Monte Carlo basics and Motivation

Rejection sampling

Importance sampling

Next time: Markov chain Monte Carlo

Enrico Fermi (1901–1954) took great delight in astonishing his colleagues with his remarkably accurate predictions of experimental results... he revealed that his "guesses" were really derived from the statistical sampling techniques that he used to calculate with whenever insomnia struck in the wee morning hours!

The beginning of the Monte Carlo method, N. Metropolis

Linear Regression: Prior

$P(\theta | \text{Data}) \propto P(\text{Data} | \theta) P(\theta)$

Posterior much more compact than prior.

Quiz

Given a (wrong) linear assumption, which explanations are typical of the posterior distribution?

A

B

C

D All of the above

E None of the above

Z Not sure

'Underfitting'

Microsoft Kinect (Shotton et al., 2011)

Eyeball modelling assumptions

Generate training data

Random forest applied to fantasies
The need for integrals

\[ p(y, |x, D) = \int d\theta \ p(y, \theta | x, D) = \int d\theta \ p(y | \theta, D) \ p(\theta | x, D) \]

A statistical problem

What is the average height of the people in this room? Method: measure our heights, add them up and divide by \( N \).

\[ E_{s} \equiv \frac{1}{S} \sum_{i=1}^{S} f(p) \text{, "intractable"?} \]

\[ = \frac{1}{3} \sum_{i=1}^{S} \int d\theta p(y | \theta, D) \int d\theta p(\theta | x, D) \]

Surveying works for large and notionally infinite populations.

Properties of Monte Carlo

Estimator: \( \int_{a}^{b} f(x) P(x) \ dx \approx f \equiv \frac{1}{S} \sum_{i=1}^{S} f(x^{(i)}) \), \( x^{(i)} \sim P(x) \)

Estimator is unbiased:

\[ E_{\mu(x^{(i)})} \equiv \frac{1}{S} \sum_{i=1}^{S} E_{\mu(x^{(i)})}[f(x)] = E_{\mu(x)}[f(x)] \]

Variance shrinks \( \propto 1/N \):

\[ \text{var}_{\mu(x^{(i)})}[f(x)] = \frac{1}{S} \sum_{i=1}^{S} \text{var}_{\mu(x)}[f(x)] = \frac{\text{var}_{\mu(x)}[f(x)]}{S} \]

"Error bars" shrink like \( \sqrt{S} \)

Aside: don’t always sample!

"Monte Carlo is an extremely bad method; it should be used only when all alternative methods are worse."

— Alan Sokal, 1996

A dumb approximation of \( \pi \)

\[ P(x, y) = \begin{cases} 1 & 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise} \end{cases} \]

\[ \pi = 4 \int \int \left( (x^2 + y^2) < 1 \right) P(x, y) \ dx \ dy \]

Simple Monte Carlo

In general:

\[ \int_{a}^{b} f(x) P(x) \ dx \approx \frac{1}{S} \sum_{i=1}^{S} f(x^{(i)}) \text{, } x^{(i)} \sim P(x) \]

Example: making predictions

\[ P(x|D) = \int P(x|\theta) p(\theta|D) \ d\theta \approx \frac{1}{S} \sum_{i=1}^{S} P(x|\theta^{(i)}) \text{, } \theta^{(i)} \sim p(\theta|D) \]

Many other integrals appear throughout statistical machine learning

Simple Monte Carlo

Want to sample to approximate expectations:

\[ \int f(x) P(x) \ dx \approx \frac{1}{S} \sum_{i=1}^{S} f(x^{(i)}) \text{, } x^{(i)} \sim P(x) \]

Alternatives to Monte Carlo

There are other methods of numerical integration!

Example: (nice) 1D integrals are easy:

\[ \text{octave:1> 4 * quadl}(\theta(x) \text{ sqrt}(1-x.^2), 0, 1, \text{tolerance}) \]

Gives \( \pi \) to 6 dp’s in 108 evaluations, machine precision in 2598.

(NB Matlab’s \text{quadl} fails at tolerance=0, but Octave works.)

In higher dimensions sometimes deterministic approximations work: Variational Bayes, Laplace, … (covered later)

Sampling simple distributions

Use library routines for univariate distributions (and some other special cases)

This book (free online) explains how some of them work

http://cg.scs.carleton.ca/~luc/rnbookindex.html
Sampling discrete values

<table>
<thead>
<tr>
<th>a</th>
<th>p=0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>b</td>
<td>p=0.5</td>
</tr>
<tr>
<td>c</td>
<td>p=0.2</td>
</tr>
</tbody>
</table>

\[ u \sim \text{Uniform}[0,1] \]
\[ u = 0.4 \Rightarrow x = b \]

Sampling from densities

How to convert samples from a Uniform[0,1] generator:

\[ p(y) \]
\[ h(y) = \int_{-\infty}^{y} p(y') \, dy' \]
\[ u \sim \text{Uniform}[0,1] \]
\[ \text{Sample, } y(u) = h^{-1}(u) \]

Although we can’t always compute and invert \( h(y) \)

Rejection sampling

Sampling from \( \pi(x) \) using tractable \( q(x) \):

\[ q(x) \geq \pi^*(x), \forall x \]
\[ \pi^*(x) = c \cdot \pi(x) \]

Importance sampling

Rewrite integral: expectation under simple distribution \( Q \):

\[
\int f(x) P(x) \, dx = \int f(x) \frac{P(x)}{Q(x)} \, Q(x) \, dx, \\
\approx \frac{1}{3} \sum_{s=1}^{S} f(x^{(s)}) \frac{P(x^{(s)})}{Q(x^{(s)})}, \quad x^{(s)} \sim Q(x)
\]

Simple Monte Carlo applied to any integral. Unbiased and independent of dimension?

Application to large problems

Approximations scale badly with dimensionality

Example: \[ P(x) = \mathcal{N}(0, 1), \quad Q(x) = \mathcal{N}(0, \sigma^2) \]

Rejection sampling:
Requires \( \sigma \geq 1 \). Fraction of proposals accepted = \( \sigma^{-D} \)

Importance sampling:
\[
\text{Var}[P(x)/Q(x)] = \left( \frac{\sigma^2}{\sigma^2 + \sigma^2} \right)^{D/2} - 1
\]
Infinite / undefined variance if \( \sigma \leq 1/\sqrt{2} \)

Summary so far

- Monte Carlo
  approximate expectations with a sample average
- Rejection sampling
  draw samples from complex distributions
- Importance sampling
  apply Monte Carlo to ‘any’ sum/integral

Next: High dimensional problems: MCMC