

Lecture 23 An Introduction to MCMC

23.1 What is Monte Carlo for?

Monte Carlo simulation, considered as mathematical experimentation, can help in statistical modelling, optimization for a function of many variables, numerical integration over a high-dimensional space, and a lot more ...

Example 23.1 (statistical modelling) See the diagram of “reconstruction cycle” in Lecture 1. Consider the “usual” procedure: Given a data set, summarize it, propose a model, fit the model (statistical inference), generate synthetic data from the fitted model (this is the simulation step!), compare the simulated data with the real data. May need to repeat the cycle, etc. Such cases are plenty: regression, time series, spatial statistics ... Why need a model? For prediction, better understanding, generalization, etc.

Example 23.2 (optimization) Maybe the last resort when calculus, discrete search, etc. cannot do it. Optimization via simulation is usually very time-consuming. Consider the 2D Ising model in statistical physics. Minimizing the energy function $U(x)$ is equivalent to finding the ground state \hat{x} , i.e. the mode of the Gibbs distribution. This can be done by simulated annealing (SA). See Liu’s book, sec. 1.3 for the Ising model, sec. 10.2 for SA.

Example 23.3 (numerical integration) Consider the integral

$$I = \int_D g(x)\pi(x)dx \tag{23.1}$$

where the function g and probability density π are defined over a bounded domain $D \subset \mathbb{R}^d$. How to compute I numerically?

If using Monte Carlo simulation, we generate iid samples $X^{(1)}, \dots, X^{(n)}$ from π and approximate I by

$$I_n = \frac{1}{n} \sum_{i=1}^n g(X^{(i)}). \tag{23.2}$$

It is well known that I_n is unbiased, i.e. its expectation $E_\pi I_n = I$, and I_n converges to I with π -probability one as $n \rightarrow \infty$ (LLN). The approximation error is of order $O(n^{-1/2})$. This follows from CLT, or can be viewed through the mean square error (MSE)

$$\sqrt{E_\pi(I_n - I)^2} = \sqrt{Var_\pi I_n} = \frac{\sigma}{\sqrt{n}},$$

where σ is the standard deviation of $g(X^{(1)})$. Note that the order $O(n^{-1/2})$ does not depend on the dimension d , although the constant σ does (variance reduction is a major issue in Monte Carlo simulation).

In contrast, the efficiency of deterministic numerical integration methods (lattice approximations) is critically affected by the dimension d . More precisely, assume the functions g and π are smooth enough and approximate I by a sum involving $g(x)\pi(x)$ evaluated at n nodal points in D . Let δ be the step size (spacing) in the lattice approximation. Then $\delta = O(n^{-1/d})$. Usually, the more terms of higher order derivatives included in a (local) approximation to $g(x)\pi(x)$ within a small neighborhood, the more accurate approximation we can obtain for the integral I . Suppose a given method (e.g. Trapezoidal rule, Simpson's rule, etc. see *Numerical Recipes in C*, 2nd edition, sec. 4.1) yields an error rate $O(\delta^k)$ for some $k > 0$ via a 1D numerical integration scheme with step size δ , then the corresponding error rate involved in computing I over D will be $O(n^{-k/d})$, which gets worse with greater d .

Therefore, Monte Carlo integration becomes more efficient than deterministic numerical methods if (and only if) the dimension d is high.

23.2 Markov chains and MCMC

Now we have a clear message: Monte Carlo methods should be the last resort when deterministic methods fail to handle high-dimensional numerical integration.

Suppose $X \sim \pi$ defined on a high-dimensional sample space \mathcal{X} . In many situations, it is even difficult to generate a single sample from π , let alone iid samples. For illustration, consider the AR(1) model and the 2D Ising model. In each case, what is the dimensionality of the model and how do we generate Monte Carlo samples?

Here is a proposed Monte Carlo strategy:

Step 1: Starting from an initial state $X_0 \in \mathcal{X}$, generate dependent samples X_1, X_2, \dots such that $X_t \xrightarrow{\mathcal{D}} X \sim \pi$ as $t \rightarrow \infty$.

Step 2: Having generated $X_t, t = 1, \dots, n$ for a sufficiently large n , discard X_0, X_1, \dots, X_{n_0} , and use $\frac{1}{n-n_0} \sum_{t=n_0+1}^n g(X_t)$ to approximate $I = \int_{\mathcal{X}} g(x)\pi(dx)$.

Definition 1 $\{X_t\}$ is called a Markov chain if for every t , the conditional distributions satisfy

$$P_{X_t}(\cdot | X_0, X_1, \dots, X_{t-1}) = P_{X_t}(\cdot | X_{t-1}). \quad (23.3)$$

$\{X_t\}$ is called a homogeneous Markov chain if for every t and every $x \in \mathcal{X}$, we have

$$P_{X_t}(\cdot|X_{t-1} = x) = P_{X_1}(\cdot|X_0 = x). \quad (23.4)$$

In what follows, we only consider homogeneous Markov chains unless mentioned otherwise. $P_{X_1}(\cdot|X_0 = x) \triangleq p(x, \cdot)$ is called the (one-step) transition probability kernel.

Note:

(a) Markov chains are among the most useful probability models in applications. π is called the *limiting distribution* of $\{X_t\}$, or a *stationary distribution*.

(b) A major question in Markov chain theory is: What conditions on the transition kernel $p(x, \cdot)$ will imply the existence and uniqueness of the limiting distribution π ? However, statisticians in the studies of *Markov chain Monte Carlo* (MCMC) are more interested in the converse: How should $p(x, \cdot)$ be designed based on a target π to obtain a fast convergence $X_t \xrightarrow{\mathcal{D}} X \sim \pi$? Furthermore, how should n_0 and n be determined to have an efficient approximation $\frac{1}{n-n_0} \sum_{t=n_0+1}^n g(X_t) \approx I$?

References:

- [1] Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer. (A great new addition to the literature of Monte Carlo methods, with a comprehensive treatment of applications in asset pricing and risk management.)
- [2] Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer. (A superb book addressing important theoretical and computational issues, also including a broad range of examples in many scientific fields.)
- [3] Ripley, B.D. (1987). *Stochastic Simulation*. Wiley. (A condensed classic book, covers many aspects of simulation, including random number generation.)
- [4] Robert, C. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer. (A more Bayesian statistics oriented good book.)
- [5] Ross, S. (2002). *Simulation* (3rd edition). Springer. (Nicely written, it covers basics and some modern topics.)
- [6] Alan Sokal's excellent lecture notes, can be downloaded from our course website.