Deep generative models for biologists

Clayton W. Seitz

January 21, 2022

Outline

Monte-Carlo Markov Chain (MCMC)

- MCMC algorithms were originally developed in the 1940's by physicists at Los Alamos
- They were interested in modeling the probabilistic behavior of collections of atomic particles
- Simulation was difficult – the normalization constant Z was not known
- The term "Monte-Carlo" was coined at Los Alamos.
- Ulam and Metropolis overcame this problem by constructing a Markov chain for which the desired distribution was the stationary distribution
- Introduced to statistics and generalized with the Metropolis-Hastings algorithm (1970) and the Gibbs sampler of Geman and Geman (1984).

Markov Chains

For a state space Ω s.t. $\mathbf{x}_t \in \Omega$. $\mathbf{x}_t$ is a Markov process if:

$$P(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{x}_{t-2}, ..., \mathbf{x}_{t-N}) = P(\mathbf{x}_t|\mathbf{x}_{t-1})$$

which is commonly called the *memoryless property*.

- $\mathbf{x}_t$ can be generally be N-dimensional
- The chain is called *homogeneous* if $T(\mathbf{x}_t|\mathbf{x}_{t-1})$ is time-invariant.
- For discrete Ω, T is a matrix of probabilities with $T_{ij} = \Pr(i \to j)$
- For continuous Ω, T is the joint probability density $T(x_t, x_{t-1})$

Markov Chains

The Chapman-Kolmogorov equation marginalizes $T(x_t, x_{t-1})$:

$$\begin{aligned} P(\mathbf{x}_t) &= \int T(x_t, x_{t-1}) d\mathbf{x}_{t-1} \\ &= \int T(x_t | x_{t-1}) P(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \end{aligned}$$

The chain satisfies *detailed balance* if

$$T(x_t, x_{t-1}) P(x_t) = T(x_{t-1}, x_t) P(x_{t-1})$$

which guarantees there is a unique stationary distribution $P_0(x_t)$

Monte-Carlo Markov Chain (MCMC)

A stationary distribution satisfies

$$P_0(\mathbf{x}_t) = \int T(x_t|x_{t-1})P_0(\mathbf{x}_{t-1})d\mathbf{x}_{t-1}$$

- If a process is Markov e.g., Brownian motion, Ornstein-Uhlenbeck, $P_0(x_t)$ is a solution to the SDE
- We can also design $T(x_t, x_{t-1})$ s.t. $P_0(x_t)$ is a distribution we cannot sample from easily such as the Ising model
- The notion of "time" in the second case is artificial
- There are several MCMC algorithms, we will focus on Gibbs MCMC

Gibbs sampling

- Suppose $p(\mathbf{x})$ is a p.d.f. or p.m.f. that is difficult to sample from directly.
- Suppose, though, that we *can* easily sample from the conditional distributions e.g., $p(x_1|x_2, ..., x_n)$.
- The Gibbs sampler proceeds as follows:
  1. set $\mathbf{x}$ to some initial starting values
  2. then sample $x_1|x_2, ..., x_n$, then sample $x_2|x_1, ..., x_n$, and so on.

Gibbs sampling

0. Set (x_0, y_0) to some starting value.
1. Sample $x_1 \sim p(x|y_0)$, that is, from the conditional distribution $X \mid Y = y_0$.
   Current state: (x_1, y_0)
   Sample $y_1 \sim p(y|x_1)$, that is, from the conditional distribution $Y \mid X = x_1$.
   Current state: (x_1, y_1)
2. Sample $x_2 \sim p(x|y_1)$, that is, from the conditional distribution $X \mid Y = y_1$.
   Current state: (x_2, y_1)
   Sample $y_2 \sim p(y|x_2)$, that is, from the conditional distribution $Y \mid X = x_2$.
   Current state: (x_2, y_2)

   $\vdots$

Repeat iterations 1 and 2, M times.

References I