第一次作业

Question 1

Let $\{X_i\}$ be i.i.d random variables with the common probability mass function

$$P\{X_1=i\}=\begin{cases}p & \text{if } i=+1\\ q & \text{if } i=-1\end{cases}$$

where $p+q=1$. Define $S_0=0$ and $S_n=X_1+\cdots+X_n$. The stochastic process $S=\{S_n,n\geq 0\}$ is called a simple random walk. Let h_n denote the probability that $S_n=1$ for the first time. We call the number of steps (the first n such that $S_n=1$) needed to reach 1 the first passage time.

(a)

Give an argument that supports the following system of recursive equations: $h_0=0,h_1=p$, and

$$h_n=\sum_{j=1}^{n-2}qh_jh_{n-j-1}\qquad n=2,3,\ldots$$

$h_0=P\{S_0=1\text{ for the first time}\}$，而$S_0=0$，故$h_0=0$。

$h_1=P\{S_1=1\text{ for the first time}\}$，事件$\{S_1=1\text{ for the first time}\}\Leftrightarrow\{X_1=1\text{ for the first time}\}$，故$h_1=p$。

下面推导$h_n=\displaystyle\sum_{j=1}^{n-2}qh_jh_{n-j-1}\quad(n=2,3,\ldots)$:

定义$g_n=P\{S_n=2\text{ for the first time}\}\quad(n=2,3,\ldots)$，可以利用全概率公式将$g_n$由$h_n$表出:

$$\begin{aligned}
g_n=&P\{S_n=2\text{ for the first time}\}\\
=&\sum_{i=1}^{n-1}P\{S_n=2\text{ for the first time}|S_i=1\text{ for the first time}\}\cdot P\{S_i=1\text{ for the first time}\}\\
=&\sum_{i=1}^{n-1}P\{\sum_{j=i+1}^{n}X_j=1\text{ for the first time}\}\cdot h_i\\
&\Updownarrow\text{由于}X_i\text{之间独立同分布}\\
=&\sum_{i=1}^{n-1}P\{\sum_{j=1}^{n-i}X_j=1\text{ for the first time}\}\cdot h_i\\
=&\sum_{i=1}^{n-1}P\{S_{n-i}=1\text{ for the first time}\}\cdot h_i\\
=&\sum_{i=1}^{n-1}h_{n-i}\cdot h_i
\end{aligned}$$

因为事件$\{S_n=1\text{ for the first time}\}\Leftrightarrow\{X_1=-1\}\bigcap\{\sum_{i=2}^{n}X_i=2\text{ for the first time}\}$，所以

$$\begin{aligned}
h_n&=P\{X_1=-1\}\cdot P\{\sum_{i=2}^{n}X_i=2\text{ for the first time}\}\quad\text{两事件相互独立}\\
&=P\{X_1=-1\}\cdot P\{\sum_{i=1}^{n-1}X_i=2\text{ for the first time}\}\quad X_i\text{之间相互独立}\\
&=q\cdot g_{n-1}\\
&=q\cdot\sum_{i=1}^{n-2}h_{n-1-i}\cdot h_i\\
&\Leftrightarrow\sum_{j=1}^{n-2}qh_jh_{n-j-1}
\end{aligned}$$

证毕。

(b)

Compute the cumulative distributions $\{\displaystyle\sum_{k=0}^{n}h_k\}$ for $p=0.45,0.50$, and 0.55, over the range $n=0:20$

利用Python编程进行计算

```python
import pandas as pd
from itertools import accumulate

max_n = 20
p_list = [0.45,0.50,0.55]
sum_p = []
for p in p_list:
    q = 1-p

    h = []
    h_0 = 0 # 计算h0
    h.append(h_0)
    h_1 = p # 计算h1
    h.append(h_1)

    for n in range(2, max_n):
        # 计算hn
        h_n = 0
        for j in range(1, n-2+1):
            # 公式中的下标是从1到n-2
            # 因此range中的参数为1,n-2+1
            h_n += q * h[j] * h[n-j-1]
        h.append(h_n)

    # 计算逐一位置的和
    result = list(accumulate(h))
    sum_p.append(result)

sum_p = pd.DataFrame(sum_p, index=['P=0.45','P=0.50','P=0.55']).T
print(sum_p)
```

输出结果如下

```
      P=0.45    P=0.50    P=0.55
0   0.000000  0.000000  0.000000
1   0.450000  0.500000  0.550000
2   0.450000  0.500000  0.550000
3   0.561375  0.625000  0.686125
4   0.561375  0.625000  0.686125
5   0.616506  0.687500  0.753507
6   0.616506  0.687500  0.753507
7   0.650618  0.726562  0.795199
8   0.650618  0.726562  0.795199
9   0.674257  0.753906  0.824092
10  0.674257  0.753906  0.824092
11  0.691810  0.774414  0.845545
12  0.691810  0.774414  0.845545
13  0.705463  0.790527  0.862233
14  0.705463  0.790527  0.862233
15  0.716445  0.803619  0.875656
16  0.716445  0.803619  0.875656
17  0.725506  0.814529  0.886729
18  0.725506  0.814529  0.886729
19  0.733130  0.823803  0.896048

```

(c)

Speculate what will happen to these cumulative distributions when $n \to \infty$, specifically, whether they will approach to 1.

为方便讨论，我们首先计算事件$\{S_n > 0\}$当$n \to \infty$时的概率分布：

根据中心极限定理有

$$\lim_{n\to\infty}\frac{S_n-nE(X)}{\sqrt{nD(X)}}\sim N(0,1)$$

因为

$$E(X)=p-q \qquad E(X^2)=1 \qquad D(x)=1-(p-q)^2=(1-p+q)(1+p-q)=4pq$$

则

$$\lim_{n\to\infty}\frac{S_n-n(p-q)}{\sqrt{4npq}}\sim N(0,1)$$

进而

$$\lim_{n\to\infty}P\{S_n>0\}=P\{\frac{S_n-n(p-q)}{\sqrt{4npq}}>-\frac{\sqrt{n}}{2}\frac{(p-q)}{\sqrt{pq}}\}$$

根据$(p-q)$符号的不同，有以下几种分类：

若$p>q$，$\lim\limits_{n\to\infty}P\{S_n>0\}=1$

若$p=q$，$\lim\limits_{n\to\infty}P\{S_n>0\}=0.5$

若$p<q$，$\lim\limits_{n\to\infty}P\{S_n>0\}=0$

因为$\forall k\neq j$，h_k,h_j是互斥事件，所以累积概率$\{\sum\limits_{k=0}^{n}h_k\}$表示事件$\bigcup\limits_{i=1}^{n}\{S_i=1\text{ for the first time}\}$发生的概率。

类似于上面的分类，我们逐一讨论当$n\to\infty$时，累积分布的情况$\{\sum\limits_{k=0}^{n}h_k\}$。

考虑概率$1-\lim\limits_{n\to\infty}\sum\limits_{k=0}^{n}h_k$，其表示事件$\overline{\bigcup\limits_{i=1}^{\infty}\{S_i=1\text{ for the first time}\}}$，该事件有如下等价：

$$\overline{\bigcup_{i=1}^{\infty}\{S_i=1\text{ for the first time}\}}\Leftrightarrow\bigcap_{i=1}^{\infty}\overline{\{S_i=1\text{ for the first time}\}}\Leftrightarrow\bigcap_{i=1}^{\infty}\{S_i\leq 0\}$$

当$p>q$时，有$\lim\limits_{n\to\infty}P\{S_n>0\}=1$，因此$P\bigcap\limits_{i=1}^{\infty}\{S_i\leq 0\}=0$，从而$\lim\limits_{n\to\infty}\sum\limits_{k=0}^{n}h_k=1$；

当$p=q$时，有$\lim\limits_{n\to\infty}P\{S_n>0\}=0.5$，因此$P\bigcap\limits_{i=1}^{\infty}\{S_i\leq 0\}=0$，从而$\lim\limits_{n\to\infty}\sum\limits_{k=0}^{n}h_k=1$；

当$p<q$时，有$\lim\limits_{n\to\infty}P\{S_n>0\}=0$，因此$P\bigcap\limits_{i=1}^{\infty}\{S_i\leq 0\}\neq 0$，从而$\lim\limits_{n\to\infty}\sum\limits_{k=0}^{n}h_k\neq 1$。

Question 2

Consider Problem Q1 again. Let $H(z)$ denote the probability generating function of $\{h_n\}$.

(a)

Show that $H(z)$ satisfies the quadratic equation in $H(z): qz(H(z))^2-H(z)+pz=0$.

先写出$H(z)$的表达式，根据概率生成函数的定义：

$$\begin{aligned}H(z)=&\sum_{n=0}^{\infty}h_nz^n\\=&h_0+h_1z+\sum_{n=2}^{\infty}h_nz^n\\=&pz+\sum_{n=2}^{\infty}h_nz^n\end{aligned}$$

等式$H(z): qz(H(z))^2 - H(z) + pz = 0$等价于$qz(H(z))^2 = H(z) - pz$.、

等式的右边为：

$$\begin{aligned} H(z) - pz &= \sum_{n=2}^{\infty} h_n z^n \qquad \text{根据Q1(a)} \\ &= \sum_{n=2}^{\infty}\sum_{j=1}^{n-2} q h_j h_{n-j-1} z^n \qquad \text{移出公因式q} \\ &= q\sum_{n=2}^{\infty}\sum_{j=1}^{n-2} h_j h_{n-j-1} z^n \end{aligned}$$

等式的左边为：

$$\begin{aligned} qz(H(z))^2 &= qz[\sum_{n=0}^{\infty} h_n z^n]^2 \\ &= qz[\sum_{0\le i,j} h_i h_j z^{i+j}] \qquad \text{移入z} \\ &= q[\sum_{0\le i,j} h_i h_j z^{i+j+1}] \end{aligned}$$

观察左右两边，即证

$$\sum_{0\le i,j} h_i h_j z^{i+j+1} = \sum_{n=2}^{\infty}\sum_{j=1}^{n-2} h_j h_{n-j-1} z^n$$

下面按照z的n次幂逐一比对系数进行证明：

当$n=1$时，左边的系数为$h_0h_0 = 0$，右边的系数为0，相等；

当$n \ge 2$时，左边的系数为$\sum\limits_{i+j+1=n} h_i h_j$，右边的系数为$\sum\limits_{j=1}^{n-2} h_j h_{n-j-1}$，有

$$\begin{aligned} \sum_{i+j+1=n} h_i h_j &= \sum_{j=0}^{n-1} h_j h_{n-j-1} \\ &= h_j h_{n-j-1}|_{j=0} + \sum_{j=1}^{n-2} h_j h_{n-j-1} + h_j h_{n-j-1}|_{j=n-1} \\ &= h_0 h_{n-1} + \sum_{j=1}^{n-2} h_j h_{n-j-1} + h_{n-1} h_0 \\ &= \sum_{j=1}^{n-2} h_j h_{n-j-1} \end{aligned}$$

证毕。

(b)

Solve the equation for $H(z)$.

$H(z): qz(H(z))^2 - H(z) + pz = 0$为一元二次方程，利用求根公式解得：

$$H(z) = \frac{1 \pm \sqrt{1-4pqz^2}}{2qz} \qquad (|z| \le 1)$$

因为$H(z) \le 1$，所以

$$H(z) = \frac{1 - \sqrt{1-4pqz^2}}{2qz} \qquad (|z| \le 1)$$

(c)

Let N denote the first passage time. Note that $P(N<\infty)=\sum_{n=1}^{\infty}h_n=H(z)|_{z=1}$.

Find $P(N<\infty)$ for the three cases: $p>q,\ p=q,\ p<q$.

首先计算出$P(N<\infty)$的表达式

$$P(N<\infty)=H(z)|_{z=1}=\frac{1-\sqrt{1-4pq}}{2q}$$

利用Python编程，绘制出不同p取值情况下的$P(N<\infty)$:

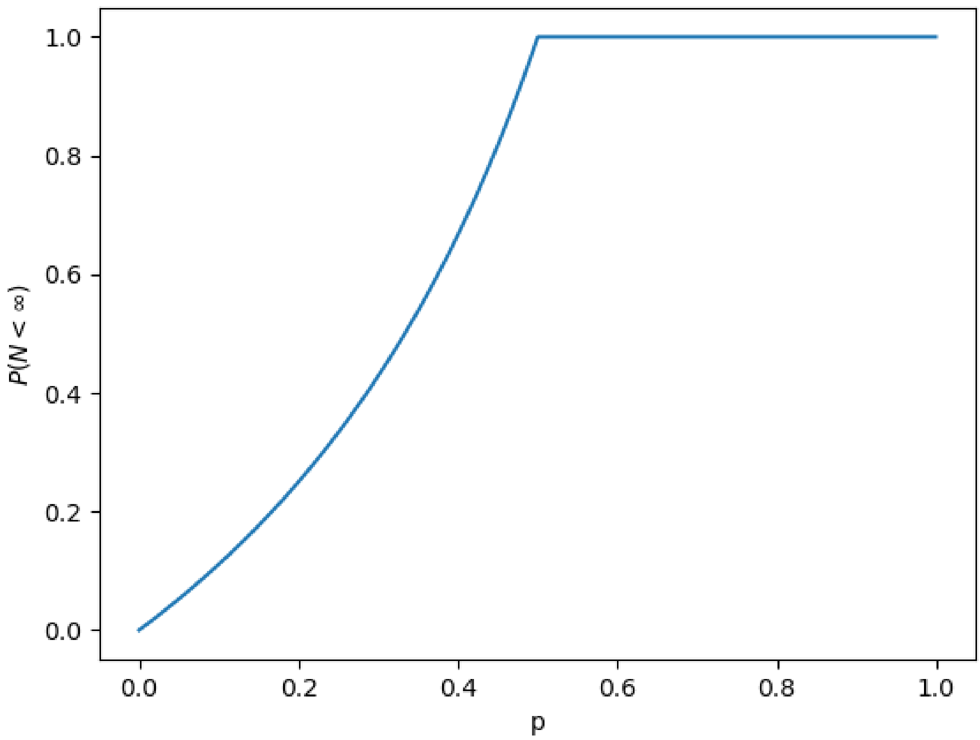
```python
import numpy as np
from matplotlib import pyplot as plt

p = np.linspace(0,1,1000)
q = 1-p

def H(p,q):
    s = 1-np.sqrt(1-4*p*q)
    m = 2*q
    return s/m

h = H(p,q)

plt.plot(p,h)
plt.xlabel('p')
plt.ylabel(r'$P( N < \infty )$')
plt.show()
```



若$p>q$，则$P(N<\infty)=1$;

若$p=q$，则$P(N<\infty)=1$;

若$p<q$，则$P(N<\infty)\neq 1$。

(d)

Compute the cumulative distribution $\{\sum_{k=0}^{n} h_k\}$ for $p = 0.45$ over the range $n = 0 : 20$ by inverting $H(z)$ numerically using MATLAB M-function `invt_pgf`.

查阅作者在其网站提供的函数，具体如下：

> https://ww2.mathworks.cn/matlabcentral/fileexchange/2265-an-introduction-to-stochastic-processes?s_tid=srchtitle

```matlab
function  invt_pgf
%
%  Inverting PGF to the time domain
%  Algorithm based on Abate and Whitt
%  "Numerical inversion of probability generation function"
%  OR Letter, Vol 12 (1992) 245-251, 1992.
%
pmf=[];
for n=1:10
ga=8; r=10^(ga/(2*n)); r=1/r; h=pi/n; u=1/(2*n*r^n); sum=0;
for k=1:n-1
   z=r*exp(i*h*k); sum=sum+((-1)^k)*e129pgf(z);
end
pn=2*sum+e129pgf(r)+(-1)^n*e129pgf(-r); pn=u*pn; pmf=[pmf pn];
end
for n=1:10
   fprintf('  p(%2.0f) = %8.4f  \n',n,pmf(n));
end

function [y]=e129pgf(z)
%
%  pgf of Example 1.2.9
%
p=0.6; q=1-p; up=(p^2)*(z^2); down=(1-q*z-p*q*z^2); y=up/down;
y=real(y);
```

将其用Python进行复写，并作出一些改进：

```python
import numpy as np
import pandas as pd
from itertools import accumulate

def pgf(z):
    '''
    定义概率生成函数
    '''
    p = 0.45
    q = 1 - p
    up = 1 - np.sqrt(1-4*p*q*z**2) # 分子
    down = 2*q*z # 分母
    y = up/down
    return np.real(y) # 获取 PGF 的实部值，以确保返回一个实数

def invt_pgf():
    '''
    反演概率生成函数到时域
    算法基于 Abate 和 Whitt 的论文
    '''
    pmf = [0] # 初始化概率质量函数数组
    max_n = 20
    for n in range(1,max_n):
        # 依据算法进行反演
        ga = 8
```

```python
        r = np.power(10, ga/(2*n))
        r = 1 / r
        h = np.pi / n;
        u = 1 / (2*n*r**n)
        sum = 0
        for k in range(1,n):
            z = r * np.exp(1j * h * k)
            sum = sum + ((-1)**k) * pgf(z)

        pn = 2 * sum + pgf(r) + (-1)**n * pgf(-r)
        pn = u * pn

        # 将反演结果添加到概率质量函数数组中
        pmf.append(pn)

    pmf = list(accumulate(pmf))
    return pmf

def Q1():
    '''
    Q1中计算累计概率的方法,用于结果进行比对
    '''
    max_n = 20
    p = 0.45
    q = 1-p

    h = []
    h_0 = 0 # 计算h0
    h.append(h_0)
    h_1 = p # 计算h1
    h.append(h_1)

    for n in range(2, max_n):
        # 计算hn
        h_n = 0
        for j in range(1, n-2+1):
            # 公式中的下标是从1到n-2
            # 因此range中的参数为1,n-2+1
            h_n += q * h[j] * h[n-j-1]
        h.append(h_n)

    # 计算逐一位置的和
    h = list(accumulate(h))
    return h

if __name__ == '__main__':
    pmf = invt_pgf()
    h = Q1()
    result = pd.DataFrame([h,pmf],index=['Q1','Q2']).T
    print(result)
```

结果如下:

```
          Q1        Q2
0   0.000000  0.000000
1   0.450000  0.450000
2   0.450000  0.450000
3   0.561375  0.561375
4   0.561375  0.561375
5   0.616506  0.616506
6   0.616506  0.616506
7   0.650618  0.650618
8   0.650618  0.650618
9   0.674257  0.674257
10  0.674257  0.674257
```

```
11  0.691810  0.691810
12  0.691810  0.691810
13  0.705463  0.705463
14  0.705463  0.705463
15  0.716445  0.716446
16  0.716445  0.716446
17  0.725506  0.725506
18  0.725506  0.725506
19  0.733130  0.733130
```

二者几乎一样，没有明显差距。

第二次作业

Question 1

Let X_1 and X_2 be independent exponential random variables, each with a rate of μ. Define

$$X_{(1)} = min(X_1, X_2), \quad X_{(2)} = max(X_1, X_2).$$

Determine the following: the expected value $E[X_{(1)}]$, the variance $Var[X_{(1)}]$, the expected value $E[X_{(2)}]$, and the variance $Var[X_{(2)}]$.

X_1, X_2的概率密度函数为

$$f_X(X=x) = \mu e^{-\mu x}, \quad x \geq 0$$

分布函数为

$$F_X(X=x) = 1 - e^{-\mu x}, \quad x \geq 0$$

根据次序统计量的概率密度函数公式计算可得：

$$\begin{aligned} f_{X_{(1)}}(X_{(1)}=x) &= 2[1-F_X(X=x)]f_X(X=x) = 2\mu e^{-2\mu x} \\ f_{X_{(2)}}(X_{(2)}=x) &= 2F_X(X=x)f(X=x) = 2\mu e^{-\mu x} - 2\mu e^{-2\mu x} \end{aligned}$$

因为$f_{X_{(1)}}(X_{(1)}=x) \sim Exp(2\mu)$，则$E[X_{(1)}] = \dfrac{1}{2\mu}$，$Var[X_{(1)}] = \dfrac{1}{4\mu^2}$。

下面计算$E[X_{(2)}]$和$Var[X_{(2)}]$：

$$\begin{aligned} E[X_{(2)}] &= \int_0^\infty x f_{X_{(2)}}(X_{(2)}=x)dx \\ &= \int_0^\infty x(2\mu e^{-\mu x} - 2\mu e^{-2\mu x})dx \\ &= 2\int_0^\infty \mu x e^{-\mu x}dx - \int_0^\infty 2\mu x e^{-2\mu x}dx \\ &= 2E[X] - E[X_{(1)}] \\ &= \frac{2}{\mu} - \frac{1}{2\mu} \\ &= \frac{3}{2\mu} \end{aligned}$$

$$
\begin{aligned}
E[X_{(2)}^2] &= \int_0^\infty x^2 f_{X_{(2)}}(X_{(2)} = x)dx \\
&= \int_0^\infty x^2(2\mu e^{-\mu x} - 2\mu e^{-2\mu x})dx \\
&= 2\int_0^\infty \mu x^2 e^{-\mu x}dx - \int_0^\infty 2\mu x^2 e^{-2\mu x}dx \\
&= 2E[X^2] - E[X_{(1)}^2] \\
&= 2(Var[X] + (E[X])^2) - (Var[X_{(1)}] + (E[X_{(1)}])^2) \\
&= \frac{4}{\mu^2} - \frac{1}{2\mu^2} \\
&= \frac{7}{2\mu^2}
\end{aligned}
$$

$$
Var[X_{(2)}] = E[X_{(2)}^2] - (E[X_{(2)}])^2 = \frac{7}{2\mu^2} - \frac{9}{4\mu^2} = \frac{5}{4\mu^2}
$$

Question 2

Consider an inventory system in which demand for an item follows a Poisson process with rate λ. At each demand occurrence only one unit of the item is needed. The system starts with S units on hand and will replenish when the inventory becomes empty. Each replenishment brings the inventory back to level S. Let $X(t)$ denote the inventory level at time t and T the length of the interval starting with the epoch at which the inventory moves up to S and ending with the epoch when the inventory reaches zero. The area under the sample path $X(t)$ is given by $A = \int_0^T X(t)dt$. Find $E[A]$. Note that if h is the cost of one unit inventory per unit time then $hE[A]$ gives the cost of holding inventory per replenishment cycle.

记$T_n, n = 1, 2, 3, ...$表示第n次需求发生的时刻，同时规定$T_0 = 0$。

记$X_n, n = 1, 2, 3, ...$表示第n次需求发生与第$n-1$次需求发生的时间间隔。

因为需求的到达过程是速率为 λ 的泊松过程，所以$X_n, n = 1, 2, 3, ...$服从参数为λ的指数分布，且相互独立。

在给定$X_1, X_2, \ldots, X_S$的条件下，根据题目中T的定义有：

$$
T = \sum_{i=1}^S X_i
$$

有$X(t)$如下：

$$
X(t) = \begin{cases} S & 0 \le t < X_1 \\ S-k & \sum_{i=1}^k X_i \le t < \sum_{i=1}^{k+1} X_i \quad k = 1, 2, \ldots, S-1 \\ 0 & \sum_{i=1}^S X_i = t \end{cases}
$$

进而可以计算A：

$$
A = \sum_{i=1}^S (S - i + 1)X_i
$$

根据双重期望定理有：

$$
E[A] = E\{E[A|X_1, X_2, \ldots X_s]\}
$$

先计算$E[A|X_1, X_2, \ldots X_s]$：

$$
\begin{aligned}
E[A|X_1, X_2, \ldots X_s] &= E[\sum_{i=1}^S (S-i+1)X_i] \qquad X_1, X_2, \ldots, X_n\text{相互独立} \\
&= \sum_{i=1}^S (S-i+1)E(X_i) \qquad X_1, X_2, \ldots, X_n\text{服从参数为}\lambda\text{的指数分布} \\
&= \frac{1}{\lambda} \cdot \sum_{i=1}^S (S-i+1) \\
&= \frac{1}{\lambda} \cdot \frac{S(S+1)}{2}
\end{aligned}
$$

再计算$E[A]$:

$$E[A]=E[\frac{1}{\lambda}\cdot\frac{S(S+1)}{2}]=\frac{1}{\lambda}\cdot\frac{S(S+1)}{2}$$

Question 3

Consider a mathematical model designed for estimating the number of COVID-19 infections. The model is based on the following assumptions:

1. Individuals get infected with COVID-19 following a Poisson process at an unknown rate λ.
2. The time from infection to symptom onset is modeled as a random variable with a known distribution G.
3. The incubation periods for different infected individuals are independent of each other.

Let $N_1(t)$ denote the number of people who show symptoms up to time t, and $N_2(t)$ denote the number of people who are COVID-19 positive but asymptomatic up to time t. It's known that an individual infected at time s will show symptoms at time t with probability $G(t-s)$ and remain asymptomatic at time t with probability $\bar{G}(t-s)$.

(1)

Given the value of $N_1(t)$ as n_1, derive an estimate $\hat{\lambda}$ for λ and use this estimate to calculate the expected value of $N_2(t)$. Assume that G follows an exponential distribution with mean μ. Provide the expression for the estimated value of $N_2(t)$.

$N_1(t)$表示到时间t已出现症状的人数，$N_2(t)$表示到时间t感染但未出现症状的人数。

定义类型1事件为出现症状，类型2事件为感染但没出现症状，那么接下来可以使用泊松过程的分解来解决问题。

考虑在s时刻被感染的人，那么他会以$G(t-s)$的概率归为类型1，以$\bar{G}(t-s)$的概率归为类型2。

根据泊松过程的分解，事件$\{N_1(t)=n\}$发生的概率为

$$P\{N_1(t)=n\}=e^{-\lambda pt}\frac{(\lambda pt)^n}{n!} \tag{1}$$

其中

$$p=\frac{1}{t}\int_0^t G(t-s)ds \tag{2}$$

事件$\{N_2(t)=m\}$发生的概率为:

$$P\{N_2(t)=m\}=e^{-\lambda qt}\frac{(\lambda qt)^m}{m!} \tag{3}$$

其中

$$q=1-p=1-\frac{1}{t}\int_0^t G(t-s)ds \tag{4}$$

在$N_1(t)=n_1$的条件下，用最大似然估计法来估计λ:

似然函数为:

$$L(\lambda)=e^{-\lambda pt}\frac{(\lambda pt)^{n_1}}{n_1!}$$

对数似然函数为:

$$\begin{aligned}\ln L(\lambda)&=-\lambda pt+n_1\ln(\lambda pt)-\ln(n_1!)\\&=-\lambda pt+n_1\ln\lambda+n_1\ln(pt)-\ln(n_1!)\end{aligned}$$

对数似然方程为:

$$
\begin{aligned}
\frac{d}{d\lambda}\ln L(\lambda) &= 0 \\
-pt + \frac{n_1}{\lambda} &= 0 \\
\lambda &= \frac{n_1}{pt}
\end{aligned}
$$

则λ的最大似然估计为$\hat{\lambda} = \frac{n_1}{pt}$。

根据泊松过程的分解，$N_2(t)$的期望为$E[N_2(t)] = \lambda qt$。

代入最大似然估计$\hat{\lambda} = \frac{n_1}{pt}$，$N_2(t)$的期望为$\hat{E}[N_2(t)] = \hat{\lambda}qt = \frac{n_1 q}{p}$。

若G服从均值为μ的指数分布，那么$G \sim Exp(1/\mu)$，即$G(X = x) = 1 - e^{-\frac{1}{\mu}x}, x \geq 0$.

那么依据(2)式计算p:

$$
\begin{aligned}
p &= \frac{1}{t}\int_0^t G(t-s)ds \\
&= \frac{1}{t}\int_0^t [1 - e^{-\frac{1}{\mu}(t-s)}]ds \\
&= 1 - \frac{\mu}{t} + \frac{\mu}{t}e^{-\frac{t}{\mu}}
\end{aligned}
$$

依据(4)式计算q:

$$
\begin{aligned}
q &= 1 - p \\
&= \frac{\mu}{t} - \frac{\mu}{t}e^{-\frac{t}{\mu}}
\end{aligned}
$$

则$N_2(t)$期望的表达式为:

$$
\hat{E}[N_2(t)] = \frac{n_1 q}{p} = n_1\left(\frac{\mu}{t} - \frac{\mu}{t}e^{-\frac{t}{\mu}}\right)/\left(1 - \frac{\mu}{t} + \frac{\mu}{t}e^{-\frac{t}{\mu}}\right) \tag{5}
$$

(2)

If $t = 16$ years, $\mu = 10$ years,and $n_1 = 220,000$, calculate the estimated number of asymptomatic infections at the 16-year mark.

将$t = 16$，$\mu = 10$，$n_1 = 220,000$，代入(5)式计算得:

```
import numpy as np
t = 16
mu = 10
n1 = 220000
q = mu/t - mu/t*np.exp(-t/mu)
p = 1 - mu/t + mu/t * np.exp(-t/mu)
res = n1*q/p
print(res)
# 218959.38204129488
```

$$
\hat{E}[N_2(t = 16)] = 218959.38204129488
$$

Question 4

Consider a Poisson process with rate λ. For $i = 1, 2$, we let S_i denote the time of the ith arrival. Use the fact that the two events $\{S_i \leq t_1, S_2 \leq t_2\}$ and $\{N(t_1) \geq 1, N(t_2) \geq 2\}$ are equivalent to find the joint distribution $F_{S_1,S_2}(t_1, t_2)$ of S_1 and S_2, where $0 < t_1 < t_2$.

$$
\begin{aligned}
F_{S_1,S_2}(t_1,t_2) &= P\{S_1 \le t_1, S_2 \le t_2\} \\
&\downarrow \text{等价事件} \\
&= P\{N(t_1) \ge 1, N(t_2) \ge 2\} \\
&\downarrow \text{条件概率} \\
&= P\{N(t_2) \ge 2 | N(t_1) \ge 1\} \cdot P\{N(t_1) \ge 1\} \\
&\downarrow \text{等价事件} \\
&= \sum_{n=1}^{\infty} P\{N(t_2) \ge 2 | N(t_1) = n\} \cdot P\{N(t_1) = n\} \\
&\downarrow n \ge 2\text{时，} P\{N(t_2) \ge 2 | N(t_1) = n\} \equiv 1 \\
&= P\{N(t_2) \ge 2 | N(t_1) = 1\} \cdot P\{N(t_1) = 1\} + \sum_{n=2}^{\infty} P\{N(t_1) = n\} \\
&\downarrow \text{事件的逆} \\
&= [1 - P\{N(t_2) < 2 | N(t_1) = 1\}] \cdot P\{N(t_1) = 1\} + \sum_{n=2}^{\infty} P\{N(t_1) = n\} \\
&\downarrow \text{等价事件} \\
&= [1 - P\{N(t_2) = 1 | N(t_1) = 1\}] \cdot P\{N(t_1) = 1\} + \sum_{n=2}^{\infty} P\{N(t_1) = n\}
\end{aligned}
$$

因为事件$\{N(t_2) = 1 | N(t_1) = 1\}$等价于事件$\{N(t_2) - N(t_1) = 0 | N(t_1) = 1\}$，则:

$$
\begin{aligned}
P\{N(t_2) = 1 | N(t_1) = 1\} &= P\{N(t_2) - N(t_1) = 0 | N(t_1) = 1\} \\
&\downarrow \text{独立增加性质} \\
&= P\{N(t_2) - N(t_1) = 0\} \\
&\downarrow \text{稳定增加性质} \\
&= P\{N(t_2 - t_1) = 0\} \\
&= e^{-\lambda(t_2 - t_1)}
\end{aligned}
$$

代回$F_{S_1,S_2}(t_1,t_2)$的计算公式得:

$$
\begin{aligned}
F_{S_1,S_2}(t_1,t_2) &= [1 - P\{N(t_2) = 1 | N(t_1) = 1\}] \cdot P\{N(t_1) = 1\} + \sum_{n=2}^{\infty} P\{N(t_1) = n\} \\
&\downarrow \text{等价事件} \\
&= [1 - e^{-\lambda(t_2 - t_1)}] \cdot \lambda t_1 e^{-\lambda t_1} + P\{N(t_1) \ge 2\} \\
&\downarrow \text{事件的逆} \\
&= \lambda t_1 e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_2} + [1 - P\{N(t_1) < 2\}] \\
&\downarrow \text{等价事件} \\
&= \lambda t_1 e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_2} + [1 - P\{N(t_1) = 0\} - P\{N(t_1) = 1\}] \\
&= \lambda t_1 e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_2} + 1 - e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_1} \\
&= 1 - e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_2}
\end{aligned}
$$

综上:

$$
F_{S_1,S_2}(t_1,t_2) = 1 - e^{-\lambda t_1} - \lambda t_1 e^{-\lambda t_2}
$$

Question 5

Ships arrive at the port according to a Poisson process with rate 5 per day. The ship's dwell time X (day) is subject to uniform distribution $U\,[0,10]$. Let $Y(t)$ denote the number of remaining ships at the port by time t.

(a)

Please describe $Y(t)$ by the response function $W_0(s,t)$.

设$N = \{N(t), t \ge 0\}$为$(0,t]$时间内船只到达港口的数量，则$N(t)$是速率为5的泊松过程。

用过滤泊松过程表达$Y(t)$:

$$
Y(t) = \sum_{n=1}^{N(t)} W(t, S_n, X_n)
$$

其中

$$
\begin{aligned}
W(t,S_n,X_n) &= W(t,\tau,y)\\
&\downarrow s=t-\tau\\
&= W_0(s,y)\\
&= \begin{cases} 1 & 0\le s\le y\\ 0 & \text{其它} \end{cases}
\end{aligned}
$$

函数$W(t,S_n,X_n)$为响应函数，t表示观察时刻，S_n或τ表示第n个船只的到达时刻，X_n或y表示第n个船只的停泊时长。

(b)

Compute $E[Y(t)]$.

过滤泊松过程的期望公式如下：

$$
X(t)=\sum_{i=1}^{N(t)} w(t,S_n,Y_n)
$$

$$
E[X(t)]=\lambda\int_0^t E[w(t,\tau,y)]d\tau
$$

其中λ为泊松过程$N(t)$的速率，$w(t,S_n,Y_n)$为响应函数。

下面计算$E[Y(t)]$：

$$
\begin{aligned}
E[Y(t)] &= \lambda\int_0^t E[W(t,\tau,y)]d\tau\\
&= 5\int_0^t E[W(t,\tau,y)]d\tau
\end{aligned}
$$

先计算$E[W(t,\tau,y)]$：

$$
\begin{aligned}
E[W(t,\tau,y)] &= E[W_0(t-\tau,y)]\\
&\downarrow t\text{是固定的}, \tau\in(0,t], y\in[0,10], \text{取点}(\tau,y)\text{的概率是}\frac{1}{10t}\\
&= \iint\limits_{0\le t-\tau\le y} 1\cdot\frac{1}{10t}d\tau dy\\
&= \begin{cases} 1-\frac{t}{20} & 0<t\le 10\\ \frac{5}{t} & t>10 \end{cases}
\end{aligned}
$$

代会$E[Y(t)]$得：

$$
\begin{aligned}
E[Y(t)] &= 5\int_0^t E[W(t,\tau,y)]d\tau\\
&= \begin{cases} 5\int_0^t(1-\frac{t}{20})d\tau & 0<t\le 10\\ 5\int_0^t\frac{5}{t}d\tau & t>10 \end{cases}\\
&= \begin{cases} 5(t-\frac{t^2}{20}) & 0<t\le 10\\ 5\cdot 5 & t>10 \end{cases}\\
&= \begin{cases} 5t-\frac{t^2}{4} & 0<t\le 10\\ 25 & t>10 \end{cases}
\end{aligned}
$$

综上：

$$
E[Y(t)] = \begin{cases} 5t-\frac{t^2}{4} & 0<t\le 10\\ 25 & t>10 \end{cases}
$$

第三次作业

Question 1

In an $M/G/\infty$ queueing system, a cleanup occurs at fixed intervals $T, 2T, 3T, \ldots$. At the start of each cleanup, all customers currently being served are forced to leave prematurely, and each of them pays a fee of C_1. It is assumed that each cleanup takes time $T/4$, and customers arriving during the cleanup are lost, with each lost customer incurring a cost of C_2.

(a)

Determine the long-term average cost per unit time.

设$C(t)$表示t时刻的期望成本，即求$\lim\limits_{t\to\infty}\dfrac{C(t)}{t}$。

设泊松过程的速率为λ。

根据泊松过程的性质，在一个清理周期$T/4$内，预计到达的顾客数量为$\lambda\cdot\dfrac{T}{4}$，则一个清理周期成本为$C_2\cdot\lambda\cdot\dfrac{T}{4}$。

若清理时间t不足一个周期$T/4$，即$0<t<T/4$，则预计到达的顾客数量为$\lambda\cdot t$，清理成本为$C_2\cdot\lambda\cdot t$。

那么$C(t)$有如下表达式：

$$C(t)=\begin{cases}0 & 0<t\le T\\ (n-1)\cdot C_2\cdot\lambda\cdot\frac{T}{4}+C_2\cdot\lambda\cdot(t-nT) & nT<t\le nT+\frac{T}{4}, n=1,2,\ldots\\ n\cdot C_2\cdot\lambda\cdot\frac{T}{4} & nT+\frac{T}{4}<t\le(n+1)T, n=1,2,\ldots\end{cases}$$

则

$$\lim_{t\to\infty}\frac{C(t)}{t}=\begin{cases}\lim\limits_{t\to\infty}C_2\lambda\cdot[1-\frac{(3n+1)T}{4t}] & nT<t\le nT+\frac{T}{4}, n=1,2,\ldots\\ \lim\limits_{t\to\infty}C_2\lambda\cdot\frac{nT}{4t} & nT+\frac{T}{4}<t\le(n+1)T, n=1,2,\ldots\end{cases}$$

因为$\lim\limits_{t\to\infty}\frac{nT}{t}=1$，则

$$\lim_{t\to\infty}\frac{C(t)}{t}=\frac{1}{4}C_2\lambda$$

则长期情况下单位时间平均成本为$\dfrac{1}{4}C_2\lambda$。

(b)

Calculate the long-term proportion of time the system spends in cleanup.

设$R(t)$表示t时刻系统用于清理的总时间，即求$\lim\limits_{t\to\infty}\dfrac{R(t)}{t}$。

类似于(a)部分进行讨论，那么$R(t)$有如下表达式：

$$R(t)=\begin{cases}0 & 0<t\le T\\ (n-1)\cdot\frac{T}{4}+(t-nT) & nT<t\le nT+\frac{T}{4}, n=1,2,\ldots\\ n\cdot\frac{T}{4} & nT+\frac{T}{4}<t\le(n+1)T, n=1,2,\ldots\end{cases}$$

则

$$\lim_{t\to\infty}\frac{R(t)}{t}=\begin{cases}\lim\limits_{t\to\infty}[1-\frac{(3n+1)T}{4t}] & nT<t\le nT+\frac{T}{4}, n=1,2,\ldots\\ \lim\limits_{t\to\infty}\frac{nT}{4t} & nT+\frac{T}{4}<t\le(n+1)T, n=1,2,\ldots\end{cases}$$

因为$\lim\limits_{t\to\infty}\frac{nT}{t}=1$，则

$$\lim_{t\to\infty}\frac{R(t)}{t}=\frac{1}{4}$$

那么长期情况下清理时间所占比例为$\dfrac{1}{4}$。

Question 2

Consider a tennis match involving players A and B. It is assumed that in each game where A starts serving, the probability of A winning is p_a, and the probability of B winning is $q_a = 1 - p_a$. Similarly, in games where B starts serving, the probability of A winning is p_b, and the probability of B winning is $q_b = 1 - p_b$. Assume that the winner of each game earns 1 point and serves in the next game.

(a)

What is the long-term proportion of points won by player A?

根据比赛规则，获胜者可以在下一局中先发球，那么我们可以根据选手先发球的次数推测选手的得分。

假设比赛在$t = 0, 1, 2, ...$时刻都举行一次，令计数过程$N = \{N(t), t > 0\}$表示$(0, t]$时间内A先发球的次数。

因为$N(t)$不包含0时刻的情况，所以不必做第一场比赛发球选手的讨论。

根据比赛规则，可以直接得出$(0, t]$时间内A选手的得分为$N(t)$，$(0, t]$时间内两选手的总积分为t取下整，即$\lfloor t \rfloor$。

则A选手得分的长期积分比例为：

$$\lim_{t \to \infty} \frac{E[N(t)]}{\lfloor t \rfloor}$$

假设事件$\{A\text{先发球}\}$发生的时间间隔为X，X服从的分布为F，均值为μ，方差为σ^2。

此时$N(t)$是一个更新过程，则根据极限的性质和基本更新定理 (elementary renewal theorem)有：

$$\lim_{t \to \infty} \frac{E[N(t)]}{\lfloor t \rfloor} = \lim_{t \to \infty} \frac{E[N(t)]}{t} = \frac{1}{\mu}$$

下面求μ：

X表示两次$\{A\text{先发球}\}$的间隔时间，其取值为离散整数，可以计算其分布列为：

X	1	2	3	4	$\cdots$	n	$\cdots$
$P\{X = x\}$	p_a	$q_a \cdot p_b$	$q_a \cdot q_b \cdot p_b$	$q_a \cdot q_b^2 \cdot p_b$	$\cdots$	$q_a \cdot q_b^{n-2} \cdot p_b$	$\cdots$

则μ为

$$\begin{aligned}
\mu &= \sum_{n=1}^{\infty} n \cdot P\{X = n\} \\
&= p_a + q_a \cdot \sum_{n=2}^{\infty} n \cdot q_b^{n-2} \cdot p_b \\
&= p_a + \frac{q_a}{q_b} \cdot \sum_{n=2}^{\infty} n \cdot q_b^{n-1} \cdot p_b \\
&= p_a - p_b \cdot \frac{q_a}{q_b} + \frac{q_a}{q_b} \cdot \sum_{n=1}^{\infty} n \cdot q_b^{n-1} \cdot p_b \\
&\downarrow \sum_{n=1}^{\infty} n \cdot q_b^{n-1} \cdot p_b \text{是几何分布} GE(p_b) \text{的期望} \\
&= p_a - p_b \cdot \frac{q_a}{q_b} + \frac{q_a}{q_b} \cdot \frac{1}{p_b} \\
&= \frac{1 + p_b - p_a}{p_b}
\end{aligned}$$

则

$$\lim_{t \to \infty} \frac{E[N(t)]}{\lfloor t \rfloor} = \lim_{t \to \infty} \frac{E[N(t)]}{t} = \frac{p_b}{1 + p_b - p_a}$$

(b)

If the serving protocol involves alternating serves between the players, meaning A serves on the first point, then B serves on the second, followed by A serving on the third, and so on, what is the long-term proportion of points won by player A under this protocol?

记$N = \{N(t), t > 0\}$为$(0, t]$时间内A选手的得分。

如果规则是两位选手交替发球，那么可以认为在t足够大的情况下，A、B选手分别发球$\frac{\lfloor t \rfloor}{2}$次。

则A选手得分的期望$E[N(t)]$为:

$$E[N(t)] = \frac{\lfloor t \rfloor}{2} \cdot p_a + \frac{\lfloor t \rfloor}{2} \cdot p_b$$

则A选手得分的长期积分比例为:

$$\lim_{t \to \infty} \frac{E[N(t)]}{\lfloor t \rfloor} = \frac{p_a + p_b}{2}$$

Question 3

Consider a TV game show with one million dollars at stake. The host selects a random number X from a uniform distribution over the interval $(0, 1)$ and divides the one million into two pots: the X million dollar pot and the $(1 - X)$ million dollar pot. You will independently select a random number Y from a uniform distribution over the interval $(0, 1)$. If $Y \leq X$, you will take the X million dollar pot; otherwise you will take the $(1 - X)$ million dollar pot.

(a)

What is your expected winning?

记W为获得的奖金，在观众选择$Y = y$的情况下:

若$X = x$落在$[y, 1)$区间，则对应的奖金为x百万美元，因为X服从$(0, 1)$上的均匀分布，所以这种情况的概率为1;

若$X = x$落在$(0, y)$区间，则对应的奖金为$1 - x$百万美元，因为X服从$(0, 1)$上的均匀分布，所以这种情况的概率为1。

则$E[W|Y = y]$为:

$$E[W|Y = y] = \int_0^y (1 - x) dx + \int_y^1 x dx = \frac{1}{2} + y - y^2$$

在更一般的情况下，观众获得奖金的平均值为:

$$E[W] = \int_0^1 [\frac{1}{2} + y - y^2] dy = \frac{2}{3}$$

(b)

What is the moral of the story?

即使在不确定性的情况下，通过谨慎的判断和选择，依然能够获得足够大的利益。

例如本例中，每次都选择$Y = 0.5$，那么会得到最大的预期收益。

Question 4

The Distribution of the Last Arrival Time before t of a Renewal Process .

We recall that $S_{N(t)} = X_1 + \cdots + X_{N(t)}$ denotes the last arrival time before time t.

(1)

Show that

$$P\{S_{N(t)} \le s\} = \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \quad 0 \le s \le t.$$

当$0 \le s \le t$时，事件$\{S_{N(t)} \le s\}$等价于事件$\{t - S_{N(t)} \ge t - s\}$，因此我们可以借用$t - S_{N(t)}$的分布来推导$S_{N(t)}$的分布。

因为$t - S_{N(t)}$是更新过程的当前寿命$A(t)$，所以直接利用书上的(3.3.8)式，其给出了$A(t)$的分布为：

$$U_t(x) = \begin{cases} F(t) - \int_0^{t-x}[1-F(t-y)]m(y)dy & if\ x < t \\ 1 & if\ x \ge t \end{cases}$$

那么

$$\begin{aligned} P\{t - S_{N(t)} \ge t - s\} &= 1 - U_t(t-s) \\ &\downarrow t - s < t \\ &= 1 - F(t) + \int_0^{t-(t-s)} [1-F(t-y)]m(y)dy \\ &= \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \end{aligned}$$

综上：

$$P\{S_{N(t)} \le s\} = P\{t - S_{N(t)} \ge t - s\} = \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \quad 0 \le s \le t$$

证毕。

(2)

What is the density of $S_{N(t)}$?

设$G_t(s)$为$S_{N(t)}$的分布函数，$g_t(s)$为$S_{N(t)}$的密度函数。

则

$$\begin{aligned} G_t(s) &= \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \quad 0 \le s \le t \\ g_t(s) &= \frac{d}{ds}G_t(s) = \frac{d}{ds}\int_0^s \bar{F}(t-y)m(y)dy = \bar{F}(t-s)m(s) \end{aligned}$$

(3)

For the Poisson process with parameter λ，find $P\{S_{N(t)} \le s\}$.

因为$P\{S_{N(t)} \le s\} = \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \quad 0 \le s \le t$，所以只要找到$\bar{F}(t)$和$m(t)$再进行计算即可。

$\bar{F}(t) = 1 - F(t)$，其中$F(t)$为相邻到达的时间间隔的分布函数。

对于速率为λ的泊松过程来说，$F(t)$是参数为λ的指数分布：$F(t) = 1 - e^{-\lambda t}, t \ge 0$。

令$f(t)$为$F(t)$的密度函数：$f(t) = \lambda e^{-\lambda t}, t \ge 0$，那么$f$的拉普拉斯变换为：$f^e(s) = \lambda/(s+\lambda)$。

因为更新密度函数$m(t)$的拉普拉斯变换与$f(t)$的拉普拉斯变换具有以下关系：$m^e(s) = \dfrac{f^e(s)}{1 - f^e(s)}$。

从而得到 $m^e(s) = \lambda/s$，再用拉普拉斯逆变换得$m(t) = \lambda, t \ge 0$。

代入$P\{S_{N(t)} \le s\}$的公式得：

$$
\begin{aligned}
P\{S_{N(t)} \leq s\} &= \bar{F}(t) + \int_0^s \bar{F}(t-y)m(y)dy \qquad && 0 \leq s \leq t \\
&= e^{-\lambda t} + \int_0^s e^{-\lambda(t-y)} \cdot \lambda dy && 0 \leq s \leq t \\
&= e^{-\lambda t} + \lambda e^{-\lambda t}(\frac{1}{\lambda}e^{\lambda s} - \frac{1}{\lambda}) && 0 \leq s \leq t \\
&= e^{\lambda(s-t)} && 0 \leq s \leq t
\end{aligned}
$$

第四次作业

Question 1

Given state space $S = \{1,2,3,4\}$, and its transition probability matrix:

$$
P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix}
$$

(a)

Give the classification of all the states with corresponding reason.

根据转移矩阵P，状态之间可到达的(accessible)关系如下：

$$
\begin{aligned}
1 \rightarrow 2 \\
2 \rightarrow 1 \\
2 \rightarrow 2 \\
3 \rightarrow 2 \\
3 \rightarrow 4 \\
4 \rightarrow 1 \\
4 \rightarrow 4
\end{aligned}
$$

因此状态之间的互通(communicate)关系如下：

$$
\begin{aligned}
1 \leftrightarrow 2 \\
1 \nleftrightarrow 3 \\
1 \nleftrightarrow 4 \\
2 \nleftrightarrow 3 \\
2 \nleftrightarrow 4 \\
3 \nleftrightarrow 4
\end{aligned}
$$

因此根据等价类(equivalence class)的概念，可以将状态空间划分为3类，分别为$\{1,2\}$、$\{3\}$、$\{4\}$。

根据闭类(closed class)的定义，因为类$\{1,2\}$无法到达类外的状态，所以$\{1,2\}$为闭类；
因为类$\{3\}$可以到达状态2和4，类$\{4\}$可以到达状态1，所以类$\{3\}$和$\{4\}$不是闭类。

各状态到自身的首次通过时间(first passage time)的分布为：

$$
\begin{aligned}
f_{11}^{(n)} &= P\{T_{11} = n\} = \begin{cases} 0 & n = 1 \\ (\frac{1}{2})^{n-1} & n \geq 2 \end{cases} \\
f_{22}^{(n)} &= P\{T_{22} = n\} = \begin{cases} \frac{1}{2} & n = 1,2 \\ 0 & n \geq 2 \end{cases} \\
f_{33}^{(n)} &= P\{T_{33} = n\} = \{0 \quad n \geq 1 \\
f_{44}^{(n)} &= P\{T_{44} = n\} = \begin{cases} \frac{2}{3} & n = 1 \\ 0 & n \geq 2 \end{cases}
\end{aligned}
$$

所以各状态到自身的到达概率(reaching probability)为：

$$f_{11} = P\{T_{11} < \infty\} = \sum_{n=1}^{\infty} f_{11}^{(n)} = 1$$
$$f_{22} = P\{T_{22} < \infty\} = \sum_{n=1}^{\infty} f_{22}^{(n)} = 1$$
$$f_{33} = P\{T_{33} < \infty\} = \sum_{n=1}^{\infty} f_{33}^{(n)} = 0$$
$$f_{44} = P\{T_{44} < \infty\} = \sum_{n=1}^{\infty} f_{44}^{(n)} = \frac{2}{3}$$

则根据常返(recurrent)的定义，状态1、2是常返态，状态3、4为瞬态。
又

$$E[T_{11}] = 1 \cdot 0 + \sum_{n=2}^{\infty} n \cdot (\frac{1}{2})^{n-1} = 3 < \infty$$
$$E[T_{22}] = 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{3}{2} < \infty$$

所以状态1、2都是正常返(positive recurrent)。

因为$p_{22}^{(1)} = \frac{1}{2} > 0$，且周期性是一个类属性，所以类$\{1, 2\}$的是非周期(aperiodic)的。

(b)

Find all the ergodic states in the example.

根据part (a)的讨论可知，有且仅有状态1和2为遍历态。

Question 2

Consider there are m black balls and m white balls in two bags, and there are m balls in each bag. Take two balls randomly from each bag at a time and swap the two balls. Define X_n as the number of black balls remained in the first bag after the n-th exchange.

(a)

Write the state space of this Markov chain for the random variable X_n.

X_n的状态空间S为：

$$S = \{0, 1, 2, \dots, m\}$$

(b)

Draw the Markov chain.

在$X_n = k, k \in S$的情况下，X_{n+1}的分布如下：

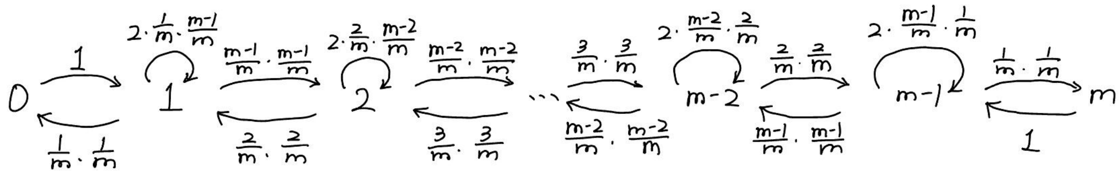
X_{n+1}	**black ball from first bag**	**white ball from first bag**
black ball from second bag	$X_{n+1} = k$	$X_{n+1} = k + 1$
white ball from second bag	$X_{n+1} = k - 1$	$X_{n+1} = k$

P	**black ball from first bag**	**white ball from first bag**
black ball from second bag	$\frac{k}{m} \cdot \frac{m-k}{m}$	$\frac{m-k}{m} \cdot \frac{m-k}{m}$
white ball from second bag	$\frac{k}{m} \cdot \frac{k}{m}$	$\frac{k}{m} \cdot \frac{m-k}{m}$

则转移矩阵为：

$$
P = \begin{array}{c|ccccccc}
 & 0 & 1 & 2 & \cdots & m-2 & m-1 & m \\
\hline
0 & 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\
1 & \frac{1}{m}\cdot\frac{1}{m} & 2\cdot\frac{1}{m}\cdot\frac{m-1}{m} & \frac{m-1}{m}\cdot\frac{m-1}{m} & \cdots & 0 & 0 & 0 \\
2 & 0 & \frac{2}{m}\cdot\frac{2}{m} & 2\cdot\frac{2}{m}\cdot\frac{m-2}{m} & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
m-2 & 0 & 0 & 0 & \cdots & 2\cdot\frac{m-2}{m}\cdot\frac{2}{m} & \frac{2}{m}\cdot\frac{2}{m} & 0 \\
m-1 & 0 & 0 & 0 & \cdots & \frac{m-1}{m}\cdot\frac{m-1}{m} & 2\cdot\frac{m-1}{m}\cdot\frac{1}{m} & \frac{1}{m}\cdot\frac{1}{m} \\
m & 0 & 0 & 0 & \cdots & 0 & 1 & 0
\end{array}
$$

则马链为：



(c)

Is there a stationary distribution? If yes, give the distribution.

下面仅考虑$m < \infty$的情况：

1. 上述马链中的各状态是互通的，因此该马链是不可约的。
2. 根据上述马链的状态图和转移矩阵，该马链是正常返的。
3. 因为上述马链中的转移矩阵$P = \{p_{ij}\}$有$p_{11} \neq 0$，即$p_{11}^{(1)} \neq 0$，所以该马链是非周期的。

因此该马链是各态遍历马链，其状态的极限分布就是唯一的稳态分布，下面进行求解。

设状态的极限分布为$\pi = \{\pi_j\}, j \in S$。针对每个状态建立平衡方程组：

$$
\begin{gathered}
\frac{1}{m^2}\pi_1 = \pi_0 \\
\frac{(m-k+1)^2}{m^2}\pi_{k-1} + \frac{(k+1)^2}{m^2}\pi_{k+1} = \frac{k^2}{m^2}\pi_k + \frac{(m-k)^2}{m^2}\pi_k \qquad (1 \le k \le m-1) \\
\frac{1}{m^2}\pi_{m-1} = \pi_m
\end{gathered} \tag{2.1}
$$

同时还有：

$$
\sum_{j=0}^{m} \pi_j = 1 \tag{2.2}
$$

方程组（2.1）可以化简为：

$$
\frac{(k+1)^2}{m^2}\pi_{k+1} = \frac{(m-k)^2}{m^2}\pi_k \qquad (0 \le k \le m-1) \tag{2.3}
$$

式（2.3）是一个递推公式，从中可以进一步推出

$$
\pi_{j+1} = [\binom{m}{j+1}]^2 \pi_0 \qquad 0 \le j \le m-1 \tag{2.4}
$$

下面运用归纳法证明：

当$j = 0$时，有$\pi_1 = [\binom{m}{1}]^2 \pi_0 = m^2 \pi_0$，满足；

当$j = k$成立时，验证$k+1$的情况，根据递推公式（2.3）有：

$$
\begin{aligned}
\pi_{k+1} &= \frac{(m-k)^2}{(k+1)^2}\pi_k \\
&= \frac{(m-k)^2}{(k+1)^2}[\binom{m}{k}]^2\pi_0 \\
&= [\frac{m!}{k!(m-k)!}\cdot\frac{(m-k)}{(k+1)}]^2\pi_0 \\
&= [\frac{m!}{(k+1)!(m-k-1)!}]^2\pi_0 \\
&= [\binom{m}{k+1}]^2\pi_0
\end{aligned}
$$

证毕。

显然，式（2.4）等价于

$$
\pi_j = [\binom{m}{j}]^2\pi_0 \qquad 1 \le j \le m \tag{2.5}
$$

又发现$\pi_0 = [\binom{m}{0}]^2\pi_0$，其满足式（2.5）$j=0$的情况，所以可以将式（2.5）扩展为

$$
\pi_j = [\binom{m}{j}]^2\pi_0 \qquad 0 \le j \le m \tag{2.6}
$$

那么就可以根据$\sum_{j=0}^{m}\pi_j = 1$计算出π_0为:

$$
\begin{aligned}
\sum_{j=0}^{m}\pi_j &= 1 \\
\pi_0\sum_{j=0}^{m}[\binom{m}{j}]^2 &= 1 \\
\pi_0 &= 1/\sum_{j=0}^{m}[\binom{m}{j}]^2 \\
\pi_0 &= 1/\binom{2m}{m}
\end{aligned}
$$

最后一个等式成立是利用了组合数平方和公式，可参考如下博客:

> http://t.csdnimg.cn/Cbc9J

将π_0代回式（2.6）得:

$$
\pi_j = [\binom{m}{j}]^2/\binom{2m}{m} \qquad 0 \le j \le m
$$

综上：状态的极限分布$\pi = \{\pi_j\}, j \in S = \{0, 1, 2, \ldots, m\}$为

$$
\pi_j = [\binom{m}{j}]^2/\binom{2m}{m} \qquad 0 \le j \le m
$$

Question 3

A DNA nucleic acid is one of four types, and the standard model for nucleic acid mutations is a Markov chain model. It assumes that for $\alpha \in (0, \frac{1}{3})$, the nucleic acid remains unchanged with probability $1-3\alpha$ within a certain period of time. And if it changes, it is equally likely to change into one of the other three nucleic acid types.

(a)

Show that:

$$
P_{1,1}^n = \frac{1}{4} + \frac{3}{4}(1-4\alpha)^n
$$

由题知，核算每过一段时间进行一次突变，以概率$1-3\alpha, \alpha \in (0, \frac{1}{3})$保持不变，若发生变化，以等可能变为其他三种类型之一。

设X_n为第n次突变后的核酸类型，则X_n的状态空间$S = \{1, 2, 3, 4\}$。

转移矩阵为，

$$P = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 1-3\alpha & \alpha & \alpha & \alpha \\ 2 & \alpha & 1-3\alpha & \alpha & \alpha \\ 3 & \alpha & \alpha & 1-3\alpha & \alpha \\ 4 & \alpha & \alpha & \alpha & 1-\alpha \end{array}$$

P为实对称矩阵，可以进行对角化。

P的特征值为$\lambda_1 = 1, \lambda_{2,3,4} = 1-4\alpha$，故存在正交矩阵$A$（$A^{-1} = A^T$）：

$$A = \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & 0 & \frac{3}{\sqrt{12}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \end{pmatrix}$$

使得

$$A^{-1}PA = \begin{pmatrix} 1 & & & \\ & 1-4\alpha & & \\ & & 1-4\alpha & \\ & & & 1-4\alpha \end{pmatrix}$$

记上式的右侧矩阵为B，则

$$\begin{aligned} P^{(n)} &= P^n = AB^nA^{-1} \\ &= \begin{pmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \\ \frac{1}{2} & 0 & 0 & \frac{3}{\sqrt{12}} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} \end{pmatrix} \begin{pmatrix} 1 & & & \\ & (1-4\alpha)^n & & \\ & & (1-4\alpha)^n & \\ & & & (1-4\alpha)^n \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{12}} & \frac{3}{\sqrt{12}} & -\frac{1}{\sqrt{12}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{4}+\frac{3}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n \\ \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}+\frac{3}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n \\ \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}+\frac{3}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n \\ \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}-\frac{1}{4}(1-4\alpha)^n & \frac{1}{4}+\frac{3}{4}(1-4\alpha)^n \end{pmatrix} \end{aligned}$$

显然有：

$$P_{1,1}^n = \frac{1}{4} + \frac{3}{4}(1-4\alpha)^n$$

证毕。

(b)

Calculate the long-term proportion of time with this chain in each state.

1. 上述马链中的各状态是互通的，因此该马链是不可约的。
2. 根据上述马链的转移矩阵，该马链是正常返的。
3. 因为上述马链中的转移矩阵$P = \{p_{ij}\}$有$p_{11} \neq 0$（因为$\alpha \in (0, \frac{1}{3})$），即$p_{11}^{(1)} \neq 0$，所以该马链是非周期的。

因此该马链是各态遍历马链，设状态的极限分布为$\{\pi_j\}, j \in S = \{1,2,3,4\}$，根据part (a)，我们已经得到了$n$步转移概率矩阵，那么自然有

$$\pi_1 = \lim_{n\to\infty} P_{1,1}^{(n)} = \frac{1}{4}$$
$$\pi_2 = \lim_{n\to\infty} P_{2,2}^{(n)} = \frac{1}{4}$$
$$\pi_3 = \lim_{n\to\infty} P_{3,3}^{(n)} = \frac{1}{4}$$
$$\pi_4 = \lim_{n\to\infty} P_{4,4}^{(n)} = \frac{1}{4}$$

因此，该马尔科夫链在每个状态下的长期时间比例都为$\frac{1}{4}$。

Question4

Consider the number of arrival taxis in a taxi station follows a Poisson process with rate $\lambda_1 = 1$ per minute; the number of arrival customers follows another Poisson process with rate $\lambda_2 = 2$ per minute. Taxis will stop if no customer is waiting there, no matter there is a taxi or not; customers will not stop if no taxi is waiting there; customers will take a taxi if the taxi is already waiting there.

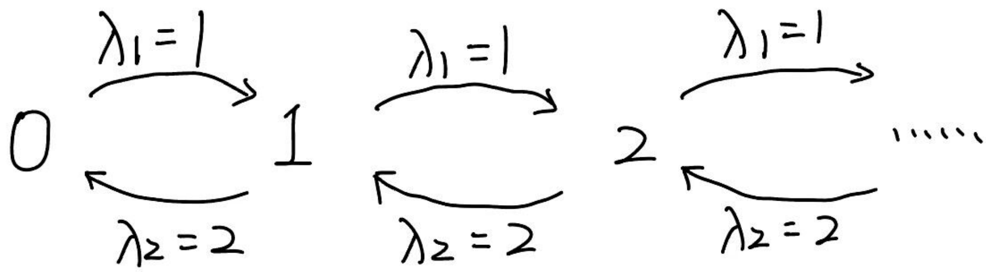
(a)

Give the flow chart and the Q matrix regarding to the number of waiting taxis.

设$\{X(t), t > 0\}$表示t时刻的出租车数量，其状态空间为$S = \{0, 1, ...\}$。

可以把本题看作一个出生消亡过程，其中出生速率为λ_1，消亡速率为λ_2。

类似于书上的图5.2，其转移图为：



其马链的无穷小算子矩阵Q为

$$Q = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots \\ 2 & -3 & 1 & 0 & \dots \\ 0 & 2 & -3 & 0 & \dots \\ 0 & 0 & 2 & -3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix}$$

(b)

Find the mean number of waiting taxis.

设$\{\pi_j\}, j \in S$，表示$X(t)$停留在状态j的长期概率，则

$$E[X(t)] = \sum_{j=0}^{\infty} j \cdot \pi_j$$

下面求解$\{\pi_j\}, j \in S$。

根据课本例5.3.2，出生消亡过程的长期概率为：

$$\pi_j = \pi_0 \rho_{j-1} \cdots \rho_0 \tag{4.1}$$

其中$\rho_j = \text{birth rate at state } j / \text{death rate at state } j+1$，在本题中$\rho_j \equiv \frac{1}{2}$。

因为$\sum_{j=0}^{\infty} \pi_j = 1$，则根据式（4.1）有：

$$\pi_0[1+\sum_{j=1}^{\infty}\rho_{j-1}\cdots\rho_0]=1$$
$$\pi_0[1+\sum_{j=1}^{\infty}(\frac{1}{2})^j]=1$$
$$\pi_0=\frac{1}{2}$$

将结果代回式（4.1）得：

$$\pi_j=(\frac{1}{2})^{j+1}\qquad j\in S$$

则

$$\begin{aligned}E[X(t)]&=\sum_{j=0}^{\infty}j\cdot\pi_j\\&=\sum_{j=0}^{\infty}j\cdot(\frac{1}{2})^{j+1}\\&=\frac{1}{2}\sum_{j=0}^{\infty}j\cdot(\frac{1}{2})^j\\&\downarrow\text{求和部分为几何分布}GE(\frac{1}{2})\text{的期望}\\&=\frac{1}{2}\cdot 2\\&=1\end{aligned}$$

Question 5

Consider the customers who enter a service system follow a Poisson process with arrival rate λ, there is only one waiter in the system, and the service time follows the exponential distribution with parameter μ. If the system is idle, the arrived customer will be served immediately. Otherwise he has to queue up. If there are already two customers waiting in the system, the new customer will leave and never come back. Let $X(t)$ be the number of customers in this system.

(1)

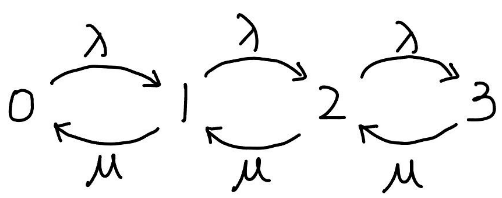
Write the state space of this Markov chain.

该马链的状态空间为$S=\{0,1,2,3\}$。

(2)

Draw the Markov chain and give the infinitesimal generator Q.

该马链的示意图为



该马链的无穷小算子矩阵为：

$$Q=\begin{pmatrix}-\lambda & \lambda & 0 & 0\\ \mu & -(\lambda+\mu) & \lambda & 0\\ 0 & \mu & -(\lambda+\mu) & \lambda\\ 0 & 0 & \mu & -\mu\end{pmatrix}$$

(3)

Describe the stable probability distribution of each state.

设$\{\pi_j\}, j \in S$为马链停留在状态j的长期概率。

建立平衡方程组：

$$\begin{cases} \lambda\pi_0 = \mu\pi_1 \\ \lambda\pi_1 = \mu\pi_2 \\ \lambda\pi_2 = \mu\pi_3 \\ \sum_{j=0}^{3}\pi_j = 1 \end{cases}$$

这类似于出生消亡过程长期概率的计算，结果为：

$$\pi_j = \pi_0(\frac{\lambda}{\mu})^j = \frac{1-\frac{\lambda}{\mu}}{1-(\frac{\lambda}{\mu})^4}(\frac{\lambda}{\mu})^j \qquad j \in S$$