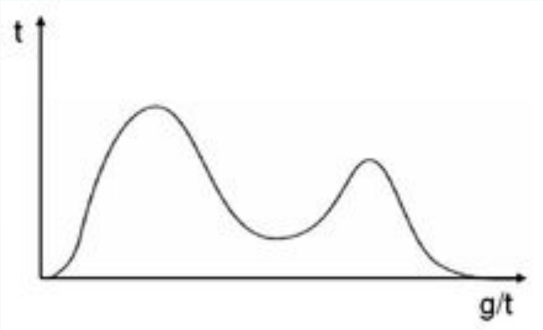
A Brief Tour of Physics-Inspired Sampling

Eric Regis, Sinho Chewi
Yale University

What is Physics-Inspired Sampling?

Sampling is the process of drawing from a probability distribution.



By "**physics-inspired sampling**", we mean algorithms that:

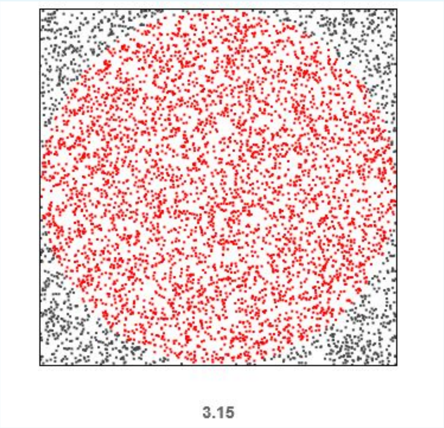
- Were invented by a team of physicists
- Were invented for the purpose of solving a physical problem
- Were inspired by physics principles

Something that physics-inspired sampling algorithms have in common is:

- Based on statistical physics principles
- Take a dynamical perspective
- Utilize stochastic processes to explore probability distributions

Markov Chain Monte Carlo

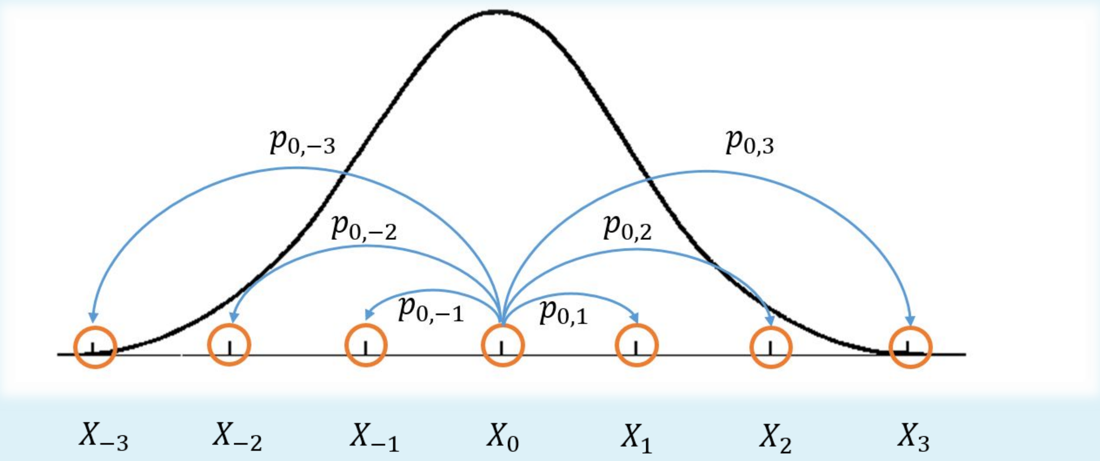
Monte Carlo methods are computational techniques that rely on repeated random sampling to solve problems that are difficult to address analytically.



$$\mathbb{E}[f(X)] \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i)$$

Markov Chain Monte Carlo (MCMC) is a method to sample from a probability distribution by constructing a Markov chain that has the target distribution as its equilibrium distribution.

It's inspired by the principle of **ergodicity**: the time average of a single system trajectory equals the ensemble average across probability space



Metropolis-Hastings

The first MCMC algorithm was **the Metropolis algorithm**. It was invented at Los Alamos Laboratory in 1953.

In the Metropolis algorithm, to determine the next state transition, there are two steps:

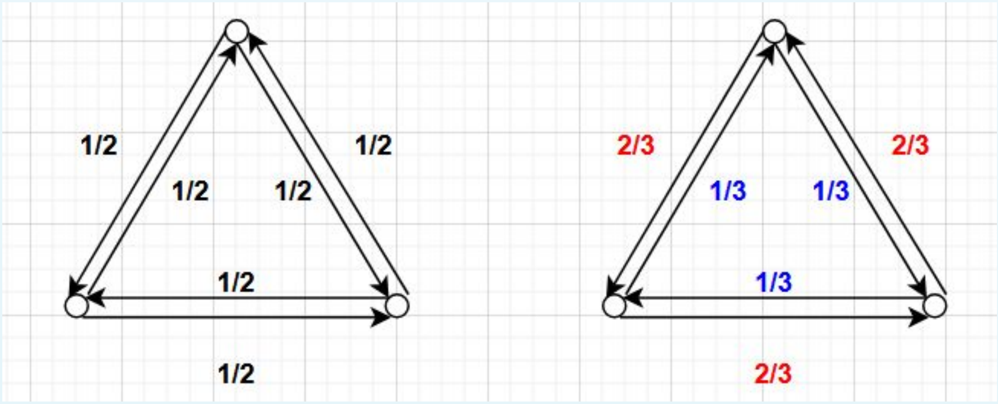
1. Sample a candidate state from a proposal kernel q.
2. Accept the proposal with probability α, which depends on whether the candidate state has higher or lower probability than the current state.

The **Metropolis-Hastings algorithm**, invented in 1970, improves on the original Metropolis algorithm by allowing the proposal kernel to be asymmetric.

$$\alpha(x,y) = \min\left\{\frac{\pi(y)q(x,y)}{\pi(x)q(y,x)}, 1\right\}$$

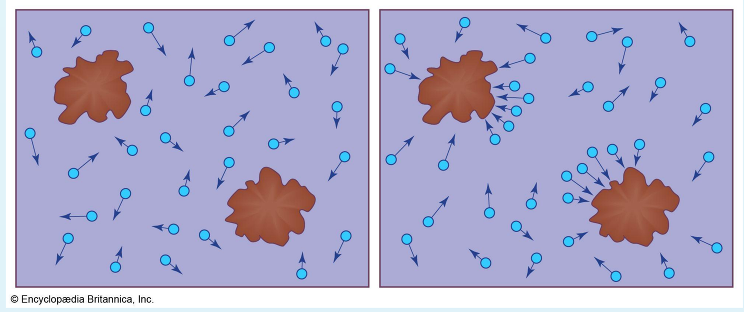
$$p(x,y) = \alpha(x,y)q(x,y)$$

This is closely related to **detailed balance**, also known as microscopic irreversibility. With detailed balance, a trajectory and its time-reversed counterpart are equally probable.



Langevin Dynamics

In 1905, Albert Einstein provided a theoretical explanation for **Brownian motion,** the random movement of particles suspended in a liquid or gas.



A **stochastic differential equation** is a mathematical equation that combines deterministic dynamics (drift term) with random fluctuations modeled by Brownian motion.

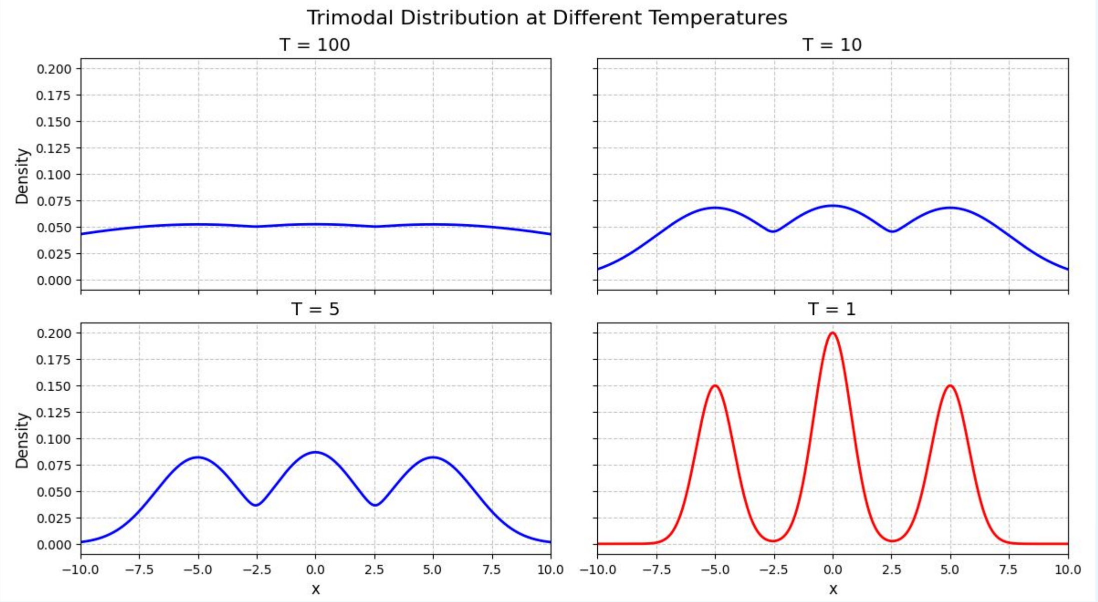
$$dX_t = \nabla \log p(x)dt + \sqrt{2}\, dW_t$$

Simulated Annealing

The **Boltzmann distribution** describes the probability distribution of microstates in a system at thermal equilibrium. A given state has a probability proportional to the exponential of its negative energy divided by temperature

$$p(x) = \frac{e^{-\frac{E(x)}{T}}}{Z}$$

Simulated annealing is a sampling procedure where you start with Markov Chain Monte Carlo (MCMC) sampling from a distribution at a higher temperature, then gradually decrease the temperature, allowing the sampling to approach the target distribution at the lowest temperature.



Hamilton Monte Carlo

Hamiltonian Monte Carlo is a modified version of Monte Carlo sampling where, rather than directly updating position based on the gradient of the potentia, the gradient of the potential updates the momentum and then the momentum updates the position. This can be implemented through the **leapfrog integrator.**

$$\mathbf{p}_n\left(t+\frac{\Delta t}{2}\right) = \mathbf{p}_n(t) - \frac{\Delta t}{2}\nabla U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_n(t)}$$

$$\mathbf{x}_n(t+\Delta t) = \mathbf{x}_n(t) + \Delta t M^{-1}\mathbf{p}_n\left(t+\frac{\Delta t}{2}\right)$$

$$\mathbf{p}_n(t+\Delta t) = \mathbf{p}_n\left(t+\frac{\Delta t}{2}\right) - \frac{\Delta t}{2}\nabla U(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_n(t+\Delta t)}$$

The algorithm is:

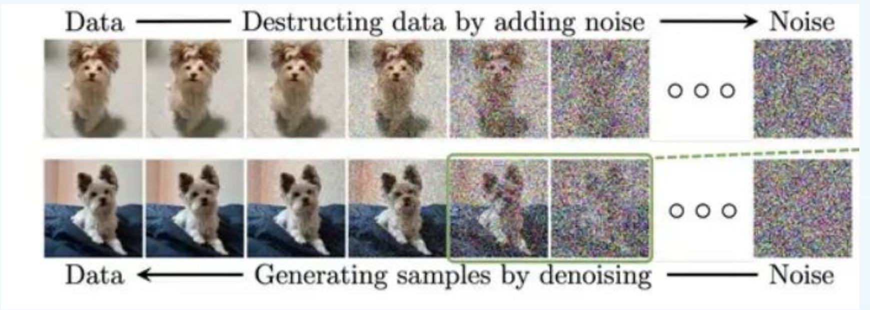
1. Randomly sample a new momentum vector.
2. Run deterministic Hamiltonian dynamics for L steps.
3. Perform a Metropolis acceptance/rejection step
4. Repeat from step 1.

Diffusion Models

Diffusion models are generative models that learn to sample from a desired distribution by reversing a diffusion process.

During training, the model gradually adds noise to images, and a neural network learns to reverse this noise.

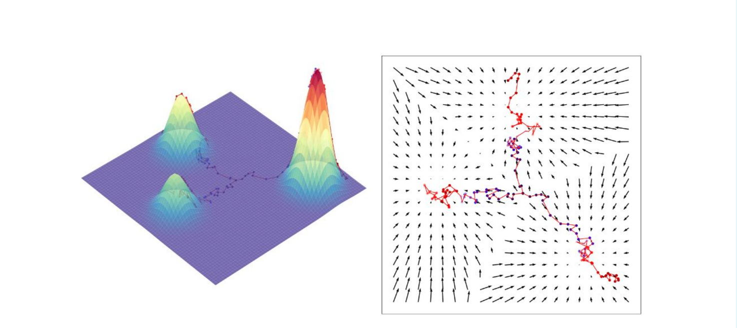
During inference, the model starts with random noise and progressively denoises it to generate samples from the distribution of natural images



Diffusion models are an example of a **score-based generative model**. A score-based generative model is a generative model that aims to the learn the **score function**: the gradient of the log probability density.

$$\text{score} = \nabla \log p(x)$$

For diffusion models, they aim to learn the score functions of the intermediary distributions, representing increasingly-blurred images..



References

1. Geman, S., & Geman, D. (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6(6), 721-741.*
2. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics, 21(6), 1087-1092.*
3. Hastings, W. K. (1970). "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika, 57(1), 97-109.*
4. Roberts, G. O., & Tweedie, R. L. (1996). "Exponential Convergence of Langevin Distributions and Their Discrete Approximations." *Bernoulli, 2(4), 341-363.*
5. Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). "Optimization by Simulated Annealing." *Science, 220(4598), 671-680.*
6. Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). "Hybrid Monte Carlo." *Physics Letters B, 195(2), 216-222.*
7. Neal, R. M. (2011). "MCMC Using Hamiltonian Dynamics." *Handbook of Markov Chain Monte Carlo, 113-162.*
8. Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics." *Proceedings of the 32nd International Conference on Machine Learning (ICML), 2256-2265.*
9. Ho, J., Jain, A., & Abbeel, P. (2020). "Denoising Diffusion Probabilistic Models." *Advances in Neural Information Processing Systems (NeurIPS), 33, 6840-6851.*