随机算法课程

实验报告

实验四：最小生成树期望权值计算

姓名：袁野

学号：1190200122

班级：1903102

评分表：（由老师填写）

最终得分：	
对实验题目的理解是否透彻：	
实验步骤是否完整、可信 ：	
代码质量 ：	
实验报告是否规范 ：	
趣味性、难度加分 ：	
特 色：	1
	2
	3

一、实验题目概述

蒙特卡洛算法是一种基于采样和统计的随机算法；在本次实验中，实现一种随机图，在随机图上实现对最小生成树的抽样过程，由抽样过程实现蒙特卡罗方法计算最小生成树权值的数学期望估计，比较估计结果的准确性。

通过计算最小生成树的数学期望估计，理解蒙特卡罗算法的过程和效用，体会由经验公式得到理论公式的过程。

二、对实验步骤的详细阐述

1、首先生成一个有 n 个点的随机完全图，所有边的权值为独立均匀分布在 [0,1] 的随机变量，使用结构体变量储存两个端点和边权进行储存，然后将这个数组返回即可，在后边的对比实验中我们要对边权的分布函数进行修改实验，所以还需要实现不同的权值分布函数，比如下面函数

```
double p1 = 0.4, p2 = 0.8, p3 = 1;
double t1 = 0.2, t2 = 0.6, t3 = 1;
double p = rnd.next(); //生成[0,1)的随机浮点数
if (p < t1) return rnd.next(p1); // 生成[0,p1)的随机浮点数
else if (p < t2) return rnd.next(p1, p2); // 生成[p1,p2)的随机浮点数
else return rnd.next(p2, p3); // 生成[p2,p3)的随机浮点数
```

这个函数生成了一个边权满足如下分布函数的完全图

$$F(x)=\begin{cases}0.5x & 0\le x< 0.4\\ x-0.2 & 0.4\le x<0.8\\ 2x-1 & 0.8\le x< 1\end{cases}$$

2、最小生成树部分采用 kruskal 算法实现，算法的基本是想为首先将所有的边按照边权从大到小进行排序，然后依次枚举这些边，借助并查集实现对点对之间连通性的维护，如果当前枚举的边的两个端点不在同一集合内，那么将该边加入到最小生成树中并在并查集上将这两个点联通，最终得到的结果就是这个图的最小生成树权值和，时间复杂度为 $O(n^2 log\ n)$。

3、为了使的期望更加准确，显然重复实验的次数越多越好，但是由于当数据规模越大时最小生成树求解做需要的时间就越长，因此我们让试验次数为 $\frac{2*10^8}{n^2}$（n 为完全图的点的数量）。这样就能保证及时完全图的规模不同，但是试验次数仍能保证尽可能多。

4、记录算法运行时间时，为了保证较小规模的完全图的运行时间太短而造成误差偏大，我采用了记录多次实验的总时间然后除以试验次数的方式进行计算。

5、对于最小生成树的权值期望的下界的计算，我们考虑取到下界的情况为当前完全图的前 $n-1$ 小的边权恰好组成一棵生成树，因此我们在实验验证的过程中只需要将生成的完全图的前 $n-1$ 小的边权求和即可。

三、实验数据

1. 实验设置

实验环境：

Ubuntu20.04.4 LTS(GNU/Linux 5.10.102.1-microsoft-standard-WSL2 x86_64)

数据：

数据直接在代码运行中生成并储存在内存中然后供后续程序计算，并未输出到磁盘。而对于数据的构造方式如第二部分所示。

我分别构建了四种不同概率分布的数据，如下：

数据一：

$$F(x) = x(0 \le x < 1)$$

数据二：

$$F(x) = \begin{cases} 0.5x & 0 \le x < 0.4 \\ x - 0.2 & 0.4 \le x < 0.8 \\ 2x - 1 & 0.8 \le x < 1 \end{cases}$$

数据三：

$$F(x) = \begin{cases} 0.5x & 0 \le x < 0.4 \\ 1.5x - 0.4 & 0.4 \le x < 0.8 \\ x & 0.8 \le x < 1 \end{cases}$$

数据四：

$$F(x) = \begin{cases} 2x & 0 \le x < 0.4 \\ 0.25x + 0.7 & 0.4 \le x < 0.8 \\ 0.5x + 0.5 & 0.8 \le x < 1 \end{cases}$$

2. 实验结果

最小生成树的权值期望与完全图的点数的关系如下图所示：

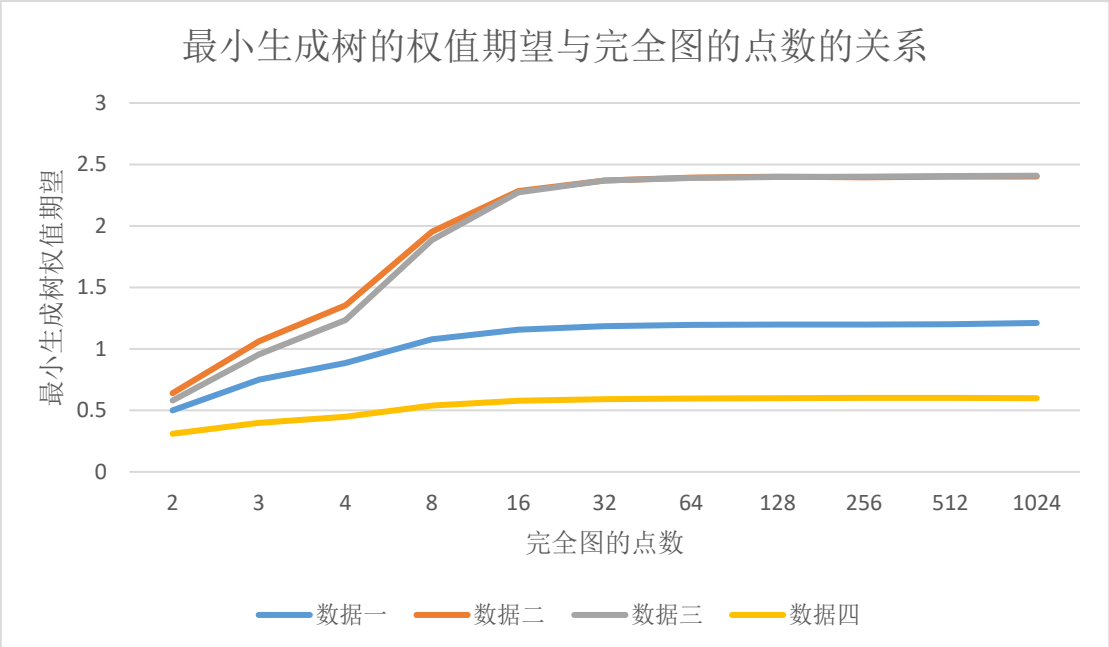

图 1

数据一下每次最小生成树的 kruskal 算法所用时间的对数与数据规模大小的关系如下所示：

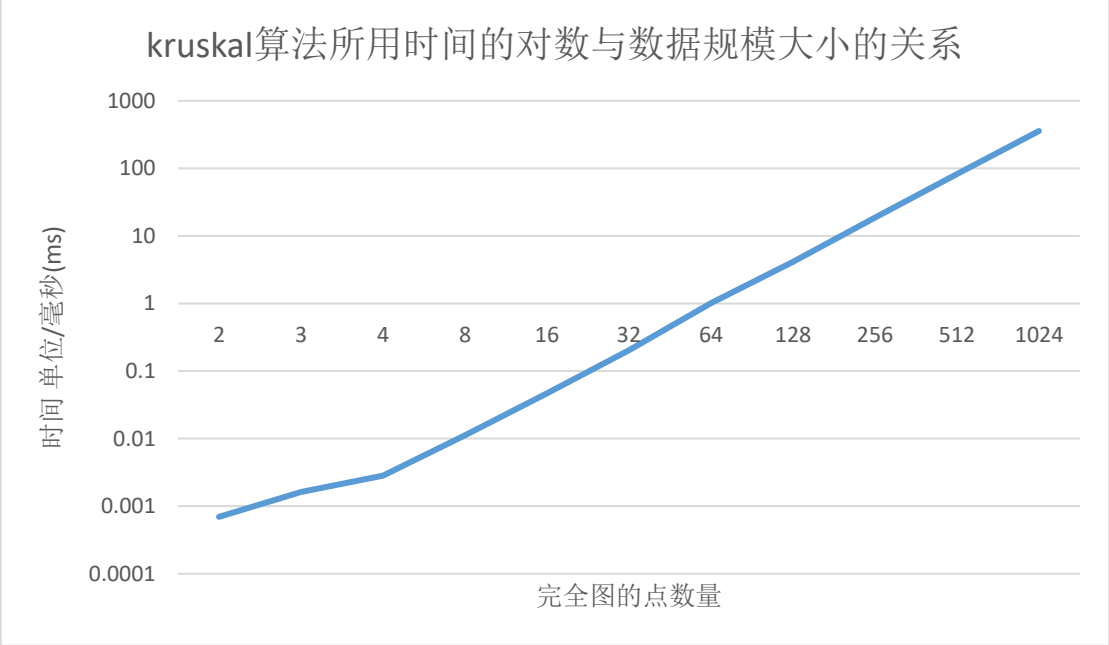

图 2

不同数据下完全图的前 $n-1$ 小的边的边权和与完全图点数的关系如下：

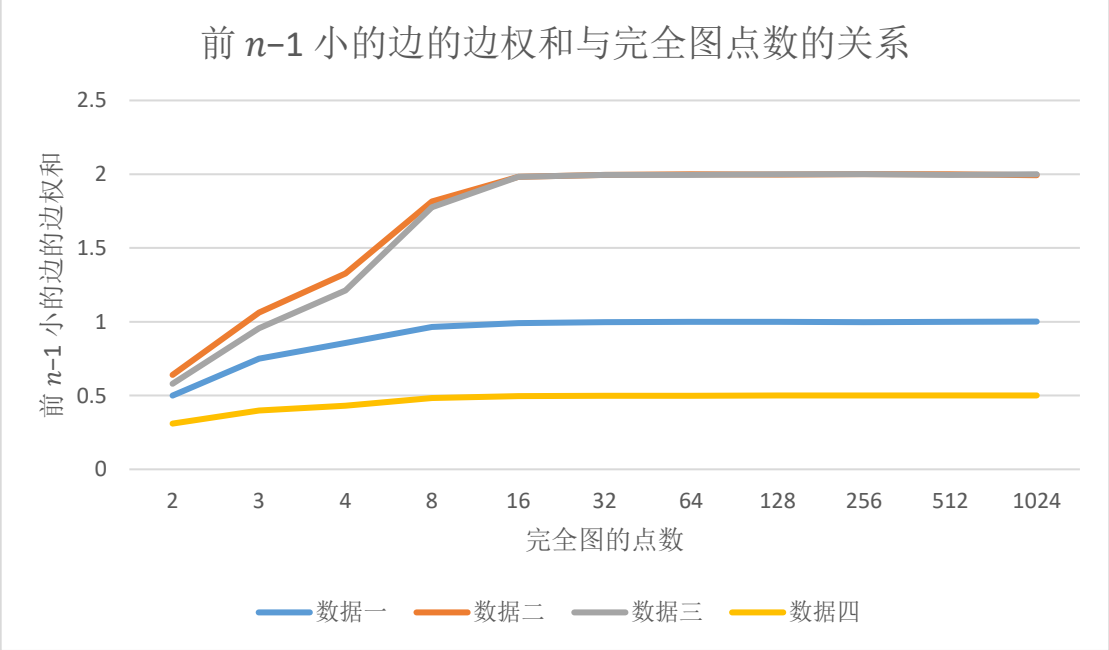

图 3

不同数据下完全图的第 $\frac{k(k-1)}{2}+1(k=1,2,3,\cdots,n-1)$ 小的边的边权和与完全图点数的关系如下：

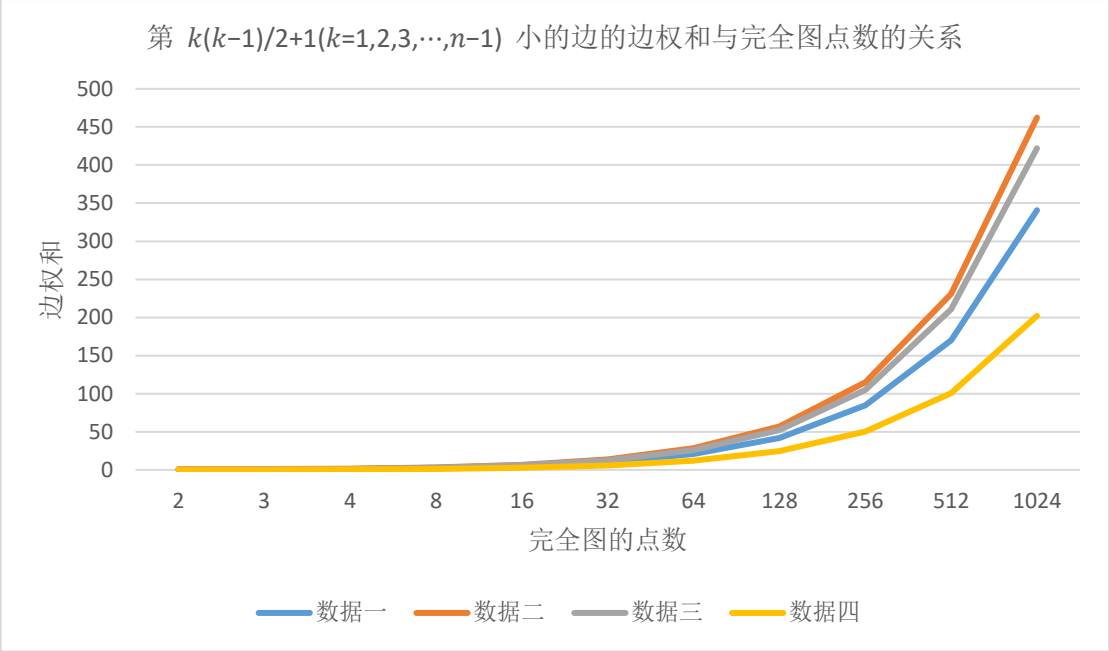

图 4

四、对实验结果的理解和分析

1、 首先我们观察 kruskal 算法的运行时间与完全图规模的关系，该算法的实际复杂度为 $O(n^2 log\ n)$，每当点数扩大两倍时，实际上运行的时间扩大四倍多，其中 $O(n^2)$ 部分扩大了四倍，$O(log\ n)$ 部分增加了一个常数。那么时间与完全图规模应该成近似线性关系，而图中将时间取了对数，且横坐标的分布也可以认为对 2 取了对数，因此图中也展示出了一条近似的直线。

2、 通过对图 1 分析我们可以发现，对于边权为[0,1)均匀分布的完全图来说，其最小生成树的权值和随着完全图点数的增加而增加，增加的速度越来越慢，最终无限趋近于 1.2 后不再增加。这是一个很有趣的现象，因为在我的直觉上最小生成树的权值应该是随着完全图规模的增加而不断增加的，虽然增加幅度会越来越小，但是最终收敛这个现象并不是很直观可以想到的。

3、 通过查阅资料可知，有这样一个定理：对于一个边权符合一定概率分布，边权的概率分布函数处处连续且在 0 出的导数为 D 的其最小生成树的期望权值总有上界，且这个上界不随 n 的增加而增加，也就是说随着 n 增大，**最小生成树的期望权值和无限趋近于一个定值 lim，这个定值 lim 为 $\frac{\xi(3)}{D}$**，其中 ξ 是黎曼函数，$\xi(3)$ 为阿培里常数。

$$\xi(3)=\sum_{i=1}^{\infty}\frac{1}{i^3}=1.202\cdots$$

根据以上的结论，当生成树的边权均满足 [0,1) 的均匀分布的话，$D=1$，因此 n 无穷大时最小生成树的权值趋近于 1.202。

4、 为了验证上述的结论，我们采用另外三种不同分布函数的数据进行测试，数据二的边权分布函数在 0 处的导数为 0.5，那么 $lim\approx 2.4$，数据三中的边权分布函数在 0 处的导数为 0.5，那么 $lim\approx 2.4$，数据四中的边权分布函数在 0 处的导数为 2，那么 $lim\approx 0.6$，最终实验之后的结果也的确如此。对比数据二和数据三的结果我们可以发现，lim 的取值确实和除零点以外的其他位置的概率分布函数是没有任何关系的。

5、 感性的理解一下上边的事实其实我们可以想到，最小生成树上的边的数量和完全图上的边的数量的比值为 $\frac{n-1}{\frac{n(n-1)}{2}}=\frac{2}{n}$，也就是说随着 n 的不断增大，这个比值会越来越小，而选出被作为最小生成树中的边权在完全图的所有边权中的分布会越来越向 0 靠近，n 足够

大的时候最小生成树中的边权的分布就可以被认为是在 0 点附近，与其他位置的概率函数无关。这个事实在下遍我们估计最小生成树的下界是也会被提到。

五、实验过程中最值得说起的几个方面

我们在这里讨论一下在一定概率分布函数下的完全图的最小生成树权值和上界和下界的期望。

1、 首先我们要考虑的是将所有的边权排序后，他们的期望排布是长什么样的。我们考虑最小的边权的期望。设 $X_i\left(1 \le i \le \mathrm{k} = \frac{n(n-1)}{2}\right)$ 为各个边的边权， $Y = min\{X_i\}$，设所有的边权服从$[0,p)$的均匀分布，那么有

$$\begin{aligned} F_Y(x) &= P(Y < x) \\ &= P(min\{X_i\} < x) \\ &= 1 - P(min\{X_i\} \ge x) \\ &= 1 - P(X_1 \ge x)P(X_2 \ge x)\cdots P(X_k \ge x) \\ &= 1 - \left(1 - F_X(x)\right)^k \end{aligned}$$

则有概率密度函数：

$$\begin{aligned} f_Y(x) &= F_Y'(x) \\ &= \frac{d}{dx}\left(1 - \left(1 - F_X(x)\right)^k\right) \\ &= k\left(1 - F_X(x)\right)^{k-1}\left(f_X(x)\right) \end{aligned}$$

则有期望：

$$\begin{aligned} E(Y) &= \int_0^p x f_Y(x)\,dx \\ &= \int_0^p xk\left(1 - F_X(x)\right)^{k-1}\left(f_X(x)\right)dx \\ &= \int_0^p \frac{xk}{p}\left(1 - \frac{x}{p}\right)^{k-1} dx \\ &= \frac{p}{k+1} \end{aligned}$$

因此最小的边权的期望为 $\frac{p}{k+1}$

我们在考虑次小的边权，出去边权最小的边以外，其余的$k-1$ 边权服从 $\left[\frac{p}{k+1}, \mathrm{p}\right)$ 的均匀分布，等价于以 $\left[0, \frac{kp}{k+1}\right)$ 的均匀分布选取$k+1$条边中的最小边权加上 $\frac{p}{k+1}$，前半部分问题与我们之前讨论的问题是等价的，结果为最小边权期望是 $\frac{p}{k+1}$，则原完全图的边权次小值期望为 $\frac{2p}{k+1}$，同理后边的边权期望依次为 $\frac{3p}{k+1}, \frac{4p}{k+1}, \cdots, \frac{kp}{k+1}$，由此我们得到了边权服从$[0,p)$的均匀分布的完全图的边权期望值从大到小的排列。对于概率密度函数分段的边权只需要进行分类讨论即可。

2、 接下来我们考虑最小生成树下界的期望。最小生成树权值取到下界时恰好由排好序的边权的前 $n-1$ 条边组成。由 1 中的结论我们可以知道这些边的边权期望，那么最小生成树权值和的下界即为前 $n-1$ 小的边的边权期望值之和。而当 n 趋近于无穷大时，前$n-1$条边的边权期望值总是无线趋近于 0 处，因此我们有以下结论：**当完全图的点数 n 无限趋近于无穷时，无论边权服从的是什么样的分布，只要其分布函数在 0 处可导且导数 $D > 0$，那么其最小生成树的下界期望情况总是与所有边权服从 $\left[0, \frac{1}{D}\right)$ 的均匀分布的完全图的最小生成树的下界期望情况等价**，那么对于 0 处导数为 D 边权分布函数的完全图，其前$n-1$ 条边的边权期望值和为

$$\sum_{i=1}^{n-1}\frac{\frac{i}{D}}{\frac{n(n-1)}{2}+1}=\frac{1}{D+\frac{2D}{n(n+1)}}$$

当 n 趋近于正无穷时边权期望值和趋近于 $\frac{1}{D}$，这与实验过程中四组不同数据的结果一致。

3、 然后我们讨论最小生成树权值和的上界情况：我们将边权排好序之后，第一条边 (a,b) 和第二条边 (b,c) 是一定会被选做最小生成树的边的，最坏情况下，第三条边是连接a和c的，此时不应加入第三条边，第四条边 (c,d)一定会被选中，第 5，6 条边分别为$(a,d),(b,d)$，显然不应被选……一般来说，每前 $\frac{i(i-1)}{2}(i=2,3,\cdots,n)$条边都会组成点数为 i 的子完全图，这种情况下我们只能选第$\frac{(i-1)i}{2}+1(i=1,2,\cdots,n\text{-}1)$ 条边组成最小生成树，这种情况下最小生成树的权值之和是最大的，由我们在 1 中的分析可知，所有的边权服从$[0,p)$的均匀分布的完全图中，最小生成树的权值和的上界的期望值为

$$\sum_{i=1}^{n-1}\frac{\frac{i(i-1)}{2}+1}{\frac{n(n-1)}{2}+1}=\frac{n^3-3n^3+8n-6}{3n^2-3n+6}$$

当 n 趋近于无穷时，最小生成树权值和的上界的期望值趋近于 $\frac{n}{3}$，与实验结果一致。另外三组数据的结果仅需要在上述步骤中加入分类讨论即可，这里不再赘述。