Monte Carlo 8/6/2022

Today: Introduction, some notation

Our goal is to compute averages of the form

$$\pi[f] = \int f(x)\, \pi(dx) \quad \text{where}$$

π is some probability distribution

x could live in a discrete space or it could live in a continuous space but we will mostly focus on continuous spaces.

In that case I may also write

$$\pi[f] = \int f(x)\, \pi(x)\, dx$$

i.e. I'm overloading the symbol π as both a distribution and a density why do we want to compute $\pi[f]$?

E.g. from statistics we may have some prior belief about the distribution of x characterized by a (usually simple) density $p(x)$

Then we make an observation Y conditioned on $X = x$

i.e. $Y \sim p(y|x)$

The distribution of X given $Y = y$ is then

$$\pi(x|y) = \frac{p(y|x)\,p(x)}{\underbrace{\int p(y|x)p(x)\,dx}_{\text{does not depend on } x}}$$

Note that 1) we only know π up to an unknown constant multiple

2) Even if we can generate a sample from $p(x)$ we cannot (usually) generate one from $\pi(x|y)$

Why not use deterministic quadrature to compute $\pi[f]$?

You should if you can!

But you can't, e.g. if $x \in \mathbb{R}^d$ for d more than a few.

Monte Carlo for d in the billions is routine.

<u>The simplest Monte Carlo estimator</u>

Suppose we <u>can</u> draw X from π <u>independently</u>.

$X^{(0)}, X^{(1)}, \ldots$ i.i.d. from π

Let $$\bar{f}_N = \frac{1}{N} \sum_{k=0}^{N-1} f(X^{(k)})$$

first notice that $E[\bar{f}_N] = \pi[f]$

i.e. $\bar{f}_N$ is an unbiased estimator of $\pi[f]$

We hope that $\bar{f}_N \to \pi[f]$

$\to$ can mean diffeent things:

At least we want that $\forall\ \delta > 0$

$$\lim_{N\to\infty} P(|\bar{f}_N - \pi[f]| > \delta) = 0$$

(convergence in probability)

We often expect that

$$\lim_{N\to\infty} E[\,|\bar{f}_N - \pi[f]|^2] = 0$$

(convergence in mean squared error)

Or $\forall\ \varepsilon$ $\exists$ a constant $\gamma(\varepsilon) > 0$ so that

(LP or concentration)

$$\lim_{N\to\infty} \frac{1}{N} \log P(|\bar{f}_N - \pi[f]| > \varepsilon) = -\gamma(\varepsilon)$$

In most cases almost sure convergence of $\bar{f}_N$ to $\pi[f]$ isn't more valuable than convergence in probability

back to $\bar{f}_N = \frac{1}{N} \sum_{k=1}^{N} f(X^{(k)})$, $X^{(1)}, \dots, X^{(N)}$ i.i.d.

What i var of $\bar{f}_N$?

$$\mathrm{Var}(\bar{f}_N) = \frac{1}{N^2} E\Big[\sum_{k=0}^{N-1} (f(X^{(k)}) - \pi[f])^2 \Big]$$

$$+ \frac{2}{N^2} \sum_{k<l} \mathrm{cov}\big(f(X^{(k)}), f(X^{(l)})\big)\Big)$$

if $\mathrm{Var}_\pi f = \mathrm{Var}(f(X)) = \sigma^2$ then

$$\mathrm{Var}(\bar{f}_N) = \frac{\sigma^2}{N}$$ ← no explicit dimension dependence, i.e. if $\|f\|_\infty \leq 1$ then $\sigma^2 \leq 1$

The second term vanishes because the $X^{(k)}$ are independent. But in most practical applications the $X^{(k)}$ will not be independent and the second term could be very large (even dimension dependent)

Chapter 2: Exact sampling techniques

In most cases we cannot generate X from π exactly in a finite number of operations.

We will cover a few important exceptions to this rule.

We assume that we can generate i.i.d. uniform $(0,1)$ R.V.s.

Inversion: suppose $X \sim \pi$ and $F(x) = P(X \leq x)$

Let $U \sim$ uniform $(0,1)$ and $Y = F^{-1}(U)$

(if F is not invertible we can use a pseudo inverse, see notes)

Then $P(Y \leq y) = P(F^{-1}(U) \leq y)$

$= P(U \leq F(y)) = F(y)$

example: (exponential R.V.)

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F^{-1}(u) = -\frac{1}{\lambda} \log(1-u)$$

So if $U \sim \text{uniform}(0,1)$ then

$$Y = -\frac{1}{\lambda} \log(1-U) \sim \mathcal{E}xp(\lambda)$$

Transformations:

Suppose we can sample $Y \sim \tilde{\pi}$
(e.g. $\tilde{\pi} = \text{uniform}(0,1)$)

But we want to sample from π.

Can we choose a map φ so that
$X = \varphi(Y) \sim \pi$?

Let's assume that φ is invertible.

Note that for any test fn f,

$$E[f(X)] = \int f(\varphi(y))\,\tilde{\pi}(y)\,dy$$

$$= \int f(x) \frac{\tilde{\pi}(\varphi^{-1}(x))}{|D\varphi(\varphi^{-1}(x))|} dx$$

$\leftarrow$ |det|

So the density of X must be

$$\frac{\tilde{\pi}(\varphi^{-1}(x))}{|D\varphi(\varphi^{-1}(x))|} \qquad \left((D\varphi)_{ij} = \frac{\partial \varphi_i}{\partial y_j} \right)$$

can we choose φ so that

$$\frac{\tilde{\pi}(\varphi^{-1}(x))}{|D\varphi(\varphi^{-1}(x))|} = \pi(x) \ ?$$

example: $U_1, U_2 \sim$ uniform $(0,1)$ and independent.

$$\phi_1(u_1, u_2) = \sqrt{-2\log u_1}\cos(2\pi u_2)$$

$$\phi_2(u_1, u_2) = \sqrt{-2\log u_1}\sin(2\pi u_2)$$

$$D\phi = \begin{bmatrix} \dfrac{-\cos(2\pi u_2)}{\sqrt{-2\log u_1}\, u_1} & -2\pi\sqrt{-2\log u_1}\sin(2\pi u_2) \\[2ex] \dfrac{-\sin(2\pi u_2)}{\sqrt{-2\log u_1}\, u_1} & 2\pi\sqrt{-2\log u_1}\cos(2\pi u_2) \end{bmatrix}$$

so $|D\phi| = \dfrac{2\pi}{u_1}$

if $(y_1, y_2) = \phi(u_1, u_2)$ then

$$x_1^2 + x_2^2 = -2\log u_1 \quad \text{or} \quad u_1 = e^{-\frac{x_1^2}{2} - \frac{x_2^2}{2}}$$

so if $X = \phi(U)$ then it's density is $\dfrac{e^{-\frac{x_1^2}{2} - \frac{x_2^2}{2}}}{2\pi}$ (i.e. X_1 and X_2 are indep $N(0,1)$)