Advanced Monte Carlo Methods
Information theory, pattern recognition, and neural networks

- Source coding (Data compression)
- Noisy-channel coding
- Inference + probabilistic methods
  - 9-10 Inference
  - 11 Clustering
  - 12 Monte Carlo methods
  - 13 Advanced Monte Carlo methods
  - 14 Variational methods

www.inference.phy.cam.ac.uk/itprnn/
www.inference.phy.cam.ac.uk/itila/
Overview

- Data compression
  - Chs 1-6, 8-10, 14
- Noisy-channel coding
  - Chs 20, 22
- Inference, data modelling
  - Clustering, pattern recognition
  - Chs 20, 22
- Probability toolbox
  - Monte Carlo methods
    - Ch 29, 30, 32
  - Variational methods
    - Ch 33
- Neural networks
  - Chs 38, 39, (& perhaps 41, 44), 42
- State-of-the-art error-correcting codes

Recommended exercises

- 29.14, 33.5, 33.7, 27.1, 22.11, 39.4, 39.5
- Handouts 2, 3 (on website)
- The 5 cards magic trick (15.6)
  - 8♦, 2♥, 10♣, 9♦

Additional reading

- Laplace's method (Ch 27)
- Ising models (Ch 31)
The course

www.inference.phy.cam.ac.uk/itprnn/

The book

www.inference.phy.cam.ac.uk/itila/
Monte Carlo methods

Simple Monte Carlo methods
- Importance sampling
- Rejection sampling

Markov-chain Monte Carlo methods
- Metropolis method
- Gibbs sampling
- Slice sampling

Reducing random-walk behaviour
- Hamiltonian Monte Carlo
- Overrelaxation

Exact sampling
$$P(x) = P^*(x) = \frac{e^{-E(x)}}{Z}$$

Can evaluate not $E$
Problem 1

samples $x^{(r)} \sim P$

\[
\langle \phi \rangle_P = \sum_x P(x) \phi(x)
\]
samples $X^{(n)} \sim P$

$\langle \phi \rangle_P = \sum_x P(x) \phi(x)$
\[ P(x) = e^{-\frac{\mathbb{E}(x, \mathbb{J})}{\mathbb{Z}(\mathbb{J})}} \]

\[ \mathbb{E}(x) = -\frac{1}{2} \sum_{m \neq n} \]
\[ E(x) = \frac{1}{2} \sum_{m \neq n} J_{mn} x_m x_n - \sum_n h_n x_n \]
\[ x \in \mathbb{Z}^2 - (1, 1)^N \]
\[ e^{-\beta E(x; I)} \]
\[ \frac{1}{Z(\beta, I)} \]
\[ E(x) = -\frac{1}{2} \sum_{m+n} J_{mn} x_m x_n - \sum_n h_n x_n \]
\[ \langle E \rangle \quad \langle E^2 \rangle \quad \langle m \rangle \quad \langle m^2 \rangle \]
\[ S = \langle \log \frac{1}{P(x)} \rangle_p \]
\[ \beta = \frac{1}{T} \]
\[ x \in \mathbb{Z}_{-1, 1} \cap \mathbb{N} \]
Monte Carlo methods

- Simple Monte Carlo methods
  - Importance sampling
  - Rejection sampling

- Markov-chain Monte Carlo methods
  - Metropolis method
  - Gibbs sampling
  - Slice sampling

- Reducing random-walk behaviour
  - Hamiltonian Monte Carlo
  - Overrelaxation

- Exact sampling

itp/exact/rc  RUNME
Problems with standard Monte Carlo methods

- Random walk behaviour
- Sensitivity to step size
- When to stop

Efficient methods
Slice sampling
Exact sampling
Problems with standard Monte Carlo methods

- Random walk behaviour
- Sensitivity to step size
- When to stop

Efficient methods
Slice sampling
Exact sampling
$$P^*(x) = \begin{cases} 1 & x \in \{0, 1, 2, 3, \ldots, 20\} \\ 0 & \text{otherwise} \end{cases}$$

\[ Q(x', x) = \begin{cases} 2^{1/2} & x' = x - 1 \\ 2^{1/2} & x' = x + 1 \end{cases} \]
25:lewis:/home/mackay/itp/metrop> ./demo.p

20 iterations
Q1: roughly how long until this chain generates a "good" sample from P?
Q2: how long until we hit a wall
Q3: how long until we've hit both walls?
Minimum # of steps to reach a wall = \( \left( \frac{\pi}{3} \right) = 10 \)

Min # of steps to visit both walls = 31 \( \left( \frac{\pi}{3} \right) \)
Distance after time $T$

$$\Delta x = \sum_{t=1}^{T} s_t$$

Neglect walls

$$\text{Var}(\Delta x) = \sum_{t=1}^{T} \langle s_t^2 \rangle = T \left( \frac{L^2}{3} \right)^2$$

So

$$\langle \Delta x^2 \rangle = L^2 \quad \text{need} \quad T = \left( \frac{L}{3} \right)^2$$
20 iterations
\[ X \in \{ \text{integers} \} \]

\[ X \in \{ 0, 1, 2, 3, \ldots, 20 \} \]

otherwise

\[ Q(x; x) = \sqrt{\frac{1}{2}} \]

\[ x' = x \]
dimensions

$T \approx \left( \frac{L}{\delta} \right)^2$ steps to get a fresh independent sample
Steps to get a fresh independent sample.
\[ \text{if } \exists \rightarrow l \]

\[ \text{Prob of accepted move} = \frac{l}{3} \left( \frac{1}{l} \right) \]
Steps to get a fresh independent sample.

Bad news

Feasible sensible values

E > l
optimal $E = l$

& resulting optimal acceptance probability $\geq 1/2$
Efficient Monte Carlo methods

Hamiltonian Monte Carlo

Overrelaxation
HMC (or Hybrid)

\[ P(x,p) = P(x) \times e^{-\frac{1}{2} \frac{p^2}{m} - E(x) + \frac{1}{2} p^2/m} \]

\[ = \frac{e^{-E(x) + \frac{1}{2} p^2/m}}{Z} \]
1. Simulate Newton's law (approximately) (for a time)
2. Accept/reject based on change in total energy
   \[ \frac{1}{2} p^2 + \frac{1}{2} m \]
3. Randomize the momentum \( \vec{p} \sim \vec{e} \)
Adler’s overrelaxation for conditional distributions that are Gaussians.
Adler's Overrelaxation
Adler's Overrelaxation
Gibbs sampling - $x_1^2$
Adler's Overrelaxation - \( x1^2 \)
Adler's over-relaxation
for conditional distributions that are Gaussians.

Ordered over-relaxation
Radford Neal
Draw $K$ times

$K \approx 20$

BUGS
Robust Monte Carlo methods

Slice sampling
Slice sampling

Radford Neal

John Skilling
$P^*(x)$
Self-terminating Monte Carlo methods

Exact sampling
How long? \( \infty \) by long.
How long?

$\infty$ by

Random number

$-\infty$ $t$ $\infty$

$X_t$ $t=0$

HTHHTTTHHH

1782
Exact sampling

Hexagon

Height: 313
Exact sampling

Hexagon

Height: 78

Height: 407
Exact sampling
Exact sampling for the Ising model
In the Delft computational physics group, Ising simulations are controlled by the red bar, the spins prefer to be parallel.
controlled by the red bar, the spins prefer to be parallel.
Introduction to the Ising model

Figure 31.4. Detail of Monte Carlo simulations of rectangular Ising models with $J = 1$. (a) Mean energy and fluctuations in energy as a function of temperature. (b) Fluctuations in energy (standard deviation). (c) Mean square magnetization. (d) Heat capacity.
Exact sampling - Ising model at Tc
Exact sampling - Ising model at $T_c$
Exact sampling - Ising model at $T_c$
Exact sampling - Ising model at Tc
Problems with standard Monte Carlo methods

- Random walk behaviour
- Sensitivity to step size
- When to stop

Efficient methods
- Slice sampling
- Exact sampling

Not revealing the normalizing constant
- Thermodynamic integration
- Reversible-jump Markov chain Monte Carlo
- The acceptance ratio method
- Umbrella sampling
- Simulated tempering
- Tempered transitions (Radford Neal)
- Annealed importance sampling (Radford Neal)
- Linked importance sampling (Radford Neal)
Skills and Neal Skilling