Statistics and Application 统计学与应用, 2024, 13(1), 169-174
Published Online February 2024 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2024.131018

Hans汉斯

基于平均值法的二重积分数值模拟与方差缩减技术研究

伍豆豆，张　彤，赵惠华，张　伟，李　宏

兰州工商学院经济学院，甘肃 兰州

收稿日期：2024年1月3日；录用日期：2024年2月22日；发布日期：2024年2月28日

摘　要

针对重积分计算困难的问题，将一重积分数值计算的平均值法推广到二重积分上。采用该方法对二重积分进行了数值算法得到了积分的估计值，在此基础上根据中心极限定理得到了二重积分的区间估计。利用分层抽样法对二重积分估计值的方差进行缩减，模拟结果表明，分层抽样法相比一般方法能使估计值的方差缩减率达到50%左右，由此可见，等比例分层抽样效果较好。

关键词

蒙特卡罗积分法，二重积分，区间估计，分层抽样，数值仿真

Research on Double Integral Numerical Simulation and Variance Reduction Technique Based on Mean Value Method

Doudou Wu, Tong Zhang, Huihua Zhao, Wei Zhang, Hong Li

School of Economics, Lanzhou Technology and Business College, Lanzhou Gansu

Received: Jan. 3rd, 2024; accepted: Feb. 22nd, 2024; published: Feb. 28th, 2024

Abstract

In order to solve the problem of difficult calculation of multiple integrals, this paper extends the mean value method of numerical calculation of single integrals to double integrals. By using this method, the estimated value of the double integral is obtained by numerical algorithm. On this basis, the interval estimation of the double integral is obtained according to the central limit theorem. The stratified sampling method is used to reduce the variance of the double integral estimate. The simulation results show that the stratified sampling method can reduce the variance of the es-

文章引用: 伍豆豆, 张彤, 赵惠华, 张伟, 李宏. 基于平均值法的二重积分数值模拟与方差缩减技术研究[J]. 统计学与应用, 2024, 13(1): 169-174. DOI: 10.12677/sa.2024.131018

timate by about 50% compared with the general method. Therefore, the effect of equal proportional stratified sampling is better.

Keywords

Monte Carlo Integral Method, Double Integral, Interval Estimation, Stratified Sampling, Numerical Simulation

1. 引言

重积分的计算问题是统计计算与数学计算的核心问题之一，其计算的难度随着维数的增加呈指数增长。在许多物理问题和高维贝叶斯统计推断中，重积分很难直接用数学运算得到其值，基于这样的背景下，各种各样的数值近似计算应用而生。

其主要研究有，聂黎明，周永权(2009) [1]采用人工鱼群算法求解二重积分，将不等距离分割方法与人工鱼群算法相结合来求解数值积分同时在积分求和式子中引入一个参数，仿真结果表明，该方法具有更高阶的精度和更快的收敛速度。韦杏琼，周永权(2009) [2]基于粒子群算法对二重积分进行了数值计算，数值积分算例表明，该算法得到的积分值精度高，自适应性强。李满枝，王洪涛等(2011) [3]采用蒙特卡洛方法对二重积分进行了数值仿真并且将矩形区域上的二重积分推广到一般区域上其算法程序结构简单，易于编程和调试。王若鹏，夏赞勋等(2012) [4]采用 MATLAB 软件对二重积分进行了计算。邓泽喜，刘晓翼(2013) [5]应用差分进化算法对二重积分进行了计算。王洪涛，李满枝等(2014) [6]采用重要函数法对二重积分进行了数值计算。孙兰敏，岳亚楠(2015) [7]讨论了直线对称区域上二重积分的计算问题。徐小勇，周凤英(2017) [8]应用 Laguerre 小波算子矩阵对二重积分进行了数值计算。张孔生(2018) [9]基于 R 软件对二重积分进行了近似计算。何洪英，张世禄(2020) [10]研究了平面坐标系上二重积分的数值算法。娄汝馨，崔嵬(2022) [11]研究了圆域上二重积分数值计算的一种构造方法。

以上的研究都是对二重积分进行了数值计算，等同于统计学中未知参数的点估计。由于点估计不一定准确与可靠，本文既对二重积分进行了点估计又对其进行了区间估计，并且采用分层抽样方法对点估计的方差进行了缩减。文章安排如下，第一部分平均值法模拟定积分的原理，第二部分积分的精度也就是积分的区间估计，第三部分二重积分的点估计与区间估计，第四部分二重积分的方差缩减技术。

2. 平均值法模拟定积分的原理

考虑定积分 $I=\int_0^1 g(x)\mathrm{d}x$，这里不在采用微积分的方法对其进行积分，而采用数值模拟的方法对其进行数值模拟。对于任意的函数 $y=g(x)$，变量 y 的数学期望可表示为：

$$E(y)=E\left[g(x)\right]=\int_{-\infty}^{\infty} g(x)p(x)\mathrm{d}x$$

其中 $p(x)$ 为变量 x 的密度函数。

假设变量 x 服从(0,1)区间上的均匀分布，则其密度函数为：

$$p(x)=\begin{cases}1, & 0<x<1\\ 0, & \text{其他}\end{cases}$$

那么变量 y 的数学期望变为：

$$E(y)=E\left[g(x)\right]=\int_0^1 g(x)\mathrm{d}x$$

通过上式得到，定积分的值和变量 y 的期望值相等。而变量 y 的期望值可以通过算术平均值来估计，即用 $\frac{1}{n}\sum_{i=1}^{n}y_i$ 作为 $E(y)$ 的估计值。根据辛钦大数定律有：

$$\lim_{n\to+\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}y_i-E(y)\right|<\varepsilon\right\}=\lim_{n\to+\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}g(x_i)-E\left[g(x)\right]\right|<\varepsilon\right\}=1$$

即样本平均值依概率收敛到变量的期望值，也就是说样本平均值是变量 y 的数学期望的一致估计量。那么定积分值可以由被积函数的平均值来近似的估计：

$$I=\int_0^1 g(x)\mathrm{d}x=E\left[g(x)\right]\approx\frac{1}{n}\sum_{i=1}^{n}g(x_i)$$

概括起来平均值法模拟定积分的算法步骤为：

1、生成(0,1)区间上的均匀分布随机数 X 共 n 个为： $x_1,x_2,\cdots,x_n$；

2、将每一个 x_i 作为自变量带入函数 $y=g(x)$ 中，可以得到 n 个函数值 $y_1,y_2,\cdots,y_n$，其中 $y_i=g(x_i)$；

3、计算变量 y 的平均值 $\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i=\frac{1}{n}\sum_{i=1}^{n}g(x_i)$；

4、根据辛钦大数定律，将 $\bar{y}$ 作为定积分 I 的近似值。

前面模拟的(0,1)区间上的定积分，对于 $I=\int_a^b g(x)\mathrm{d}x$ 是否可以采用平均值法对其进行模拟。答案是肯定的，首先对积分进行变量变换。

令

$y=\frac{1}{b-a}(x-a)$，则当 x 从 a 变到 b 时，y 从 0 变到 1，并且 $x=(b-a)y+a$， $\mathrm{d}x=(b-a)\mathrm{d}y$ 有

$$I=\int_a^b g(x)\mathrm{d}x=\int_0^1 g\left[(b-a)y+a\right](b-a)\mathrm{d}y$$

令 $h(y)=g\left[(b-a)y+a\right](b-a)$ 有

$$I=\int_a^b g(x)\mathrm{d}x=\int_0^1 h(y)\mathrm{d}y$$

通过上面可以看出，采用变量变换的方法将(a,b)区间上的定积分变为(0,1)区间上的定积分。因此，可以采用平均值法对其进行模拟计算。这样得到的仅仅是定积分的点估计，这个点估计的精度如何，接下来对定积分作区间估计，以此来讨论其精度。

3. 积分的精度

上述定积分的估计值 $\frac{1}{n}\sum_{i=1}^{n}g(x_i)$ 是随抽样而变化的随机变量，那么它服从什么分布？根据林德贝格－勒维中心极限定理可得该估计值的分布如下：

$$\frac{1}{n}\sum_{i=1}^{n}g(x_i)\sim N\left(\int_0^1 g(x)\mathrm{d}x,\frac{\sigma_g^2}{n}\right)$$

其中 σ_g^2 表示随即变量 $g(x_i)$ 的方差，总体的方差可能为已知，也可能为未知，当未知时用样本方差 s_g^2 代替 σ_g^2。所以在构造定积分的置信区间的时候分总体方差已知与未知两种情况，构造的置信区间分别如下：

σ_g 已知：$\left[\frac{1}{n}\sum_{i=1}^{n}g(x_i)-Z_{1-\frac{\alpha}{2}}\frac{\sigma_g}{\sqrt{n}},\frac{1}{n}\sum_{i=1}^{n}g(x_i)-Z_{1-\frac{\alpha}{2}}\frac{\sigma_g}{\sqrt{n}}\right]$

σ_g 未知：$\left[\frac{1}{n}\sum_{i=1}^{n}g(x_i)-t_{1-\frac{\alpha}{2}}(n-1)\frac{s_g}{\sqrt{n}},\frac{1}{n}\sum_{i=1}^{n}g(x_i)+t_{1-\frac{\alpha}{2}}(n-1)\frac{s_g}{\sqrt{n}}\right]$

其中 $Z_{1-\frac{\alpha}{2}}$ 为标准正态分布分位数，$t_{1-\frac{\alpha}{2}}(n-1)$ 为自由度 n-1 的 t 分布的分位数。

4. 二重积分的点估计与区间估计

数值积分法在多元积分中的应用显得更加突出，因为多数多重积分采用分部积分法或换元法很难积的。因此将上述一元函数的数值积分法推广到二元函数上。假设二重积分如下，

$$\theta=\int_0^1\int_0^1 g(x,y)\mathrm{d}x\mathrm{d}y$$

根据平均值法模拟上述定积分的关键是估计下面的数学期望。

$$\theta=E\left[g(U_1,U_2)\right]$$

其中 U_1,U_2 是相互独立的(0,1)上的均匀分布随机数。

因此产生 n 个样本观测值 $\{(x_i,y_i),i=1,2,\cdots,n\}$，将每一个观测值作为自变量代入函数解析式得到函数值，然后计算 n 个函数值的平均值即可作为二重积分的估计值。即 $\frac{\sum_{i=1}^{n}g(x_i,y_i)}{n}$ 作为积分估计值。

算例，采用平均值法对二重积分 $\int_0^1\int_0^1 \mathrm{e}^{(x+y)^2}\mathrm{d}x\mathrm{d}y$ 作出估计。

Table 1. Changes of estimates and variances of double integrals with sample size
表 1. 二重积分的估计值与方差随样本量的变换情况表

样本容量	100	500	1000	1500	2000	2500	3000	3500	4000
估计值	4.29	5.02	4.97	4.95	4.88	5.03	4.94	4.98	4.77
方差	0.26	0.077	0.035	0.024	0.017	0.015	0.012	0.010	0.008

从表 1 可以看出，二重积分估计值的方差随着样本量的增加在减小，但是当样本容量为 2000 时，再增大样本容量方差的缩减幅度降低。表 1 是二重积分的点估计，但是点估计不一定准确、可靠。下面对二重积分进行区间估计，这里将一重积分的区间估计推广到二重积分，在大多数情况下总体的方差是未知的，所以采用第二种情况得到二重积分置信度为 $1-\alpha$ 的置信区间为：

$$\left[\frac{1}{n}\sum_{i=1}^{n}g(x_i,y_i)-t_{1-\frac{\alpha}{2}}(n-1)\frac{s_g}{\sqrt{n}},\frac{1}{n}\sum_{i=1}^{n}g(x_i,y_i)+t_{1-\frac{\alpha}{2}}(n-1)\frac{s_g}{\sqrt{n}}\right]$$

其中 s_g 为样本的标准差，即 $s_g=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left[g(x_i,y_i)-\frac{1}{n}\sum_{i=1}^{n}g(x_i,y_i)\right]^2}$

接下来在 R 软件中编程得到 $\int_0^1\int_0^1 \mathrm{e}^{(x+y)^2}\mathrm{d}x\mathrm{d}y$ 的置信水平为 $1-\alpha=0.95$ 的置信区间。

从图 1 可以看出，当样本容量为 2000 时重积分的模拟值达到稳定状态，随着样本容量的增加重积分的精度没有发生显著性的改变。表 2 给出分布样本容量下重积分的置信下限与上线。

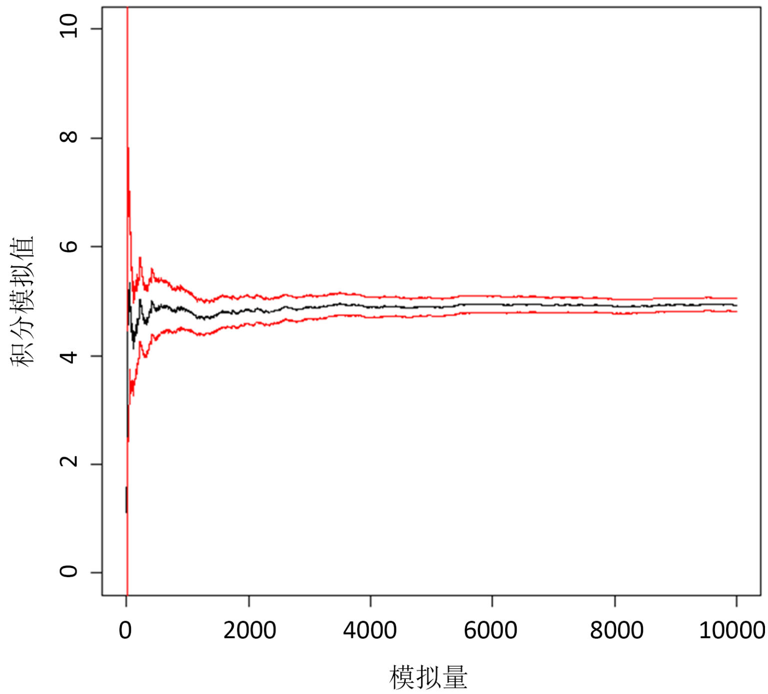

Figure 1. The trend of the simulated value of the double integral with the sample size
图 1. 重积分的模拟值随样本容量的变化趋势图

Table 2. Lists the lower confidence limits and upper limits of multiple integrals under partial sample size
表 2. 部分样本量下重积分的置信下限与上线表

样本容量	2000	4000	6000	8000	10000
置信下限	4.588	4.706	4.784	4.776	4.813
置信上线	5.100	5.077	5.087	5.036	5.047

5. 二重积分的方差缩减技术

本文采用分层抽样法对重积分估计值的方差进行缩减。分层抽样又称为类型抽样或分类抽样，首先将总体划分为若干层，然后在每一层中独立进行抽样，总的样本由各层样本组成，总体参数则根据各层样本参数的汇总作出估计，这种抽样就称为分层抽样，所得样本称为分层样本。设总的样本量为 n，从 L 个子总体中所抽取的样本量分别为 $n_1, n_2, \cdots, n_L$，则有

$$n_1 + n_2 + \cdots + n_L = n$$

对于重积分 $\int_0^1 \int_0^1 \mathrm{e}^{(x+y)^2} \mathrm{d}x\mathrm{d}y$，首先将积分区间等分为 4 层，每层概率 $p_i = 0.25$，并且按照等比例方式划分样本量。

一般方法：产生均匀分布随机数 $\{(x_i, y_i), i = 1, 2, \cdots, n\}$，代入 $\bar{g} = \frac{\sum_{i=1}^{n} g(x_i, y_i)}{n}$ 作为重积分的估计值。估计量的方差为 $\sigma^2 = \frac{\sigma_g^2}{n}$，$\sigma_g^2$ 通常用样本方差 s_g^2 代替，其中样本方差 $s_h^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[g(x_i, y_i) - \bar{g} \right]^2$。

采用分层等比例抽样方法，模拟量为 2000，分四层，即 $L = 4$，按照等比例方式，每层抽取样本量 500 个，每层概率 $p_i = 0.25$。这意味着分层变量 X, Y 的分布为(见表 3)。

Table 3. Distribution table of uniformly distributed random variables
表 3. 均匀分布随机变量的分布表

X	$\left[0,\frac{1}{4}\right]$	$\left[\frac{1}{4},\frac{2}{4}\right]$	$\left[\frac{2}{4},\frac{3}{4}\right]$	$\left[\frac{3}{4},1\right]$
Y	$\left[0,\frac{1}{4}\right]$	$\left[\frac{1}{4},\frac{2}{4}\right]$	$\left[\frac{2}{4},\frac{3}{4}\right]$	$\left[\frac{3}{4},1\right]$
p_i	0.25	0.25	0.25	0.25
$n_i = np_i$	500	500	500	500

下面在 R 软件中对该重积分进行模拟并且比较一般方法和分层抽样方法估计量的方差。一般方法估计量的方差为 0.01778808，分层抽样方法估计量的方差为 0.009119134，方差缩减率为 48.73%。由此可见，等比例分层抽样效果较好。

6. 结论

本文对于重积分的计算困难问题，将一重积分的平均值模拟法推广到二重积分上。采用该方法对二重积分进行了点估计与区间估计，并且得到了点估计与区间估计随样本量的变化规律。在点估计过程中采用分层抽样法对方差进行了缩减，分层抽样方法相比一般方法，估计量的方差缩减了 50%左右，说明估计量更有效。但是本文仅仅研究了矩形区域上的重积分，可以将该方法推广到一般区域上。在方差缩减方面也可以尝试使用对偶变量法、控制变量法、条件期望法与重要抽样等方法。这些技术与方法还需要更进一步的研究。

基金项目

本文由兰州工商学院 2023 年创新创业项目(S202313511020)资助。

参考文献

[1] 聂黎明, 周永权. 用人工鱼群算法求解二重数值积分[J]. 计算机工程与应用, 2009, 45(10): 34-37.
[2] 韦杏琼, 周永权. 基于粒子群算法的二重积分方法研究[J]. 计算机工程与设计, 2009, 30(20): 4705-4708.
[3] 李满枝, 王洪涛, 苗俊红. 二重积分的 Monte-Carlo 数值仿真[J]. 计算机仿真, 2011, 28(5): 114-117.
[4] 王若鹏, 夏赞勋, 谢鹏燕, 张鹏. 基于 MATLAB 的二重积分计算方法[J]. 高等数学研究, 2012, 15(2): 61-63.
[5] 邓泽喜, 刘晓翼. 求解二重积分问题的差分进化算法[J]. 计算机仿真, 2013(1): 319-322.
[6] 王洪涛, 李满枝, 沈有建. 二重积分的重要函数法模拟[J]. 海南广播电视大学学报, 2014(3): 143-149.
[7] 孙兰敏, 岳亚楠. 关于直线对称区域上二重积分的计算[J]. 衡水学院学报, 2015, 17(1): 8-11.
[8] 许小勇, 周凤英. 应用 Laguerre 小波算子矩阵求二重数值积分[J]. 计算机工程与应用, 2017, 53(16): 45-49.
[9] 张孔生. 基于 R 软件的二重积分近似计算[J]. 滁州学院学报, 2018, 20(5): 29-30.
[10] 何洪英, 张世禄. 坐标平面上的二重积分的数值算法[J]. 绵阳师范学院学报, 2020, 39(11): 11-17.
[11] 娄汝馨, 崔嵬. 圆域上二重积分数值计算的一种构造方法[J]. 保定学院学报, 2022, 35(6): 104-108.