2．4 蒙特卡洛计算中减少方差的技巧

蒙特卡洛求积分的方差为

$$\sigma^2 = V\{f\}/n \quad .$$

其中$V\{f\}$为被积函数f的方差。

公式反映出增加随机点数n时，蒙特卡洛计算的精度可以得到改善，但是精度提高非常缓慢。因此用增加蒙特卡洛计算的随机投点数来提高精度总是耗费大量的机时。

另一个减少计算结果误差的途径是减少f的方差$V\{f\}$。

重要的减少方差$V\{f\}$的技巧。

一、 分层抽样(stratified sampling)

数学上，分层抽样是基于黎曼积分的特性：

$$I = \int_0^1 f(x)dx = \int_0^a f(x)dx + \int_a^1 f(x)dx \text{，} \qquad 0 < a < 1$$

将积分区域划分成小区域是在数值积分中常用的技巧。

蒙特卡洛的分层抽样技巧的抽样步骤：

(1) 将积分区间(或空间)划分为不相交的子区间(或子空间)；然后在第i个子区间（或子空间）内抽取n_i个随机点。

(2) 如果将子区间长度（或子空间体积）记为$\{i\}$，我们将子区间（或子空间）内所有点上的函数值乘上权重因子$\{i\}/n_i$之后迭加起来，就得到该积分在这个子区间的积分估计值；

(3) 将所有子区间的积分值迭加起来，就得到在整个区间的积分估计值。这样得到积分的估计值的方差为：

$$V\{\bar{I}\}=\sum_j\left(\frac{\{j\}}{n_j}\right)^2\sum_{i=1}^{n_j}V\{f(x_{ij})\}=\sum_j\frac{\{j\}^2}{n_j}\sigma_j^2$$

如果适当选择子区间$\{i\}$的大小和随机点数n_i，就可以使计算结果的方差得以减小。这里选择$\{i\}$和n_i的关键是要了解被积函数f在子区间内的特性。如果$\{i\}$的划分和n_i的选择都不适当，也可能造成更大的误差。

如果我们不管被积函数的特性,而简单地将积分区域划分成相等的子区间$\{i\}$，并在各子区间内抽取相同数量的随机点数n_i。这种处理方法称为均匀分层抽样法。

求一维定积分的问题,比较一下用分层抽样法和用原始蒙特卡洛方法计算得到的方差。

设所求积分为：

$$I=\int_0^1 f(x)dx \quad .$$

数学上可以写成

$$I=\int_0^1 f(x)dx\equiv\int_0^1 g(x)f_1(x)dx \quad .$$

在[0,1]区间插入 J 个点，其中$0=x_0<x_1<\cdots<x_J=1$。令

$$\begin{cases} p_j=\int_{x_{j-1}}^{x_j} f_1(x)dx, \\ \bar{f}_j(x)=\begin{cases} f_1(x)/p_j, x_{j-1}\le x<x_j, \\ 0,\text{其它} \end{cases} \\ I_j=\int_{x_{j-1}}^{x_j} g(x)\bar{f}_j(x)dx,(j=1,2,\cdots,J) \end{cases}$$

在上面的公式中，显然有关系式

$$I=\sum_{j=1}^{J}p_jI_j \quad .$$

如果用分层抽样蒙特卡洛方法计算的积分值，在第j个子区间上

以 $\bar{f}_j(x)$ 分布密度函数抽取 n_j 个简单子样 $x_{ij}(j=1,2,\cdots,J)$，则积分的无偏估计值为

$$\bar{I}_J=\sum_{j=1}^{J}p_jI_j=\sum_{j=1}^{J}p_j\left(\frac{1}{n_j}\sum_{i=1}^{n_j}g(x_{ij})\right).$$

令第 j 区间积分的方差为 σ_j^2，根据方差的定义我们有关系式

$$\sigma_j^2=V\{g(x_{ij})\}=\int_{x_{j-1}}^{x_j}g^2(x)\bar{f}_j(x)dx-I_j^2 \quad .$$

则得到分层抽样计算结果的方差 $V\{\bar{I}_J\}$ 为：

$$V\{\bar{I}_J\}=\sum_{j=1}^{J}p_j^2\frac{1}{n_j^2}\cdot\sum_{i=1}^{n_j}V\{g(x_{ij})\}=\sum_{j=1}^{J}\frac{p_j^2}{n_j}\sigma_j^2 .$$

如果用通常的原始蒙特卡洛方法计算，以分布密度函数 $f_1(x)$ 抽取 N 个简单子样，则积分的无偏估计值为：

$$\bar{I}=\frac{1}{N}\sum_{i=1}^{N}g(x_i)$$

它的方差为：

$$V\{\bar{I}\}=\frac{1}{N^2}\sum_{i=1}^{N}V\{g(x_i)\}\equiv\frac{\sigma_g^2}{N}$$

其中 σ_g^2 又可以表示为

$$\sigma_g^2\equiv\int_0^1[g(x)-I]^2f_1(x)dx=\sum_{j=1}^{J}\int_{x_{j-1}}^{x_j}[g(x)-I]^2f_1(x)dx$$

$$=\sum_{j=1}^{J}p_j\int_{x_{j-1}}^{x_j}[g(x)-I_J+I_J-I]^2\bar{f}_j(x)dx$$

$$=\sum_{j=1}^{J}p_j\int_{x_{j-1}}^{x_j}\left[(g(x)-I_j)^2+2(I_j-I)(g(x)-I_j)+(I_j-I)^2\right]\bar{f}_j(x)dx$$

$$=\sum_{j=1}^{J}p_j\sigma_j^2+\sum_{j=1}^{J}p_j(I_j-I)^2 \quad .$$

设分层抽样法的总抽样数为 N ，我们有

$$N=n_1+n_2+\cdots+n_J \quad .$$

比较这两种方法计算出的结果的方差，我们有

$$V\{\bar{I}\}-V\{\bar{I}_J\}=\frac{1}{N}\left[\sum_{j=1}^{J}p_j\sigma_j^2+p_j(I_j-I)^2\right]-\sum_{j=1}^{J}\frac{p_j^2}{n_j}\sigma_j^2$$

$$=\sum_{j=1}^{J}p_j\left(\frac{1}{N}-\frac{p_j}{n_j}\right)\sigma_j^2+\frac{1}{N}\sum_{j=1}^{J}p_j(I_j-I)^2 \text{.}$$

公式的右边第二项显然是大于零的量。第一项的正负则是取决于分层抽样时子区间的划分和子区间内的抽样点数n_j。

如果上式的值大于零，则分层抽样计算积分的方差小于采用原始蒙特卡洛方法的方差。若取$\frac{p_j}{n_j}=\frac{1}{N}$，即$n_j=Np_j$，此时公式(2.4.14)中第一项为零，公式（2.4.14）总是大于零。这就意味着按比例的分层抽样的方差比原始蒙特卡洛方法小。这样的分层抽样方法具有实用意义。

如果采用均匀分层抽样方法，将[0，1]区间分成 J 个相等的子区间，每个子区间内抽取的点数$n_j=\frac{N}{J}$，并且这些点是均匀分布的，即$f_1(x)=1, p_j=\frac{1}{J}$，这时公式(2.4.14)中的第一项也为零，因而(2.4.14)式的值总是正的。

由此我们也可以看出：均匀分层抽样法是一个减小方差的保险方法。

二、 重要抽样法(importance sampling)

重要抽样法的原理起源于数学上的变量代换方法的思想，即

$$\int_0^1 f(x)dx=\int_0^1\frac{f(x)}{g(x)}g(x)dx=\int_0^1\frac{f(x)}{g(x)}dG(x) \text{.}$$

此时随机点的选择不再是均匀的，而是以分布函数$G(x)$分布的。新的被积函数为$f(x)$乘以权重$1/g(x)$。公式（2.4.15）中$g(x)=\frac{dG(x)}{dx}$。这里$g(x)$称为偏倚分布密度函数。该方法使原本对$f(x)$的抽样，变成由另一个分布密度函数$f^*(x)\equiv\frac{f(x)}{g(x)}$中产生简单子样，并附带一个权重$g(x)$。这种方法也称为偏倚抽样法。

公式右边积分中被积函数的方差为$V\{f/g\}$。如果$g(x)$选择恰当，并使它在积分域内的函数曲线形状与f接近，则该方差可以变得很小。

函数$g(x)$应当满足如下条件：

（1）$g(x)$应当是个分布密度函数。

（2）$f(x)/g(x)$不应在积分域内起伏太大，使之尽量等于常数，以保证方差$V\{f/g\}$比$V\{f\}$小。

（3）分布密度函数$g(x)$所对应的分布函数$G(x)$能够比较方便地解析求出。

（4）能方便地产生在积分域内满足分布函数$G(x)$分布的随机点。

如能按上述条件找到函数$g(x)$，我们就可以依下列步骤求积分：

(1) 根据分布密度函数$g(x)$产生随机点 x．例如采用反函数法。

(2) 求出各抽样点 x 的函数值$f(x)/g(x)$，并将所有点上的该函数值迭加起来，再除以抽样点数n就得到积分结果。

也可以采用$w\equiv f(x)/g(x)$作为分布密度函数，利用舍选法来舍去

或接受个随机点的 x 的值。用此方法时，应至少可以事先判断出 w 的最大值。当然最好能从 $f(x)/g(x)$ 的函数中，推导出 $w_{\max}$。

上述讨论可以很容易地推广到更高维的积分中。但是要注意如下两个方面的问题：

- 在产生随机向量 $\vec{x}$ 的所有分量后，再用舍选法往往更快，效率更高。
- 在计算 $f(\vec{x})/g(\vec{x})$ 值之前，做随机变量 $x_1,x_2,\cdots,x_N$ 到 $y_1,y_2,\cdots,y_N$ 的变换有时是很有用的。这时需要将雅可比行列式 $\left|\partial(x_1,x_2,\cdots,x_N)/\partial(y_1,y_2,\cdots,y_N)\right|$ 包括在权重因子内。

重要抽样法无疑是蒙特卡洛计算中最基本和常用的技巧之一。它无论在提高计算速度和增加数值结果的稳定性方面都有很大的潜力。

局限性：

（1）　能寻找出某分布密度函数 $g(x)$，并能解析求出其对应的分布函数 $G(x)$ 的情况并不多。当然我们也可以用数值计算方法求出 $G(x)$，但通常这样处理不灵活，运算速度也慢，而且结果也不准确。

（2）　当所选择的 $g(x)$ 在某点函数值为零或很快趋于零时（如高斯分布），这时在该点的数值计算是十分危险的。其方差 $V\{f/g\}$ 可能趋于无穷大。即使是在某点上函数 $g(x)$ 不为零，但其值很小时，方差 $V\{f/g\}$ 也可能很大。这一问题采用通常的从样本点估计方差的方法却不一定能检查出来。这种情况会使计算结果不稳定。

三、 控制变量法（相关抽样法）(control variates)

控制变量法利用数学上积分运算的线性特性：

$$\int f(x)dx = \int [f(x)-g(x)]dx + \int g(x)dx$$

选择函数 $g(x)$ 时要考虑到：$g(x)$ 在整个积分区间都是容易精确算出，并且在上式右边第一项的运算中对 $(f-g)$ 积分的方差应当要比第二项对 f 积分的方差小。

在应用这种方法时，在重要抽样法中所遇到的，当 $g(x)$ 趋于零时，被积函数 $(f-g)$ 趋于无穷大的困难就不再存在，因而计算出的结果稳定性比较好。该方法也不需要从分布密度函数 $g(x)$，解析求出分布函数 $G(x)$。 由此我们可以看出选择 $g(x)$ 所受到的限制比重要抽样法要小些。

四、对偶变量法(antithetic variates)

通常在蒙特卡洛计算中采用互相独立的随机点来进行计算。

对偶变量法中却使用相关联的点来进行计算。它利用相关点间的关系可以是正关联的，也可以是负关联的这个特点。

两个函数值 f_1 和 f_2 之和的方差为

$$V\{f_1+f_2\}=V\{f_1\}+V\{f_2\}+2E\{(f_1-E\{f_1\})(f_2-E\{f_2\})\}$$.

如果我们选择一些点，它们使 f_1 和 f_2 是负关联的。这样就可以使上式所示的方差减小。当然这需要对具体的函数 f_1 和 f_2 有充分的了解。

例　已知 $f(x)$ 是一个单调递增的函数，现求积分

$$I=\int_0^1 f(x)dx\ .$$

解　首先，按通常的方法在积分域[0，1]区间上产生均匀分布的随机点集 $\{x_i\}$。计算对应每个 x_i 点的函数 $[f(x_i)+f(1-x_i)]/2$ 的值，再将所有点上的函数值迭加起来，除以总的随机点数，则得到(2.4.18)式的积分值。即

$$I\approx\frac{1}{N}\sum_{i=1}^{N}[f(x_i)+f(1-x_i)]/2\quad .$$

这种做法与通常的蒙特卡洛计算中将 $f(x_i)$ 的值迭加起来不相同。由于 $f(x)$ 的单调递增性，$[f(x_i)+f(1-x_i)]/2$ 的值应当比单个点的函数值 $f(x_i)$ 更接近于常数。因而方差也小些。

这实际上是采用了 $f(x)$ 和 $f(1-x)$ 的积分期望值的平均值作为结果。由于采用相同的随机数列 $\{x_i\}$，使得 $f(x)$ 和 $f(1-x)$ 两个函数高度负关联，因而方差比 $f(x)$ 和 $f(1-x)$ 两者各自积分的方差之和要小。