2.6 随机游走

随机游走也是一种基于运用[0，1]区间的均匀分布随机数序列来进行的计算。

醉汉行走问题

醉汉开始从一根电线杆的位置出发（其坐标为$x=0$，x坐标向右为正，向左为负），假定醉汉的步长为l，他走的每一步的取向是随机的，与前一步的方向无关。如果醉汉在每个时间间隔内向右行走的一步的几率为p，则向左走一步的几率为$q=1-p$。我们记录醉汉向右走了n_R步，向左走了n_L步，即总共走了$N=n_R+n_L$步。那末醉汉在行走了N步以后，离电线杆的距离为$x=(n_R-n_L)l$，其中$-Nl\le x\le Nl$。然而我们更感兴趣的是醉汉在行走N步以后，离电线杆的距离为x的概率$P_N(x)$。

下面便是醉汉在走了N步后的位移和方差的平均值（$<x_N>,<\Delta x_N^2>$）的计算公式。

$$<x_N>=\sum_{x=-Nl}^{Nl}xP_N(x),$$

$$<\Delta x_N^2>=<x_N^2>-<x_N>^2,$$

其中 $$<x_N^2>=\sum_{x=-Nl}^{Nl}x^2P_N(x).$$

公式中的求平均是指对N步中所有可能的行走过程的平均。

$<x_N>=(p-q)Nl$，　　$<\Delta x_N^2>=4pqNl^2$.

注意到在向左、向右对称的情况下，即$p=q=1/2$，得到$<x_N>=0$。

查点法和蒙特卡洛方法

在查点法中，对给定的行走总步数N及总位移x，要求把游走时可能的每一步的坐标和几率都确定下来。这是可以用概率理论精确计算的。

例如，对于$N=3$，$l=1$的醉汉一维行走问题，由概率理论可以得到$P_3(x=-3)=q^3$，$P_3(x=-1)=3pq^2$，$P_3(x=1)=3p^2q$，$P_3(x=3)=q^3$，由此可以算出

$$<x_3>=\sum xP_3(x)=-3q^3-3pq^2+3p^2q+3p^3=3(p-q),$$

$$<x_3{}^2>=\sum x^2P_3(x)=9q^3+3pq^2+3p^2q+9p^3=12pq+[3(p-q)]^2.$$

则 $$<\Delta x_3^2>=<x_3^2>-<x_3>^2=12pq.$$

查点法只有在总步数N较小时才可以使用。N比较大时用起来

就比较困难了。

蒙特卡洛方法就可以克服在游走中的这个困难，具有更广泛的可操作性。蒙特卡洛方法可以对许多步的游走过程进行抽样，例如$N \sim 10^2 - 10^5$。我们可以按照正确的概率，对确定的N产生出各种可能的行走样本。原则上只要我们增加抽样的个数，要达到较高的精度总是可能的。

随机游走的蒙特卡洛方法求解泊松型微分方程

$$\begin{cases} \dfrac{\partial^2 \phi}{\partial x^2} + \dfrac{\partial^2 \phi}{\partial y^2} = q(x,y) \\ \phi|_\Gamma = F(s) \end{cases}$$ ， Γ为求解区域 D 的边界，s 为边界Γ上的点。

这里我们采用等步长h的正方形格点划分的差分法。在区域 D 内的任意正则内点 0 (其相邻的节点都在区域 D 内)的函数值可以用周围四个邻近点 1，2，3，4 上的函数值来表示。如同在第四章中将要介绍的，这个表达式有如下差分方程表示

$$\phi_0 = \frac{1}{4}\left(\phi_1 + \phi_2 + \phi_3 + \phi_4 - h^2 q_0\right) .$$

其中q_0是在区域 D 的正则内点 0 上的函数$q(x,y)$的值。公式右边的系数 1/4 可以解释为概率。即我们有

$$\phi_0 = \sum_{j=1}^{4} W_{0,j}\phi_j - \frac{h^2}{4} q_0 , \quad \sum_{j=1}^{4} W_{0j} = 1 , \quad W_{0,j} = \frac{1}{4}, (j = 1,2,3,4) .$$

游走的判据是：选定一个[0，1]区间的均匀分布的随机数ξ，若满足条件$\xi \le \frac{1}{4}$，我们选定下一个游走到达点为第 1 点；若满足条件$\frac{1}{4} < \xi \le \frac{1}{2}$，选游走到的下一个点为 2 点；若满足条件$\frac{1}{2} < \xi \le \frac{3}{4}$，选定游走到下一个点为 3 点；$\xi$在其他的情况下，我们则选游走到第 4 点。

如果我们按上面的判据选择了 0 点周围四个点中之一m点，则 0 点函数ϕ_0的估计值为$\eta_o = \phi_m - \frac{h^2}{4} q_0$；

从m点上又按判据选择周围四个点中的n点时，m点函数ϕ_m的估计值为$\eta_m = \phi_n - \frac{h^2}{4} q_m$，此时 0 点函数$\phi_0$的估计值也可以写为

$\eta_o = \phi_n - \frac{h^2}{4}(q_0 + q_m)$，……。

按上面的原则和步骤，如果从 0 点开始进行游走并记下该

点函数值$q_0=q_o^{(1)}$；在第 j 步游走到第 j 点时，记下该点 q(x,y)的函数值$q_j^{(1)}$；直到该游走到第$J^{(1)}$步，到达边界Γ的$s^{(1)}$点时，停止该次游走，记下边界上这点的函数值$F(s^{(1)})$。此时我们可以得到 0 点上的函数ϕ_0的一个估计值

$$\eta_0^{(1)}=F(s^{(1)})-\frac{h^2}{4}\sum_{j=0}^{J^{(1)}}q_j^{(1)} \quad .$$

如此反复从 0 点开始进行 N 次上述的随机游走,我们得到一个函数ϕ_0的估计值序列

$$\left\{\eta_0^{(1)},\eta_0^{(2)},\ldots\eta_0^{(n)},\ldots\eta_0^{(N)}\right\},$$

其中

$$\eta_0^{(n)}=F(s^{(n)})-\frac{h^2}{4}\sum_{j=0}^{J^{(n)}}q_j^{(n)}, \qquad n=1,2,...,N.$$

则 0 点的函数ϕ_0的期望值为

$$\overline{\phi}_0=E\{\eta_0\}\approx\frac{\sum_{n=1}^{N}\eta_0^{(n)}}{N}=\frac{\sum_{n=1}^{N}\left[F(s^{(n)})-\frac{h^2}{4}\sum_{j=0}^{J^{(n)}}q_j^{(n)}\right]}{N}.$$

这个计算出的ϕ_0值的估计值序列的方差为

$$\sigma^2=\frac{N}{N-1}\left[<\eta_0^2>-E\{\eta_0^2\}\right].$$

这种随机游走的做法，实际上是个人为的概率过程。它是一个具有吸收壁的随机游走。

上面这种方法可以推广应用到更一般的二维、三维的椭圆形方程的求解。在所需求解方程的边界条件特别复杂，而我们所需求解的仅仅是系统中的若干点的函数值时，该方法是可供选择的有效方法。

在随机游走的蒙特卡洛方法中，有一种最常用方法称为 Metropolis 方法。它是前面介绍过的重要抽样法的一个特殊情况。采用此方法可以产生任意分布的随机数，包括无法归一化的分布密度函数。

以一维的 Metropolis 方法为例,它所采用的游走规则是选择一个从x点游走到x'点的“过渡几率”$w(x\to x')$，使得它在游走中所走过的点$x_0,x_1,x_2.....$的分布收敛到系统达到平衡时的分布$f(x)$。要达到这样分布的重要抽样，就需要对过渡几率$w(x,x')$的选择加上适当的限制。

可以证明：只要游走所选的“过渡几率”满足如下的细致平

衡条件，

$$f(x)w(x\to x')=f(x')w(x'\to x)\quad .$$

就可以达到平衡时的分布为 $f(x)$ 这样的目的。

实际上满足细致平衡条件只是一个充分条件，并不是一个必要条件。该条件并不能唯一地确定过渡几率 $w(x\to x')$。所以，过渡几率 $w(x\to x')$ 的选择具有很大的自由度。选取不同的过渡几率就是不同的游走方法。

Metropolis 方法采用一个简单的选择过渡几率的方法，即

$$w(x\to x')=\min\left[1,\frac{f(x')}{f(x)}\right].$$

具体操作：

(1)首先选取一个试探位置，假定该点位置为 $x_{try}=x_n+\eta_n$，其中 η_n 为在间隔 $[-\delta,\delta]$ 内均匀分布的随机数。

(2)计算 $r=\frac{f(x_{try})}{f(x_n)}$ 的数值。

(3)如果不等式 $r\geq 1$ 满足（由公式(2.6.15)可以知道：此时 $w(x_n\to x_{try})=1,\quad w(x_{try}\to x_n)=1/r$），那就接受这一步游走，并取 $x_{n+1}=x_{try}$。然后返回（1）开始对游走到 x_{n+2} 点的试探。

(4)如果 $r<1$（此时，$w(x_n\to x_{try})=r$，$w(x_{try}\to x_n)=1$），那么就再另产生一个[0，1]区间均匀分布的随机数 ξ。

(5)如果此时 $\xi\leq r$，那么也还接受这步游走，并取这步游走所到达的点为 $x_{n+1}=x_{try}$。然后返回到步骤(1)，开始下一步到达 x_{n+2} 点的游走。

(6)如果此时 $\xi>r$，就拒绝游走到 x_{try} 这一点，即仍留在 x_n 点的位置不变。

(7)返回到步骤(1)，重新开始对游走到 x_{n+1} 点的具体位置的又一次试探。

采用这样的游走过程时，只有在产生了大量的点 $x_0,x_1,x_2.....$ 后，才能得到收敛到满足分布 $f(x)$ 的点集。

如何选择 δ 的大小，以提高游走的效率?

δ 选得太大，那么绝大部分试探的步子都将会被舍弃，就很难达到平衡分布；

δ 取得太小，那么绝大部分试探步子都会被接受，这同

样难以达到所要求的平衡分布。

根据实际应用中的经验，选取δ的一个粗略标准应当是：选择适当δ大小的原则是要在游走的试探过程中，有 1/3 到 1/2 的试探步子将被接受。

按照这样的标准选择得到的δ，就可以大大提高游走的效率。

进行这样的随机游走，从哪一点出发才可以比较快地达到平衡分布呢?

原则上讲，从任何一个初始位置出发均可达到平衡分布，但是为了尽快地达到平衡分布，我们最好是要选择一个合适的初始位置，这个初始位置应当是在游走范围内所要求的几率分布密度$f(x)$最大的区域。