Problem 3

FFR110 Computational biology

Max Sedersten
maxse@student.chalmers.se

Axel Johansson
axejoh@student.chalmers.se

March 13, 2023

Solution proposal

Task 1

The equation for the SIS model is given as

$$\begin{cases} \frac{dI}{dt} = \frac{\alpha}{S+I}SI - \beta I \\ \frac{dS}{dt} = -\frac{\alpha}{S+I}SI + \beta I \end{cases} . \tag{1}$$

Task 1a)

In order to determine the steady-state of the system, the equations outlined in model (1) were equated to zero. This process yielded a straightforward solution whereby $I^* = 0$ and any S. This solution means that there are no infected individuals within the population, suggesting an absence of the disease. The non trivial solution if found using $N = S + I$ in combination with

$$0 = \frac{\alpha}{S+I}S - \beta = \{I = N - S\} = \frac{\alpha}{N + \cancel{S} - \cancel{S}}S - \beta \quad \Longrightarrow S^* = N\frac{\beta}{\alpha}. \tag{2}$$

Rewriting

$$\frac{\alpha}{S+I}S - \beta = 0 \quad \text{as} \quad I = S\frac{(\alpha - \beta)}{\beta} \tag{3}$$

and inserting S^* we obtain

$$I^* = N\frac{\beta}{\alpha}\frac{\alpha - \beta}{\beta} = N\frac{\alpha - \beta}{\alpha}. \tag{4}$$

Hence the non trivial solution is

$$S^* = N\frac{\beta}{\alpha} \text{ and } I^* = N\frac{\alpha - \beta}{\alpha}. \tag{5}$$

Linear stability analysis is performed by analyzing the effect of a small perturbation around the steady state. By using the fact that $S = N - I$, we derive the following equation only based on I:

$$\frac{dI}{dt} = \frac{\alpha}{N}(N - I) - \beta I \tag{6}$$

$$
\begin{aligned}
&I = I^\star + \eta(t) \Rightarrow \\
\Rightarrow \cancel{\frac{dI^\star}{dt}} + \frac{d\eta(t)}{dt} &\approx \frac{\alpha}{N}(I^\star + \eta(t))(N - I^\star - \eta(t)) - \beta(I^\star + \eta(t)) \Rightarrow \\
\Rightarrow \frac{d\eta(t)}{dt} &= I^\star \left(\alpha - \alpha\frac{I^\star}{N} - \beta\right) + \eta(t)\left(a - \frac{2\alpha I^\star}{N} - \beta\right) \Rightarrow \\
&\Rightarrow \left[I^\star = N - \frac{\beta}{\alpha}N\right] \Rightarrow \\
&\Rightarrow \frac{d\eta(t)}{dt} = N\left(1 - \frac{\beta}{\alpha}\right)\cdot 0 + \eta(\beta - \alpha).
\end{aligned}
$$

Conclusion: $\beta < \alpha$

Task 1b)

Given the expression

$$\frac{dI}{dt} = \sum_{n=0}^{\infty} \underbrace{nb_{n-1}P_{n-1}}_{(1)} + \underbrace{nd_{n+1}P_{n+1}}_{(2)} - \underbrace{n(b_n + d_n)P_n}_{(3)} \tag{7}$$

The (3) term has correct index. But the first and second needs to be rewritten. Split the sum and for the (1) part let $n' = n - 1 \implies n = n' + 1$,

$$(1): \quad \sum_{n=0}^{\infty} nb_{n-1}P_{n-1} = \sum_{n'=-1}^{\infty}(n'+1)b_{n'}P_{n'} = \sum_{n=0}^{\infty}(n+1)b_nP_n. \tag{8}$$

In the third equal we have used the fact that the probability to have -1 individual is 0 and renamed n' as n. For part (3) do same as before but let $n' = n + 1 \implies n = n' - 1$ to get

$$(2): \quad \sum_{n=0}^{\infty} nd_{n+1}P_{n+1} = \sum_{n'=1}^{\infty}(n'-1)d_{n'-1}P_{n'} = \sum_{n=0}^{\infty}(n-1)d_nP_n - \cancelto{0}{d_0P_0}. \tag{9}$$

The product d_0P_0 cancels to zero because $d_0 = \beta \cdot 0$. Let us now write it on the same line

$$\sum_{n=0}^{\infty}\left[(n+1)b_n + (n-1)d_n - n(b_n + d_n)\right]P_n = \sum_{n=0}^{\infty}(b_n - d_n)P_n = \sum_{n=0}^{\infty}\left(\alpha n(1 - \frac{n}{N}) - \beta n\right)P_n. \tag{10}$$

But when $N \to \infty$ the fraction $n/N \to 0$ giving us

$$\frac{dI}{dt} = (\alpha - \beta)\sum_{n=0}^{\infty} nP_n = (\alpha - \beta)I. \tag{11}$$

We are told to compare the stochastic model with the deterministic. The theory predicts that they will be the same when population size goes to infinity. From first part in equation (1) and with $N = S + I$ we get

$$\frac{dI}{dt} = \frac{\alpha}{N}(N - I)I - \beta I = \alpha I - \cancelto{0}{\frac{\alpha I^2}{N}} - \beta I = (\alpha - \beta)I \tag{12}$$

which is the same as expected. This is because when population become infinite the system reaches a quasi-steady state because the variance becomes very small. The population does persists in the deterministic model but eventually goes extinct in the stochastic model due to fluctuations. However this is probability is very small for infinite population as the time untill extinction goes as $T_{ext} \sim e^N$. This wipe out is observed in later tasks.

Task 1c)

The figures 1-3 are the probability distribution with log scale. A line was fitted to the data which verified that the probability decay exponentially with exponent $-\lambda t$. We solved this using SciPy's exponential fit. Histograms are created from 10 000 iterations with timestep 1/1000.

The average time until an infection or recovery is equal to $1/\lambda$. This was confirmed by taking the mean of the values. Larger lambda means that the slope is steeper i.e. more of the events take place earlier and thus shorter time until an event. This makes sence because when b and d are large, drawing a random number smaller than them becomes easier. The scales in the figures also demonstrate that as lambda increases, the time decreases.

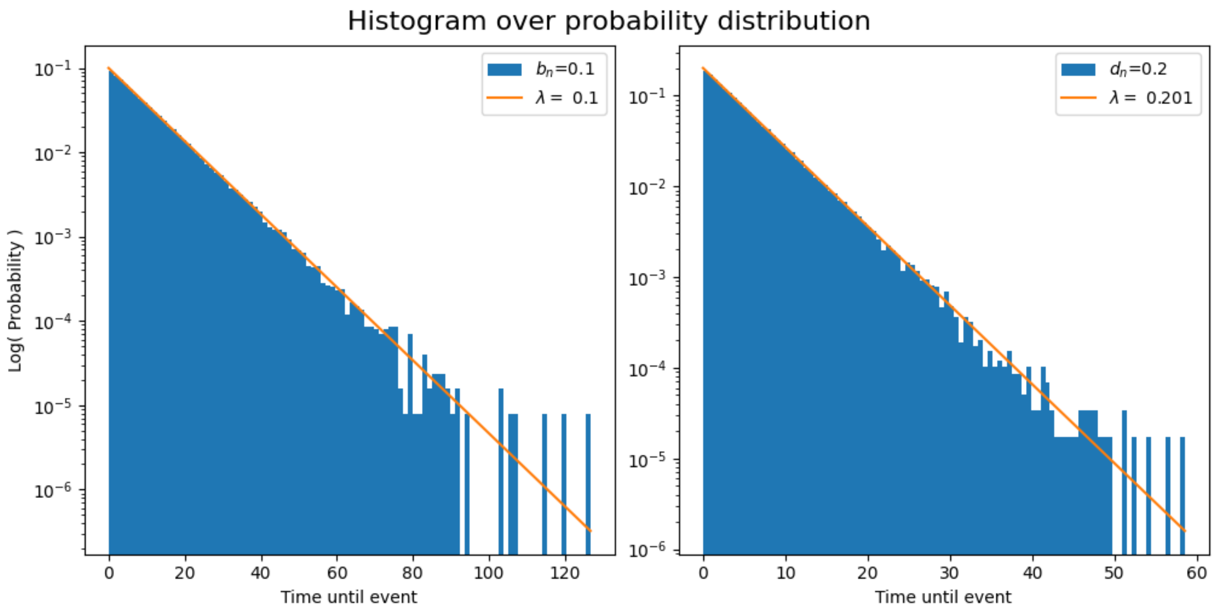

Figure 1: $b_n = 0.1$, $d_n = 0.2$:
Result: $\lambda_b \approx 0.1$, $\lambda_d \approx 0.201$

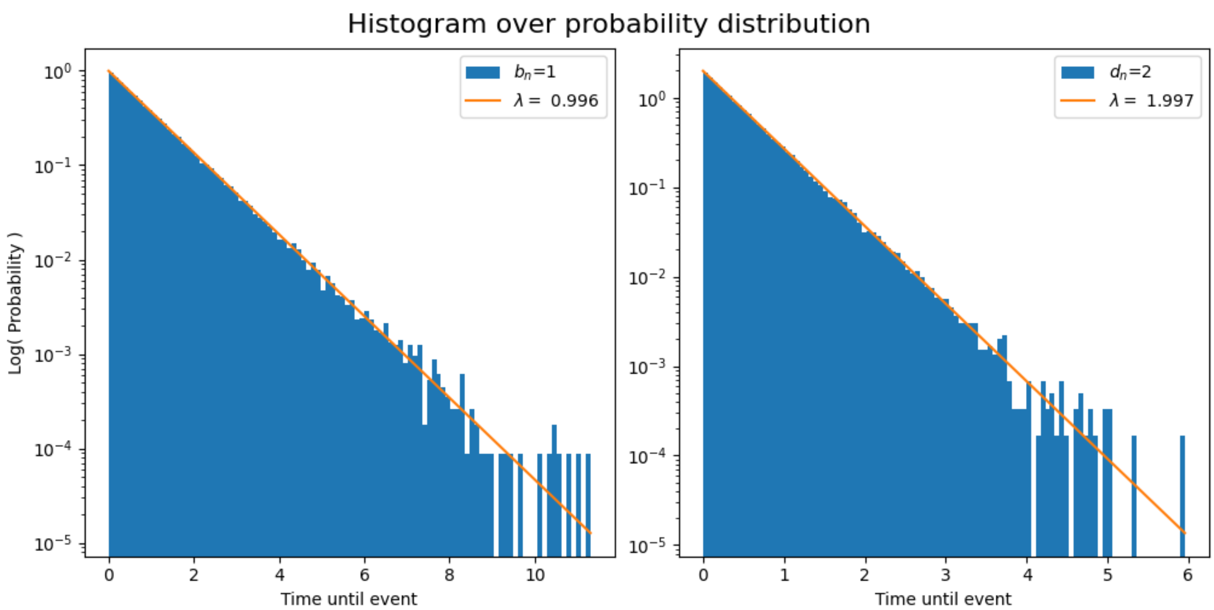

Figure 2: $b_n = 1$, $d_n = 2$:
Result: $\lambda_b \approx 0.996$, $\lambda_d \approx 1.997$

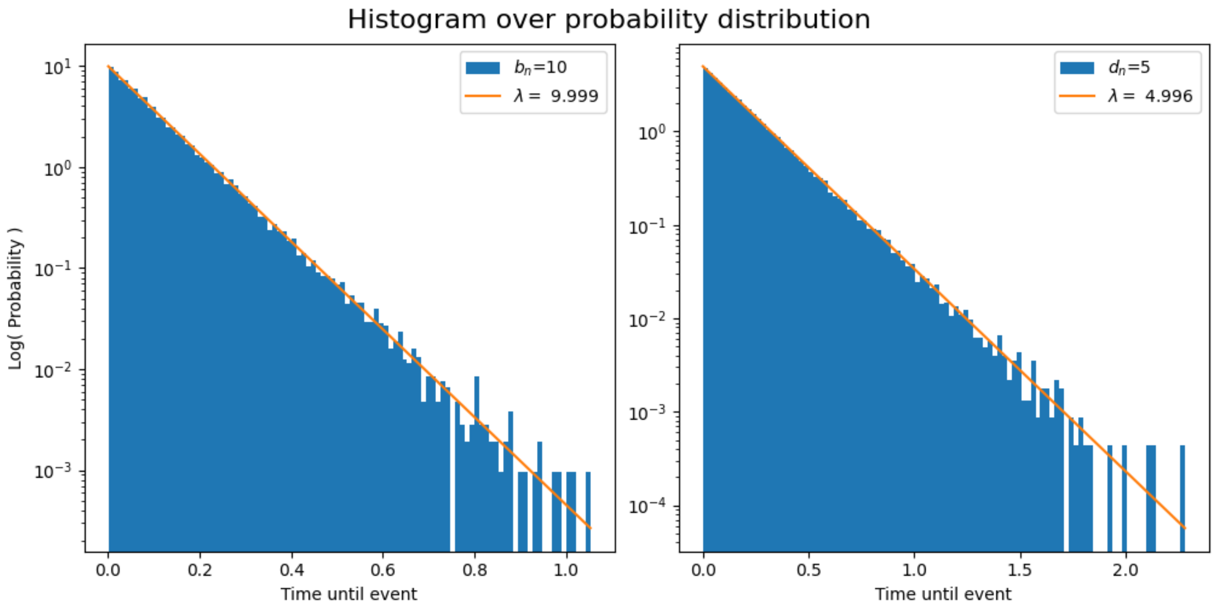

Figure 3: $b_n = 10$, $d_n = 5$:
Result: $\lambda_b \approx 9.999$, $\lambda_d \approx 4.996$

Task 1d)

The parameters choosen were the following:

$\alpha = 1.1$
$\beta = 1$
$N = 1000$

T_{ext} is calculated in the following way:

$$r_0 = \frac{\alpha}{\beta}$$
$$S(0) = \ln(r_0) - \left(1 - \frac{1}{r_0}\right)$$
$$T_{ext} = e^{NS(0)}$$
$$\Rightarrow T_{ext} \approx 81.5$$

In the beginning (t_1), the distribution approximates a gaussian distribution well, as can be seen in figure 4. In the middle scenario (t_2) when the time is close to the expected extinction time, the curve starts to become more lopsided which is shown in 5-6. This is due to the fact that the disease has gone extinct in a few of the runs, more in the later. To illustrate this further the minus logarithm of the probability is plotted in figure 7. Here we can see that the Graph is more tail-heavy towards the left side, especially for higher t.

Histogram for $t_1 \sim 0.25 T_{ext} < T_{ext}$

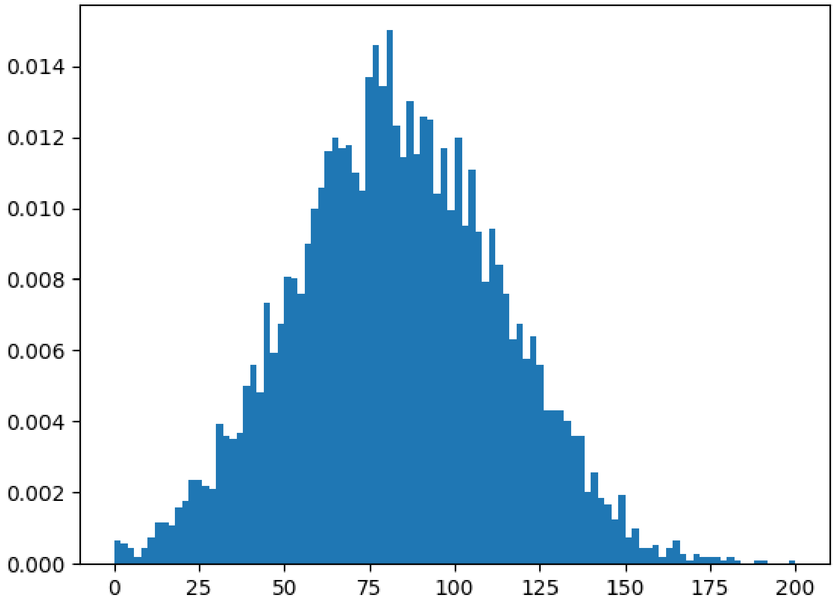

Figure 4: The figure shows the histogram over 6000 runs with 100 bins at value t_1.

Histogram for $t_2 \sim 0.9 \cdot T_{ext}$

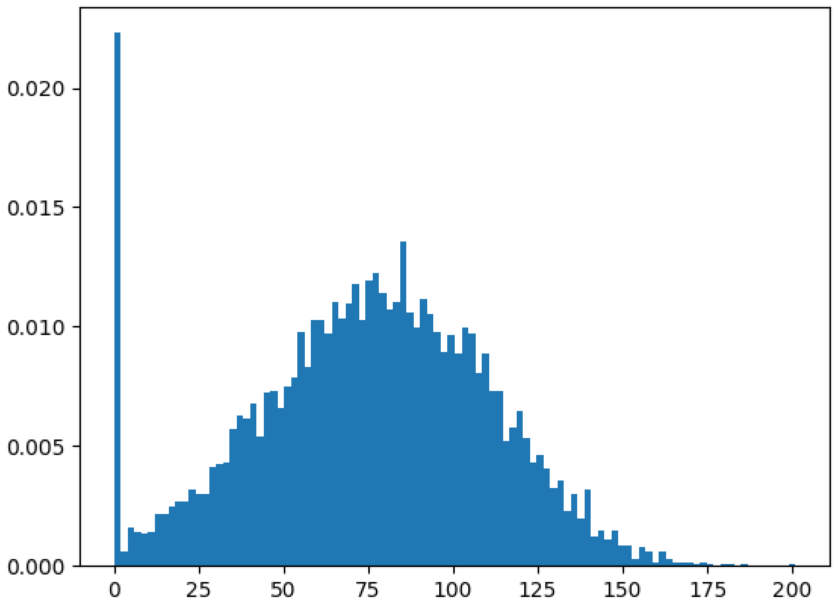

Figure 5: The figure shows the histogram over 6000 runs with 100 bins at value t_2.

Histogram for $t_3 \sim 1.1 \cdot T_{ext} > T_{ext}$

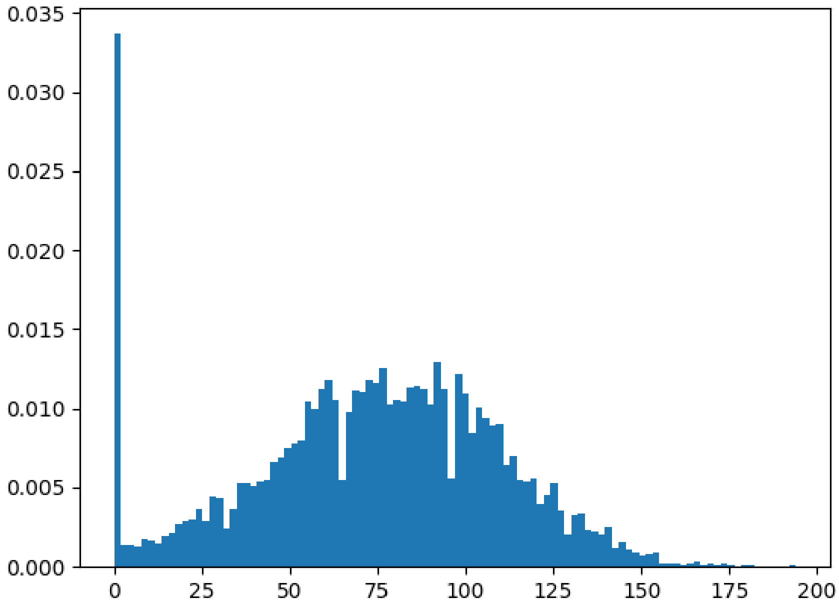

Figure 6: The figure shows the histogram over 6000 runs with 100 bins at value t_3. As can be seen the population seems to

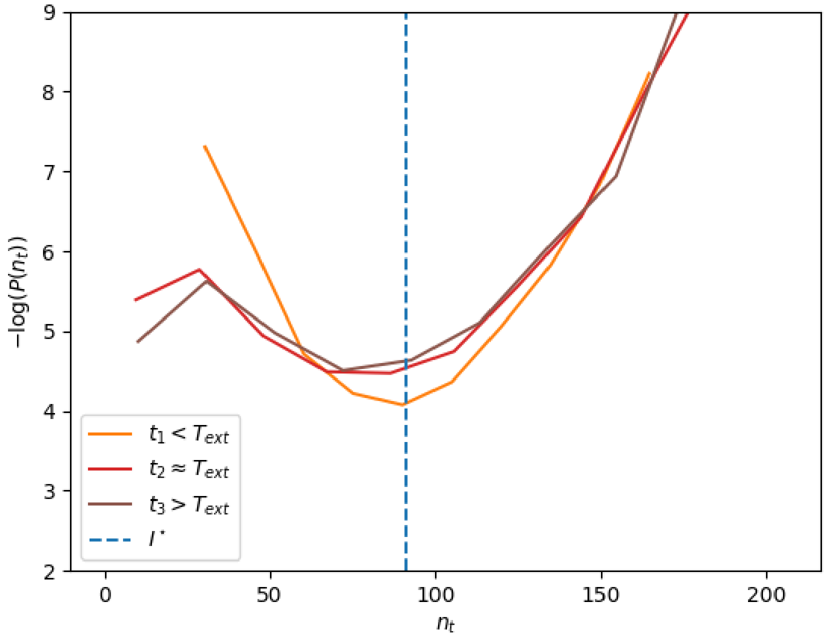

Figure 7: $-\log(P(t))$ for: $t = \{t_1, t_2, t_3\}$. As mentioned the figure shows a heavyside to the left.

Code

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import expon

bd_vals = [
    (0.1, 0.2),
    (1, 2),
    (10, 5)]
iterations = int(1e5)
b_t = np.zeros((len(bd_vals), iterations))
d_t = np.zeros((len(bd_vals), iterations))
dt = 1e-2

# Task d)
N = 10000
alpha = 0.015
beta = 0.01
time_max = 5000
number_of_runs = 3000
population = np.zeros([number_of_runs, int(time_max/dt)], dtype=int)
population[:, 0] = N * (1-beta/alpha)
time_observing = np.array([.2, .5, .9])*int(time_max/dt)


def bn(run, index):
    return alpha*population[run, index]*(1-population[run, index]/N)


def dn(run, index):
    return beta*population[run, index]


def sample_random_exp(a):
    return np.random.exponential(scale=1/a, size=1)


def time_to_event(a):
    t = 0
    while True:
        r = np.random.rand()
        t += dt
        if r < a * dt:
            return t


def simulation_c(b, d, index):
    for n in range(iterations):
        b_t[index, n] = time_to_event(b)
        d_t[index, n] = time_to_event(d)
```

```
def simulation_d():
    for run in range(number_of_runs):
        time, index_old = 0, 0
        while time < time_max:
            if population[run, index_old] == 0:
                population[run, index_old:] = population[run, index_old]
                break
            tb_sample = sample_random_exp(bn(run, index_old))
            td_sample = sample_random_exp(dn(run, index_old))
            time += min(tb_sample, td_sample)
            index = int(time/dt)
            if index > np.size(population, axis=1)-1:
                population[run, index_old:] = population[run, index_old]
                break
            population[run, index_old:index] = population[run, index_old]
            if tb_sample < td_sample:
                population[run, index] = population[run, index_old] + 1
            else:
                population[run, index] = population[run, index_old] - 1
            index_old = index
    plt.plot(population[0,:])
    plt.show()


def plot_histogram():
    for pl in range(3):
        plt.hist(population[:, int(time_observing[pl])], bins=100, density=True)
        plt.show()


def plot(index, b, d):
    fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(10, 5), constrained_layout=True)

    _, bins_b, _ = ax1.hist(b_t[index, :], bins=100, log=True, density=True, label=f'$b_n$={b
    params_b = expon.fit(b_t[index, :], floc=0)
    Y_b = expon.pdf(bins_b, *params_b)
    ax1.plot(bins_b, Y_b, label=f'$\lambda=$ {round(1/params_b[1], 3)}')
    ax1.set_ylabel('Log( Probability )')
    ax1.set_xlabel('Time until event')

    _, bins_d, _ = ax2.hist(d_t[index, :], bins=100, log=True, density=True, label=f'$d_n$={d
    params_d = expon.fit(d_t[index, :], floc=0)
    Y_d = expon.pdf(bins_d, *params_d)
    ax2.plot(bins_d, Y_d, label=f'$\lambda=$ {round(1/params_d[1], 3)}')

    ax2.set_xlabel('Time until event')

    fig.suptitle('Histogram over probability distribution', fontsize=16)
    ax1.legend()
    ax2.legend()
```

```
    plt.show()


def main():
    for index in range(len(bd_vals)):
        b, d = bd_vals[index]
        simulation_c(b, d, index)
        plot(index, b, d)
    simulation_d()
    plot_histogram()


if __name__ == '__main__':
    main()
```

Task 2

Task 2a)

We know that

$$P(S_n=j|T_c)=\frac{(\mu T_c)^j}{j!}e^{\mu T_c}, \tag{13}$$

and that

$$P(T_j)=\lambda_j e^{-\lambda_j T_j}, \quad \lambda_j=\frac{\binom{j}{2}}{N}, \quad T_c=\sum_{j=2}^{n} jT_j. \tag{14}$$

In the task $j=0$, so from equation (19) we get

$$P(S_n=0|T_c)=\langle e^{\mu T_c}\rangle=\langle e^{\mu\sum_{j=2}^{n} jT_j}\rangle=\left\langle\prod_{j=2}^{n} e^{\mu jT_j}\right\rangle=\prod_{j=2}^{n}\langle e^{\mu jT_j}\rangle \tag{15}$$

which can be rewritten in terms of an integral to

$$\prod_{j=2}^{n}\int_0^\infty dT_j P(T_j)e^{-\mu jT_j}=\prod_{j=2}^{n}\int_0^\infty dT_j\lambda_j e^{-T_j(\lambda_j+\mu j)}=\prod_{j=2}^{n}-\frac{\lambda_j}{\lambda_j+\mu j}\left[e^{-T_j(\lambda_j+\mu j)}\right]_0^\infty. \tag{16}$$

Inserting 0 gives a -1 and ∞ gives zero. Together with the expression for lambda from equation (14) gives

$$\prod_{j=2}^{n}\frac{\frac{\binom{j}{2}}{N}}{\frac{\binom{j}{2}}{N}+\mu j}=\prod_{j=2}^{n}\frac{\binom{j}{2}}{\binom{j}{2}+\frac{\theta j}{2}}=\left\{\binom{j}{2}=\frac{j(j-1)}{2}\right\}=\prod_{j=2}^{n}\frac{(j-1)}{(j-1)+\theta}. \tag{17}$$

The product can be expanded and rewritten to our final expression as,

$$P(S_n=0|T_c)=\prod_{j=2}^{n}\frac{(j-1)}{(j-1)+\theta}=\frac{(n-1)!}{(1+\theta)(2+\theta)\dots(n-1+\theta)}=\frac{(n-1)!\theta!}{(n-1+\theta)!}. \tag{18}$$

Task 2b)

Like subtask a) we have the probability given as

$$P(S_n=j|T_c)=\frac{(\mu T_c)^j}{j!}e^{\mu T_c}. \tag{19}$$

But now $n=2$ so $T_c=\sum_{j=2}^{2} jT_j=2T_2$ hence we get

$$P(S_2=j|T_c)=\left\langle\frac{(\mu 2T_2)^j}{j!}e^{-\mu 2T_2}\right\rangle=\int_0^\infty P(T_2)e^{\mu 2T_2}\frac{(\mu 2T_2)^2}{j!}dT_2. \tag{20}$$

Here we used that the expected value can be computed with an integral over our variable T_2. From equation (14) we know $P(T_2)$ with $\lambda_2=1/N$. So the expression becomes

$$\int_0^\infty\frac{1}{N}e^{-\frac{T_2}{N}}e^{-\mu 2T_2}\frac{(\mu 2T_2)^j}{j!}dT_2 \tag{21}$$

Now make a substitution $T_2/N=H$ and $dT_2=NdH$. Also using the fact that $\theta=2\mu N$ leaves us with

$$\int_0^\infty dHe^{-H(1+\theta)}\frac{(\theta H)^j}{j!}=\theta^j\int_0^\infty dHe^{-H(1+\theta)}\frac{H^j}{j!}=\frac{\theta^j}{(\theta+1)^{j+1}}=\frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^j. \tag{22}$$

Here we used the fact that the integral is solved with integration of parts j times to eliminate H^j. So $-(1+\theta)$ falls down from the exponent j times. The minus cancel from the one in integration of parts.