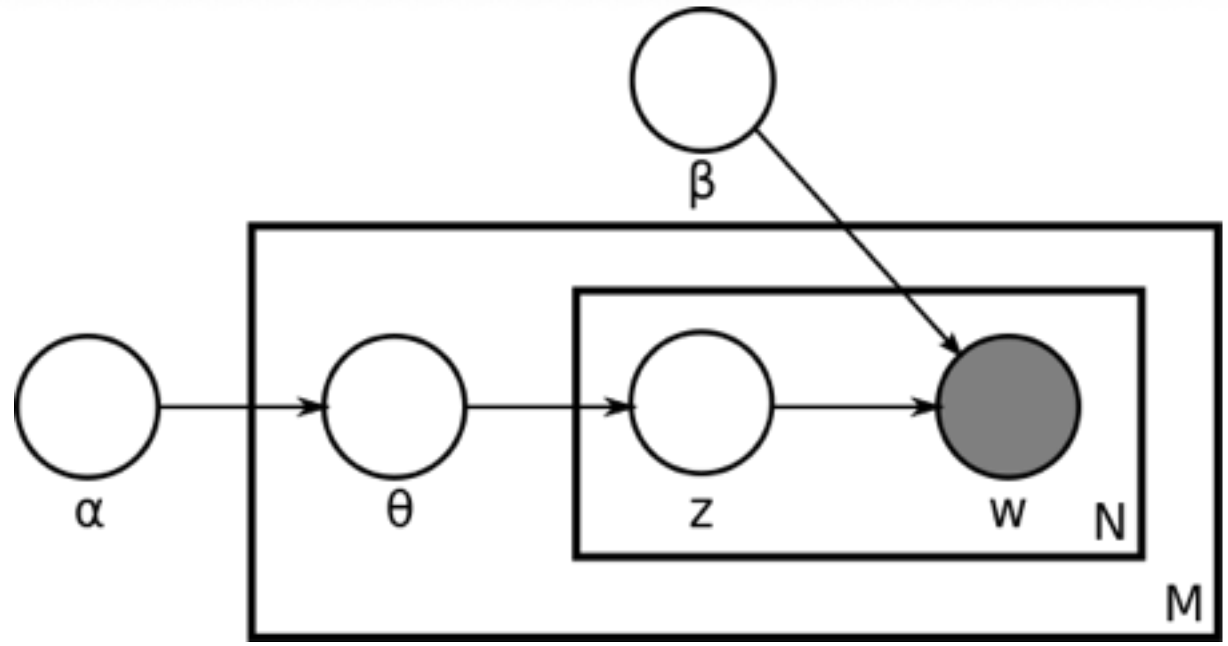
6.867
Sampling

Fall 2016

Inference is hard



$$P(\boldsymbol{Z}, \boldsymbol{W}; \alpha, \beta) = \prod_{j=1}^{M} \frac{\Gamma\left(\sum_{i=1}^{K} \alpha_i\right)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(\cdot)}^{i} + \alpha_i)}{\Gamma\left(\sum_{i=1}^{K} n_{j,(\cdot)}^{i} + \alpha_i\right)} \times \prod_{i=1}^{K} \frac{\Gamma\left(\sum_{r=1}^{V} \beta_r\right)}{\prod_{r=1}^{V} \Gamma(\beta_r)} \frac{\prod_{r=1}^{V} \Gamma(n_{(\cdot),r}^{i} + \beta_r)}{\Gamma\left(\sum_{r=1}^{V} n_{(\cdot),r}^{i} + \beta_r\right)}.$$

Sampling from $P(Z|W, \alpha, \beta)$ not easy!

LDA model: https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation

Key Points, Topics

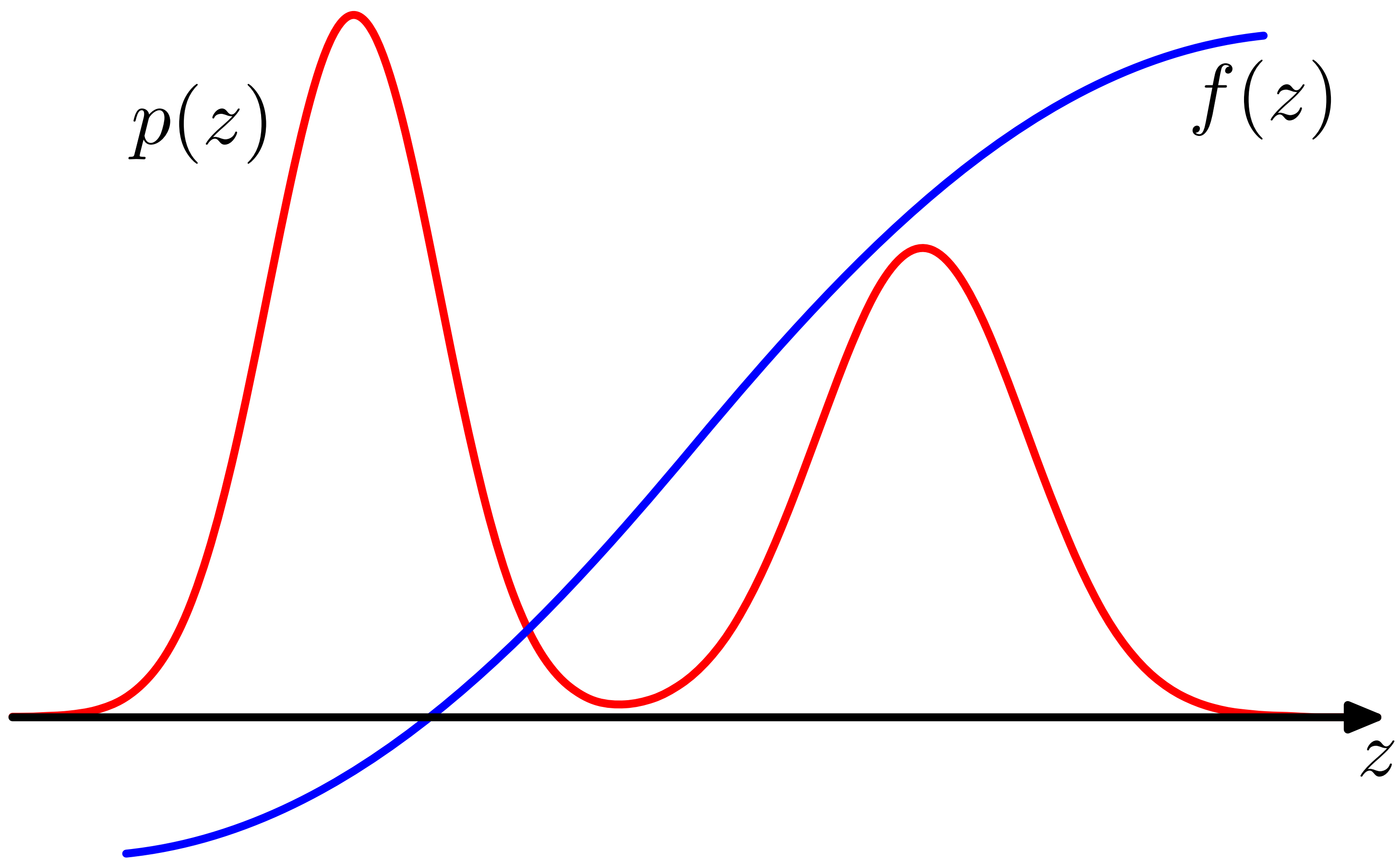
Exact inference is hard

Drawing samples from high-dim P(x) hard
even if we can easily evaluate $P^*(x) \propto P(x)$

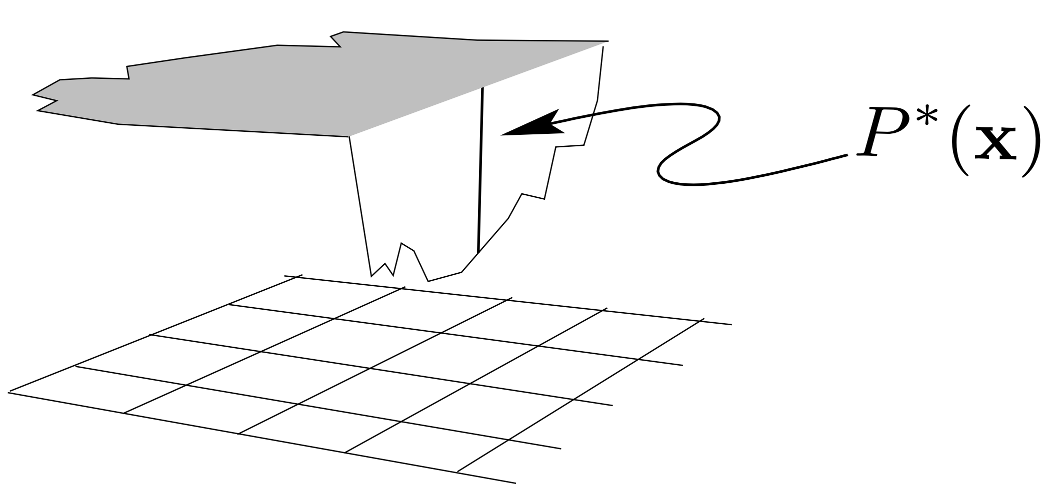
Topics

Importance Sampling

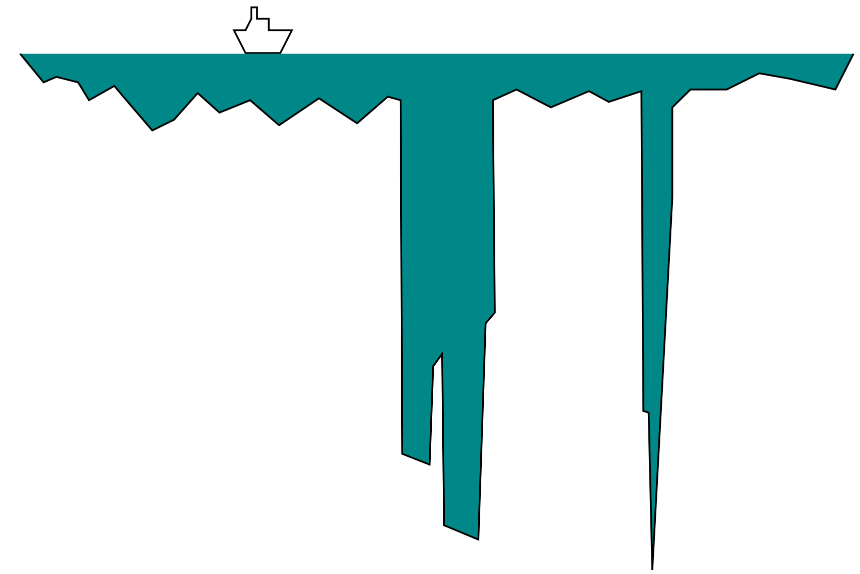
MCMC Methods
- Metropolis
- Metropolis-Hastings
- Gibbs Sampling
- Mixing time



Difficulty of estimating E[f]



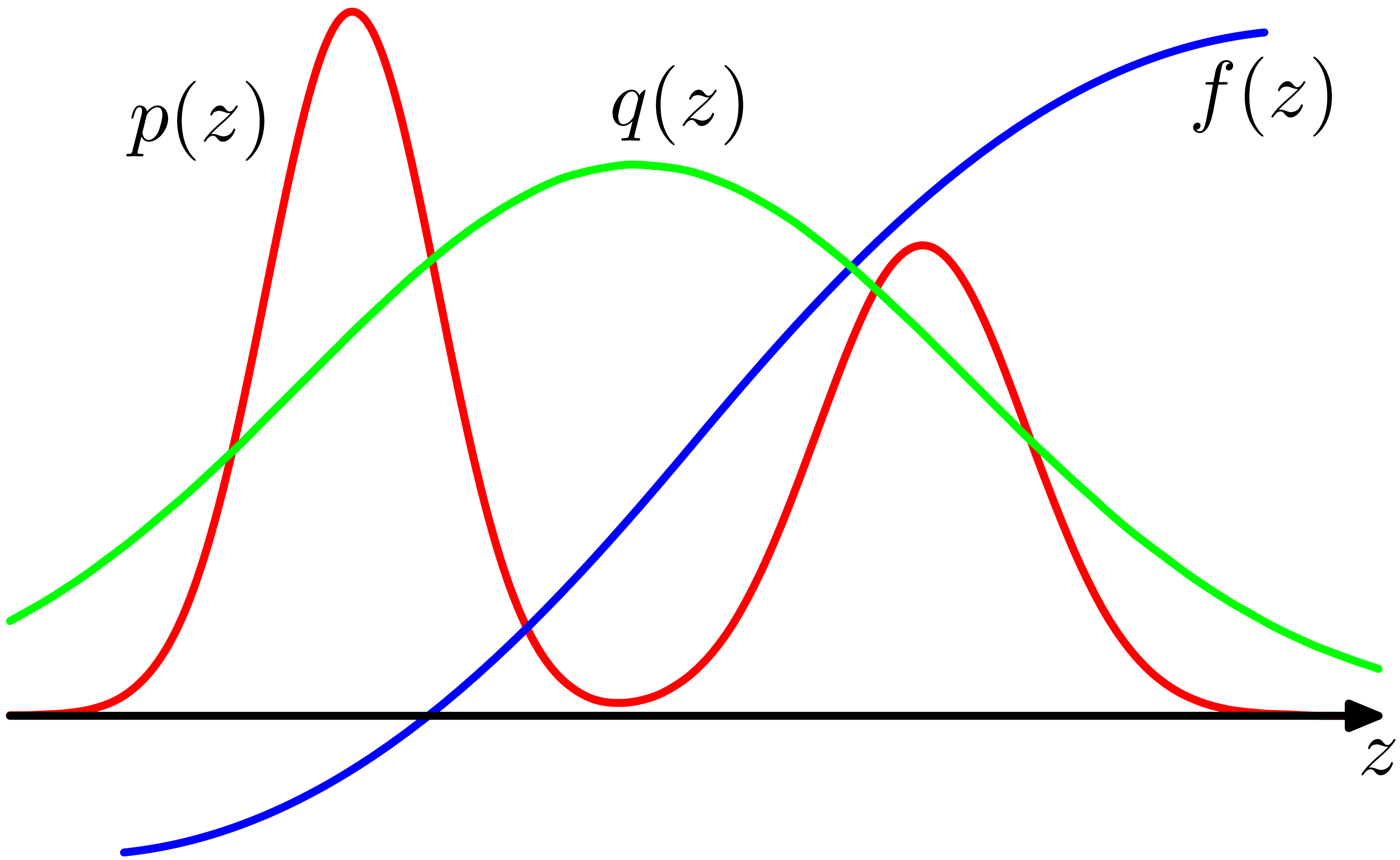
depth at (x,y) is P*



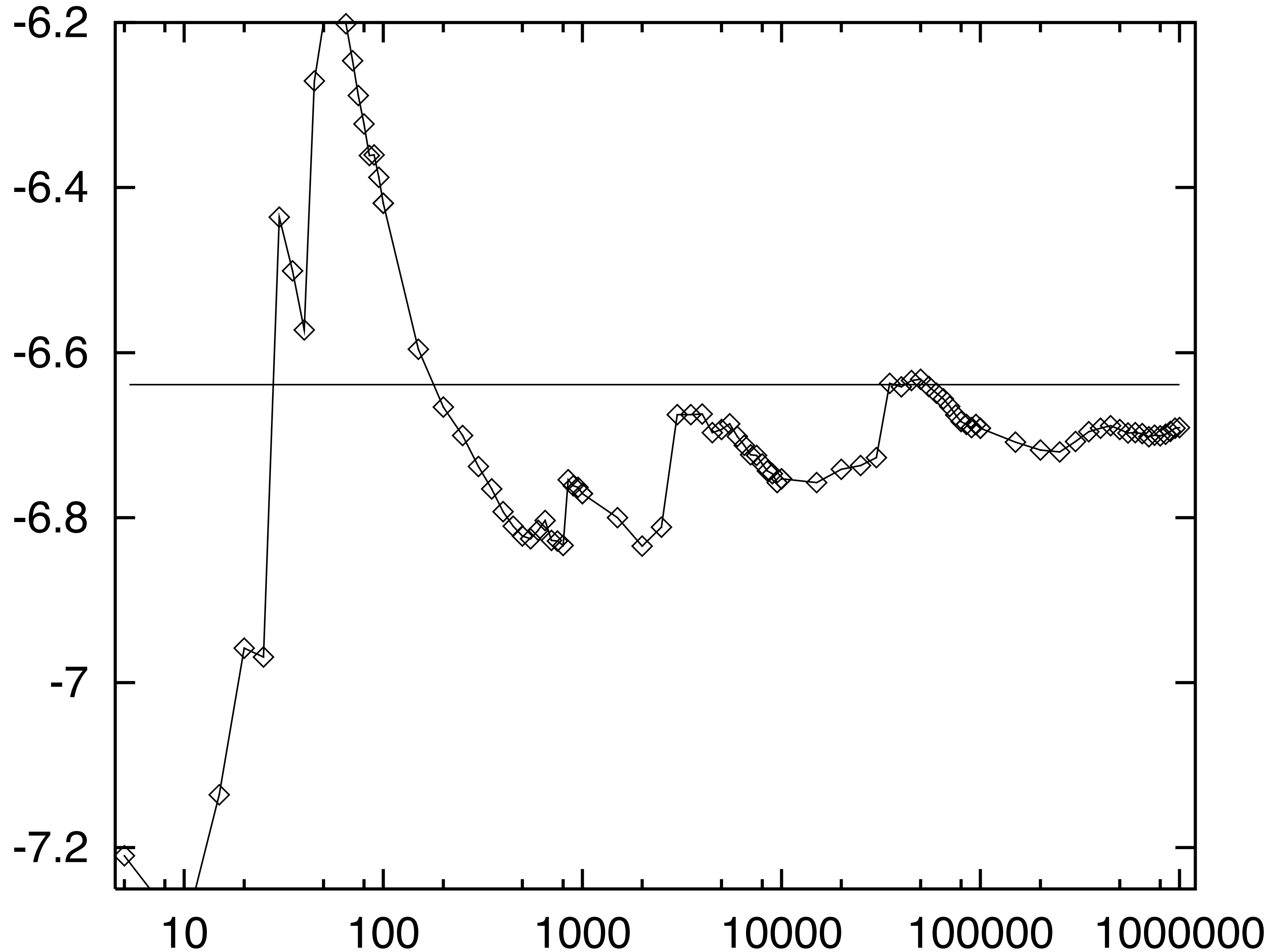
Task: draw random water samples to estimate plankton conc.

$$\Phi = \frac{1}{Z} \int P^*(x)\phi(x)dx$$

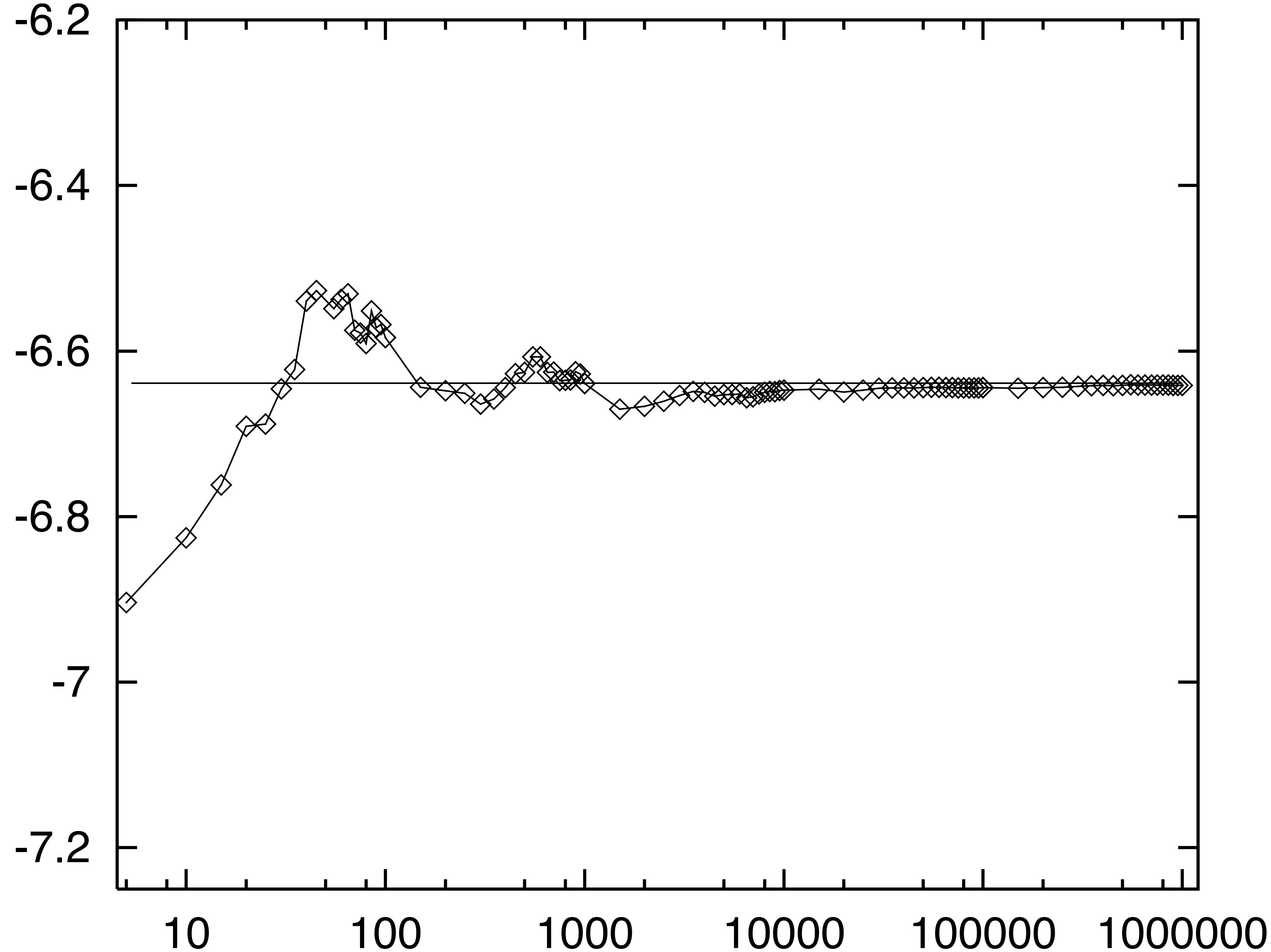
Go to any desired loc, measure P* and concentration (phi) there



Simpler proposal distribution q(z)



Gaussian proposal (straight line is true distr)



Cauchy proposal (straight line is true distr)

Metropolis-Hastings

Assume we can evaluate p*(z)

1. Tentative state z' generated from q(z'|z(t))

2. $$a = \frac{p^*(z')}{p^*(z^{(t)})} \frac{q(z^{(t)}; z')}{q(z'; z^{(t)})}$$

3. If a ≥ 1, accept new state **else** accept with prob. **a**

4. If accepted, z(t+1)=z' **else** z(t+1)=z(t)

For any positive q, i.e., q(z;z')> 0, as $t \to \infty$ prob. distribution of z(t) tends to p*(z)/Z

Gibbs Sampling

Assume: Sampling from conditional is "easy"; operate as (cyclic) coordinate descent

$$z_1^{(t+1)} \sim p(z_1 | z_2^{(t)}, z_3^{(t)}, \ldots, z_n^{(t)})$$

$$z_2^{(t+1)} \sim p(z_2 | z_1^{(t+1)}, z_3^{(t)}, \ldots, z_n^{(t)})$$

$$z_n^{(t+1)} \sim p(z_n | z_1^{(t+1)}, z_2^{(t+1)}, \ldots, z_{n-1}^{(t+1)})$$

Does it work? Why?

Markov Chains

Initial state: $p^{(0)}(z)$

Transition probability: $T(z'; z)$

Update: $$p^{(t+1)}(z') = \int T(z'; z) p^{(t)}(z) dz$$

Required properties:

$$p^*(z') = \int T(z'; z) p^*(z) dz \qquad \text{(Stationarity)}$$

$$p^{(t)}(z) \to p^*(z'), \quad \text{as } t \to \infty \qquad \text{(Ergodicity)}$$

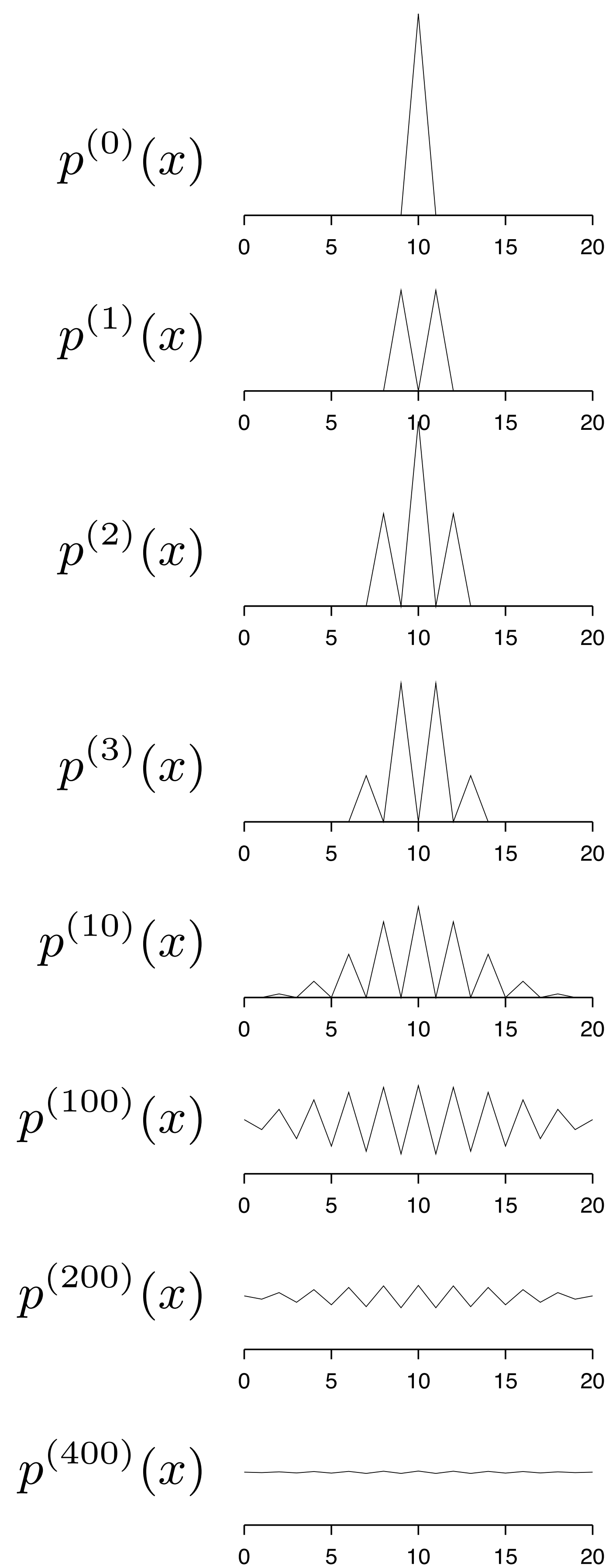
$$
\mathbf{A} = \begin{bmatrix}
\tfrac12 & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \cdot & \tfrac12 \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \tfrac12 & \tfrac12
\end{bmatrix}
$$

$$
P(x) = \begin{cases} 1/21 & x \in \{0, 1, 2, \ldots, 20\} \\ 0 & \text{otherwise.} \end{cases}
$$

$$P(x) = \begin{cases} 1/21 & x \in \{0, 1, 2, \ldots, 20\} \\ 0 & \text{otherwise.} \end{cases}$$

Proposal

$$Q(x'; x) = \begin{cases} 1/2 & x' = x \pm 1 \\ 0 & \text{otherwise.} \end{cases}$$



Credit: MacKay, 2005

What's the stationary distribution?

$$
\mathsf{A} = \begin{pmatrix} & & & 1 \\ & & \frac{1}{2} & \frac{1}{2} \\ & \iddots & & \vdots \\ \frac{1}{n} & \cdots & & \frac{1}{n} \end{pmatrix}
$$

Useful links

MCMC without the bs

Chapter 29, of David MacKay's book:
"Information Theory, Inference, and Learning Algorithms"

Andrieu, de Freitas, Doucet, Jordan.
"Introduction to MCMC for Machine Learning"

See viz. at: http://setosa.io/ev/markov-chains/