11：Monte Carlo积分

许传奇 PB16021546

1 题目

用Monte-Carlo方法计算如下定积分，并讨论有效位数。

$$\int_0^2 \sqrt{x+\sqrt{x}}dx$$

$$\int_0^{\frac{9}{10}} dx \int_0^{\frac{4}{5}} dy \int_0^{\frac{9}{10}} dz \int_0^{2} du \int_0^{\frac{13}{10}} dv (6-x^2-y^2-z^2-u^2-v^2)$$

2 原理与算法

2.1 原理

2.1.1 平均值法

求解定积分时，有积分的平均值定理，如下图所示：

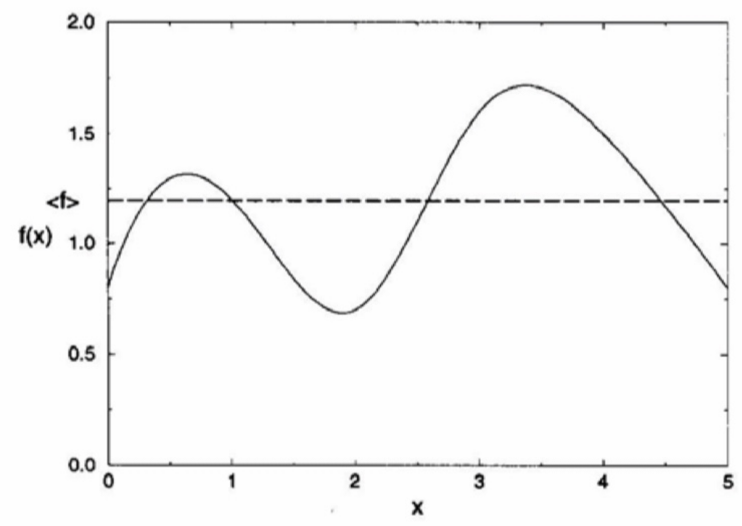


图 1: 积分的平均值定理

根据积分的平均值定理：

$$\int_a^b f(x)dx = (b-a)\langle f\rangle \tag{1}$$

用Monte-Carlo方法对函数f进行抽样，可以得到平均值：

$$\langle f\rangle \approx \frac{1}{N}\sum_{i=1}^{N} f(x_i) \tag{2}$$

因此，用Monte-Carlo方法对定积分区间内的自变量进行随机抽样，就能计算出定积分的近似值：

$$\int_a^b f(x)dx \approx \frac{(b-a)}{N}\sum_{i=1}^{N} f(x_i) \tag{3}$$

2.1.2 多重定积分

对于多变量函数的定积分，可以直接将上述的单变量情形推广至多变量的情形：

$$\int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \cdots \int_{a_n}^{b_n} dx_n f(x_1, x_2, \ldots, x_n) = \frac{1}{N}\left[\prod_{j=1}^{n}(b_j - a_j)\right]\int_{i=1}^{N} f(x_{1_i}, x_{2_i}, \ldots, x_{n_i}) \tag{4}$$

2.1.3 提取法和重要抽样法

由于σ_s正比于σ_f，因此将被积函数变为较平坦的函数可以有效提高Monte-Carlo方法的精度，因此采用几种抽样技巧可以减小方差，包括提取法和重要抽样法：

1. 提取法：

   对于一维积分，设我们能够构造一个与被积函数$f(x)$形状相似的函数$g(x)$，且它们的积分值已知：

$$|f(x) - g(x)| < \epsilon, \int_a^b g(x)dx = J \tag{5}$$

   就可以运用Monte-Carlo方法求一个更平坦的函数的被积函数，再得到定积分，即：

$$\int_a^b f(x)dx = \int_a^b [f(x) - g(x)]\, dx + J \tag{6}$$

2. 重要抽样方法：

   与提取法中的减法相比，实用中更为有效的重要重阳发中采用的是除法。

   设有一个几率分布$g(x)$与$f(x)$形状相似，且有：

$$\frac{f(x)}{g(x)} \sim 1, \int_a^b g(x)dx = 1 \tag{7}$$

   则积分可写成：

$$\int_a^b f(x)dx = \int_a^b \frac{f(x)}{g(x)} g(x)dx \tag{8}$$

   如果x在$[a, b]$区间内的随机抽样不是均匀选取的，而是按照几率分布$g(x)$选取的，则重要抽样方法下的Monte-Carlo积分为：

$$\int_a^b f(x)dx = \langle\frac{f}{g}\rangle \approx \frac{1}{N}\sum_{i=1}^{N}\frac{f(x_i)}{g(x_i)} \tag{9}$$

2.1.4 中心极限定理与误差

概率论中的大数法则和中心极限定理是Monte-Carlo方法应用于统计计算的基础。

1. 大数法则：

   如随机量序列f_i有期待值μ存在，则：

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}f_i\to\mu \tag{10}$$

2. 中心极限定理：

   当N有限时，平均值$\langle f\rangle\approx\frac{1}{N}\sum_{i=1}^{N}f(x_i)$式满足：

$$P\left\{\left|\frac{\langle f\rangle-\mu}{\sigma_f/\sqrt{N}}\right|<\beta\right\}\to\Phi(\beta) \tag{11}$$

   其中的$\Phi(\beta)$是Gauss正态分布，因此可得：

$$\sigma_S=|\langle f\rangle-\mu|\propto\frac{\sigma_f}{\sqrt{N}}=\frac{1}{\sqrt{N}}\sqrt{\langle f^2\rangle-\langle f\rangle^2} \tag{12}$$

3. 有效数字：

   可以认为有效数字位数是用Monte Carlo方法计算出来的定积分和定积分的精确值之间的误差的最大位数，因此误差反映了有效位数的大小。误差越小，有效数字个数越大；误差越大，有效数字个数越小。

2.2 算法

2.2.1 单变量定积分

在$[0,1]$上抽取均匀的ζ，并将其扩充到$[0,2]$上，根据之前原理部分的叙述，得到积分的近似值：

$$\int_0^2\sqrt{x+\sqrt{x}}dx=\frac{2}{N}\sum_{i=1}^{N}\sqrt{x_i+\sqrt{x_i}} \tag{13}$$

2.2.2 多变量定积分

在$[0,1]$ 上抽取均匀的$\xi_x,\xi_y,\xi_z,\xi_u,\xi_v$ ，并将其扩充到$[0,\frac{9}{10}]$，$[0,\frac{4}{5}]$，$[0,\frac{9}{10}]$，$[0,2]$，$[0,\frac{13}{10}]$上，根据之前原理部分的叙述，得到积分的近似值：

$$\int_0^{\frac{9}{10}}dx\int_0^{\frac{4}{5}}dy\int_0^{\frac{9}{10}}dz\int_0^{2}du\int_0^{\frac{13}{10}}dv(6-x^2-y^2-z^2-u^2-v^2)=\frac{1.6848}{N}\sum_{i=1}^{N}(6-x_i^2-y_i^2-z_i^2-u_i^2-v_i^2) \tag{14}$$

2.2.3 有效数字的讨论

使用Mathematica计算两个定积分的精确值，如下图所示：

```
In[1]:= N[Integrate[Sqrt[x + Sqrt[x]], {x, 0, 2}], 8]
Out[1]= 2.6895213

In[2]:= N[Integrate[6 - x^2 - y^2 - z^2 - u^2 - v^2,
          {x, 0, 9/10}, {y, 0, 4/5}, {z, 0, 9/10}, {u, 0, 2},
          {v, 0, 13/10}], 8]
Out[2]= 5.6440800
```

图 2: Mathematica计算得到的精确值

则误差的绝对值为：

$$|error| = |Integral - Sampling| \tag{15}$$

由原理部分的叙述，在中心极限定理的限制下，有：

$$\sigma_S \propto \frac{\sigma_f}{\sqrt{N}} \tag{16}$$

上式说明，误差减小到原来的十分之一，即有效数字增加一位时，需要的抽样点数应该是之前的100倍。

为了探究有效数字位数与点数是否满足以上关系，本次作业采用以下方法：

将误差打印到文件中，以$\ln N$为横坐标，以$\ln|\sigma_S|$为纵坐标作图，并进行线性拟合。

按照上述分析，得到的直线斜率因该是-0.5，将拟合直线的曲率与理论之-0.5进行比较，看是否在误差范围内可以认为两者一致。

若两者一致的话，我们就可以得到结论：若要使有效数字增加一位，需要将抽样点数扩大100倍。

3 源文件使用说明

编译并运行“11Integral.cpp”，将弹出命令行。

命令行上会显示抽样点个数Num、计算的积分值Integral和误差Error。

同时，会将得到的误差输出到文件“中。

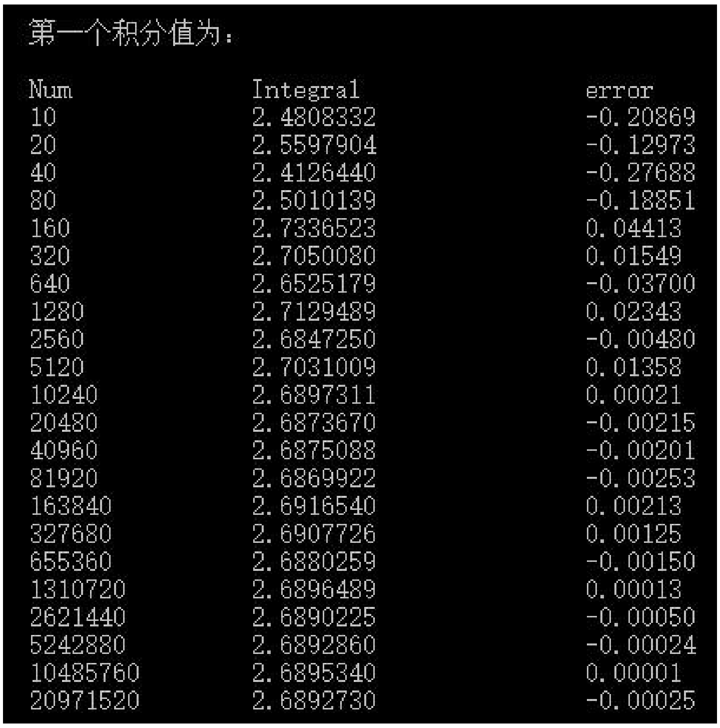
编译并运行“plot.py”即可绘制拟合直线图。

4 计算结果及具体分析

随机数生成器是16807随机数生成器，种子值由C语言自带的随机数生成。

4.1 单变量定积分

计算的结果如图所示：


第一个积分值为：

Num	Integral	error
10	2.4808332	-0.20869
20	2.5597904	-0.12973
40	2.4126440	-0.27688
80	2.5010139	-0.18851
160	2.7336523	0.04413
320	2.7050080	0.01549
640	2.6525179	-0.03700
1280	2.7129489	0.02343
2560	2.6847250	-0.00480
5120	2.7031009	0.01358
10240	2.6897311	0.00021
20480	2.6873670	-0.00215
40960	2.6875088	-0.00201
81920	2.6869922	-0.00253
163840	2.6916540	0.00213
327680	2.6907726	0.00125
655360	2.6880259	-0.00150
1310720	2.6896489	0.00013
2621440	2.6890225	-0.00050
5242880	2.6892860	-0.00024
10485760	2.6895340	0.00001
20971520	2.6892730	-0.00025

图 3: 计算得到不同随机点个数Num下的积分值和误差

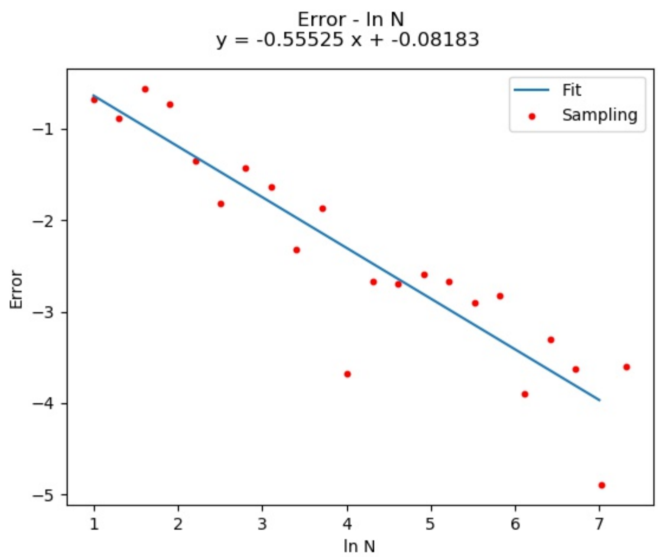


图 4: 误差的绝对值拟合得到的曲线

第一个积分值为：

Num	Integral	error
10	2.2327495	-0.45677
20	2.5895773	-0.09994
40	2.6411044	-0.04842
80	2.7190174	0.02950
160	2.7737812	0.08426
320	2.6923646	0.00284
640	2.6548978	-0.03462
1280	2.6814997	-0.00802
2560	2.6548001	-0.03472
5120	2.6950895	0.00557
10240	2.6998890	0.01037
20480	2.6951734	0.00565
40960	2.6871284	-0.00239
81920	2.6929308	0.00341
163840	2.6911401	0.00162
327680	2.6873335	-0.00219
655360	2.6898769	0.00036
1310720	2.6888092	-0.00071
2621440	2.6903603	0.00084
5242880	2.6887702	-0.00075
10485760	2.6894893	-0.00003
20971520	2.6894941	-0.00003

图 5: 计算得到不同随机点个数Num下的积分值和误差

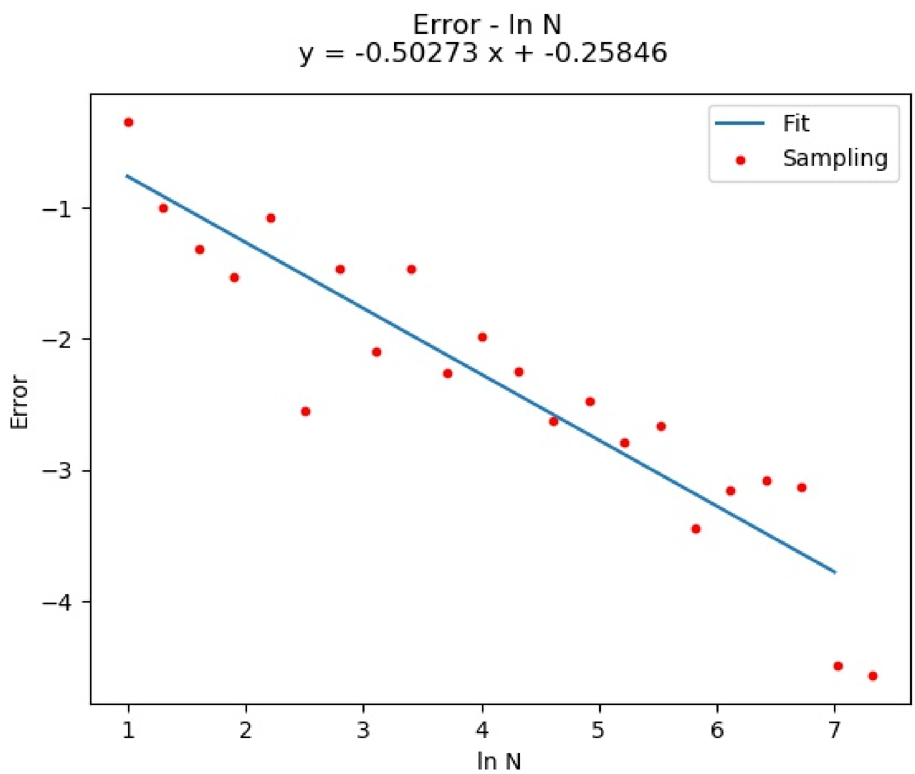
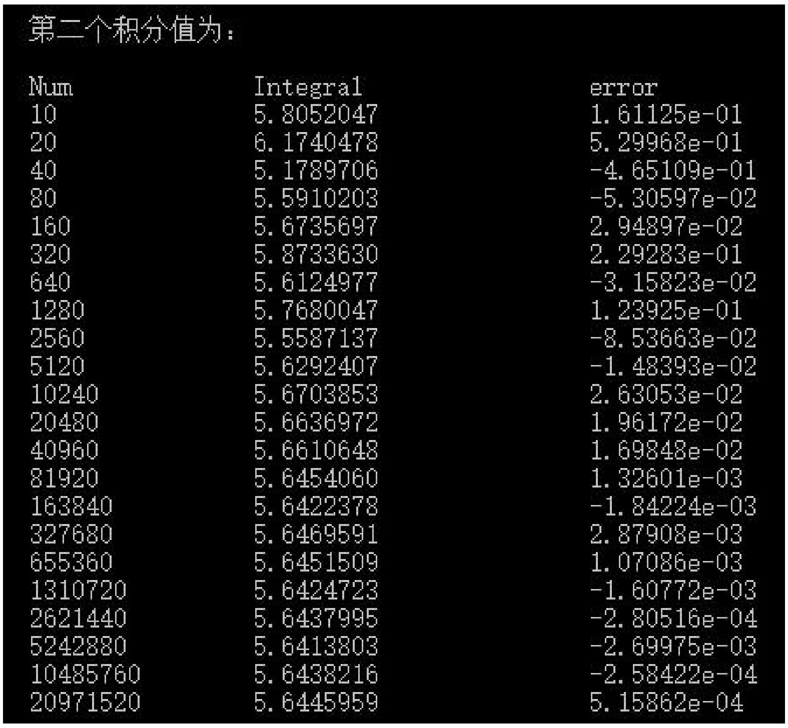


图 6: 误差的绝对值拟合得到的曲线

4.2 多变量定积分

计算的结果如图所示：


第二个积分值为：

Num	Integral	error
10	5.8052047	1.61125e-01
20	6.1740478	5.29968e-01
40	5.1789706	-4.65109e-01
80	5.5910203	-5.30597e-02
160	5.6735697	2.94897e-02
320	5.8733630	2.29283e-01
640	5.6124977	-3.15823e-02
1280	5.7680047	1.23925e-01
2560	5.5587137	-8.53663e-02
5120	5.6292407	-1.48393e-02
10240	5.6703853	2.63053e-02
20480	5.6636972	1.96172e-02
40960	5.6610648	1.69848e-02
81920	5.6454060	1.32601e-03
163840	5.6422378	-1.84224e-03
327680	5.6469591	2.87908e-03
655360	5.6451509	1.07086e-03
1310720	5.6424723	-1.60772e-03
2621440	5.6437995	-2.80516e-04
5242880	5.6413803	-2.69975e-03
10485760	5.6438216	-2.58422e-04
20971520	5.6445959	5.15862e-04

图 7: 计算得到不同随机点个数Num下的积分值和误差

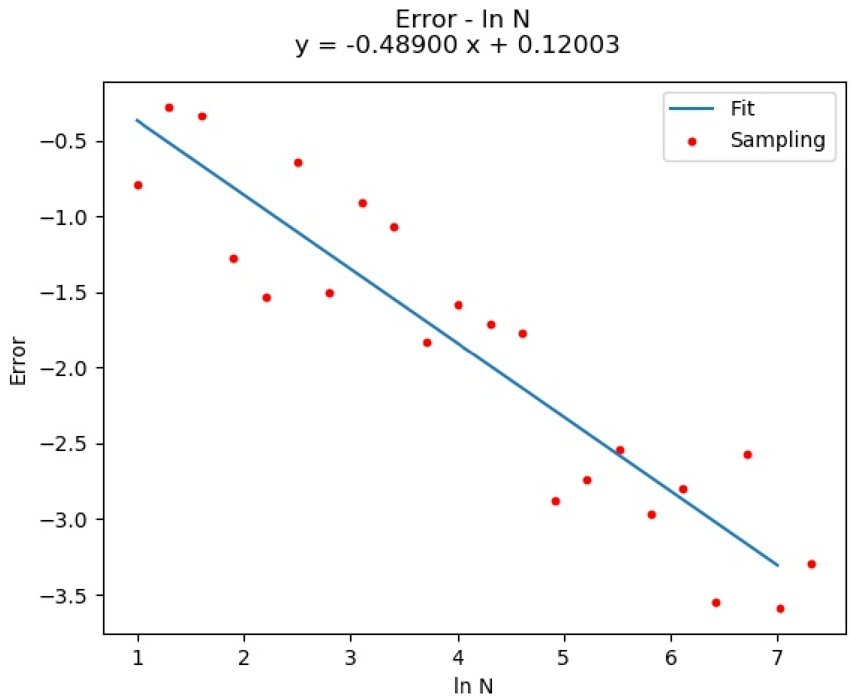


图 8: 误差的绝对值拟合得到的曲线

第二个积分值为：

Num	Integral	error
10	6.7760311	1.13195e+00
20	6.2873538	6.43274e-01
40	5.9025919	2.58512e-01
80	5.7449308	1.00851e-01
160	6.0349239	3.90844e-01
320	5.6271885	-1.68915e-02
640	5.4868658	-1.57214e-01
1280	5.6356242	-8.45576e-03
2560	5.6232065	-2.08735e-02
5120	5.6363260	-7.75400e-03
10240	5.6899297	4.58497e-02
20480	5.6496015	5.52149e-03
40960	5.6498133	5.73334e-03
81920	5.6427888	-1.29124e-03
163840	5.6395282	-4.55178e-03
327680	5.6432560	-8.23977e-04
655360	5.6411818	-2.89815e-03
1310720	5.6460752	1.99525e-03
2621440	5.6419256	-2.15437e-03
5242880	5.6432967	-7.83251e-04
10485760	5.6430917	-9.88315e-04
20971520	5.6438248	-2.55221e-04

图 9: 计算得到不同随机点个数Num下的积分值和误差

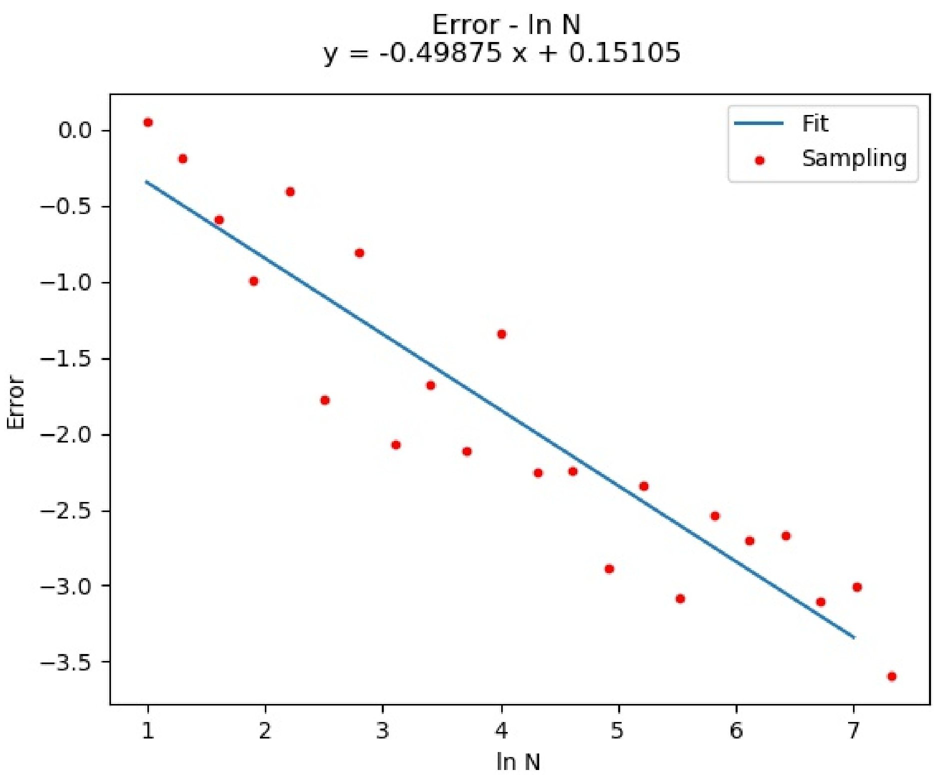


图 10: 误差的绝对值拟合得到的曲线

5 讨论

5.1 算法的改进

本题算法较简单，直接按照算法部分叙述编程即可。

本题采用的是平均值法计算定积分，但这两个函数都不是非常平整的，可以通过原理中叙述的提取法或重要抽样方法对其进行抽样，可以有效减小误差。

即选取合适的$g(x)$使$f(x)-g(x)$更平整，从而减小误差。

$$\int_a^b f(x)dx = \int_a^b [f(x)-g(x)]\,dx + J \tag{17}$$

如图所示：

```
Plot[{Callout[Sqrt[x + Sqrt[x]]],
绘图   标注      平方根     平方根
  Callout[Sqrt[x + Sqrt[x]] - Sqrt[x]]}, {x, 0, 2}]
  标注     平方根     平方根       平方根
```

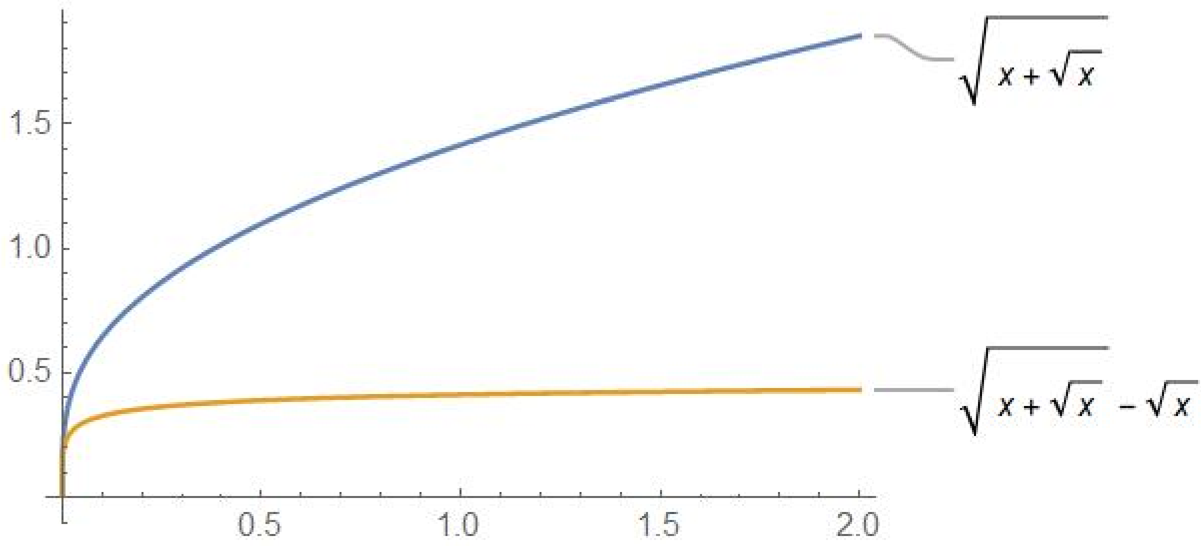


图 11: 提取法修正图

对于第一个定积分，运用提取法，设$g(x)=\sqrt{x}$，两者之差在区间内斜率非常小，可以有效减小误差，提高有效数字位数。

且其积分比较好求：

$$\int_0^2 g(x)dx = \int_0^2 \sqrt{x}dx = \frac{2}{3}2^{\frac{3}{2}} \approx 1.8856 \tag{18}$$

同理，运用重要抽样方法：

$$\int_a^b f(x)dx = \int_a^b \frac{f(x)}{g(x)} g(x)dx \tag{19}$$

也可以实现将函数变得更平整，如图所示：

```
Plot[{Callout[Sqrt[x + Sqrt[x]]],
 绘图  标注    平方根   平方根
  Callout[Sqrt[x + Sqrt[x]] / x^{1 / 3}]}, {x, 0, 2}]
  标注     平方根   平方根
```

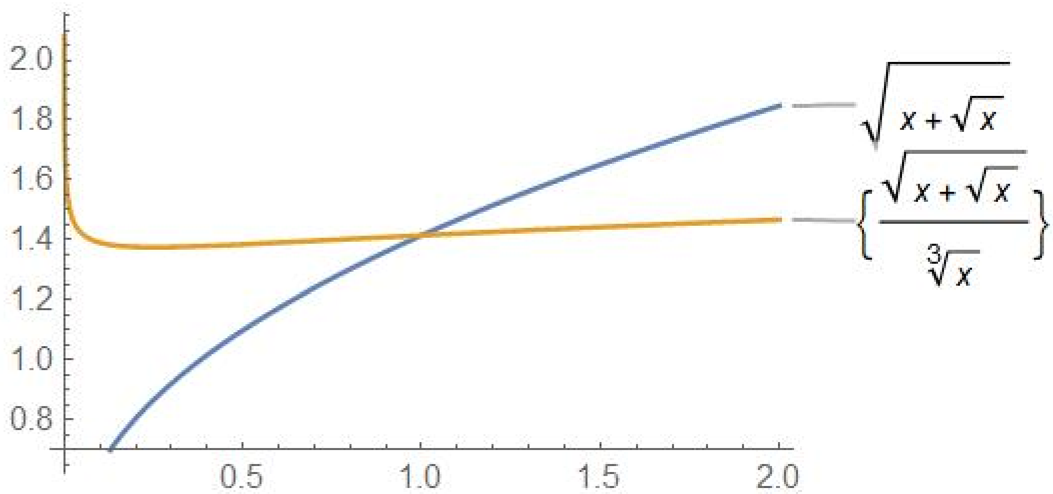


图 12: 重要抽样法修正图

设$g(x) = \sqrt[3]{x}$，将其作为除数，也可以得到更为平整的曲线。其积分也容易求出：

$$\int_0^2 g(x)dx = \int_0^2 \sqrt[3]{x}dx = \frac{3}{4}2^{\frac{4}{3}} \approx 2.5198 \tag{20}$$

5.2 结果的讨论

从结果可以看出：总体上，随着抽样点数的增加，得到的积分值的误差逐渐减小。

两个积分得到的拟合直线的斜率接近-0.5，与之前叙述一样，可以认为要是有效数字增大一位，需要将抽样点数扩大100倍。

此题判断存在一点缺陷：通过直接比较拟合直线斜率与理论值来判断是否符合我们的推断，但由于是统计规律，我们只能在一定程度上确定我们的观点，即需要用到假设检验的办法对我们的结果进行判断。不过鉴于命令行中的误差确实基本上呈这种规律，此题直接下定论而没有进行验证。