Markov Chain Monte Carlo (MCMC)

- MCMC — silver bullet of probabilistic modeling
- Learn how exploit specifics of your problem to speed up MCMC
- Understand the limitations

Monte Carlo

Monte Carlo methods were invented in the Manhattan project.

Monte Carlo

Estimate expected values by sampling

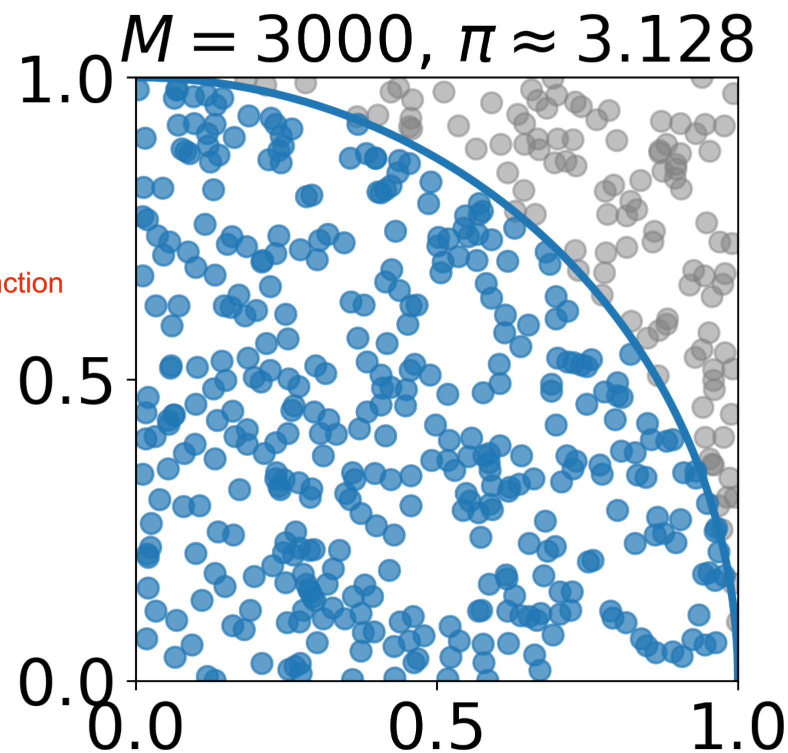
We can use MC to calculate value of pi!

$$\frac{\pi}{4} = \mathbb{E}\left[x^2 + y^2 \leq 1\right]$$

vvv notice this is the indicator function

$$\approx \frac{1}{M}\sum_{s=1}^{M}[x_s^2 + y_s^2 \leq 1]$$

$$x_s, y_s \sim \mathcal{U}(0,1)$$



Monte Carlo

Estimate expected values by sampling

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

$$x_s \sim p(x)$$

almost as if you 'discretize' what would normally be an integral - as we do an expectation of a continuous variable. To account for the probabiliy of each point x in the support, we choose x_s probabilisticly (following pdf of p(x)).

Monte Carlo

Why do we need to estimate expected values?

- Full Bayesian inference (see Week 1)

$$p(y \mid x, Y_{\text{train}}, X_{\text{train}})$$
$$= \int p(y \mid x, w) p(w \mid Y_{\text{train}}, X_{\text{train}}) dw$$
$$= \mathbb{E}_{p(w \mid Y_{\text{train}}, X_{\text{train}})} p(y \mid x, w)$$

Aha. This can be used by using monte carlo!

$$p(w \mid Y_{\text{train}}, X_{\text{train}}) = \frac{p(Y_{\text{train}} \mid X_{\text{train}}, w) p(w)}{Z}$$

Monte Carlo

Why do we need to estimate expected values?

- Full Bayesian inference (see Week 1)

- M-step of EM-algorithm (see Week 2)

$$\max_{\theta} . \mathbb{E}_q \log p(X, T \mid \theta)$$