Lecture 15 Random variables

Weinan E[1,2] and Tiejun Li[2]

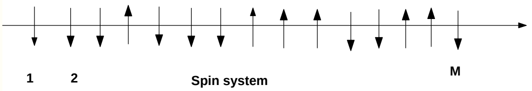
[1]Department of Mathematics,
Princeton University,
weinan@princeton.edu

[2]School of Mathematical Sciences,
Peking University,
tieli@pku.edu.cn
No.1 Science Building, 1575

Outline

High dimensional quadrature in statistical physics

Ising model for mean field ferromagnet modeling



High dimensional quadrature in statistical physics

- Ising model in statistical physics

  Define the Hamiltonian

$$H(\sigma) = -J \sum_{<ij>} \sigma_i \sigma_j,$$

  where $\sigma_i = \pm 1$, $< ij >$ means to take sum w.r.t all neighboring spins $|i - j| = 1$. The internal energy per site

$$U_M = \frac{1}{M} \sum_{\sigma} H(\sigma) \frac{\exp\{-\beta H(\sigma)\}}{Z_M},$$

  where $Z_M = \sum_\sigma \exp\{-\beta H(\sigma)\}$ is the partition function and $\beta = (k_B T)^{-1}$.

- Total number of configuration states: 2^M

High dimensional quadrature in statistical physics

- 统计物理中经常要处理如下的典型问题，求

$$\langle A \rangle \triangleq \frac{1}{Z} \int_{R^{6N}} A(c) e^{-\beta H(c)} dc$$

这里$Z = \int_{R^{6N}} e^{-\beta H(c)} dc$是所谓配分函数(partition funtion)，$\beta = (k_B T)^{-1}$，k_B是Boltzmann常数，T是绝对温度，$dc = dx_1 \cdots dx_N dp_1 \cdots dp_N$，$N$是所考虑体系的粒子数.

- 通常的数值积分方法不再可用。

Stochastic simulations

- Biological network

  Suppose there are N_s species of molecules $S_i,\ i = 1, \ldots, N_s$, and M_R reaction channels $R_j,\ j = 1, \ldots, M_R$. x_i is the number of molecules of species S_i. Then the state of the system is given by

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_{N_s}).$$

  Each reaction R_j is characterized by a rate function $a_j(x)$ and a vector ν_j that describes the change of state due to reaction (after the $j-th$ reaction, $\boldsymbol{x} \to \boldsymbol{x} + \nu_j$). In shorthand denote

$$R_j = (a_j, \nu_j)$$

  How to simulate this biological process?

Numerical solution of stochastic differential equations

- In mathematical economics, the Merton's model for asset price

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where S_t is the asset price, W_t is the standard Brownian motion.

- In the Langevin equation for Brownian particles

$$dx_t = v_t dt$$

$$dv_t = -\frac{\gamma}{m} v_t dt - \frac{1}{m} \nabla V(x_t) dt + \sqrt{2k_B T \gamma} dW_t$$

where $V(x)$ is the potential, γ is the viscosity, m is the mass.

Outline

Monte Carlo方法的基本思想

- 为了说明思想，以下面简单的一维积分问题为例：

$$I(f) = \int_0^1 f(x)dx \tag{1}$$

- 传统的计算方法，如梯形法(trapzoidal rule)：

$$I(f) \approx \left[\frac{1}{2}f(x_0) + \sum_{i=1}^{N-1} f(x_i) + \frac{1}{2}f(x_N)\right]h \tag{2}$$

这里$h = \frac{1}{N}$，$x_i = ih$ $(i = 0, 1, \ldots, N)$。众所周知，梯形法的精度为$O(h^2) = O(N^{-2})$

Basic random variables (discrete case)

- Bernoulli distribution:

$$P(X) = \begin{cases} p, & X = 1, \\ q, & X = 0. \end{cases}$$

where $p > 0, q > 0, p + q = 1$. The mean and variance are

$$\mathbb{E}X = p, \mathrm{Var}(X) = pq.$$

If $p = q = \frac{1}{2}$, it is the well-known fair-coin tossing game.

- Binomial distribution $B(n, p)$:

n independent experiments of Bernoulli distribution X_k,

$X := X_1 + \ldots + X_n$, then

$$P(X = k) = C_n^k p^k q^{n-k}.$$

The mean and variance are

$$\mathbb{E}X = np, \mathrm{Var}(X) = npq.$$

Basic random variables (continuous case)

- Uniform distribution $\mathcal{U}[0,1]$:

$$p(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

The mean and variance are

$$\mathbb{E}X = \frac{1}{2}, \mathrm{Var}(X) = \frac{1}{12}.$$

- Exponential distribution:$(\lambda > 0)$

$$p(x) = \begin{cases} 0 & \text{if } x < 0 \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \end{cases}$$

The mean and variance are

$$\mathbb{E}X = \frac{1}{\lambda}, \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

Basic random variables (continuous case)

- Normal distribution(Gaussian distribution)($N(0,1)$):

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

or more generally $N(\mu, \sigma)$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean (expectation), σ^2 is the variance.

- High dimensional case $(N(\mu, \Sigma))$

$$p(x) = \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}} e^{-(\boldsymbol{X}-\mu)^T\Sigma^{-1}(\boldsymbol{X}-\mu)}$$

where μ is the mean, Σ is a symmetric positive definite matrix, which is the covariance matrix of $\boldsymbol{X}$. $\det\Sigma$ is the determinant of Σ.

Monte Carlo方法的基本思想

- Monte Carlo方法把积分$I(f)$看作某个随机变量的函数的数学期望$I(f) = \mathbb{E}f(X)$,这里X是$[0,1]$均匀分布的随机变量。
- Monte Carlo方法：

$$I(f) \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \triangleq I_N(f) \tag{3}$$

这里x_i $(i = 1, 2, \ldots, N)$为独立同$[0,1]$区间均匀分布的随机变量（以后简记为$i.i.d.$ $\mathcal{U}[0,1]$）。

- 根据概率论中的弱大数定律，

$$I_N(f) \to I(f).$$

且

$$|I_N(f) - I(f)| \sim O(N^{-\frac{1}{2}}).$$

Monte Carlo方法的基本思想

- 考虑在$\mathbb{R}^d$空间的超立方体$\Omega=[0,1]^d$中的积分

$$I(f)=\int\cdots\int_{\Omega} f(\mathbf{x})p(\mathbf{x})d\mathbf{x} \tag{4}$$

$p(x)$满足$\int p(\mathbf{x})d\mathbf{x}=1$和$p(\mathbf{x})\geq 0$。

- 如果在每个坐标方向上将$[0,1]$区间等距剖分n等分来计算，此时的精度为$O(n^{-2})$，需要计算被积函数值和加法运算各$N=n^d$次。如果采用Monte Carlo方法，产生M个$i.i.d$随机向量序列$\mathbf{X}_1,\cdots,\mathbf{X}_M$，令$I_M\approx\frac{1}{M}\sum_{i=1}^{M} f(\mathbf{X_i})$ 来近似I收敛阶为$O(M^{-\frac{1}{2}})$，需要计算被积函数值和加法运算各M次。
- 对于同一个算例，若计算量相同，即$M=n^d$，则$n=M^{1/d}$.
当$d>4$的时候，$n^{-2}>M^{-\frac{1}{2}}$，Monte Carlo方法精度比梯形公式精度高;
当$d<4$的时候，梯形公式比Monte Carlo方法精度高。

Outline

Generation of uniform distribution $\mathcal{U}(0,1)$

- (1) Von Neumann的“平方取中”法(midsquare):
  在计算机发明的早期，Von Neumann等人为使用伪随机数提出了“平方取中”法。例如，

$$3333 \to 11108889$$

$$1088 \to 1183744$$

$$8374 \to 70123876$$

$$1238 \to \cdots$$

  显然，这一方法最大循环长度不超过10^4，而且其统计结果并不理想，但是这一算法就已在早期核反应计算中用到。

Generation of uniform distribution $\mathcal{U}(0,1)$

- (2) 线性同余法(linear congruential algorithm):
  在$\mathcal{U}[0,1]$的伪随机数发生器中，最通用的是所谓线性同余法，它们取如下的形式：
  $$X_{n+1} = aX_n + b (\text{mod } m) \quad (5)$$
  这里a, b, m是事先取定的自然数。衡量伪随机数的好坏一个重要标准是所谓最大循环长度(cycle length)，对线性同余法有下述定理：
- **定理:** 如果a, b, m的选择使得
  (i) b与m互素;
  (ii) $(a-1)$是m的任一奇数因子的倍数;
  (iii) 如果$m|4$，则$(a-1)|4$;
  那么上述伪随机数发生器的最大循环长度为m，即满长度。
  满足上述定理的一个自然的选择为：
  $$m = 2^k, \quad a = 4c+1, \quad b\text{为奇数}$$

Generation of uniform distribution $\mathcal{U}(0,1)$

- (3) 神奇的“16807”
  1969年，Lewis, Goodman和Miller提出了下述发生器：

$$X_{n+1} = aX_n \pmod{m} \tag{6}$$

  并且取$a = 7^5 = 16807, m = 2^{31} - 1 = 2147483647$. Shrage给出了一个在计算机上高效实现上述乘法同余的算法。这样得到的伪随机数发生器循环长度可达到2.1×10^9!
- L'Ecuyer采用所谓Bays-Durham洗牌算法给出了一个更为强大的随机数发生器，其循环长度达到约2.3×10^{18}！在Numerical Recipe中，给出了这一算法的具体实现程序ran2().该书作者声称，如果有人能给出使用上述算法而导致系统性失败的案例，将付款1000美元!

Generation of uniform distribution $\mathcal{U}(0,1)$

- Generation of $\mathcal{U}(0,1)$ with MATLAB
- Histograms

Law of Large Numbers (LLN)

- **Theorem**

  (Weak Law of Large Numbers, WLLN) If $\mathbb{E}|X_i| < +\infty$, *then*

$$\frac{S_n}{n} \longrightarrow \eta$$

  in probability.

- **Theorem**

  (Strong Law of Large Numbers, SLLN)

$$\frac{S_n}{n} \longrightarrow \eta \quad \textit{a.s.}$$

  if and only if $\mathbb{E}|X_i| < +\infty$

Central Limit Theorem (CLT)

- **Theorem**

  (Central Limit Theorem, CLT) Assume that $\mathbb{E}X_i^2 < +\infty$ *and let* $\sigma^2 = \mathrm{Var}(X_i)$. *Then*

  $$\frac{S_n - n\eta}{\sqrt{n\sigma^2}} \to N(0,1)$$

  in the sense of distribution.

Generation of general RVs

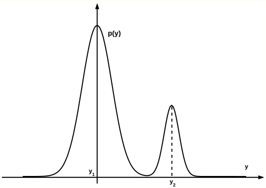
- (1) 变换法：(Transformation method)

  命题:设随机变量Y的分布函数为$F(y)$，即

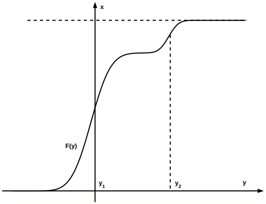
$$P\{Y \leq y\} = F(y) \tag{7}$$

  如随机变量$X \sim \mathcal{U}[0,1]$，则$Y = F^{-1}(X)$就满足所要求的分布。

- 几何解释



Y分布密度示意图



Y分布函数示意图

Generation of exponentially distributed RVs

- (i) 指数分布：

$$p(y) = \begin{cases} 0 & y \leq 0 \\ \lambda e^{-\lambda y} & y \geq 0 \end{cases} \tag{8}$$

其分布函数$F(y) = \int_0^y p(z)dz = 1 - e^{-\lambda y}$，从而$F^{-1}(x) = -\frac{1}{\lambda}\ln(1-x), x \in (0,1)$。

- 由变换法，指数分布的随机变量可由公式

$$Y_i = -\frac{1}{\lambda}\ln(1 - X_i) \quad i = 1, 2, \ldots \tag{9}$$

产生，这里$X_i \sim \mathcal{U}(0,1)$。

Generation of exponentially distributed RVs

- Generation of exponentially distributed RVs
- Histograms
- Central Limit Theorem

Generation of normally distributed RVs

- (ii) 正态分布：

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{10}$$

分布函数为

$$F(x) = \int_{-\infty}^{x} p(y)dy = \frac{1}{2} + \frac{1}{2}\mathrm{erf}(\frac{x}{\sqrt{2}}) \tag{11}$$

这里$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$为误差函数(error function)。因此$F^{-1}(x) = \sqrt{2}\mathrm{erf}^{-1}(2x-1)$。但是变换法此时在计算机上不易实现，因为$\mathrm{erf}^{-1}$难于计算！这也表明了变换法的局限性。为生成标准正态分布，我们介绍著名的Box-Muller方法。

Generation of normally distributed RVs

- (2) Box-Muler方法（标准正态分布）：
  为生成标准正态分布随机变量，考虑在微积分中求$\int_{-\infty}^{+\infty} e^{-x^2}dx$的技巧：

$$\left(\int_{-\infty}^{+\infty} e^{-x^2}dx\right)^2 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-(x^2+y^2)}dxdy = \int_{0}^{+\infty}\int_{0}^{2\pi} e^{-r^2}rdrd\theta = \pi \qquad (12)$$

  即将一个一维积分变成二维积分之后再采取极坐标换元策略。
- Box-Muller方法也采用这一思想。令$(x_1, x_2) = (r\cos\theta, r\sin\theta)$，则：

$$\frac{1}{2\pi}e^{-\frac{x_1^2+x_2^2}{2}}dx_1dx_2 = \frac{1}{2\pi}e^{-\frac{r^2}{2}}rdrd\theta = \left(\frac{1}{2\pi}d\theta\right)\cdot\left(e^{-\frac{r^2}{2}}rdr\right) \qquad (13)$$

  这样将一个二维正态分布的生成转变为对θ和r的生成。密度$\frac{1}{2\pi}$对应于θ方向的$\mathcal{U}[0,2\pi]$，而密度$e^{-\frac{r^2}{2}}r$对应于r方向的分布函数$F(r) = \int_0^r e^{-\frac{s^2}{2}}sds = 1 - e^{-\frac{r^2}{2}}$，这正好可使用变换法！

Generation of normally distributed RVs

- 于是二维正态分布的随机数产生可先选取相互独立的随机数$X_1, X_2 \sim \mathcal{U}[0,1]$，然后利用

$$\begin{cases} Y_1 &= \sqrt{-2\ln X_1}\cos(2\pi X_2) \\ Y_2 &= \sqrt{-2\ln X_1}\sin(2\pi X_2) \end{cases} \tag{14}$$

产生(Y_1, Y_2)。该算法在Numerical Recipe中有程序。

- Generation of Gaussian RVs with MATLAB
- Histogram

Homework assignment

- Generate the uniform distribution, exponential and Gaussian random variables and test the Weak Law of Large Numbers and Central Limit Theorem with MATLAB.

References

- R.E. Caflish, Monte Carlo and Quasi-Monte Carlo methods, Acta Numerica, Vol. 7, 1-49, 1998.
- 汪仁官, 初等概率论, 北京大学出版社.