An introduction to Markov Chain Monte Carlo (MCMC)

February 3, 2022

John Bulava
Zeuthen Data Science Seminar (ZDSS)

Where is MCMC useful?

- Deterministic integration algorithms for low dimensions ($n \lesssim 3$): quadrature, trapeziodal, etc.

- Monte Carlo sampling for intermediate dimensionality ($3 \lesssim n \lesssim 10$): VEGAS, MISER, etc.

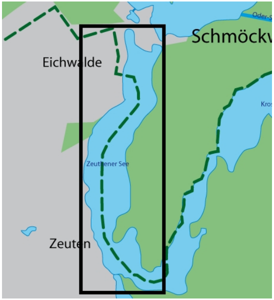
- MCMC great for large dimensional integrals ($10 \lesssim n \lesssim ??$).

- Today we will solve an $n = 1728$ dimensional integral with MCMC.

- 'integral' = sampling some large space of configurations

Random numbers do integrals: area of some shape S

- Inscribe S in rectangle R of area $L \times W$
- Generate random points uniformly in R
- Count: inside S or not?
- Repeat!



$$\frac{A_{\text{shape}}}{LW} = \lim_{N_{\text{tot}} \to \infty} \frac{N_{\text{inside}}}{N_{\text{tot}}} = \int \mathrm{d}^2 x\, f(x)\, p(x) = \langle f(X) \rangle$$

$$f(x) = \begin{cases} 1, & x \in S \\ 0, & x \notin S \end{cases}, \qquad p(x) = \begin{cases} \frac{1}{LW}, & x \in R \\ 0, & x \notin R \end{cases}$$

Importance sampling: capture some features of integrand

$$I = \int_0^3 \mathrm{d}x \frac{\mathrm{e}^{-x}}{1+\frac{x}{9}} \approx 0.873109 = \langle f(U) \rangle = \langle g(Y) \rangle$$

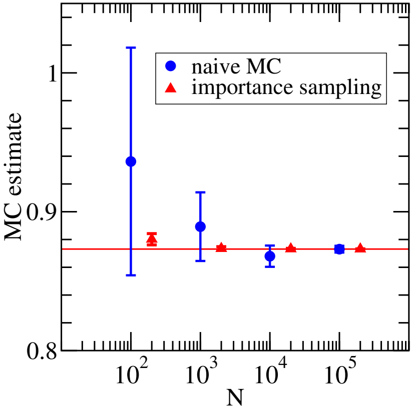
$$f(x) = \frac{3\mathrm{e}^{-x}}{1+\frac{x}{9}}, \qquad g(x) = \frac{1-\mathrm{e}^{-3}}{1+\frac{x}{9}}$$

$$p_{\mathrm{U}}(x) = \begin{cases} \frac{1}{3}, & x \in [0,3] \\ 0, & \text{otherwise} \end{cases}, \qquad p_{\mathrm{Y}}(x) = \begin{cases} \frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-3}}, & x \in [0,3] \\ 0, & \text{otherwise} \end{cases}$$

- integral written as either uniform (U) or exponential (Y) random variable.

Example from C. Morninstar (hep-lat/0702020)

- Better to use exponential random variate!
- Importance sampling reduces the variance.



Markov Chain Monte Carlo

- What if $p(x)$ is not easy to generate?
- Construct Markov Chain with $p(x)$ as limiting distribution
- Motivating example: scalar field configurations on a periodic L^3 lattice

$$\phi_x \in \mathbb{R}, \quad I = \langle m(\phi)^2 \rangle = \int \prod_x \mathrm{d}\phi_x \, m(\phi)^2 \, p(\phi),$$

$$m(\phi) = \frac{1}{\sqrt{V}} \left| \sum_x \phi_x \right|, \qquad p(\phi) = \frac{\mathrm{e}^{-S(\phi)}}{Z},$$

$$S(\phi) = \sum_x \left[-2\kappa \sum_{\mu=1,2,3} \phi_x \phi_{x+\hat{\mu}} + \phi_x^2 + \lambda(\phi_x^2 - 1)^2 \right]$$

- Easy to sample $p(\phi)$ if $\kappa = \lambda = 0$.

Markov Processes:

А. А. Марков (1886).

- A random process is a sequence of random variables: $\{X_t\}, \quad t = 0, 1, 2, \ldots$

- Markov processes satisfy the Markov property:

$$P(X_t = x_t | X_0 = x_0, ..., X_{t-1} = x_{t-1}) = P(X_t = x_t | X_{t-1} = x_{t-1})$$

- Homogeneous Markov processes (HMC's) also have

$$P(X_t = x | X_{t-1} = y) = P(X_{t'} = x | X_{t'-1} = y) = M_{xy},$$

where M_{xy} is the Markov matrix.

Fundamental limit theorem:

- Stationary distribution: $\sum_y M_{xy} p_y = p_x$
- A Markov Chain is ergodic if it is:
  - irreducible: all states are accessible from all others
  - aperiodic: no 'cycle' transitions which loop in a pattern
  - positive recurrent: won't 'run away' to infinity
- An ergodic chain has a unique, universal stationary distribution

$$p_x = \lim_{t\to\infty} (M^t)_{xy}, \quad \forall y$$

- Will approach the limiting distribution ('thermalize', 'equilibriate') for any starting configuration.

Defintions: assume a chain in equilibrium. Consider some $Y_t = f(X_t)$, with $\langle Y_t \rangle = \mu$.

- Autocovariance:

$$R_Y(|t-s|) = \langle (Y_t - \mu)(Y_s - \mu) \rangle$$

- Autocorrelation: $\rho_Y(t) = R_Y(t)/R_Y(0)$

- Integrated autocorrelation time:

$$\tau_{\mathrm{int},Y}(\tau) = \frac{1}{2} + \sum_{t=1}^{\tau} \rho_Y(t), \qquad \tau_{\mathrm{int},Y} \equiv \tau_{\mathrm{int},Y}(\infty)$$

Recall the Central Limit Theorem (CLT): if σ, $\mu < \infty$, then distribution of the sample mean S_n approaches a gaussian

$$\lim_{n\to\infty} p_{S_n}(x) = \mathrm{Gauss}(\mu, \frac{\sigma}{\sqrt{n}}; x)$$

The CLT is modified in the presence of autocorrelations!

- Need additional condition: $\tau_{\mathrm{int}} < \infty$
- CLT works as before but with modified variance

$$\tilde{\sigma}^2 = \sigma^2 + 2\sum_{t=1}^{\infty} R(t) = 2\tau_{\mathrm{int}}\sigma^2$$

- 'Effective statistics' $\tilde{n} = n/(2\tau_{\mathrm{int}})$. Large autocorrelations are a problem!

What we know:

- Ergodic chains have a unique stationary distribution
- Sampling from a chain in equilibirium behaves according to the CLT.
- Confidence intervals, statistical errors reliably estimated using the CLT accounting for autocorrelation

But how do we construct a chain with the desired limiting distribution?!?

Detailed balance: easy way to calculate stationary distribution $\sum_x M_{xy} p_y = p_x$

- Finding eigenvectors of the Markov matrix M is difficult.
- If p_x satisfies detailed balance:

$$\frac{p_x}{p_y} = \frac{M_{xy}}{M_{yx}}, \qquad \forall x, y$$

  it is a stationary distribution.

- Need also that the chain is ergodic for Fundamental Limit theorem to hold.

Common strategy: the Metropolis-Hastings algorithm

- Split application of M (update) into two steps: proposal and acceptance.
- Propose a change according to h_{xy}, accept this change with probability a_{xy}.
- Markov (transition) matrix is now:

$$M_{xy} = \begin{cases} h_{xy} a_{xy}, & x \neq y \\ 1 - \sum_{k \neq x} h_{ky} a_{ky}, & x = y \end{cases}$$

- Detailed balance is satisfied if:

$$a_{xy} = \min\left(1, \frac{p_x h_{yx}}{p_y h_{xy}}\right)$$

- Assume a reversible proposal $h_{yx} = h_{xy}$:

$$a_{xy} = \min\left(1,\ \frac{p_x}{p_y}\right)$$

- Assume $p_x = \frac{1}{Z}\mathrm{e}^{-S(x)}$:

$$a_{xy} = \min\left(1,\ \mathrm{e}^{-\Delta S}\right), \qquad \Delta S = S(x) - S(y)$$

- If $S(y) \leq S(x)$, always accept.
- If $S(y) > S(x)$, sometimes accept.

Proposal must change state optimally:

- If too large: $\Delta S >> 0$, low acceptance rate, large autocorrelation.
- If too small: $\Delta S \approx 0$, high acceptance rate, large autocorrelation

Back to scalar field example. Recall:

$$p(\phi) = \frac{\mathrm{e}^{-S(\phi)}}{Z}, \quad S(\phi) = \sum_x \left[-2\kappa \sum_{\mu=1,2,3} \phi_x \phi_{x+\hat{\mu}} + \phi_x^2 + \lambda(\phi_x^2 - 1)^2 \right]$$

- Need a proposal step $h(\phi' \leftarrow \phi)$ that is
  - reversible: $h(\phi' \leftarrow \phi) = h(\phi \leftarrow \phi')$
  - Doesn't give large $\Delta S = S(\phi') - S(\phi)$
- Local proposals are most efficient, but not generally applicable
- Global proposal: Hybrid Monte Carlo (HMC)
  (Duane, Kennedy, Pendelton, Roweth '87)

Hybrid Monte Carlo (HMC):

- Add 'conjugate' degrees of freedom: $\pi_x \in \mathbb{R}$

- Define the 'Hamiltonian' of a combined configuration:

$$H(\phi, \pi) = \frac{1}{2} \sum_x \pi_x^2 + S(\phi) = K(\pi) + V(\phi)$$

- Draw the π_x individually from a unit normal distribution.

- Evolve (ϕ, π) according to Hamilton's equations

$$\frac{d\phi_x}{dt} = \frac{\partial K}{\partial \pi_x}, \qquad \frac{d\pi_x}{dt} = -\frac{\partial V}{\partial \phi_x}$$

  for some trajectory length τ.

- Accept/Reject using $\Delta H = H(\phi', \pi') - H(\phi, \pi)$ with $\phi' = \phi(\tau)$ and $\pi' = \pi(\tau)$.

Numerical integration of Hamilton's equations:

- Divide trajectory length into N_τ steps $\epsilon = \tau / N_\tau$

- Simple but effective integrator (Leapfrog):

$$I_{\mathrm{LF}}(\epsilon) = I_1\left(\frac{\epsilon}{2}\right) I_2\left(\epsilon\right) I_1\left(\frac{\epsilon}{2}\right)$$

  where

$$I_1(\epsilon)\begin{pmatrix}\phi \\ \pi\end{pmatrix} = \begin{pmatrix}\phi + \epsilon\pi \\ \pi\end{pmatrix},$$
$$I_2(\epsilon)\begin{pmatrix}\phi \\ \pi\end{pmatrix} = \begin{pmatrix}\phi \\ \pi + \epsilon F(\phi)\end{pmatrix}$$

  with $F_x(\phi) = -\frac{\partial S}{\partial \phi_x}$.

- Reversible and symplectic scheme means small ΔH.

- These steps can be generalized to create other reversible, symplectic integrators. (Omelyan, Mryglod, Folk, '02)

- Second-order OMF (OMF2):

$$I_{\mathrm{OMF2}}(\epsilon) = I_1(\xi\epsilon)\, I_2\left(\frac{\epsilon}{2}\right) I_1(\{1-2\xi\}\epsilon)\, I_2\left(\frac{\epsilon}{2}\right) I_1(\xi\epsilon)$$

Incredibly small errors result from $\xi = 0.1931833275037836$.

- Fourth-order OMF (OMF4):

$$I_{\mathrm{OMF4}}(\epsilon) = I_1(\xi\epsilon)\, I_2\left(\frac{1-2\lambda}{2}\epsilon\right) I_1(\chi\epsilon)\, I_2(\lambda\epsilon) \times$$
$$I_1((1-2\xi-2\chi)\epsilon)\, I_2(\lambda\epsilon)\, I_1(\chi\epsilon)\, I_2\left(\frac{1-2\lambda}{2}\epsilon\right) I_1(\xi\epsilon)$$

Incredibly small(er) errors result from

$$\xi = 0.1931833275037836,$$
$$\lambda = -0.02094333910398989,$$
$$\chi = 1.235692651138917.$$

Summary:

- MCMC is great for importance sampling large-dimensional integrals with a non-seperable pdf
- Theorems allow for rigorous (in principle!) convergence and error estimates
- Metropolis-Hastings propose/accept algorithm easily ensures the desired limiting distribution
- HMC provides a global proposal step with a reasonable acceptance rate.