Monte Carlo Sampling

Damon McDougall

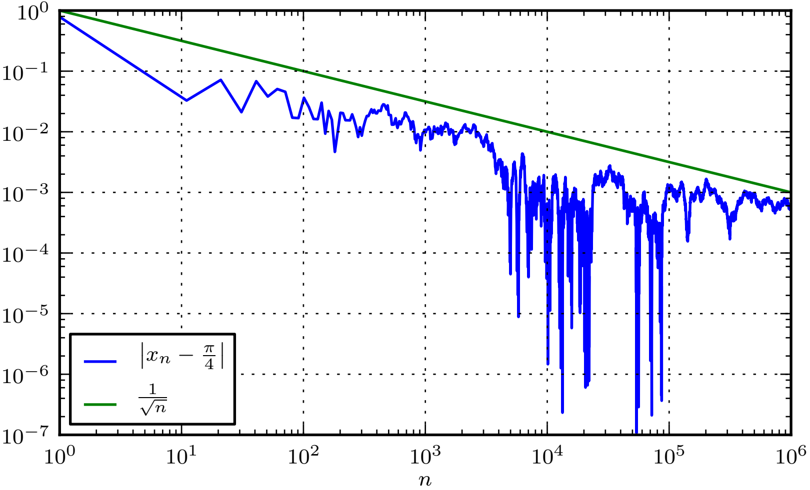
Mathematics Institute, University of Warwick, United Kingdom

15th June 2011

Sampling

1. Set $x_0 = 0, a_0 = 0$.
2. Sample $u \sim U(0,1) \times U(0,1)$.
3. Set
$$a_{n+1} = \begin{cases} a_n + 1 & \|u\| \leq 1 \\ a_n & \textit{otherwise} \end{cases}$$
4. Set $x_{n+1} = \frac{a_{n+1}}{n+1}$.
5. Go to 2.
6. Repeat ad nauseam.

Sampling



10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^0
10^1
10^2
10^3
10^4
10^5
10^6
n
$\left|x_n - \frac{\pi}{4}\right|$
$\frac{1}{\sqrt{n}}$

Sampling

- What have we done? Let $D := \{y \in \mathbb{R}^2 \mid y_j \geq 0, \|y\| \leq 1\}$, we have

$$\lim_{N\to\infty} \sum_{i=1}^{N} x_i \stackrel{?}{=} \int_D \mathrm{d}A$$

- Could we use this idea to approximate other integrals? For example

$$\frac{\int_0^\infty \frac{x^2}{\tanh(x)+\cosh(x)} \,\mathrm{d}x}{\int_0^\infty \frac{x}{\tanh(x)+\cosh(x)} \,\mathrm{d}x}$$

Sampling

Idea:

- Treat

$$f(x) = \begin{cases} \frac{x}{Z(\tanh(x)+\cosh(x))} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

  like a *probability distribution*. I.e., $Z = \int_0^\infty \frac{x}{\tanh(x)+\cosh(x)} \, \mathrm{d}x$.

- Note that if $X \sim f(x)$ then

$$\mathbb{E}(X) = \int_0^\infty x f(x) \, \mathrm{d}x = \frac{1}{Z} \int_0^\infty \frac{x^2}{\tanh(x) + \cosh(x)} \, \mathrm{d}x.$$

- Ideally we would like to draw samples from $f(x)$. How?

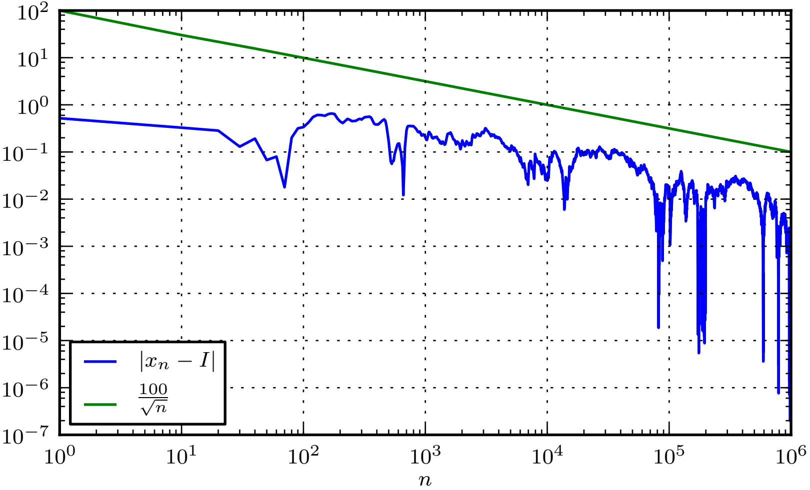
Sampling

1. Sample $x_0 \sim \mathcal{N}(0, \gamma^2)$
2. Construct $z = x_n + w$ where $w \sim \mathcal{N}(0, \gamma^2)$.
3. Compute
$$\alpha(x_n, z) = 1 \wedge \frac{z(\tanh(x_n) + \cosh(x_n))}{x_n(\tanh(z) + \cosh(z))}$$
4. Set
$$x_{n+1} = \begin{cases} z & \text{with probability } \alpha(x_n, z) \\ x_n & \text{with probability } 1 - \alpha(x_n, z) \end{cases}$$
5. Go to 2.

The x_n are now samples from $f(x) = \frac{x}{\tanh(x) + \cosh(x)}$.

$$\boxed{\frac{1}{N} \sum_{i=1}^{N} x_i \xrightarrow{N \to \infty} \int_0^\infty \frac{x^2}{\tanh(x) + \cosh(x)} \, \mathrm{d}x} \ ?$$

Sampling



10^2
10^1
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^0
10^1
10^2
10^3
10^4
10^5
10^6
n
$|x_n - I|$
$\frac{100}{\sqrt{n}}$

Sampling

What have we done?

- Generated samples from a distribution.
- Computed quantities from the samples.

There are a few points:

- It looks like computed quantities converge. Do they?
- The rate of convergence looks suspiciously well guessed.
- How did you generate samples from a nontrivial distribution so easily?
- Can we generalise this idea of sampling to other problems?

Convergence

Does it converge?

Theorem (Weak law of large numbers)

Let $\{X_i\}$ be a sequence of i.i.d random variables with $\mathbb{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$, then $\forall \epsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}\left(|S_N - \mu| > \epsilon\right) = 0$$

Note: The convergence is in probability.

Theorem (Markov's inequality)

If Y is any random variable and $a > 0$ then

$$\mathbb{P}(|Y| \geq a) \leq \frac{\mathbb{E}(|Y|)}{a}.$$

Proof.

We use the fact that $|y|p(y) \geq 0$ to obtain

$$\begin{aligned}\mathbb{E}(|Y|) = \int_{-\infty}^{\infty} |y|p(y)\,\mathrm{d}y \geq \int_{a}^{\infty} |y|p(y)\,\mathrm{d}y \geq a\int_{a}^{\infty} p(y)\,\mathrm{d}y \\ = a\mathbb{P}(|Y| \geq a).\end{aligned}$$

$\square$

Convergence

Theorem (Chebyshev's inequality)

If Y is a random variable with $\mathbb{E}(Y) = \mu$ and $Var(Y) = \sigma^2$ then $\forall k > 0$,

$$\mathbb{P}(|Y - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Proof.

Use Markov's inequality to obtain

$$\mathbb{P}(|Y - \mu| \geq k\sigma) = \mathbb{P}(|Y - \mu|^2 \geq (k\sigma)^2) \leq \frac{\mathbb{E}((Y - \mu)^2)}{(k\sigma)^2} = \frac{\sigma^2}{(k\sigma)^2} = \frac{1}{k^2}.$$

$\square$

Convergence

Theorem (Weak law of large numbers)

$\{X_i\}$ i.i.d, $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$. Let $S_N = \frac{1}{N}\sum_{i=1}^{N} X_i$, then

$$\forall \epsilon > 0, \ \lim_{N \to \infty} \mathbb{P}\left(|S_N - \mu| > \epsilon\right) = 0$$

Proof.

Use $\mathrm{Var}(S_N) = \mathrm{Var}(\frac{1}{N}\sum_{i=1}^{N} X_i) = \frac{1}{N^2}\sum_{i=1}^{N} \mathrm{Var}(X_i) = \frac{\sigma^2}{N}$. Use $\mathbb{E}(S_N) = \mu$. Chebshev says that

$$\forall k > 0, \ \mathbb{P}\left(|S_N - \mu| > k\sqrt{\mathrm{Var}(S_N)}\right) \leq \frac{1}{k^2}.$$

So take $k = \epsilon / \sqrt{\frac{\sigma^2}{N}}$ to get

$$\mathbb{P}\left(|S_N - \mu| > \epsilon\right) \leq \frac{1}{k^2} = \frac{\sigma^2}{N\epsilon^2} \xrightarrow{N \to \infty} 0.$$

$\square$

Markov chain Monte Carlo

- The distribution $f(x) = \frac{x}{Z(\tanh(x)+\cosh(x))}$ is not a *standard* distribution.
- Can we draw samples from any *arbitrary* distribution?
    - Need a formula for the *target* distribution, $\pi(x)$.
    - Need a formula for the *proposal* distribution, $q(x, z)$.
- The target in the example was $f(x)$.
- The proposal in the example was $q(x, \cdot) = \mathcal{N}(x, \gamma^2)$.
- What does the previous algorithm look like in the general case?

Markov chain Monte Carlo

1. Draw $z \sim q(x_n, \cdot)$
2. Compute
$$\alpha(x_n, z) = 1 \wedge \frac{\pi(z)q(z, x_n)}{\pi(x_n)q(x_n, z)}$$
3. Set
$$x_{n+1} = \begin{cases} z & \text{with probability } \alpha(x_n, z) \\ x_n & \text{with probability } 1 - \alpha(x_n, z) \end{cases}$$
4. Go to 1.

The x_n are now samples from $\pi(x)$.
They may be used to approximate any integrals under this density.
Where do the x_n live? Could they live in a weird space, like H^2?

Markov chain Monte Carlo

1. Define $\Phi(u) = \|u(x) - \sin(x)\|_{L^2}^2$
2. Sample $u_0 \sim \mathcal{N}\left(0, \left(-\frac{\partial^2}{\partial x^2}\right)^{-2}\right)$
3. Construct $z = u_n + w$ where $w \sim \mathcal{N}\left(0, \left(-\frac{\partial^2}{\partial x^2}\right)^{-2}\right)$
4. Compute
$$\alpha(u_n, z) = 1 \wedge \exp(\Phi(u_n) - \Phi(z))$$
5. Set
$$u_{n+1} = \begin{cases} z & \text{with probability } \alpha(u_n, z) \\ u_n & \text{with probability } 1 - \alpha(u_n, z) \end{cases}$$
6. Go to 3.

The prior

- How do we construct a ‘random function’ with, say, two derivatives

The prior

- How do we construct a 'random function' with, say, two derivatives

$$\xi(x) = \sum_{k\in\mathbb{Z}\setminus\{0\}} \frac{\qquad\qquad}{\qquad\qquad} \exp(2\pi i k x),$$

The prior

- How do we construct a ‘random function’ with, say, two derivatives

$$\xi(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{}{(4\pi^2 k^2)} \exp(2\pi i k x),$$

The prior

- How do we construct a 'random function' with, say, two derivatives

$$\xi(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{\gamma_k}{(4\pi^2 k^2)} \exp(2\pi i k x),$$

The prior

- How do we construct a 'random function' with, say, two derivatives

$$\xi(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{\gamma_k}{(4\pi^2 k^2)} \exp(2\pi i k x), \quad \gamma_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, I_2)$$

The prior

- How do we construct a 'random function' with, say, two derivatives

$$\xi(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{\gamma_k}{(4\pi^2 k^2)^{\alpha/2}} \exp(2\pi i k x), \quad \gamma_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, I_2)$$

- $\xi \in H^{\alpha}$

The prior

- How do we construct a ‘random function’ with, say, two derivatives

$$\xi(x) = \sum_{k \in \mathbb{Z} \setminus \{0\}} \frac{\gamma_k}{(4\pi^2 k^2)^{\alpha/2}} \exp(2\pi i k x), \quad \gamma_k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, I_2)$$

- $\xi \in H^\alpha$
- Notation for the distribution of functions drawn in this manner is

$$\mathcal{N}\left(0, \left(-\frac{\partial}{\partial x^2}\right)^{-\alpha}\right)$$

Bloody great. Sampling functions. Is this rubbish even useful?

An application

Suppose you have a model for some process,

$$\begin{aligned} \frac{\partial v}{\partial t} &= \frac{\partial^2 v}{\partial x^2} \quad x \in \Omega, \\ v(0,t) &= h(x) \quad x \in \partial\Omega, \end{aligned}$$

but you don't know the initial condition.
Suppose you also have observations

$$\begin{aligned} y_{jk} &= v(x_j, t_k) + \eta_{jk}, \quad \eta_{jk} \sim \mathcal{N}(0, \sigma^2), \\ \rightsquigarrow \quad y &= \mathcal{G}(v_0) + \eta, \quad \eta \sim \mathcal{N}(0, B). \end{aligned}$$

How do you find v_0?

An application

There is no single initial condition. The answer is

$$\mathbb{P}(v_0|y) = \frac{\mathbb{P}(y|v_0)\mathbb{P}(v_0)}{\mathbb{P}(y)}$$
$$\propto \underbrace{\mathbb{P}(y|v_0)}_{\text{known}}\underbrace{\mathbb{P}(v_0)}_{\text{known}}.$$

$\mathbb{P}(y|v_0) = \exp\left(-\frac{1}{2}\|\mathcal{G}(\cdot) - y\|_B^2\right)$. Can take $\mathbb{P}(v_0)$ to be Gaussian.

How do we sample this distribution?

An application

1. Define $\Phi(u) = \frac{1}{2}\|\mathcal{G}(u) - y\|_B^2$
2. Sample $u_0 \sim \mathbb{P}(v_0)$
3. Construct $z = u_n + w$ where $w \sim \mathbb{P}(v_0)$
4. Compute
$$\alpha(u_n, z) = 1 \wedge \exp(\Phi(u_n) - \Phi(z))$$
5. Set
$$u_{n+1} = \begin{cases} z & \text{with probability } \alpha(u_n, z) \\ u_n & \text{with probability } 1 - \alpha(u_n, z) \end{cases}$$
6. Go to 3.

Note: This is *exactly* the same algorithm as the previous case, with a different Φ.

Summary

I have told you

- how to approximate hard integrals using samples;
- important convergence properties;
- a benefit over using standard numerical integration techniques;
- how to draw samples from hard distributions;
- how to do all of the above on an infinite dimensional space;
- an application of MCMC to the real world.

Thank you