Week summary

Monte Carlo methods to approximate expected values

Week summary

Monte Carlo methods to approximate expected values

How to sample from distribution known up to normalization constant?

Week summary

Monte Carlo methods to approximate expected values

How to sample from distribution known up to normalization constant?

Two MCMC approaches:

- Gibbs sampling – reducing multidimensional sampling to a sequence of 1d
- Metropolis Hastings – rejection sampling for Markov Chains (gives more freedom)

Monte Carlo vs Variational Inference

Monte Carlo

$$\mathbb{E}_{p(x)} f(x) \approx \frac{1}{M} \sum_{s=1}^{M} f(x_s)$$

$$x_s \sim p(x)$$

Unbiased estimate (larger M => better accuracy)

Monte Carlo vs Variational Inference

Monte Carlo

$$\mathbb{E}_{p(x)} f(x) \approx \frac{1}{M} \sum_{s=1}^{M} f(x_s)$$

$$x_s \sim p(x)$$

Unbiased estimate (larger M => better accuracy)

$$\mathbb{E}_{p(x)} \frac{1}{M} \sum_{s=1}^{M} f(x_s) = \mathbb{E}_{p(x)} f(x)$$

Monte Carlo vs Variational Inference

Monte Carlo

$$\mathbb{E}_{p(x)} f(x) \approx \frac{1}{M} \sum_{s=1}^{M} f(x_s)$$

$$x_s \sim p(x)$$

Unbiased estimate (larger M => better accuracy)

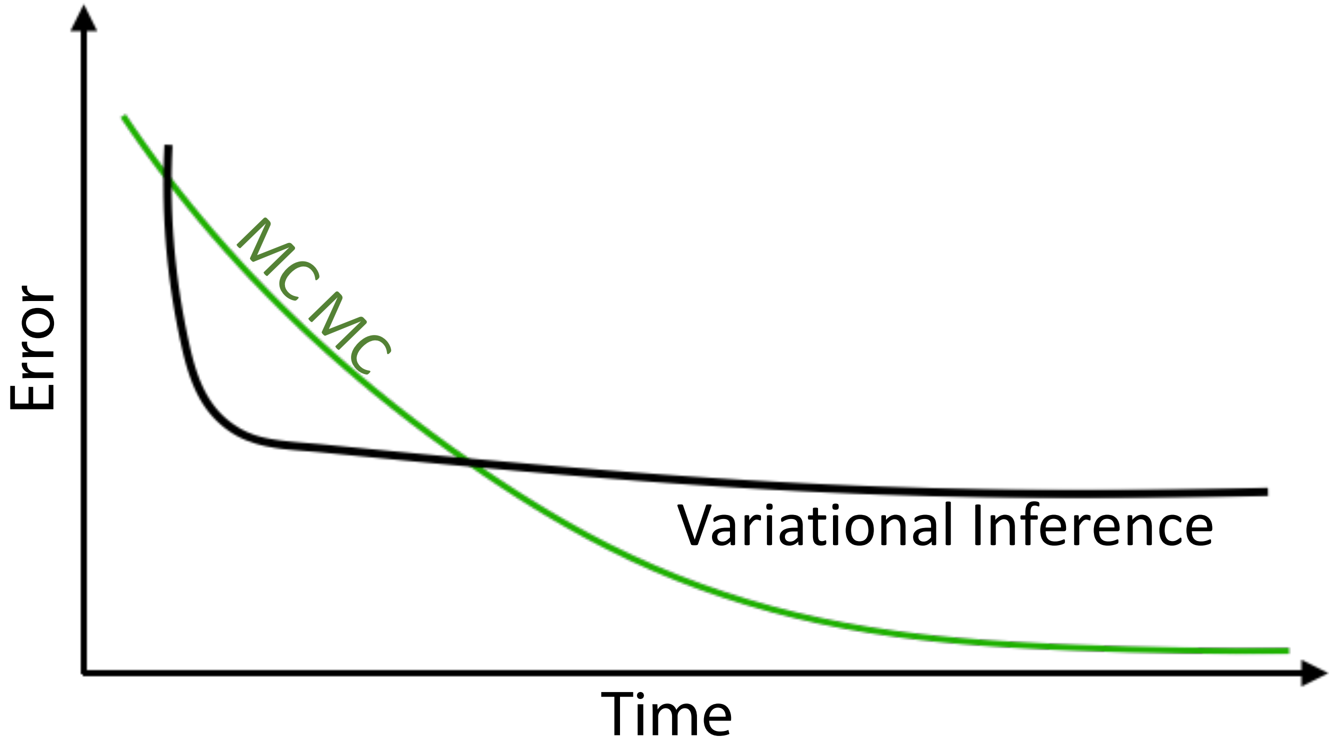
Variational Inference (week 3)

$$p(x) \approx q(x)$$

$$\mathbb{E}_{p(x)} f(x) \approx \mathbb{E}_{q(x)} f(x)$$

Monte Carlo vs Variational Inference

Schematic illustration



Methods

Best

- Full inference $p(T, \theta | X)$

Worst

Methods

Best

- Full inference $p(T, \theta | X)$
    - Mean field $q(T)q(\theta) \approx p(T, \theta | X)$

Worst

Methods

Best

- Full inference $p(T, \theta | X)$
  - Mean field $q(T)q(\theta) \approx p(T, \theta | X)$
  - MCMC $T_s, \Theta_s \sim p(T, \Theta \mid X)$

Worst

Methods

Best

- Full inference $p(T, \theta | X)$
    - Mean field $q(T)q(\theta) \approx p(T, \theta | X)$
    - MCMC $T_s, \Theta_s \sim p(T, \Theta \mid X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$

Worst

Methods

Best

- Full inference $p(T, \theta | X)$
  - Mean field $q(T)q(\theta) \approx p(T, \theta | X)$
  - MCMC $T_s, \Theta_s \sim p(T, \Theta \mid X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$
  - Variational EM $q_1(T_1) \ldots q_d(T_d),\ \theta = \theta_{\mathrm{MP}}$

Worst

Methods

Best

- Full inference $p(T, \theta | X)$
    - Mean field $q(T)q(\theta) \approx p(T, \theta | X)$
    - MCMC $T_s, \Theta_s \sim p(T, \Theta \mid X)$
- EM algorithm $q(T),\ \theta = \theta_{\mathrm{MP}}$
    - Variational EM $q_1(T_1) \ldots q_d(T_d),\ \theta = \theta_{\mathrm{MP}}$
    - MCMC EM $T_s \sim p(T \mid \Theta, X), \Theta = \Theta_{\mathrm{MP}}$

Worst

Summary of Markov Chain Monte Carlo

Pros

- Easy to implement
- Easy to parallelize
- Unbiased estimates (wait more => more accuracy)

Cons

- Usually slower than alternatives