

Lecture 11, Tuesday w6, 2014-10-21

I don't like putting lots of text on slides. These auxiliary notes give an *overview* of what's covered in each lecture, with pointers to more detail.

Monte Carlo

Monte Carlo is jargon for computational methods based on statistical sampling.

Motivation 1: Drawing samples from a simulation, or a probabilistic model, can give insight into the properties of a model. Sometimes simulations are a source of data for machine learning methods. I gave an example from Microsoft. High-energy physics is an example field where modelling simulations is common.

Motivation 2: Samples from a distribution can easily be used to approximate expectations under that distribution. When the expectation of interest is a high-dimensional integral, Monte Carlo is sometimes the only good approach.

Distributions that appear in physics and machine learning are often of the form $p(x) = p^*(x)/Z$, where we can evaluate $p^*(x)$ at any setting x , but we can't feasibly compute the normalizing constant Z . For example, the posterior over parameters of a complicated model $p(\theta|D) \propto p(D|\theta)p(\theta)$ has this form.

Rejection Sampling

One of the core ideas used to sample from many one-dimensional (or very low-dimensional) distributions on a computer. If any of the details are unclear from the lecture sketch, work through the short description of how it works in any of the three readings below.

We don't need Z , but we do need to be able to form an upper-bound of $p^*(x)$. (You should be able to explain why.)

Importance Sampling

Change a sum or integral to be an expectation under a convenient distribution $q(x)$, by dividing and multiplying by $q(x)$. Full but short descriptions are available in the three sources listed below.

There's a version where we don't need Z , but our estimator becomes biased. However, we don't need to upper-bound $p^*(x)$, and we don't reject any samples.

It's easy to accidentally construct an estimator with high variance using importance sampling, if $q(x)$ is small where the integrand is large. The MacKay reading has the best treatment of this issue. In these cases using the same $q(x)$ in rejection sampling (if possible) would lead to a lot of rejections.

High dimensions

Monte Carlo can estimate expectations under high-dimensional distributions, where other numerical techniques don't work well. However, typical samples from the distribution need to be representative of the integral. The importance sampling trick doesn't work well in high-dimensions, as it's hard to get a convenient distribution $q(x)$ to sample in the right places. Rejection sampling also doesn't work in high dimensions. So how do we sample from high-dimensional distributions of the form $p(x) = p^*(x)/Z$? An answer is "Markov chain Monte Carlo" (MCMC), which we'll cover in the next lecture.

Recommended reading

Part of the introduction to my PhD thesis pp20–24 is a terse summary of a large part of this lecture: <http://homepages.inf.ed.ac.uk/imurray2/pub/07thesis/>

MacKay's textbook Chapter 29 pp357–365 has more than enough detail for our purposes. You may need to skip over parts about the Ising model. Free online: <http://www.inference.phy.cam.ac.uk/mackay/itila/book.html>

Alternatively read Murphy Chapter 23 up to and including 23.4.2, pp846–822. (You can safely skip the Box-Muller method and adaptive rejection sampling.)

If you work properly through either MacKay or Murphy, you'll be in good shape.

Non-examinable extras

For interest: the term *Monte Carlo* is sometimes reserved for methods that give random wrong answers, that get more accurate the longer you run them. In contrast, *Las Vegas* algorithms give the right answer, but take a random amount of time to do it. Both terms are references to cities famous for gambling.

The earlier pages of my thesis pp14–19 give a high-level overview of high-dimensional distributions and summations that can appear in machine learning. Those doing PMR will learn about graphical models.

References alluded to in the slides:

- Beginnings of Monte Carlo: <http://lib-www.lanl.gov/la-pubs/00326866.pdf>
- The figure containing the motion tracking synthetic data was taken from <http://research.microsoft.com/apps/pubs/default.aspx?id=145347>
- Sokal's quote, which I've taken out of context, is from his notes on Monte Carlo methods, where he gives the more complete view I alluded to in class: <http://www.stat.unc.edu/faculty/cji/Sokal.pdf>

My thesis contains more references.