

2. 5 实用蒙特卡洛计算复合技术

在蒙特卡洛方法应用中减小方差的基本技术：重要抽样法，分层抽样法，控制变量法和对偶变量法。然而，单独使用这四种减小方差的技巧仍然有其局限性。

人们发展了一些复合蒙特卡洛计算技术，如适应性蒙特卡洛方法和多道蒙特卡洛抽样方法等。这些蒙特卡洛技巧对于被积函数在积分范围内具有多个尖峰的情况，特别具有实用价值。

一、 适应性蒙特卡洛方法(adaptive Monte Carlo method)

适应性蒙特卡洛方法是一种在执行过程中通过试探，了解被积函数习性，然后有针对性地采用蒙特卡洛技巧来减少方差的算法。

采用此方法的子程序有利帕格(G. P. Lepage)的 VEGAS[5]。它是用于计算多重积分的子程序，广泛地应用在高能物理领域。VEGAS 编程的基本思想是将重要抽样法和分层抽样法结合到迭代算法之中，该算法能够做自动调整，将对被积函数的计算集中到被积函数值最大的区间。

以一维定积分为例，VEGAS 程序一开始处于试探阶段，即将积分区间 $[0,1]$ 划分为正交子区间，并在每一个子区间中进行积分；然后按照各个子区间积分得到的结果来调整子区间大小以备下一次迭代计算，调整子区间大小的原则是按照该子区间对总积分贡

献的大小来确定，贡献大的子区间调整得更小一些，贡献小的子区间调整得更大一些。

VEGAS 程序用这种方式实际上就是采用了重要抽样法。它采用阶梯函数来近似子区间的最佳几率密度函数。该最佳几率密度函数定义

$$p_0(x) = \frac{f(x)}{\int_0^1 f(x) dx}.$$

对于高维（d 维）积分问题，由于存贮的需要，我们必须采用分离变量的分布密度函数

$$p(u_1, u_2, \dots, u_d) = p(u_1) \cdot p(u_2) \cdot \dots \cdot p(u_d).$$

最后通过若干次迭代，达到在要求的精度下，各子区间（或子空间）的积分估计值都相等，则我们就找到了优化的子区间（或子空间）。在调整子区间（或子空间）过程中为了避免子区间（或子空间）剧烈的变化，子区间（子空间）大小的调整通常有一个衰减项。在程序的第二阶段，子区间（子空间）网格就固定下来了，并通过通常的蒙特卡洛方法得到在这些优化子区间（子空间）中迭代计算高精度的积分结果。每次迭代积分都得到一个估计值 $E\{I_j\}$ 和一个方差 $V\{I_j\}$ ：

$$E\{I_j\} = \frac{1}{N_j} \sum_{n=1}^{N_j} \frac{f(x_n)}{p(x_n)}, \quad V_j\{I_j\} = \frac{1}{N_j} \sum_{n=1}^{N_j} \left(\frac{f(x_n)}{p(x_n)} \right)^2 - E^2\{I_j\}.$$

其中 N_j 表示在第二阶段第 j 次迭代 ($j = 1, 2, \dots, m$) 积分的随机点个数。

在该阶段第 j 次积分值对总积分的贡献权重应当为

$$w_j = \frac{N_j}{V\{I_j\}}$$

将每次迭代的积分值乘上与投点数 N_j 和方差相关的权重，然后累加起来求平均。则总的积分估计值为

$$E\{I\} = \frac{\sum_{j=1}^m w_j E\{I_j\}}{\sum_{j=1}^m w_j} = \frac{\left(\sum_{j=1}^m \frac{N_j E\{I_j\}}{V\{I_j\}}\right)}{\left(\sum_{j=1}^m \frac{N_j}{V\{I_j\}}\right)}.$$

此外，VEGAS 程序返回时给出每个自由度(per degree of freedom)的 χ^2 为

$$\chi^2 / dof = \frac{1}{m-1} \sum_{j=1}^m \frac{(E\{I_j\} - E\{I\})^2}{V\{I_j\}}.$$

这个结果可以作为检验各种估计值是否一致。我们期望 χ^2 / dof 不要比 1 大很多。

按照上面的思想，VEGAS 程序计算高维积分的步骤可以概括如下：

(1) 将积分区域（或空间）划分为大量不相交的子区间（或子空间）。原则上可以任意划分，但为了方便起见，往往采用均匀划分的办法。

(2) 用原始蒙特卡洛方法估计每个子区间（或子空间）上的积分值，再将各个积分值迭加起来作为整个积分域上的估计值。

(3) 调整子区间（或子空间）的边界，使得被积函数在子区间（或子空间）内的积分估计值大致相等。

(4) 重复 (1) - (3) 的过程，利用原始蒙特卡洛方法计算每次迭代的积分估计值，直到在要求达到的精度下，各子区间（或子空间）的积分估计值都相等。此时才将得到的子区间（或子空间）

固定下来。以上为程序计算的第一阶段。在这一阶段，投点个数可以少一些，并不记这个阶段的积分结果（因为一般方差都很大）。

(5) 最后，采用蒙特卡洛方法，按照公式(2.5.3)计算各子区间（或子空间）积分值和方差，然后利用公式(2.5.5)将每次迭代计算的积分值加权累加平均得到该积分在总区间的积分估计值，用公式(2.5.6)计算每个自由度的 χ^2 。这就得到该积分的数值计算结果。这是程序计算的第二阶段。

二、多道蒙特卡洛抽样方法(multi-channel Monte Carlo method)

我们仍然以前面一维定积分为例，如果被积函数 $f(x)$ 在被积区间有尖峰，则采用原始的蒙特卡洛积分的结果误差肯定是很大的。当然，我们可以采用变量变换，按照重要抽样方法，将被积函数的峰值特性吸收到随机点的分布函数中。这种方法多少会使积分结果精度得到改善。但是，可能会有下面的情况：被积函数 $f(x)$ 在被积区间的不同区域有多个不同的尖峰。在这种情况下，往往不可能找到一个变量代换，它既吸收了被积函数所有的峰值特性，又比较容易进行按特定分布的随机数抽样。多道蒙特卡洛方法就是针对被积函数 $f(x)$ 具有多个尖峰情况下的计算方法。

它的基本思想是源于蒙特卡洛方法的迭加原则加上重要抽样法。该方法应用的前提是每个峰结构变换成的近似抽样分布密

度函数形式已经知道。每个峰的变换称为一个道。如果对应第 i 个道抽样，它所选择的分布密度函数为 $h_i(x)$ 。根据分布密度函数的正定和归一化，我们有： $\int_0^1 dx h_i(x) = 1$ ， $(i=1, \dots, m)$ ，其中 i 为道数。令 α_i 为非负的实数，并且满足

$$\sum_{i=1}^m \alpha_i = 1,$$

由于我们定义

$$h(x) = \sum_{i=1}^m \alpha_i h_i(x)$$

这就表明：对分布密度函数 $h(x)$ 抽样时，可以分别对 $h_i(x)$ 抽样。但是选择对第 i 个道的分布密度函数抽样的几率为 α_i 。明显地被积函数 $f(x)$ 与分布密度函数 $h(x)$ 具有同样的多峰值结构。利用关系式(2.5.8)，我们将积分 I 如做如下形式推导：

$$\begin{aligned} I &= \int_0^1 f(x) dx = \int_0^1 \left(\frac{f(x)}{h(x)} \right) h(x) dx = \sum_{i=1}^m \int_0^1 \left(\frac{f(x)}{h(x)} \right) \alpha_i h_i(x) dx \\ &= \sum_{i=1}^m \alpha_i \int_0^1 \left(\frac{f(x)}{h(x)} \right) dH_i(x). \end{aligned}$$

此时被积函数 $f(x)/h(x)$ 中已经没有原来所有的峰值特性了，这些峰值特性已经被分布密度函数 $h(x)$ 抵消。这就是我们多道蒙特卡罗方法计算积分的基本公式。从公式中可以看出：我们可以通过对各个道 i ，按分布密度函数 $h_i(x)$ （对应分布函数为 $H_i(x)$ ）产生随机数 x_{n_i} 。例如具体做抽样时可以用反函数法：

$$x_{n_i} = H_i^{-1}(\xi_{n_i})$$

实践中，我们按照离散型变量抽样法，以 α_i 为取分布密度函数 $h_i(x)$

抽样的概率。若固定总投点数为 N , 计算第 i 道时所投点次数大约是 $N_i \approx \alpha_i N$ 。采用这样的方法 , 可以得到积分的蒙特卡洛估计值为

$$E\{I\} = \frac{1}{N} \sum_{i=1}^m \sum_{n_i=1}^{N_i} \left(\frac{f(x_{n_i})}{h(x_{n_i})} \right).$$

该蒙特卡洛积分的误差期望值为

$$\sqrt{\frac{W(\alpha) - I^2}{N}}$$

其中 $W(\alpha)$ 定义为

$$W(\alpha) = \sum_{i=1}^m \alpha_i \int_0^1 \left(\frac{f(x)}{h(x)} \right)^2 dH_i(x).$$

该方法计算的关键之一是确定参数 α_i 。实践中往往通过调整参数 α_i , 使 $W(\alpha)$ 达到最小值来优化参数 α_i 的选择。理论上积分计算值 I 应与参数 α_i 值的选择无关 , 因此在积分过程中我们可以改变 α_i 的值 , 而不会影响积分值的估计。当然这是会影响计算结果的误差的。

建议首先采用一组初始参数 $\{\alpha'_i\}$, 在按照上面介绍的步骤进行几百次蒙特卡洛计算后 , 再计算

$$W_i(\alpha') = \int_0^1 \left(\frac{f(x)}{h(x)} \right)^2 h_i(x) dx$$

最后按照下面公式重新决定参数 α_i 。

$$\alpha_i = \frac{\alpha'_i (W(\alpha')_i)^\beta}{\sum_{j=1}^m \alpha'_j (W(\alpha')_j)^\beta}$$

根据经验 , 建议上面公式中的参数 β 值取在 $[1/4, 1/2]$ 之间。