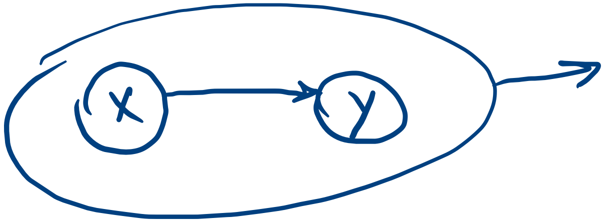
Identifiability



$$X = U_x \qquad U_y \perp U_x$$

$$Y = f(X, U_y)$$

$$Y = U_y \qquad U_x \perp U_y$$

$$X = g(Y, U_x)$$

$$D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$$

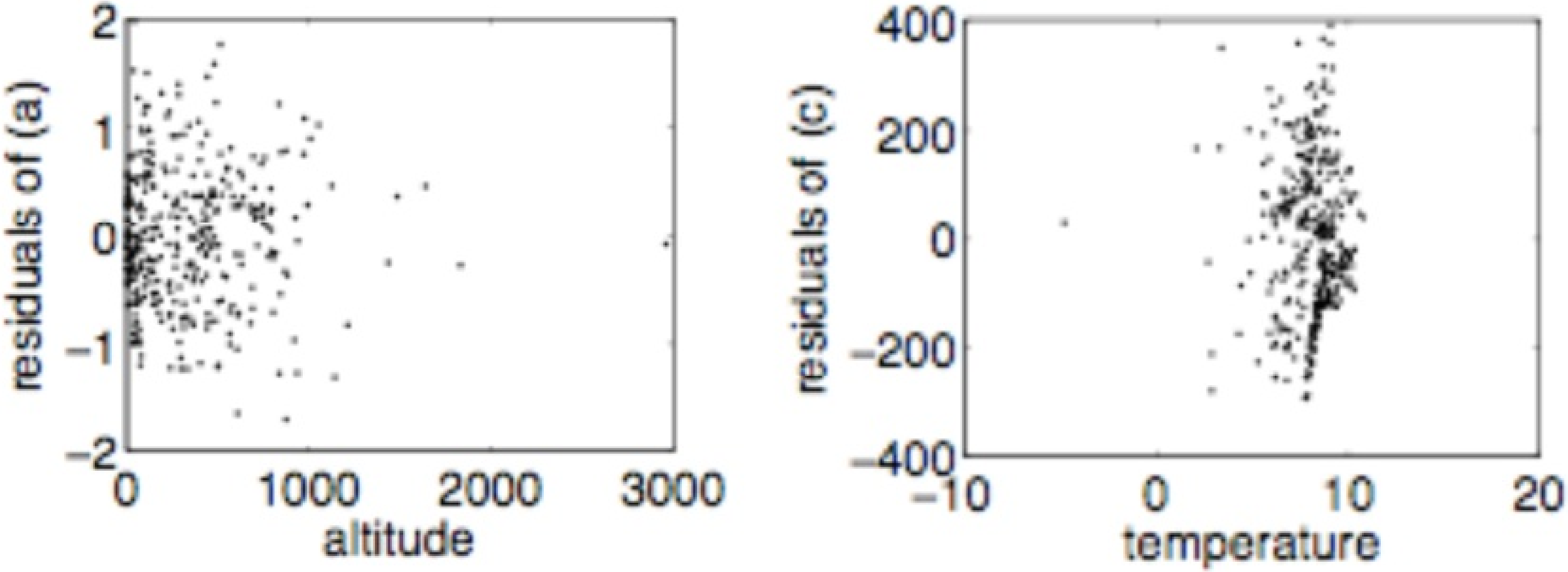
SCM: ANM

$$X = U_x$$

$$Y = f(X) + U_y \qquad U_x \perp U_y$$

$$Y = U_y$$

$$X = g(Y) + U_x \qquad U_x \perp U_y$$



Our independence tests detect strong dependence.
Hence the method prefers the correct direction

altitude → temperature

Approximate inference

- Approximate inference techniques
  - Deterministic approximation
    - Variational algorithms
  - Stochastic simulation / sampling methods

Exact

Approximate

Deterministic

VI

Stochastic

Sampling

$$P(y \mid x) = ?$$

Sampling-Based Estimation

$x \sim \underline{P(x)}$

$\{x_1, \ldots, x_n\}$

$E[x] = ?$

$E[x] \simeq \frac{1}{n} \sum_i x_i$

$$\boxed{E[f(x)] \simeq \frac{1}{n} \sum_{i=1}^{n} f(x_i)}$$

$\rightarrow \boxed{P(X = x_0)}$

MCMC

Monte Carlo methods

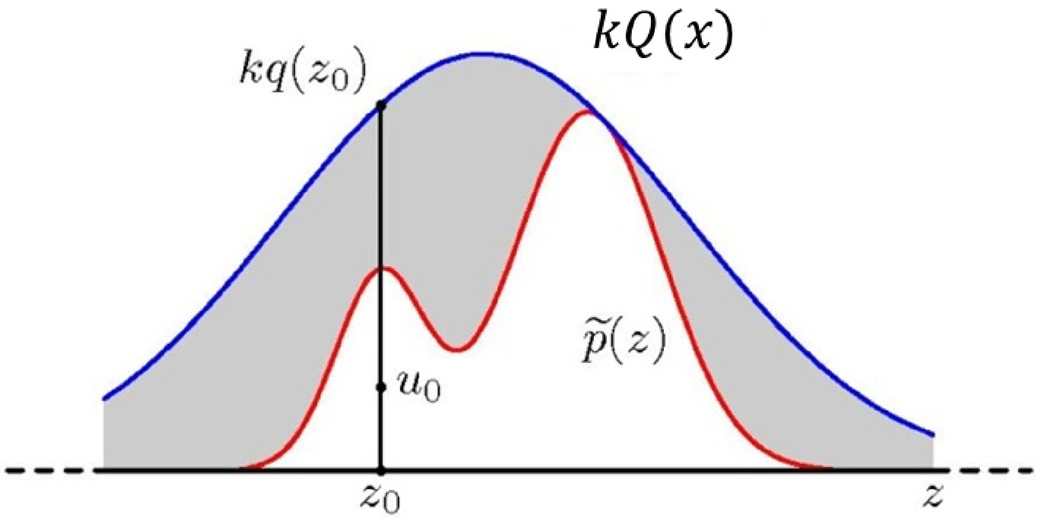
- Using a set of samples to find the answer of an inference query
  - expectations can be approximated using sample-based averages

Rejection Sampling

proposal distribution

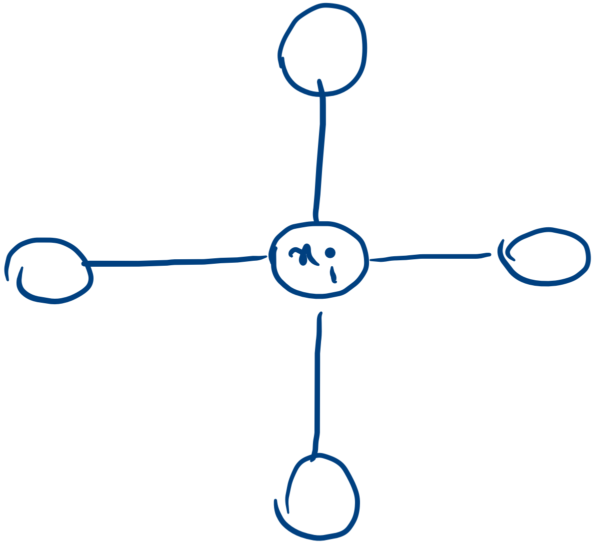
$p(x)$ $\qquad$ $Q(x)$

$$p(x) = \frac{\tilde{p}(x)}{Z}$$



$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{100} \end{bmatrix}$$

$\tilde{p}(x)$



$$\Phi(x_i, x_j) = e^{+\alpha x_i x_j}$$

$\uparrow \quad \downarrow$

$$p(x) = \frac{\tilde{p}(x)}{Z}$$

$$Z = ?$$

$$x \sim Q(x)$$

$$p(\text{Accept}) = \boxed{\frac{\tilde{p}(x)}{K\,Q(x)}} \leq 1$$

$$\boxed{K\,Q(x) \geq \tilde{p}(x)}$$

$$D = \{x_1, x_2, \ldots, x_n\}$$

$$q(x) = p(x)$$

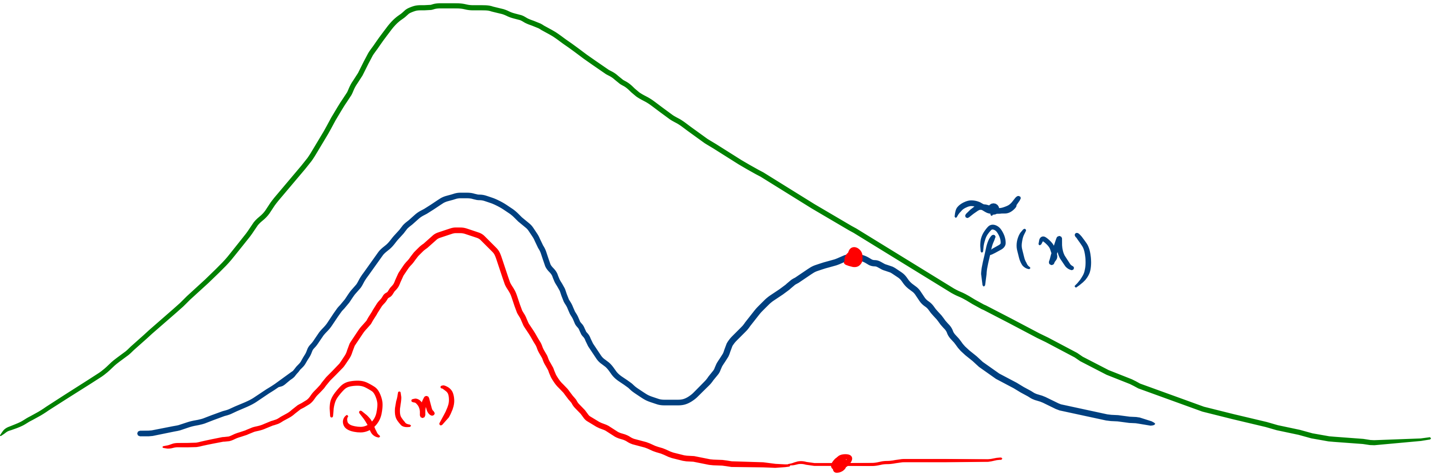
$$q(x) = \frac{Q(x)\,\frac{\tilde{p}(x)}{\cancel{K}\cancel{Q(x)}}}{\int \cancel{Q(x)}\,\frac{\tilde{p}(x)}{\cancel{K}\cancel{Q(x)}}\,dx} = \frac{\tilde{p}(x)}{\int \tilde{p}(x)\,dx} = \frac{\tilde{p}(x)}{Z} = p(x)$$

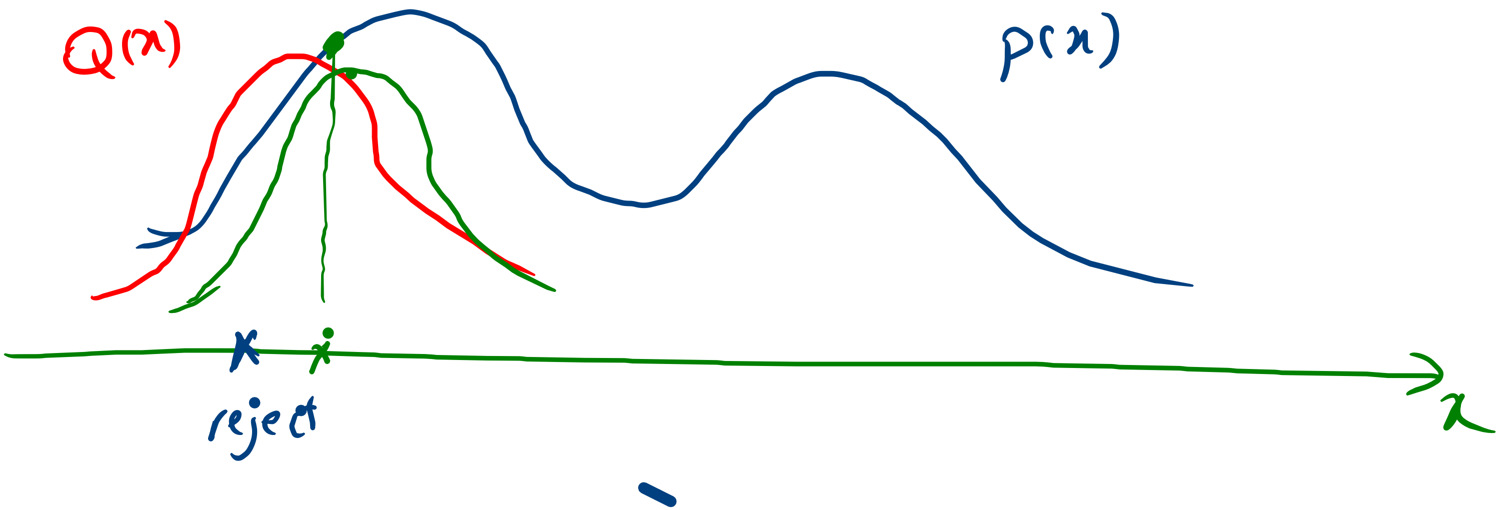
$$K\,Q(x) \geq \tilde{p}(x)$$

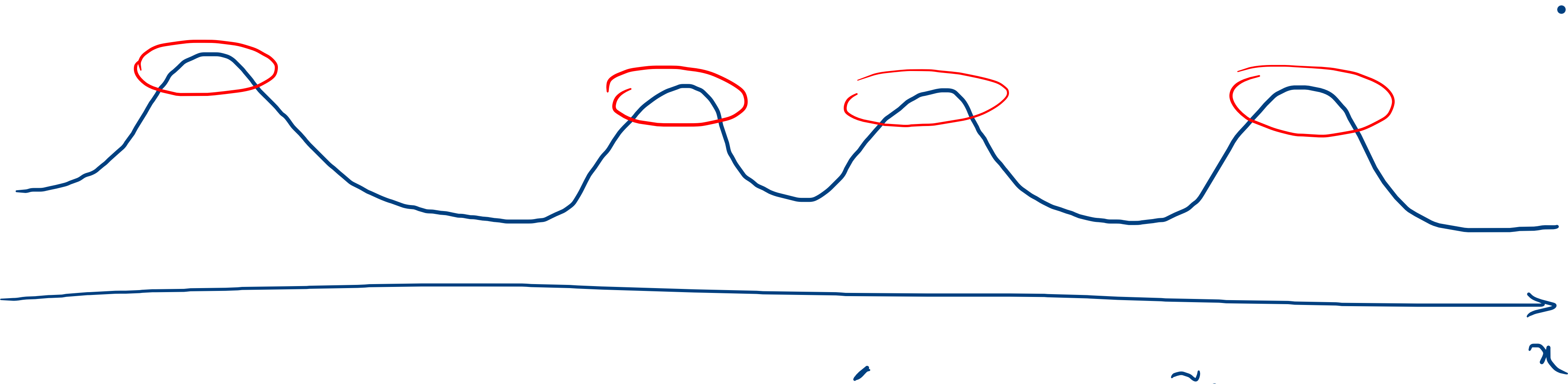
$$\frac{\tilde{p}(x)}{K\,Q(x)}$$

10000

MCMC







Z' $\quad$ $\tilde{p}(x)$

$f(x)$

$$Q(x) = \frac{\tilde{p}(x)}{Z'}$$

$$Q(x) = \frac{\tilde{p}(x)}{Z} = p(x)$$

MCMC → Markov chain Monte Carlo

- MCMC algorithms feature adaptive proposals
  - Instead of $Q(\boldsymbol{x}')$, they use $Q(\boldsymbol{x}'|\boldsymbol{x})$ where $\boldsymbol{x}'$ is the new state being sampled, and $\boldsymbol{x}$ is the previous sample

Gibbs sampling