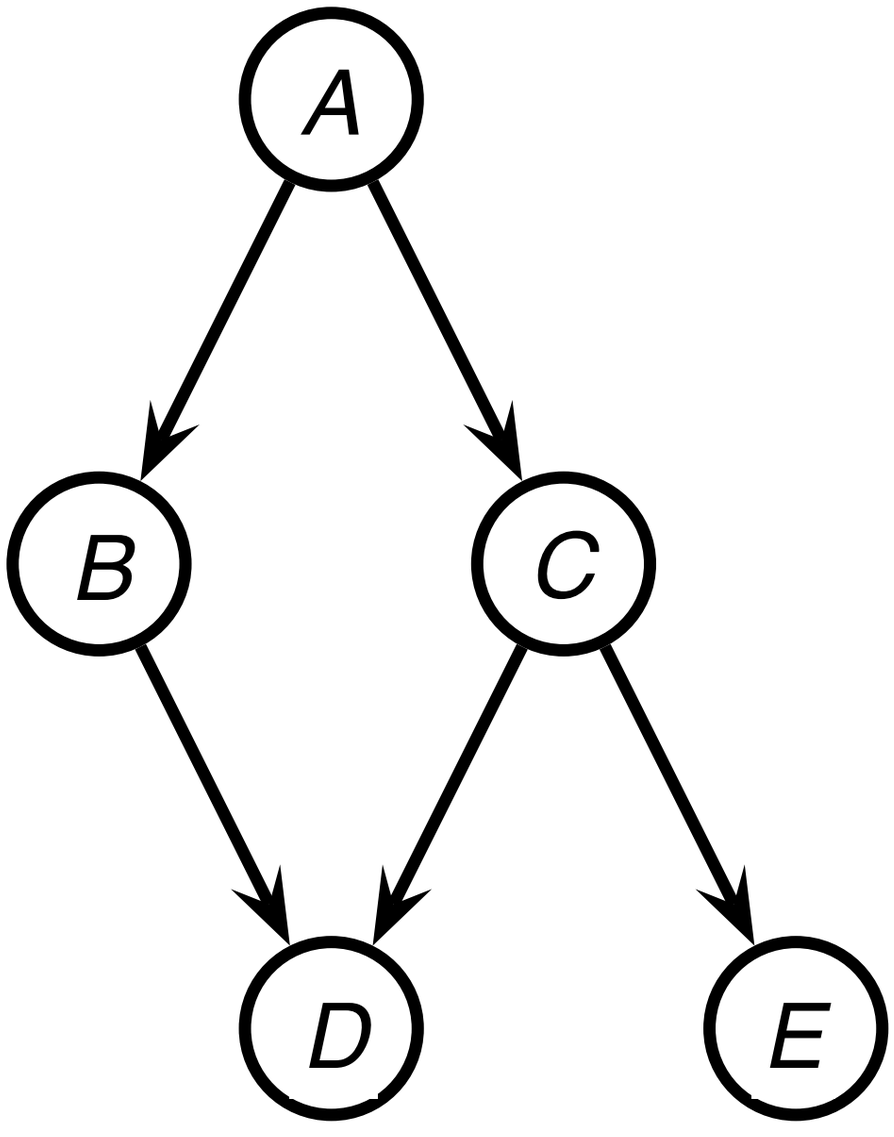
$$x \sim q(x)$$

$$u \sim U(0,1)$$

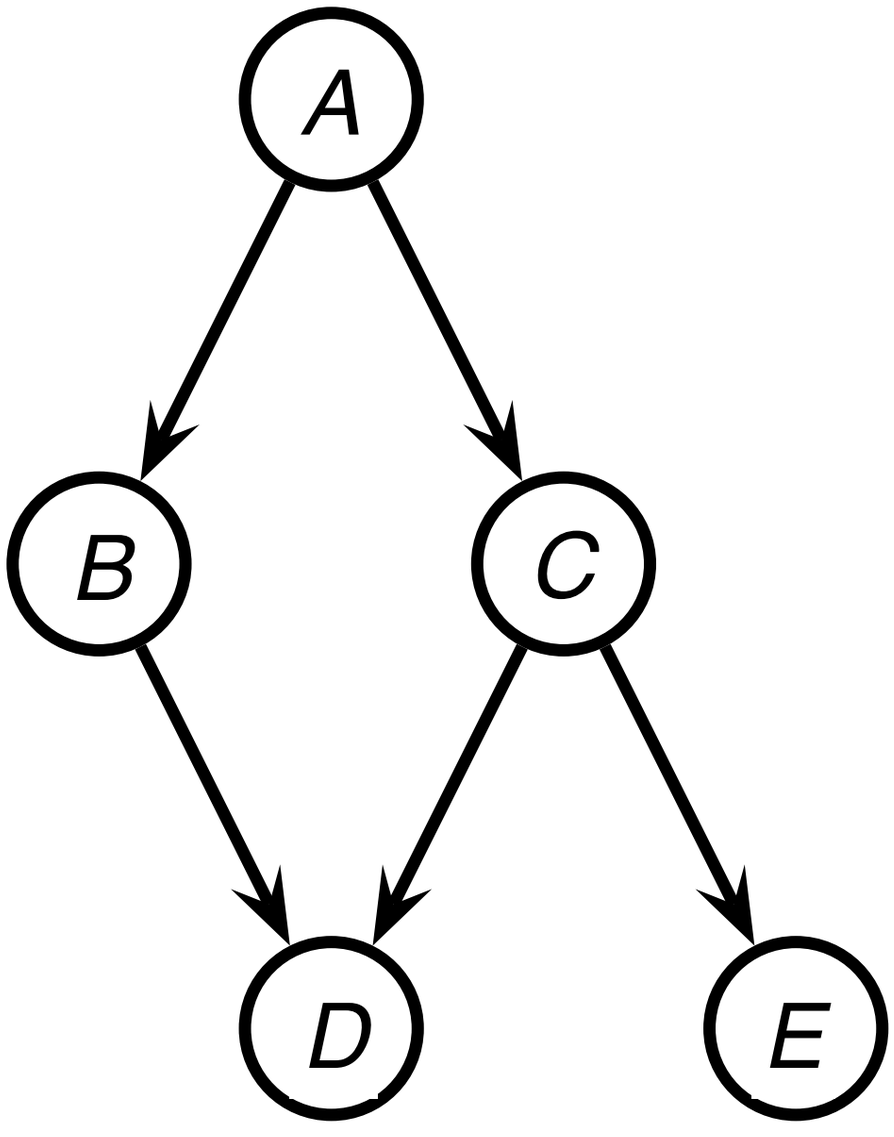
$$\text{Accept if } u < \frac{\tilde{p}(x)}{Mq(x)}$$

Rejection sampling for multivariate distributions

Sampling from a Bayesian network



Importance weighting



Inference using sampling

Exercise

Propose a way to sample from a Beta(2,2) distribution.

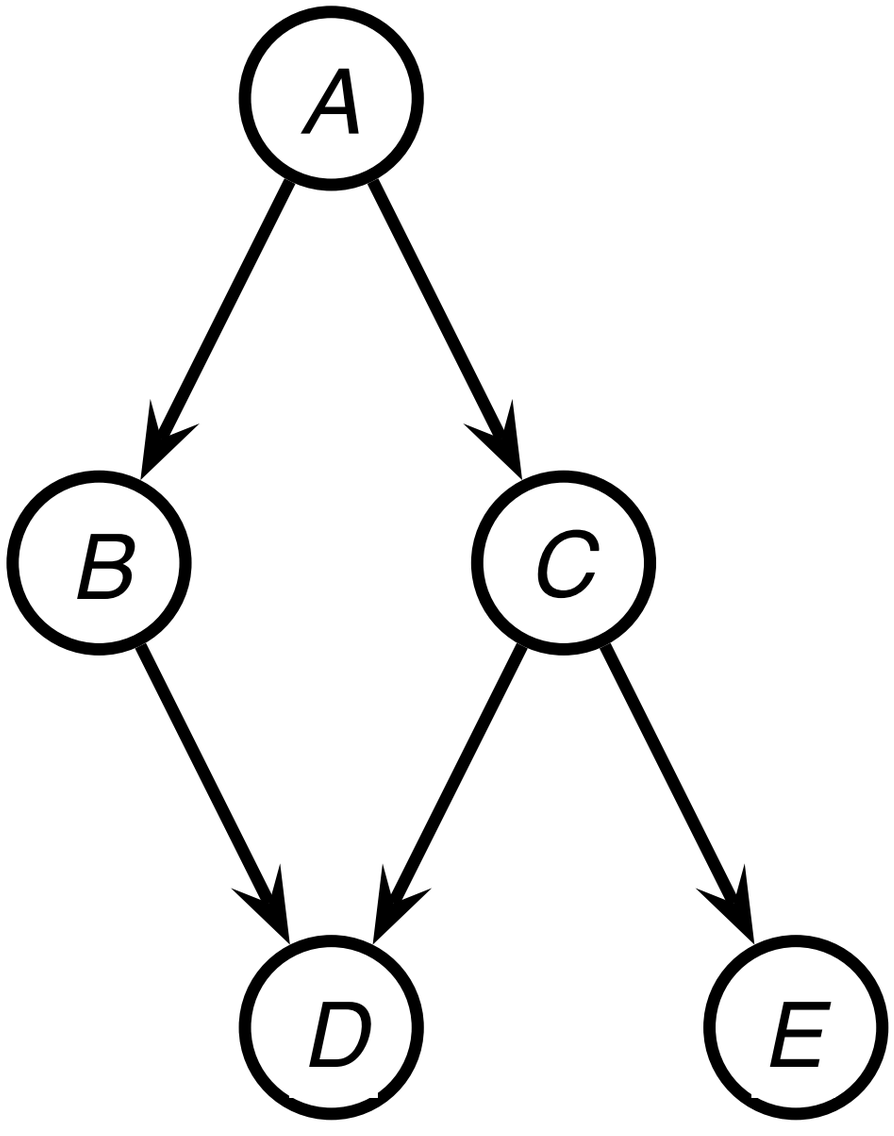
Propose a way to sample the label from a NB model where the features (and parameters) are observed.

Exercise

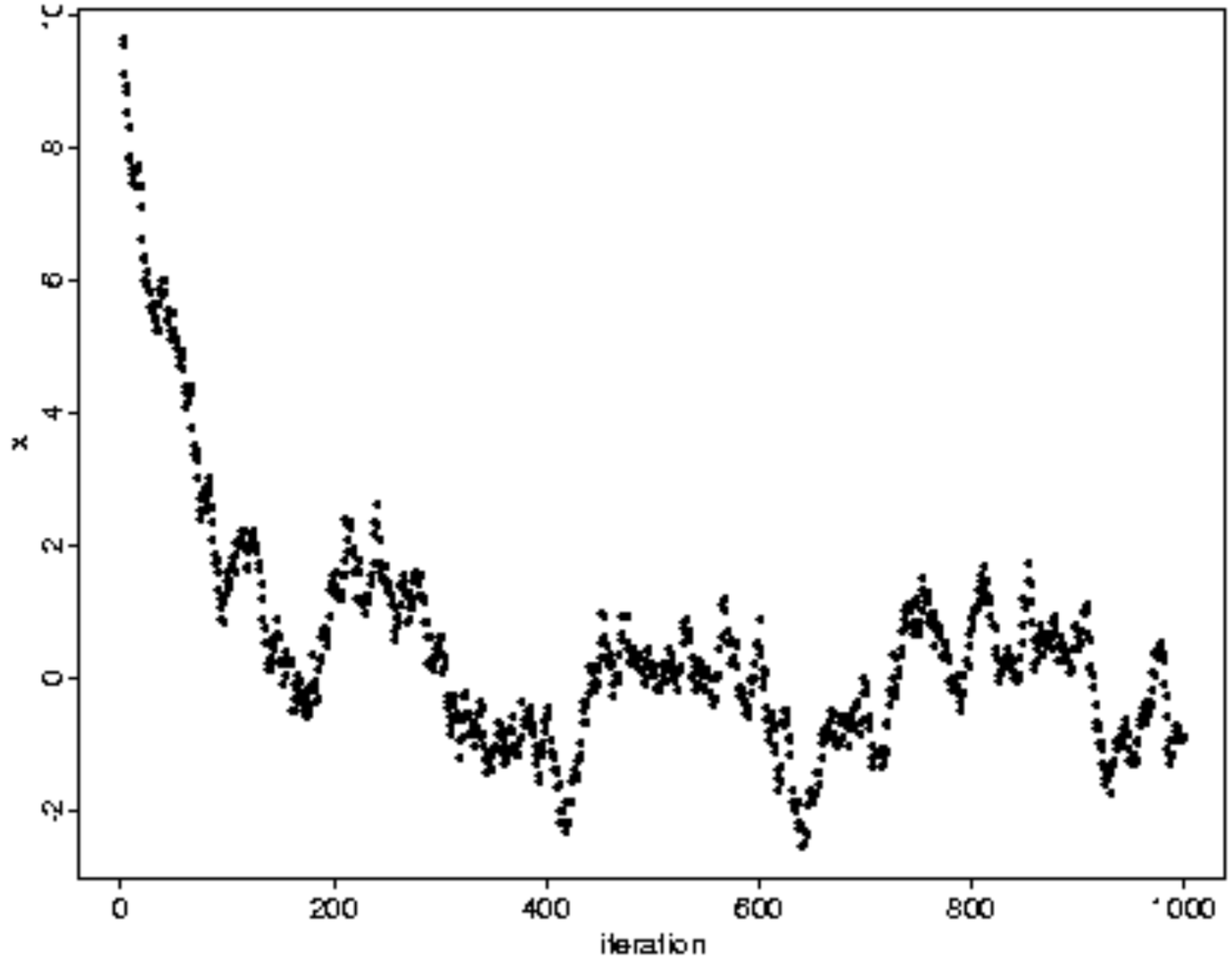
Given only a fair coin, sample from a weighted coin with 1/3 probability of heads.

Chapter 24: Gibbs sampling and Markov chain Monte Carlo (MCMC)

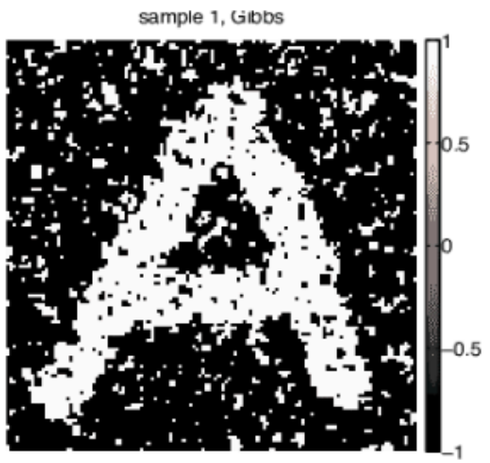
Gibbs sampling

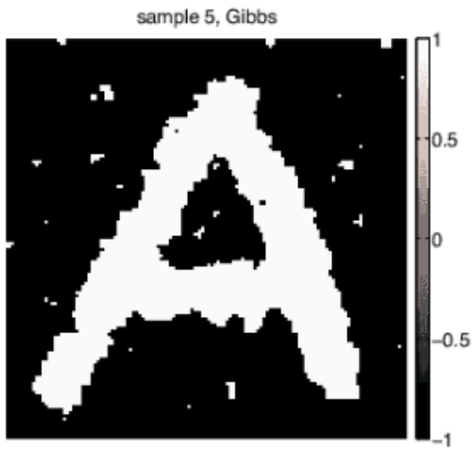


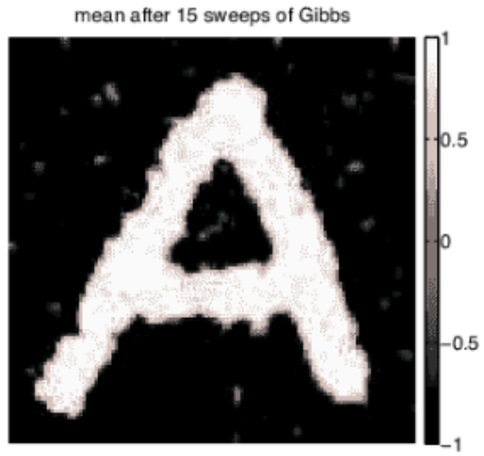
Burn-in

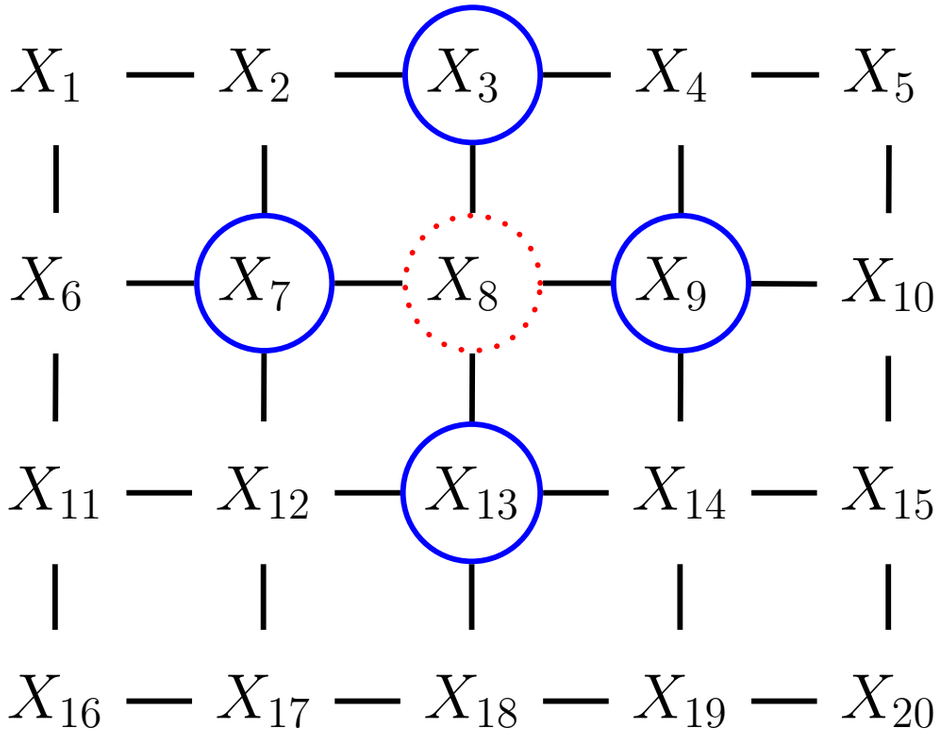


Gibbs sampling for the Ising model



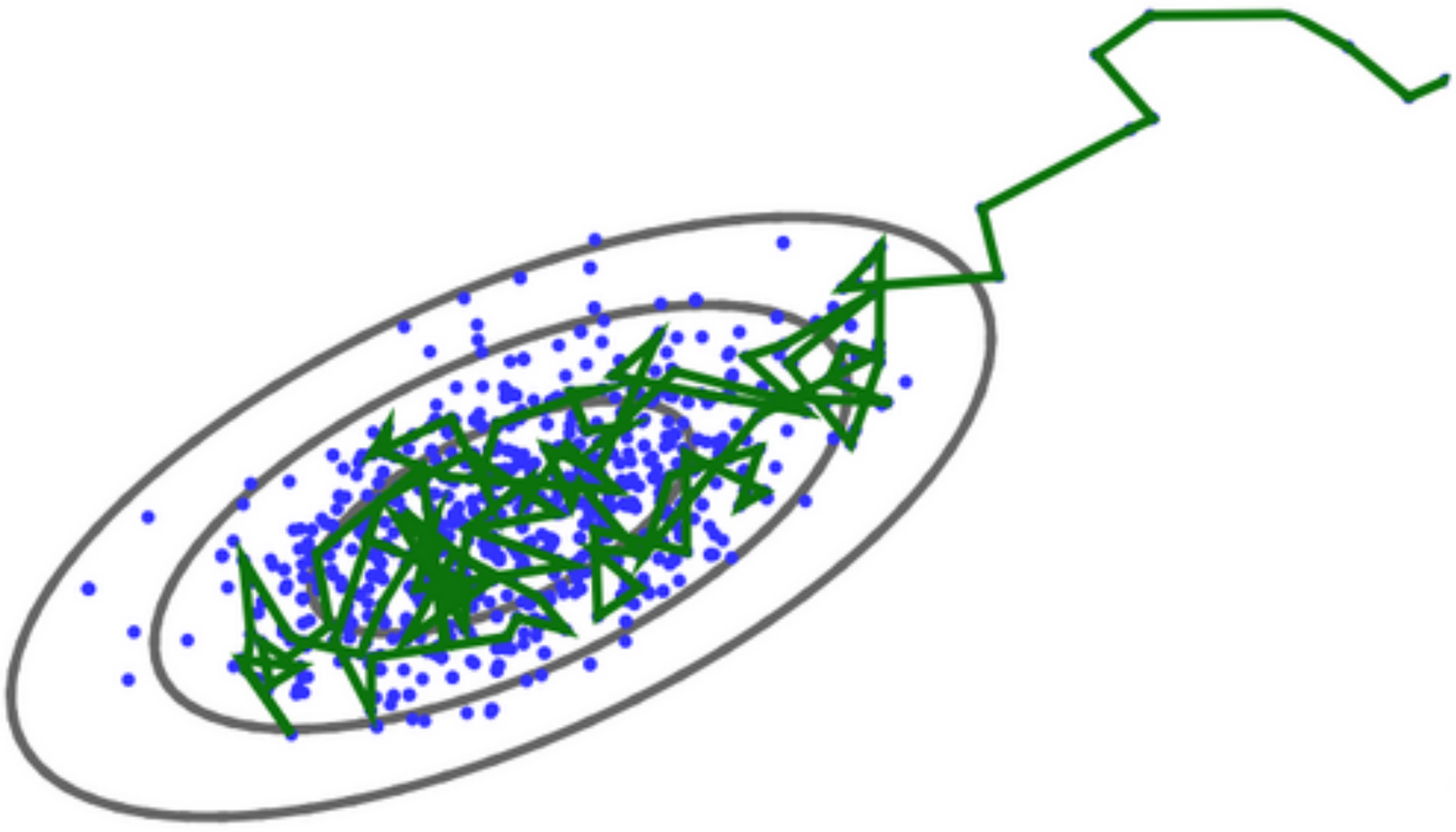






$$p(x_t|\mathbf{x}_{-t}, \boldsymbol{\theta}) \propto \prod_{s \in \text{nbr}(t)} \psi_{st}(x_s, x_t)$$

Metropolis Hastings algorithm



Proposal distribution and accept probability

Hastings correction for assymetric proposal distributions

$$\alpha = \frac{p^*(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{p^*(\mathbf{x})q(\mathbf{x}'|\mathbf{x})} = \frac{p^*(\mathbf{x}')/q(\mathbf{x}'|\mathbf{x})}{p^*(\mathbf{x})/q(\mathbf{x}|\mathbf{x}')}$$

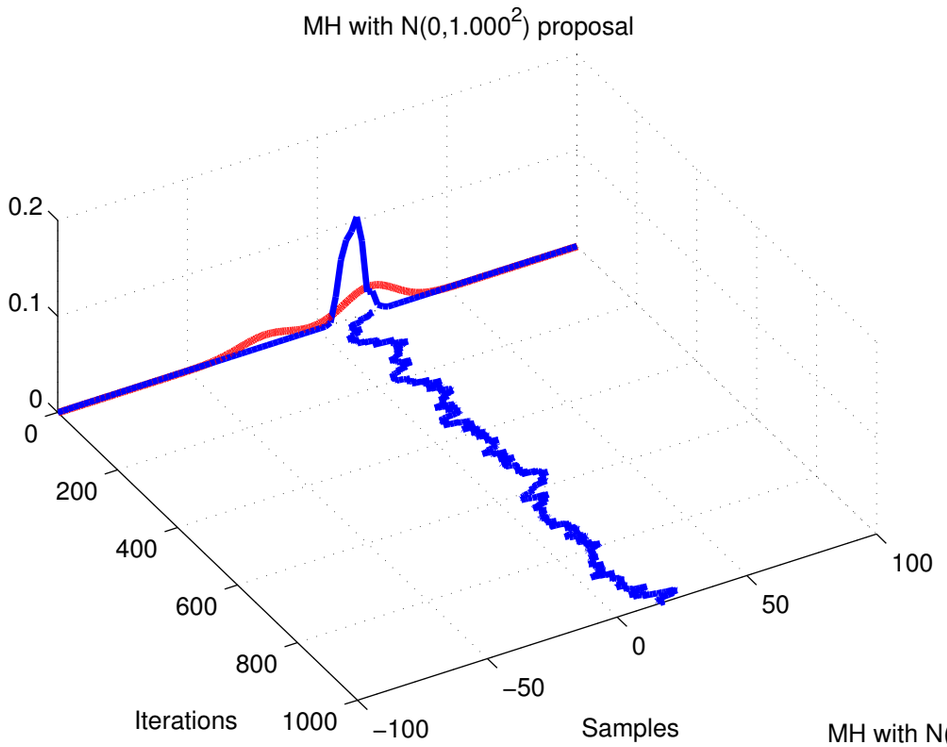
Gibbs sampling is a special case of MH

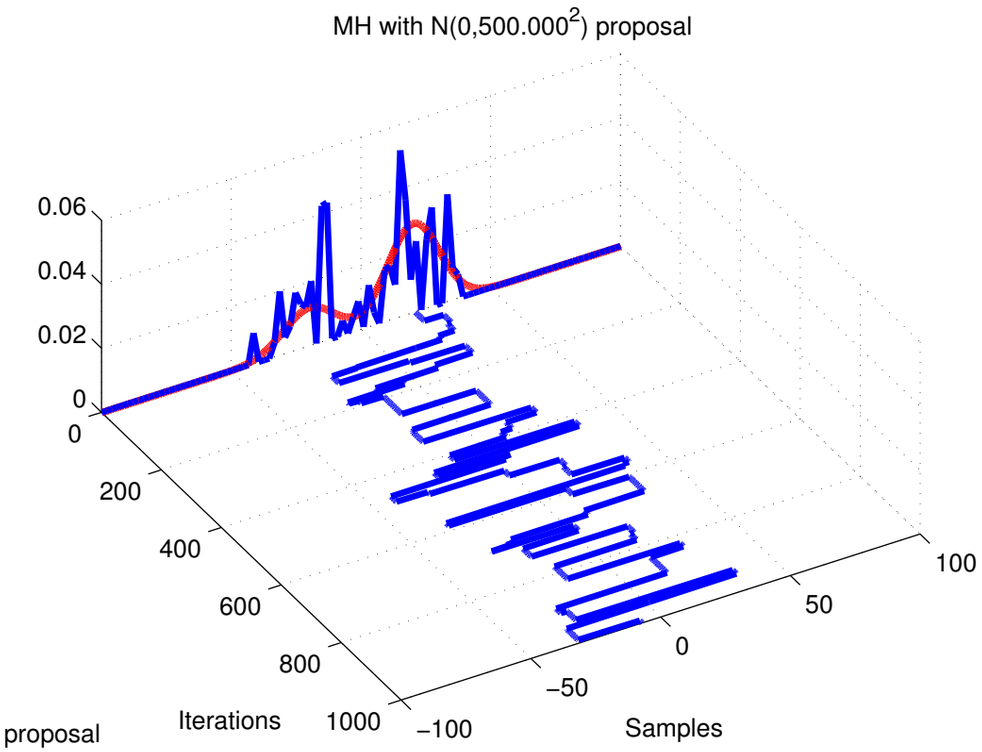
$$q(\mathbf{x}'|\mathbf{x}) = p(x'_i|\mathbf{x}_{-i})\mathbb{I}(\mathbf{x}'_{-i} = \mathbf{x}_{-i})$$

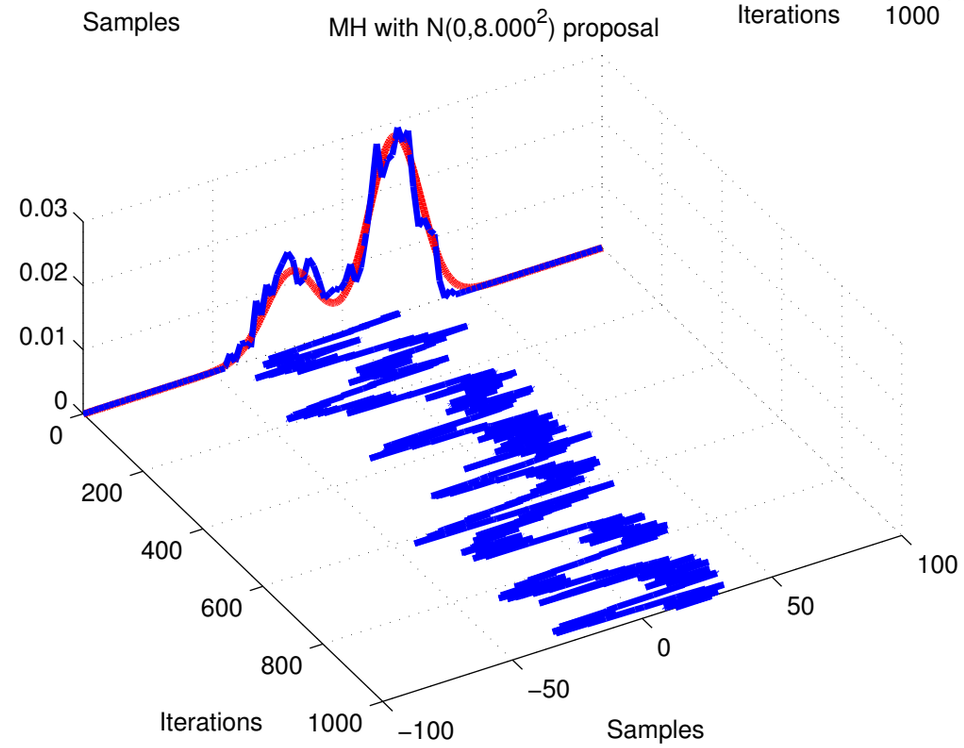
$$\begin{aligned}
\alpha &= \frac{p(\mathbf{x}')q(\mathbf{x}|\mathbf{x}')}{p(\mathbf{x})q(\mathbf{x}'|\mathbf{x})} = \frac{p(x'_i|\mathbf{x}'_{-i})p(\mathbf{x}'_{-i})p(x_i|\mathbf{x}'_{-i})}{p(x_i|\mathbf{x}_{-i})p(\mathbf{x}_{-i})p(x'_i|\mathbf{x}_{-i})} \\
&= \frac{p(x'_i|\mathbf{x}_{-i})p(\mathbf{x}_{-i})p(x_i|\mathbf{x}_{-i})}{p(x_i|\mathbf{x}_{-i})p(\mathbf{x}_{-i})p(x'_i|\mathbf{x}_{-i})} = 1
\end{aligned}$$

Proposal distributions

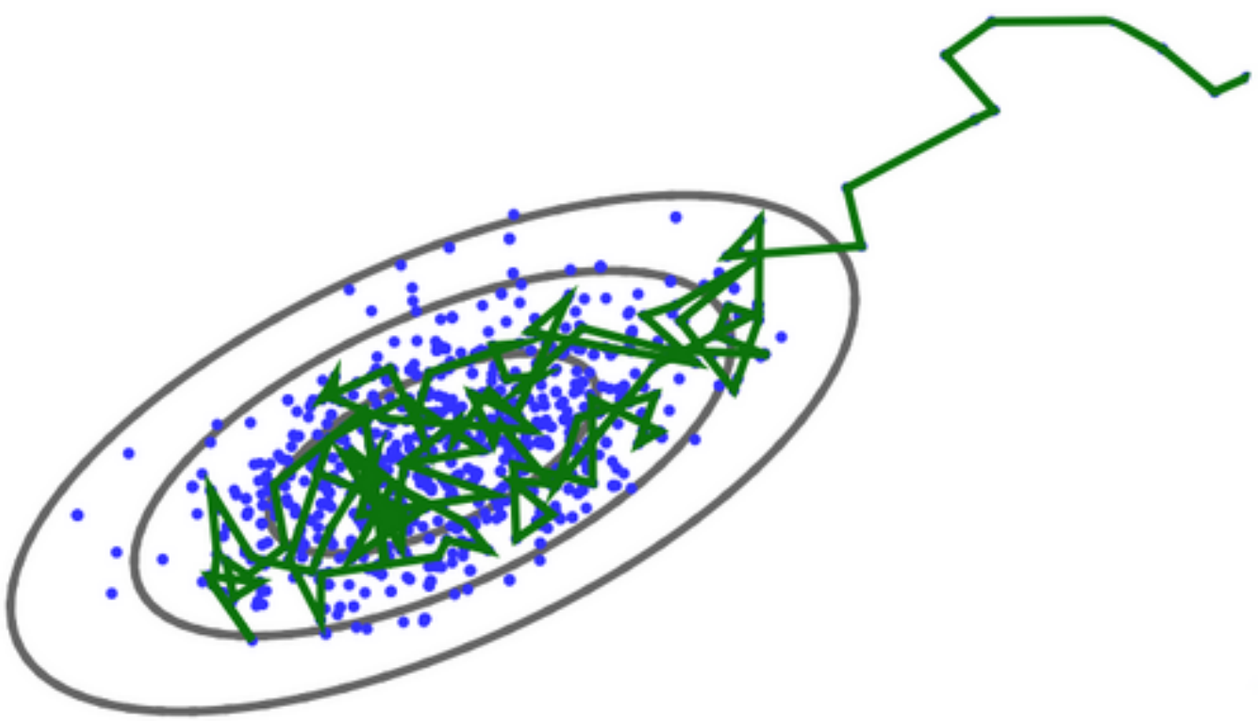
Proposal distributions

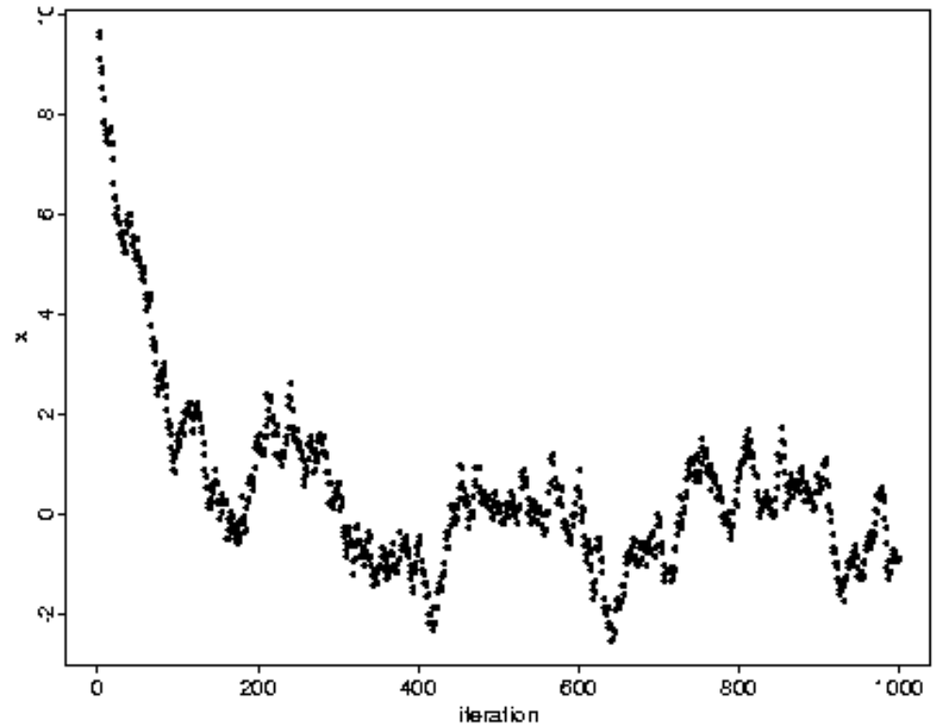






Burn-in





Diagnosing convergence in practice

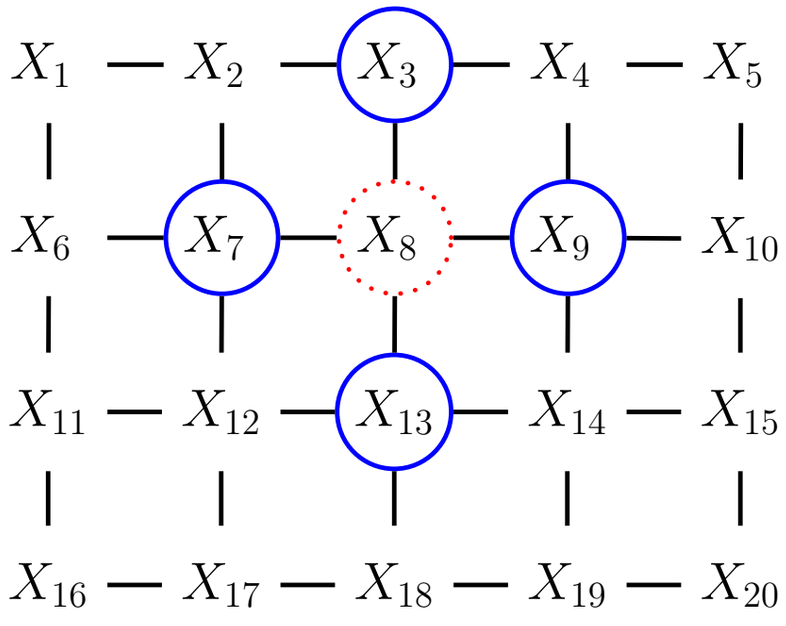
Annealing algorithms

Exercise

Suppose $x_i \in \{-1, +1\}$

and $\phi(x_s, x_t) = \exp(J x_s x_t)$

Derive an expression for $p(x_t | x_{-t})$



$$p(x_t | \mathbf{x}_{-t}, \boldsymbol{\theta}) \propto \prod_{s \in \mathrm{nbr}(t)} \psi_{st}(x_s, x_t)$$

Exercise

Show how to apply Gibbs sampling to a univariate mixture of Gaussians.