

随机采样黎曼和：定积分数值算法

武国宁¹, 孙娜¹

(1. 中国石油大学-北京 理学院, 北京 102200)

[摘要] 定积分黎曼和需要从分割后的每一个小区间上任取一点 $\xi_i \in [x_{i-1}, x_i]$, 如果将积分的区间 n 等份, 若 ξ_i 为第 i 个小区间上服从均匀分布随机变量 X_i 的取值, 则黎曼和为随机变量的函数。在大数定律的启发下, 本文研究了当分割给定后, 前 n 次黎曼和的平均值的变化趋势。试验表明: 前 n 次黎曼和的平均值随着试验次数的增加趋于真实的积分值; 随着分割份数的增加前 n 次黎曼和的方差趋于零; 随着试验次数增加, 随机黎曼和的直方图呈正态分布。与蒙塔卡洛方法分析对比表明: 该方法具有较高的收敛速度和精度。

[关键词] 黎曼和; 随机变量; 大数定律; 中心极限定理; 蒙塔卡洛方法

[中图分类号] 0174.034 **[文献标识码]** C **[文章编号]**

1 引言

在一些工程模型求解中, 往往遇到以下问题: (1) 实际问题中有些函数无具体的表达式, 只知道函数在一些离散点的函数值。例如函数来自实验采样; (2) 有些函数的原函数不能够用基本初等函数表达, 例如椭圆积分等; (3) 一些函数的原函数需要采用函数级数或者函数的无限乘积来表示。还有的函数的原函数需要借助于特殊函数来表示, 例如含参变量积分中的伽马, 贝塔函数等等。以上这些问题使得数值积分具有了重要的意义。常见的数值积分方法有: (1) 利用函数在一些点的加权和来表示定积分。例如数值积分求解中的梯形公式、辛普森公式和高斯节点积分法等等^[1-2]。这些积分方法主要针对一元函数求解, 若求解的函数为多元函数, 虽然可以采用 Fubini 定理将重积分转换为一元函数积分。但是, 随着维数的增加, 计算量曾指数增长。(2) 另外一种数值积分法为蒙特卡洛方法^[3-4]。该方法采用随机投点的方式, 通过计算落入积分区域与包含积分区域的一个规则区域点的比例来计算定积分, 该方法提高了多元函数数值积分的计算速度, 应用较为广泛。

定积分定义为函数黎曼和的极限^[5-6], 对积分区域分割后, 黎曼和与在每个小区间上如何取点相关。若把每个小区间上的选点, 看作一个在该小区间上服从均匀分布的一个随机变量的取值, 则黎曼和为随机变量函数的函数。大数定律说明: 对于独立重复试验, 频率稳定于概率。当分割给定, 我们可以通过分析黎曼和的重复试验来研究函数的定积分。数值试验表明: 当分割份数给定, 随着试验次数的增多, 黎曼和的平均值收敛于函数的定积分; 随着分点个数的增加, 随机黎曼和呈现正态分布 (中心极限定理)。其均值随着试验次数的增多收敛于积分值, 其标准差随着试验次数的增加趋于零。

2 随机采样黎曼和

[收稿日期]: [修改日期]

[基金项目] 中国石油大学-北京, 教改项目

[作者简介] 武国宁 (1978-), 男, 博士, 副教授, 从事应用数学研究. Email: wuguoning@163.com

在微积分中，定积分的定义如下：

定义 1

设 $f(x)$ 为定义在 $[a, b]$ 上的有界函数，在 $[a, b]$ 中任取分点 $\{x_i\}_0^n$ ，做成一种划分：

$$P: a = x_0 < x_1 < x_2 < \dots < x_n = b$$

并任意取点 $\xi_i \in [x_{i-1}, x_i]$ 。记小区间的长度为 $\Delta x_i = x_i - x_{i-1}$ ，并令 $\lambda(P) = \max_{1 \leq i \leq n} \Delta x_i$ ，若极限

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n f(\xi_i) \Delta x_i,$$

存在，且极限值既与划分 P 无关，又与 ξ_i 的取法无关，则称 $f(x)$ 在 $[a, b]$ 黎曼可积。其极限值 I 称为 $f(x)$ 在 $[a, b]$ 上的定积分，记为：

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n f(\xi_i) \Delta x_i = I = \int_a^b f(x) dx.$$

在上述定义中，任意取点 $\xi_i \in [x_{i-1}, x_i]$ 这一要求，可以认为 $[x_{i-1}, x_i]$ 上服从均匀分布的一个随机变量 X_i 的取值。当函数 $f(x)$ 和划分 P 给定时，函数的黎曼和

$$\sigma(f; \xi, P) = \sum_{i=1}^n f(\xi_i) \Delta x_i,$$

可以看作随机变量 $X_i = \xi_i$ 的表达式（一次随机试验），这里 $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ 。

为分析黎曼和

$$\sigma(f; \mathbf{X}, P) = \sum_{i=1}^n f(X_i) \Delta x_i$$

作为随机向量 $\mathbf{X} = (X_1, X_2, \dots, X_n)$ 的函数，随着试验次数增加和随着分割份数增加的规律，我们可以从两个方面展开：

- (1) 当分割给定，通过增加重复试验的次数研究黎曼和的稳定性（频率稳定于概率）；
- (2) 当分割加细，随机变量增多时，黎曼和是否近似于正态分布（中心极限定理）。

为了回答以上问题，我们进行以下数值实验：讨论积分 $\int_0^1 x^2 dx$ ，这里我们把 $[0, 1]$ 区间 n 等份：

$$P: \frac{0}{n} < \frac{1}{n} < \frac{2}{n} < \dots < \frac{i-1}{n} < \frac{i}{n} < \dots < \frac{n}{n} = 1$$

假设 X_i 为区间 $[\frac{i-1}{n}, \frac{i}{n}]$ 上服从均匀分布的一个随机变量： $X_i \sim U([\frac{i-1}{n}, \frac{i}{n}])$ 。黎曼和

$$\sigma(f; \mathbf{X} = \xi^1, P) = \sum_{i=1}^n f(\xi_i^1) \Delta x_i$$

为一次随机试验，这里 $\xi^1 = (\xi_1^1, \xi_2^1, \dots, \xi_n^1)$ 。类似于投币试验，我们定义（频率）如下：

$$p_m = \frac{1}{m} \sum_{j=1}^m \sigma(f; \xi^j, P) = \frac{1}{m} \sum_{j=1}^m \sum_{i=1}^n f(\xi_i^j) \Delta x_i,$$

这里 $\xi^j = (\xi_1^j, \xi_2^j, \dots, \xi_n^j)$ 表示第 j 次随机试验各均匀分布随机变量的取值, p_m 表示 $[0, 1]$ 区间 n 等份后前 m 次随机试验黎曼和的平均值。

图 1 为 $n = 10$ (10 等份), 前 100 次试验黎曼和的平均值 p_{100} ; 图 2 为 $n = 10$ (10 等份), 前 10^4 次试验黎曼和的平均值。从图中可以看出, 随着试验次数的增加, 黎曼和的平均值趋于真实的积分值 $1/3$ 。这里验证了频率稳定于概率这一规律。

图 3 (a) 为 $n = 10$ (10 等份), 试验 10^4 次黎曼和的分布直方图; (b) 中上曲线为每次试验黎曼和的值的分布, 下面曲线为黎曼和与真值 $1/3$ 的误差; (c) 为 $n = 10$ (10 等份), 试验次数为 10^5 次黎曼和分布直方图; (d) 图中上面曲线为每次试验黎曼和值的分布, 下面为黎曼和与真实值 $1/3$ 的误差; (e) 为 $n = 100$ (100 等份), 试验次数为 10^4 次黎曼和分布直方图; (f) 图中上方曲线为每次随机试验黎曼和的值, 下方曲线为试验值和真实值 $1/3$ 的差。从图中可以看出黎曼和近似于正态分布, 该分布的中心为积分的精确值。随着试验次数的增加分布函数愈发趋于一个正态分布。当分割的份数为 100 时, 其计算的精度达到了较大的提高 (误差限由 0.02 减小为 0.001 左右)。

下面分析积分值随分割份数的变化趋势。图 4 为等份小区间的份数分别为 10, 20, 30, ..., 500。对于每个给定的分割份数, 试验次数均为 10^4 次。我们分析黎曼和平均值的变化趋势。从图中可以看出: 随着分割次数的增加, 黎曼和的均值收敛于真实的积分值。为了分析随机黎曼和与真实积分值的偏离程度, 图 5 分析了针对不同分割份数黎曼和的标准差。从图中可以看出: 随着分割次数的增加, 黎曼和的方差趋于很小的值。

图 6 为采用蒙塔卡洛随机试验对积分 $\int_0^1 x^2 dx$ 模拟得到的图形, 其中投点个数为 10^4 次。该方法得到的积分的近似值为 0.3300。

为了对比随机黎曼和平均值与蒙塔卡洛方法的收敛速度和精度, 我们做了以下对比试验: 对于随机黎曼和, 分割的份数为 10 份 ($n = 10$), 试验的次数分别为 10, 20, 30, ..., 10^3 次, 计算其平均值 $p_{10}, p_{20}, \dots, p_{10^3}$ 。对于蒙塔卡洛方法, 投点的个数分别为: 100, 200, 300, ..., 10^4 , 计算其近似值。这样保证了每次对比试验中两种方法取点个数相同。图 7、图 8、图 9、图 10 分别为定积分

$$\int_0^1 x^2 dx, \int_2^{10} \frac{1}{\ln x} dx, \int_0^1 \frac{\sin x}{x} dx, \int_0^1 e^{-x^2} dx$$

采用以上两种数值算法计算结果对比示意图。图中蓝色曲线为蒙塔卡洛方法, 红色曲线为随机黎曼和的平均值。从图中可以看出: 在相同点数的情况下, 随机黎曼和方法具有较快的收敛速度。

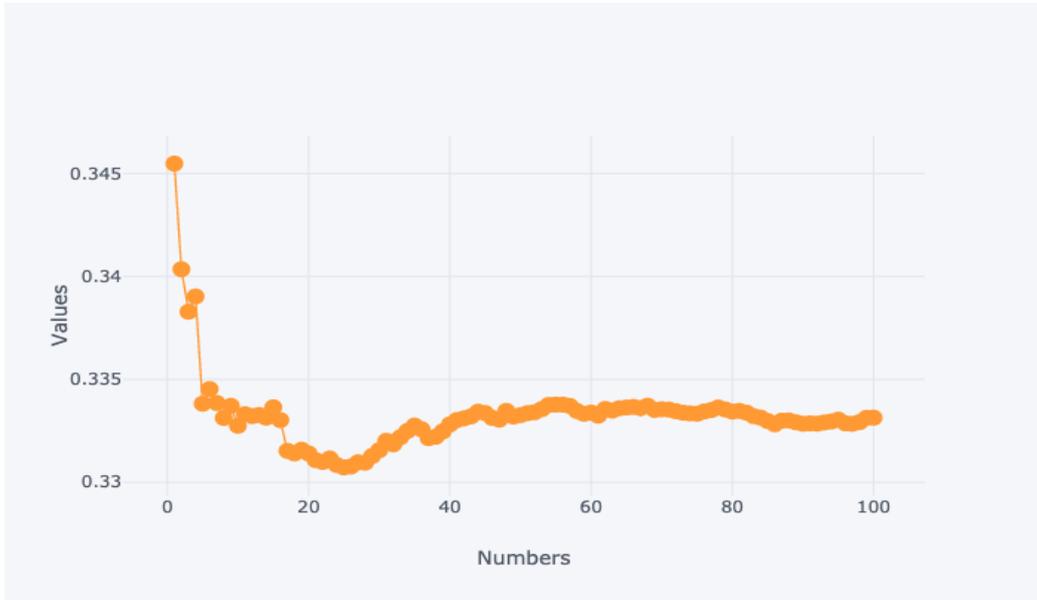


图 1：前 100 次试验黎曼和平均值随试验次数的变化趋势。

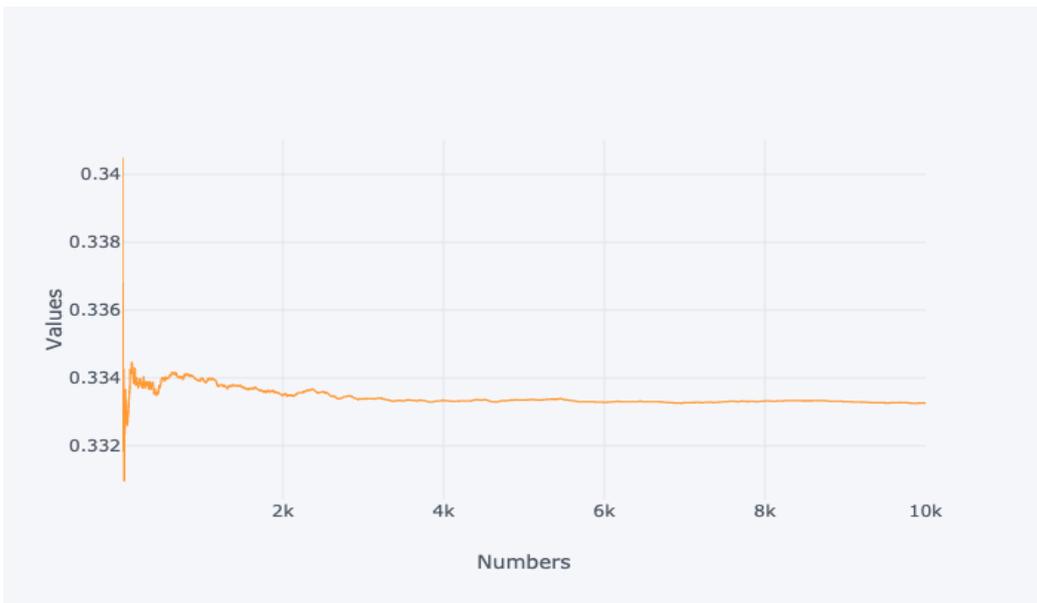
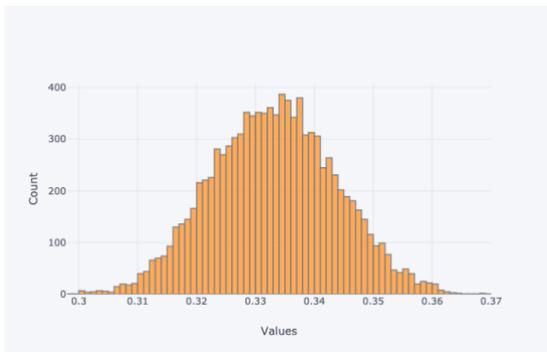
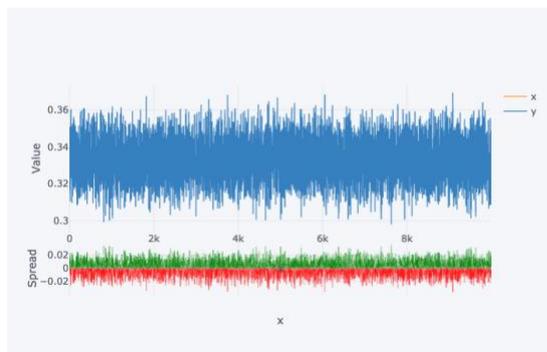


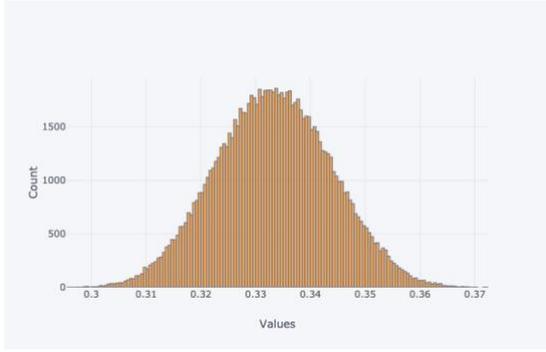
图 2：前 10⁴ 次试验黎曼和平均值随试验次数的变化趋势。



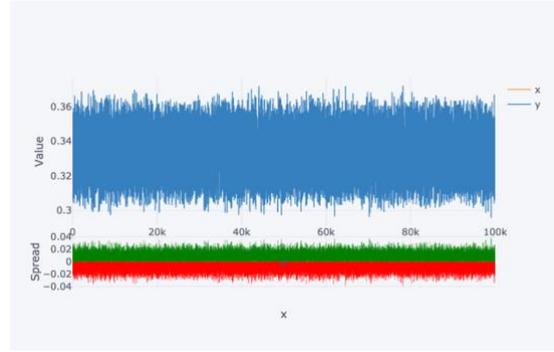
(a)



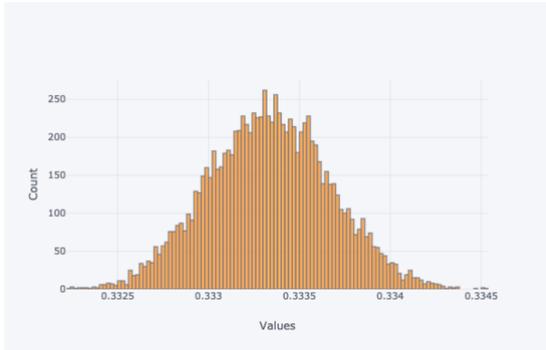
(b)



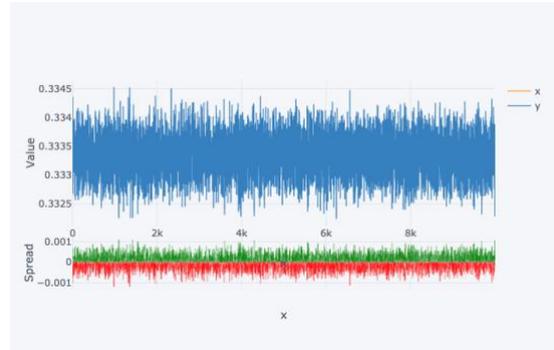
(c)



(d)



(e)



(f)

图 3: (a) $n=10$, 前 10^4 次试验黎曼和分布直方图; (b) 上曲线, 前 10^4 次试验黎曼和变化趋势, 下面曲线为前 10^4 次试验黎曼和与真实积分值的误差; (c) $n=10$, 前 10^5 次试验黎曼和分布直方图; (d) 上曲线为前 10^5 次试验黎曼和变化趋势, 下曲线为前 10^5 次试验黎曼和与真实积分值的误差; (e) $n=100$, 前 10^4 次试验黎曼和分布直方图; (f) 上曲线为前 10^4 次试验黎曼和变化趋势, 下曲线为前 10^4 次试验黎曼和与真实积分值的误差.

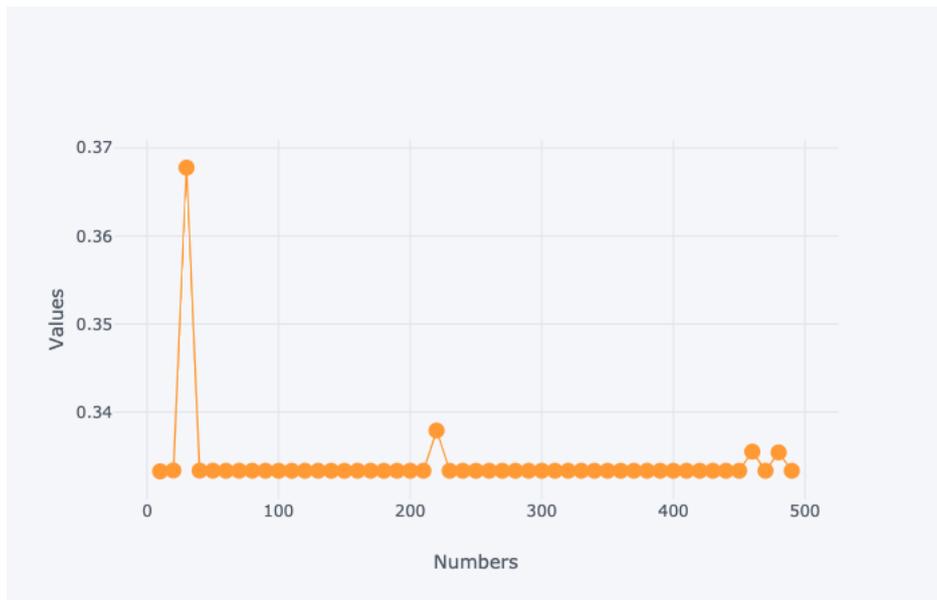


图 4: 分割份数 $n=10,20,30,\dots,500$, 10^4 次随机黎曼和平均值的变化趋势.

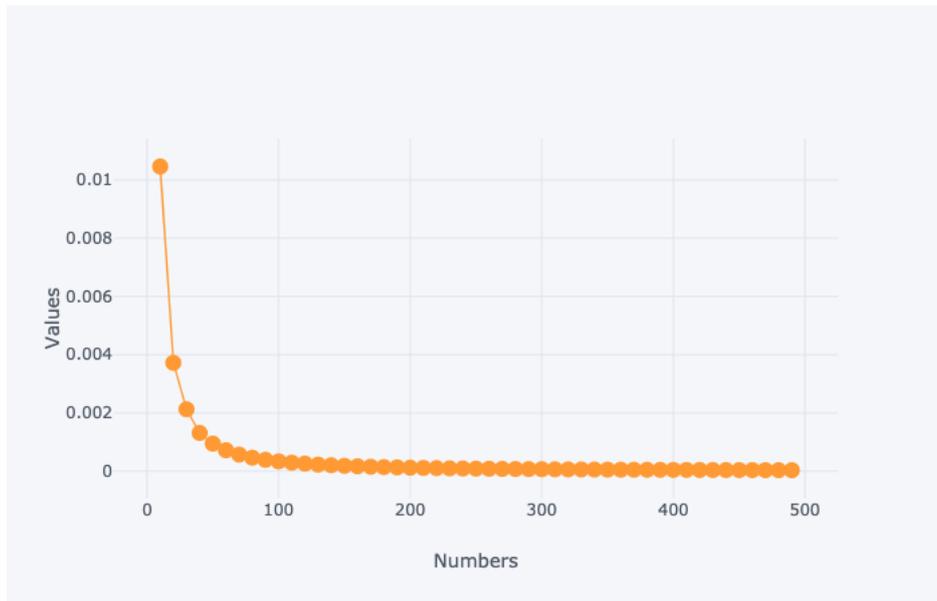


图 5: 分割份数 $n=10,20,30,\dots,500$, 10^4 次随机黎曼和标准差的变化趋势.

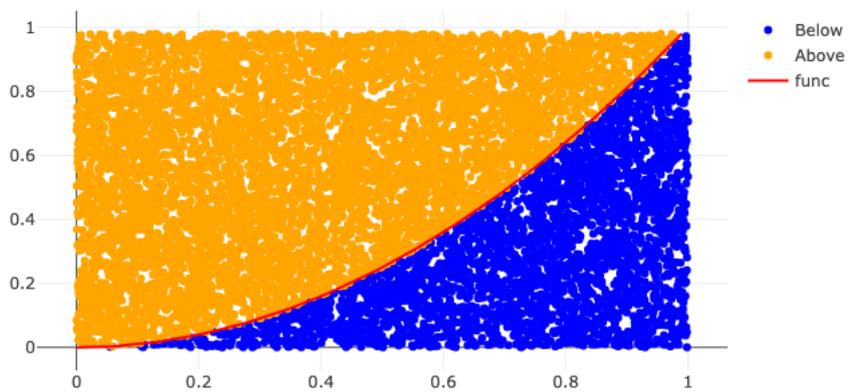


图 6: 蒙特卡洛随机投点试验, 投点个数为 10^4 个, 积分近似值为 0.3300.

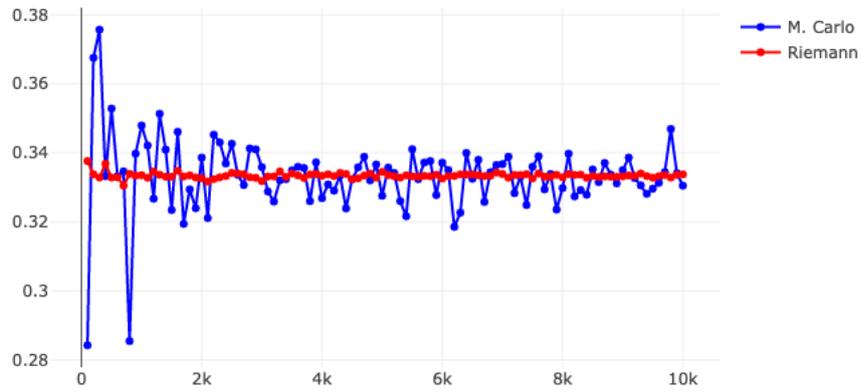


图 7: 随机黎曼和与蒙特卡洛数值积分方法对比图, 积分为 $\int_0^1 x^2 dx$.

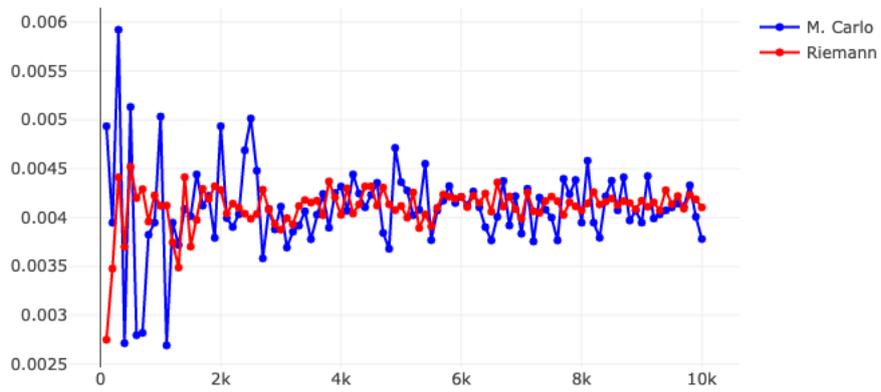


图 8: 随机黎曼和与蒙特卡洛数值积分方法对比图, 积分为 $\int_2^{10} \frac{1}{\ln x} dx$.

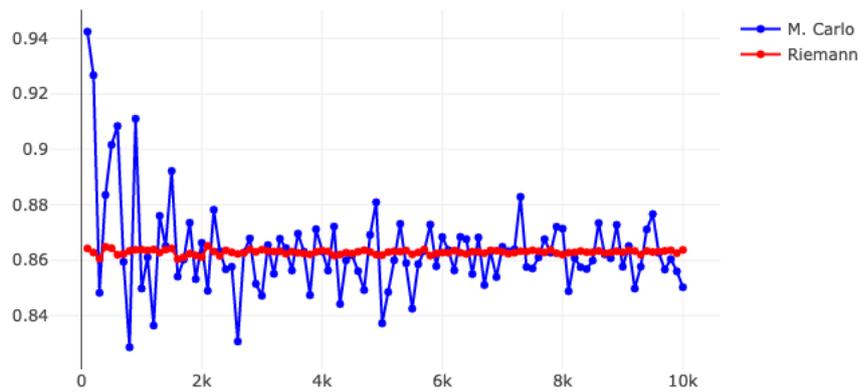


图 9: 随机黎曼和与蒙特卡洛数值积分方法对比图, 积分为 $\int_0^1 \frac{\sin x}{x} dx$.

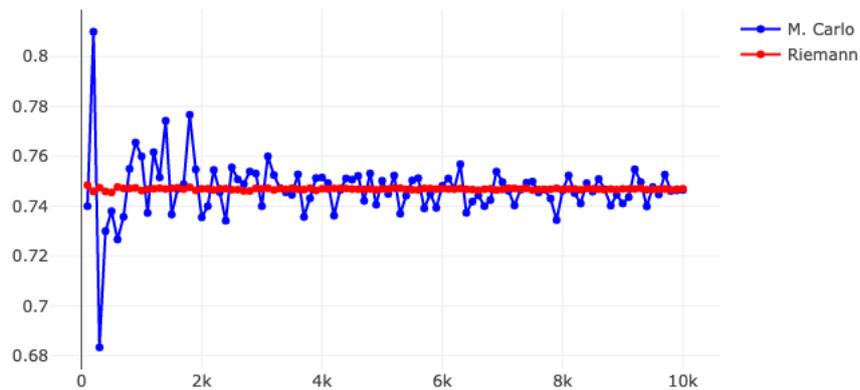


图 10: 随机黎曼和与蒙特卡洛数值积分方法对比图, 积分为 $\int_0^1 e^{-x^2} dx$.

3 理论分析

设 I 为 μ 测度空间中的一个集合, P_n 为一个对 I 的分划, 该分划将 I 分割为 n 个 μ 可测集 $\{I_{n1}, I_{n2}, \dots, I_{nn}\}$, 且每个子集的测度为 $\frac{1}{n}$. 假设对于固定的 n , 取 n 个相互独立且为子区间 I_{nk} 均匀分布的随机变量 $\{X_{n1}, X_{n2}, \dots, X_{nn}\}$, 有以下定理:

定理 1

设 $f \in L^2(\mu)$, 则依概率 1 有以下结果^[7]:

$$\frac{1}{n} \sum_{k=1}^n f(X_{nk}) \rightarrow \int_I f d\mu, (n \rightarrow \infty)$$

反之，如果 $\{X_{n1}, X_{n2}, \dots, X_{nm}\}$ 相互独立且 f 为 μ 可测，但 $f \notin L^2(\mu)$ 。则依概率 1

$$\frac{1}{n} \sum_{k=1}^n f(X_{nk})$$

无界。

4 结 论

本文将黎曼和在每个小区间上的取值看作均匀分布的随机变量，从数值试验的角度探讨了黎曼和的分布。试验表明：当分割给定，前 n 随机试验黎曼和的均值收敛于函数的定积分，其方差随着分割份数的增加而减小。对比蒙塔卡洛方法，说明该方法具有较好的收敛速度和精度，从而为定积分的数值计算提供了一个新的求解思路。不足之处在于，没有从理论上证明随着分割次数的增加，随机黎曼和趋于正态分布。

[参 考 文 献]

- [1] 王能超. 数值分析简明教程[M]. 2版. 武汉: 华中科技大学出版社, 2002:43-59.
- [2] 克莱茵. 古今数学思想[M]. 上海: 上海科学技术出版社, 1979.
- [3] Stewart J M. Python for scientists[M]. Cambridge University Press, 2017.
- [4] Robert C P, Casella G. Monte Carlo Statistical Methods[M]. Springer, 2009.
- [5] 陈纪修, 於崇华, 金路. 数学分析, 上册[M]. 2版. 北京: 高等教育出版社, 2004:272-280.
- [6] 丁韞, 杨晓春. 了解定积分-求和[J]. 大学数学, 2013, 29 (5): 144-147.
- [7] Pruss A R. Randomly sampled Riemann sums and complete convergence in the law of large numbers for a case without identical distribution. Proceeding of the American Mathematical Society, 1996, 124(3):919-929.

Randomly Sampled Riemann Sums: Numerical Method for Definite Integral

*WU Guo-Ning*¹, *SUN Na*¹

(1. China University of Petroleum, Beijing, 102200, China)

Abstract: Riemann sum of definite integral needs to take a point ξ_i from each subinterval after division. If the lengths of the subintervals are equal, and ξ_i is the value of a uniformly distributed random variable on the i th subinterval, then the Riemann sum is a function of the random variables. Inspired by the law of large numbers, we study the change trend of the average value of the first n Riemann sums after the division is given. The tests show that the average value of the first n Riemann sums tends to the true integral value as the number of trials increases; the variance of the first n Riemann sums tends to zero as the number of segmentation increases. The histogram of random Riemann sums is normally distributed. Analysis and comparison with Monte Carlo method show that the method has high convergence speed and accuracy.

Key words: Riemann sum; random variable; law of large numbers; central limit theorem; Monte Carlo method