ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ

2η ΟΜΑΔΑ ΑΣΚΗΣΕΩΝ

ΘΟΔΩΡΗΣ ΑΡΑΠΗΣ – EL18028

7 ΜΑΙΟΥ, 2023

Ονοματεπώνυμο: Θοδωρής Αράπης	**ΑΜ:** el18028

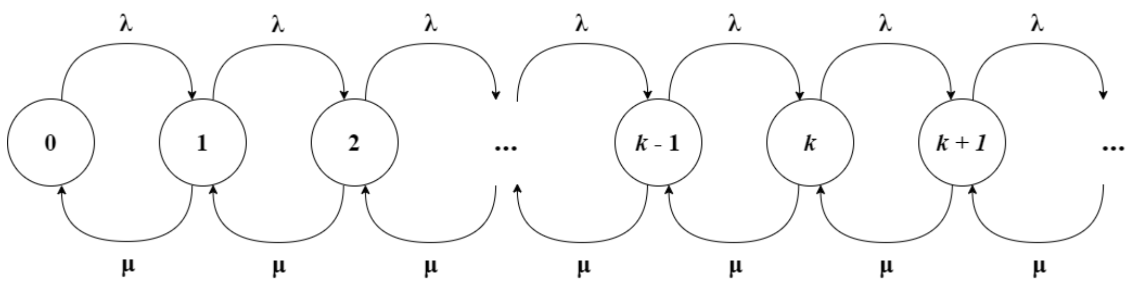
Θεωρητική μελέτη της ουράς M/M/1

(α)

Γνωρίζουμε από την θεωρία ότι για μια ουρά $Markov\ M/M/1$ (απείρου μεγέθους), όπου οι αφίξεις στην ουρά ακολουθούν κατανομή Poisson με παράμετρο $\boldsymbol{\lambda}\ \pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma/sec$ και οι εξυπηρετήσεις ακολουθούν εκθετική κατανομή με παράμετρο $\boldsymbol{\mu}\ \pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma/sec$, η συνθήκη οριακής ισορροπίας για να είναι η ουρά εργοδική είναι:

$$\rho = \frac{\lambda}{\mu} < 1\ Erlang$$

Το διάγραμμα ρυθμού μεταβάσεων της παραπάνω ουράς είναι:



Αξιοποιώντας τις Εξισώσεις Λεπτομερούς Ισορροπίας (Detailed Balance Equations), υπολογίζουμε τις εργοδικές πιθανότητες των καταστάσεων του συστήματος:

- $\lambda P_0 = \mu P_1 \Longrightarrow \boldsymbol{P_1} = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$

- $\lambda P_1 = \mu P_2 \Longrightarrow \boldsymbol{P_2} = \left(\frac{\lambda}{\mu}\right) P_1 = \rho P_1 = \rho^2 P_0, \quad \acute{\alpha}\rho\alpha\ \varepsilon\pi\alpha\gamma\omega\gamma\iota\kappa\acute{\alpha}$:
$$\boldsymbol{P_k = \rho^k P_0}$$

- $P_0 + P_1 + \cdots + P_k + \cdots = 1 \Longrightarrow P_0\left(1 + \rho + \cdots + \rho^k + \cdots\right) = 1 \xrightarrow{0<\rho<1} P_0\left(\frac{1}{1-\rho}\right) = 1,$
Οπότε προκύπτουν οι ζητούμενες εργοδικές πιθανότητες:

$$\boldsymbol{P_0 = (1 - \rho)}$$

$$\kappa\alpha\iota\ \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ \kappa\alpha\tau\acute{\alpha}\sigma\tau\alpha\sigma\eta:$$

$$\boldsymbol{P_k = (1 - \rho)\rho^k}$$

(β)

Σύμφωνα με την θεωρία, ο μέσος χρόνος καθυστέρησης (για άπειρη ουρά $Markov\ M/M/1$, όπου $P[blocking] = 0$) δίνεται από τον τύπο $Little$:

$$E(T) = \frac{E[n(t)]}{\gamma} = \frac{E[n(t)]}{\lambda} = \frac{\left(\frac{\rho}{1-\rho}\right)}{\lambda} \Longrightarrow$$

$$\boldsymbol{E(T) = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu(1-\rho)} = \frac{1}{\mu-\lambda}}$$

(γ)

Θα χρησιμοποιήσουμε τη σχέση για τις εργοδικές πιθανότητες που υπολογίσαμε στο ερώτημα (α).

$$\Gamma\iota\alpha\ k = 57\ \pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma:$$

$$P_{57} = (1-\rho)\rho^{57} > 0, \qquad 0 < \rho < 1$$

Συνεπώς, υπάρχει πιθανότητα να υπάρξουν 57 πελάτες κάποια χρονική στιγμή στο σύστημα. Η πιθανότητα αυτή εξαρτάται από την τιμή του $\rho = \frac{\lambda}{\mu}$, όσο μικραίνει το ρ τόσο μικραίνει η πιθανότητα. Πιο ειδικά, ισχύει:

$$\frac{dP(\rho)}{d\rho} = \rho^{56}(57 - 58\rho) > 0\ \ \gamma\iota\alpha\ \ \rho < \frac{57}{58}\ (\sigma\eta\mu\varepsilon\acute{\iota}o\ \mu\acute{\varepsilon}\gamma\iota\sigma\tau\eta\varsigma\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha\varsigma)$$

Το ρ μπορεί να μειωθεί είτε μειώνοντας τον ρυθμό αφίξεων λ και κρατώντας σταθερό τον ρυθμό εξυπηρέτησης μ, είτε αυξάνοντας το μ και κρατώντας σταθερό το λ, είτε αυξάνοντας το λ και μειώνοντας το μ την ίδια στιγμή.

<u>Ανάλυση ουράς M/M/1 με Octave</u>

(α)

Όπως εξηγήσαμε και στην προηγούμενη άσκηση, για να είναι εργοδικό το σύστημα θα πρέπει $\rho = \frac{\lambda}{\mu} < 1$, δηλαδή $\lambda < \mu$. Συνεπώς οι επιτρεπτές τιμές για τον ρυθμό εξυπηρέτησης είναι το διάστημα $(5, 10]$ $(\pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma/sec)$.

(β)

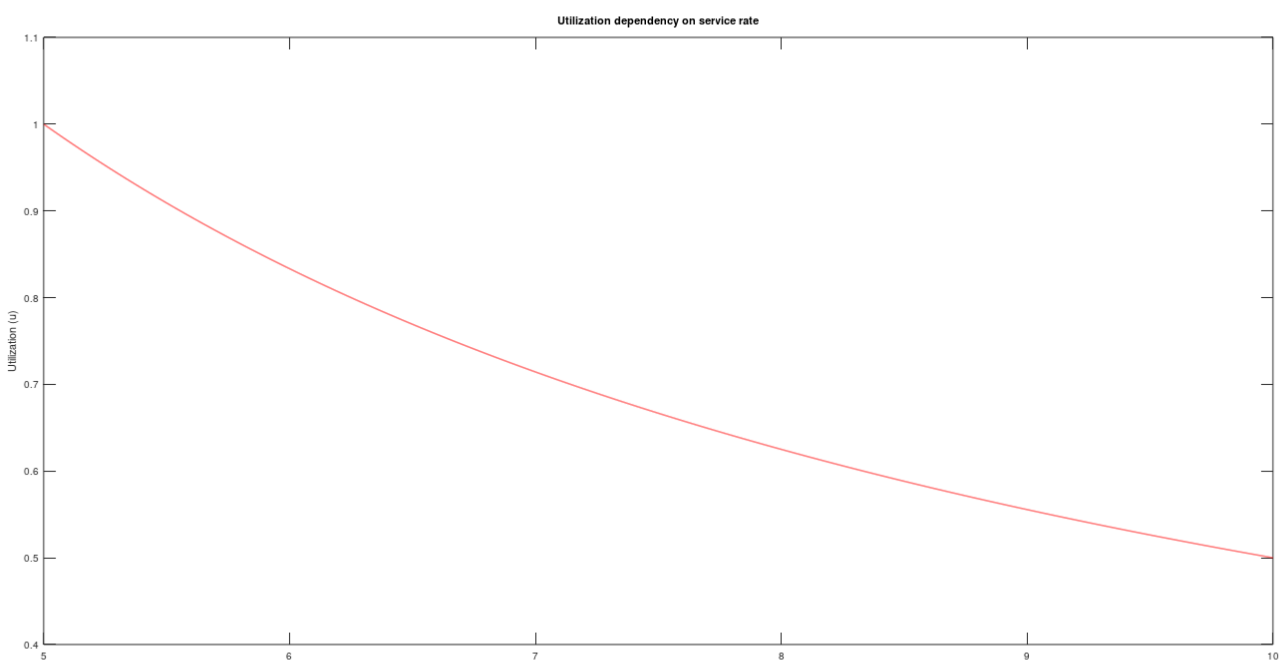
- **Βαθμός χρησιμοποίησης (utilization) ως προς το ρυθμό εξυπηρέτησης**.

Όπως γνωρίζουμε, για το σύστημα $M/M/1$ (άπειρη ουρά) έχουμε $P[blocking] = 0$, συνεπώς για τον βαθμό χρησιμοποίησης u ισχύει:

$$u = \frac{\gamma}{\mu} = \frac{\lambda}{\mu} = \rho \quad (\acute{\varepsilon}\nu\tau\alpha\sigma\eta\ \varphi o\rho\tau\acute{\iota}o\upsilon)$$

Αναμένουμε λοιπόν οι τιμή του u να μειώνεται καθώς αυξάνεται το μ από 5.001 σε 10, με σταθερό $\lambda = 5$, από αρχική τιμή περίπου 1 $\left(u = \frac{\lambda}{\mu_{max}} = \frac{5}{5.001} \approx 1\right)$, μέχρι την τιμή 0.5 $\left(u = \frac{\lambda}{\mu_{max}} = \frac{5}{10} = 0.5\right)$.

Οι ζητούμενη γραφική είναι:

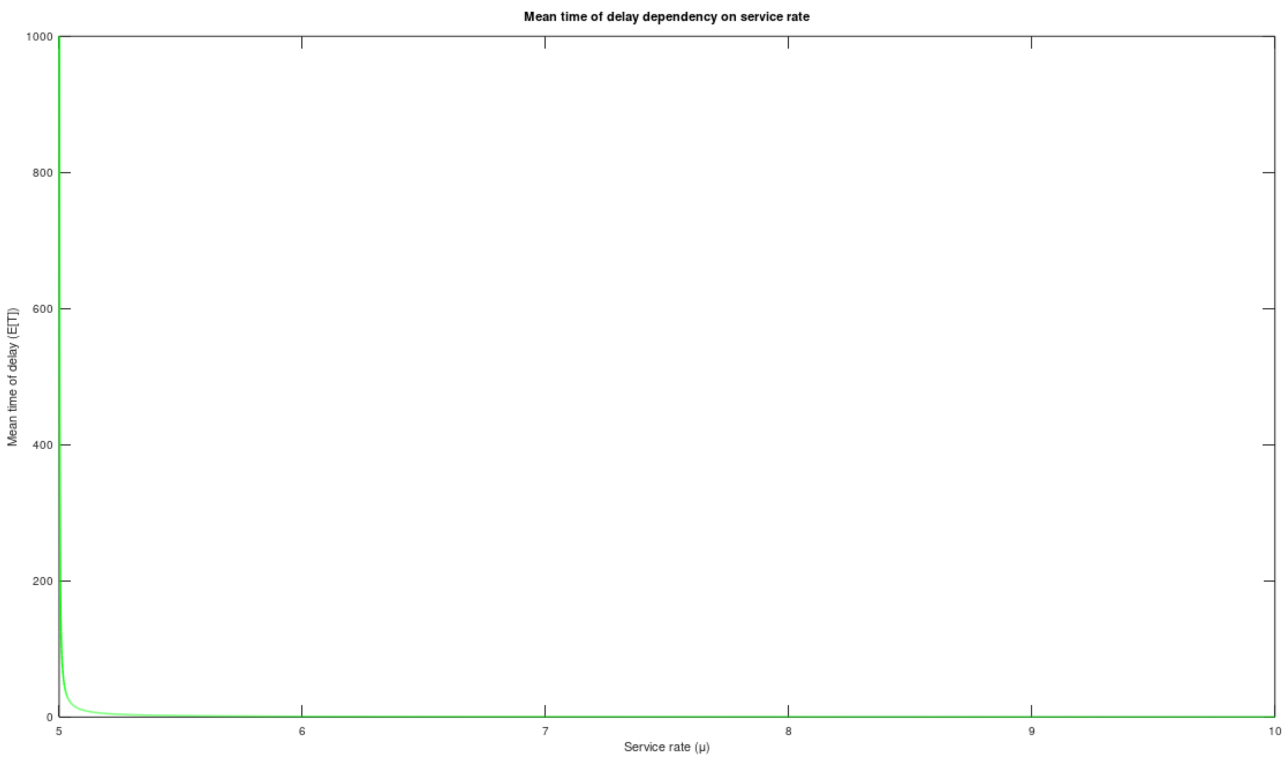


- **Μέσος χρόνος καθυστέρησης του συστήματος $E(T)$ ως προς το ρυθμό εξυπηρέτησης.**

Υπολογίσαμε στην προηγούμενη άσκηση τον μέσο χρόνο καθυστέρησης του συστήματος $E(T)$ ως προς το ρυθμό εξυπηρέτησης:

$$E(T) = \frac{1}{\mu - \lambda}$$

Στην περίπτωσή μας έχουμε $\lambda = 5$ και $\mu \in (5, 10]$, (άρα $\mu > \lambda$) και επομένως περιμένουμε να μειώνεται ο μέσος χρόνος καθυστέρησης καθώς αυξάνεται το μ. Πράγματι:

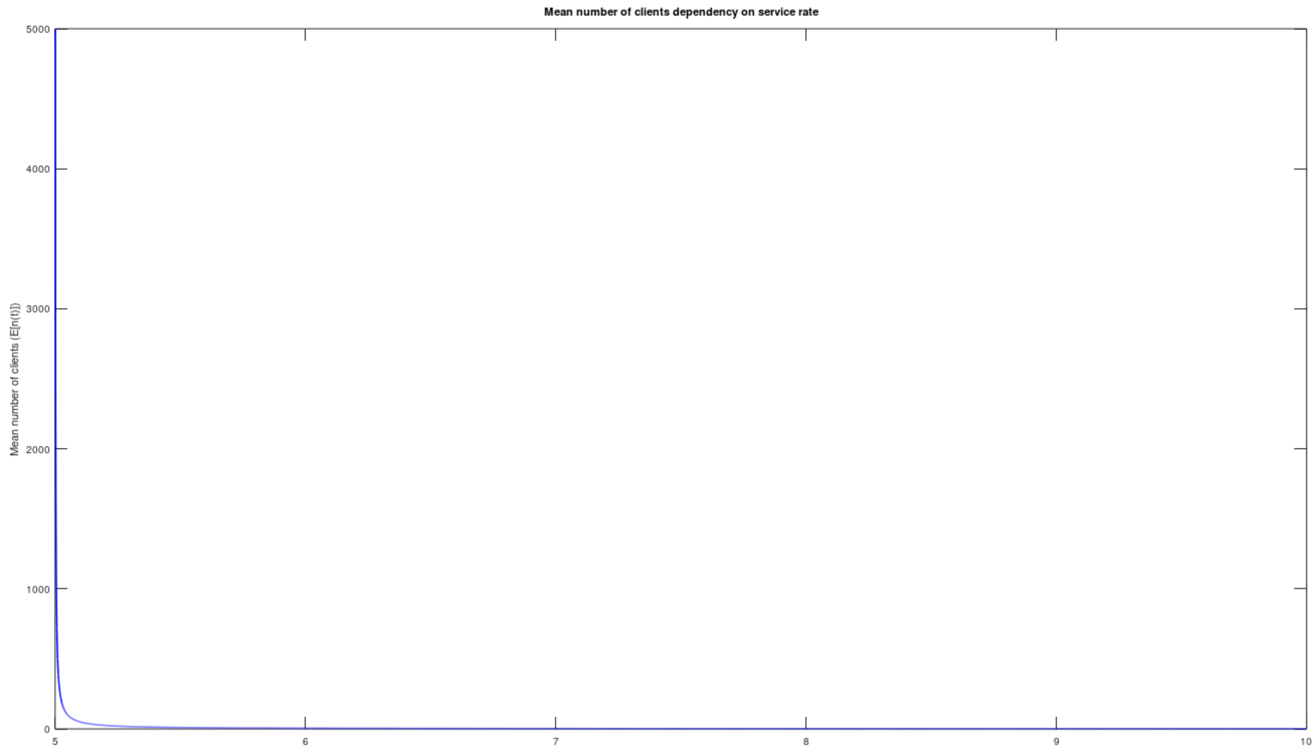


- **Μέσος αριθμός πελατών στο σύστημα ως προς το ρυθμό εξυπηρέτησης.**

Εδώ ισχύει όμοια με πριν:

$$E[n(t)] = \frac{\rho}{1-\rho} = \frac{\left(\frac{\lambda}{\mu}\right)}{1-\left(\frac{\lambda}{\mu}\right)} = \frac{\lambda}{\mu-\lambda},$$

Οπότε προκύπτει παρόμοια γραφική με προηγουμένως:

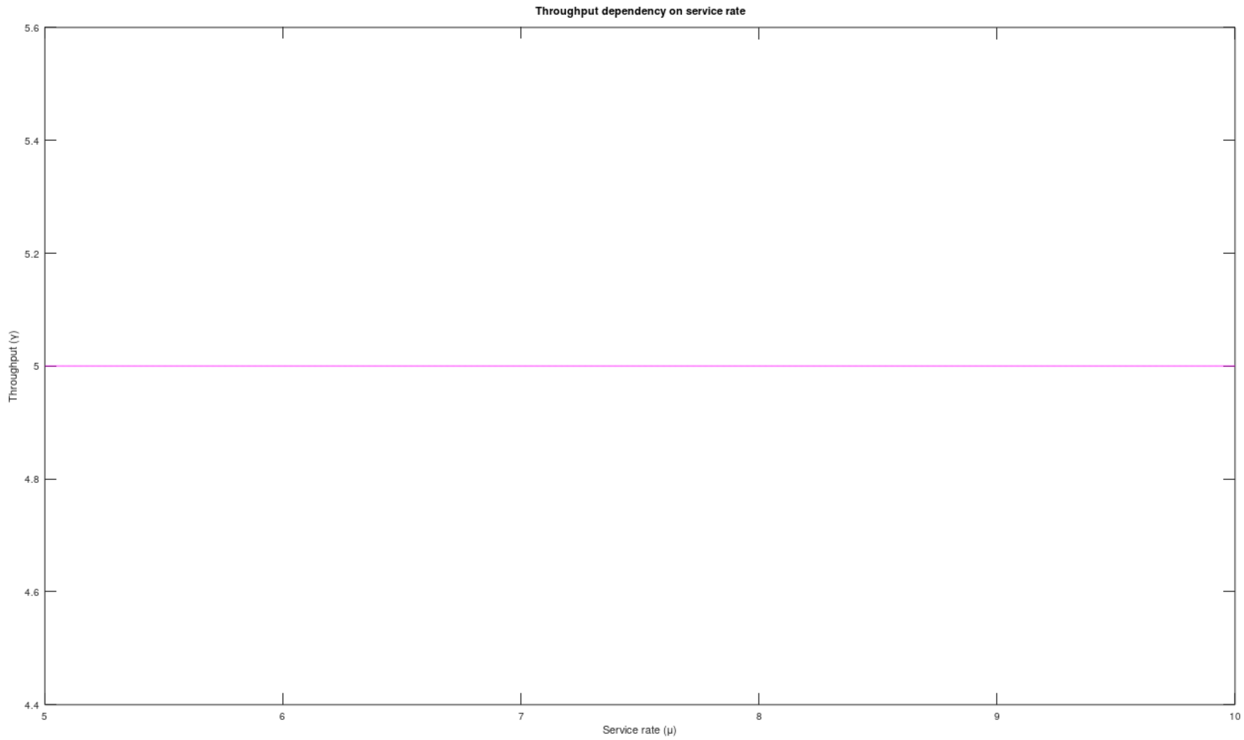


- **Ρυθμαπόδοση (throughput) πελατών ως προς το ρυθμό εξυπηρέτησης.**

Γνωρίζουμε από την θεωρία ότι για το throughput γ ισχύει:

$$\gamma = \lambda(1 - P[blocking]) \xRightarrow{M/M/1 \longrightarrow P[blocking]=0} \boldsymbol{\gamma = \lambda}$$

Παρατηρούμε ότι η ρυθμαπόδοση γ είναι ανεξάρτητη του ρυθμού εξυπηρέτησης μ. Επομένως αναμένουμε η γραφική να έχει μορφή οριζόντιας ευθείας παράλληλης στον άξονα x, η οποία θα συμπίπτει στην τιμή 5 του άξονα y:



Ο κώδικας που χρησιμοποιήσαμε για την παραγωγή των γραφικών παραστάσεων είναι ο ακόλουθος:

```
clc;
clear all;
close all;

# Ανάλυση ουράς M/M/1 με Octave
# Task b

mu_values = 5.001 : 0.001 : 10;
lambda_rate = 5;

lambda_values = lambda_rate * ones(1, numel(mu_values));

[u, et, ent, g] = qsmm1(lambda_values, mu_values);

# Βαθμός χρησιμοποίησης (utilization) ως προς το ρυθμό εξυπηρέτησης.

figure('Name', 'Utilization dependency on service rate');
plot(mu_values, u, "r", 'LineWidth', 1.2);
title('Utilization dependency on service rate');
xlabel('Service rate (\mu)');
ylabel('Utilization (u)');

```

```
# Μέσος χρόνος καθυστέρησης του συστήματος E(T) ως προς το ρυθμό εξυπηρέτησης.

figure('Name', 'Mean time of delay dependency on service rate');
plot(mu_values, et, "g", 'LineWidth', 1.2);
title('Mean time of delay dependency on service rate');
xlabel('Service rate (\mu)');
ylabel('Mean time of delay (E[T])');

# Μέσος αριθμός πελατών στο σύστημα ως προς το ρυθμό εξυπηρέτησης.

figure('Name', 'Mean number of clients dependency on service rate');
plot(mu_values, ent, "b", 'LineWidth', 1.2);
title('Mean number of clients dependency on service rate');
xlabel('Service rate (\mu)');
ylabel('Mean number of clients (E[n(t)])');

# Ρυθμαπόδοση (throughput) πελατών ως προς το ρυθμό εξυπηρέτησης

figure('Name', 'Throughput dependency on service rate');
plot(mu_values, g, "m", 'LineWidth', 1.2);
title('Throughput dependency on service rate');
xlabel('Service rate (\mu)');
ylabel('Throughput (\gamma)');
```

(γ)

Κοιτώντας το διάγραμμα του μέσου χρόνου καθυστέρησης $E(T)$ ως προς το ρυθμό εξυπηρέτησης, καθώς και το αντίστοιχο διάγραμμα για το utilization, συμπεραίνουμε ότι προκειμένου να μειώσουμε το χρόνο καθυστέρησης, διατηρώντας παράλληλα το utilization όσο πιο υψηλό γίνεται, μια καλή επιλογή είναι να θέσουμε $\mu = 6\ πελάτες/sec$. Περαιτέρω αύξηση του μ θα οδηγούσε σε σημαντική μείωση του utilization και αμελητέα μείωση του χρόνου καθυστέρησης, γεγονός που δεν μας συμφέρει εν προκειμένω.

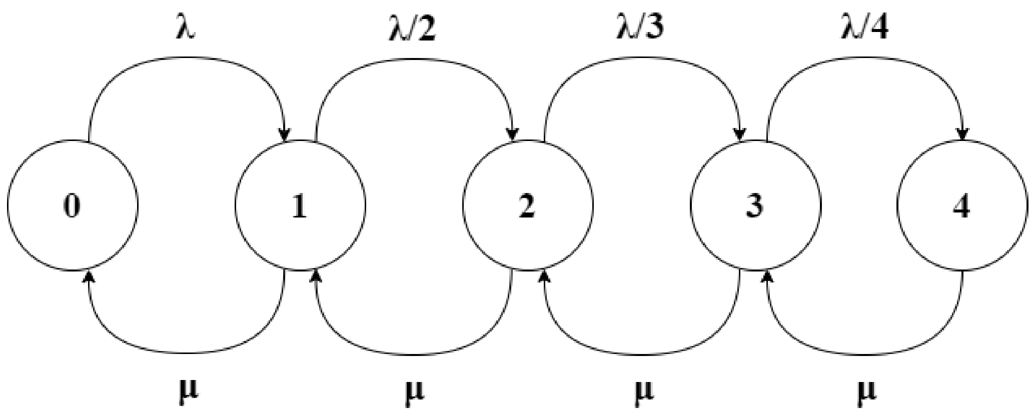
(δ)

Στο ερώτημα (β) παρατηρήσαμε ότι για ουρές Markov M/M/1 η ρυθμαπόδοση πελατών είναι ανεξάρτητη του ρυθμού εξυπηρέτησης.

<u>Διαδικασία γεννήσεων θανάτων (birth-death process): εφαρμογή σε σύστημα M/M/1/K</u>

(α)

Σχεδιάζουμε αρχικά το διάγραμμα ρυθμού μεταβάσεων για το σύστημα αναμονής $M/M/1/4$ με $\lambda_i = \frac{\lambda}{(i+1)}$ για $i = 0, 1, 2, 3$ και $\mu_i = \mu$ για $i = 1, 2, 3, 4$ όπως φαίνεται παρακάτω:



Στη συνέχεια, αξιοποιώντας τις Detailed Balance Equations, υπολογίζουμε τις πιθανότητες $P_0,\ P_1,\ P_2,\ P_3,\ P_4$ συναρτήσει των λ και μ:

$$\lambda_{k-1} P_{k-1} = \mu_k P_k \Longrightarrow P_k = \frac{\lambda_{k-1}}{\mu_k} \cdot P_{k-1}$$

Άρα για $k = 1, 2, 3, 4$:

- $\boldsymbol{k = 1}$: $\quad P_1 = \left(\frac{\lambda}{\mu}\right) P_0$
- $\boldsymbol{k = 2}$: $\quad P_2 = \left(\frac{\lambda}{2\mu}\right) P_1 \Longrightarrow P_2 = \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 P_0$
- $\boldsymbol{k = 3}$: $\quad P_3 = \left(\frac{\lambda}{3\mu}\right) P_2 \Longrightarrow P_3 = \frac{1}{6}\left(\frac{\lambda}{\mu}\right)^3 P_0$
- $\boldsymbol{k = 4}$: $\quad P_4 = \left(\frac{\lambda}{4\mu}\right) P_3 \Longrightarrow P_4 = \frac{1}{24}\left(\frac{\lambda}{\mu}\right)^4 P_0$

Αρκεί τώρα να βρούμε το P_0, το οποίο μπορούμε να επιτύχουμε κανονικοποιώντας τις παραπάνω εργοδικές πιθανότητες ως εξής:

$$P_0 + P_1 + P_2 + P_3 + P_4 = 1 \Longrightarrow$$

$$P_0 + \left(\frac{\lambda}{\mu}\right) P_0 + \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 P_0 + \frac{1}{6}\left(\frac{\lambda}{\mu}\right)^3 P_0 + \frac{1}{24}\left(\frac{\lambda}{\mu}\right)^4 P_0 = 1 \Longrightarrow$$

$$P_0\left(1 + \left(\frac{\lambda}{\mu}\right) + \frac{1}{2}\left(\frac{\lambda}{\mu}\right)^2 + \frac{1}{6}\left(\frac{\lambda}{\mu}\right)^3 + \frac{1}{24}\left(\frac{\lambda}{\mu}\right)^4\right) = 1 \xRightarrow{\boldsymbol{\frac{\lambda}{\mu}=\frac{1}{2}}}$$

$$P_0\left(1 + \left(\frac{1}{2}\right) + \frac{1}{2}\left(\frac{1}{2}\right)^2 + \frac{1}{6}\left(\frac{1}{2}\right)^3 + \frac{1}{24}\left(\frac{1}{2}\right)^4\right) = 1 \Longrightarrow$$

$$P_0\left(1 + \frac{1}{2} + \frac{1}{8} + \frac{1}{48} + \frac{1}{384}\right) = 1 \Longrightarrow$$

$$1.6484375 \cdot P_0 = 1 \Longrightarrow$$

$$\boldsymbol{P_0} \approx \mathbf{0.6066} \text{ ή } \boldsymbol{P_0} \approx \mathbf{60.66}\%$$

Οπότε προκύπτουν:

- $\boldsymbol{P_1} = \frac{P_0}{2} \approx 30.33\%$
- $\boldsymbol{P_2} = \frac{P_0}{8} \approx 7.58\%$
- $\boldsymbol{P_3} = \frac{P_0}{48} \approx 1.26\%$
- $\boldsymbol{P_4} = \frac{P_0}{384} \approx 0.16\%$

Τέλος, όσον αφορά την πιθανότητα απώλειας, θα ισούται με την πιθανότητα $\boldsymbol{P_4} = \mathbf{0.16}\%$, ($P[blocking]$), καθώς όταν το σύστημα βρίσκεται στην κατάσταση όπου υπάρχουν τέσσερις πελάτες στην ουρά, ο πέμπτος θα απορριφθεί μιας και η ουρά χωράει μέχρι 4 πελάτες.

(β)

i) Η μήτρα του ρυθμού μεταβάσεων είναι η ακόλουθη:

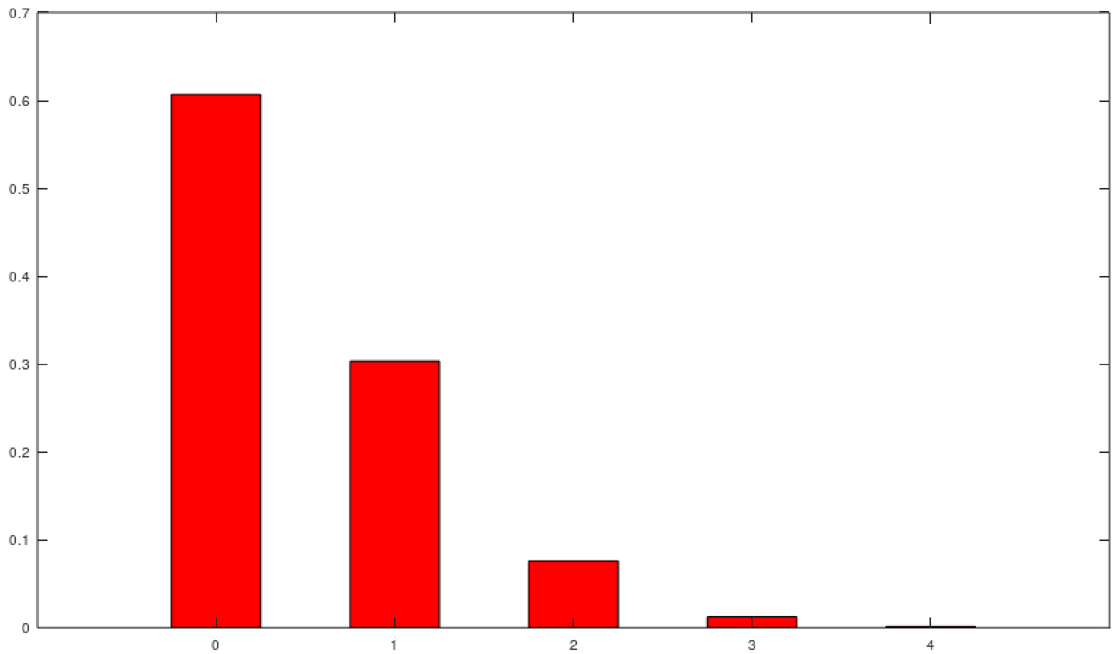
```
The Transition Matrix of our M/M/1/4 system is:
transition_matrix =

   -5.0000    5.0000         0         0         0
   10.0000  -12.5000    2.5000         0         0
         0   10.0000  -11.6667    1.6667         0
         0         0   10.0000  -11.2500    1.2500
         0         0         0   10.0000  -10.0000
```

ii) Όπως βλέπουμε ακολούθως, οι εργοδικές πιθανότητες που υπολογίσαμε μέσω Octave ταυτίζονται με αυτές που υπολογίσαμε θεωρητικά προηγουμένως:

```
The ergodic probabilities of our system are:
probabilities =

   60.6635   30.3318    7.5829    1.2638    0.1580
```



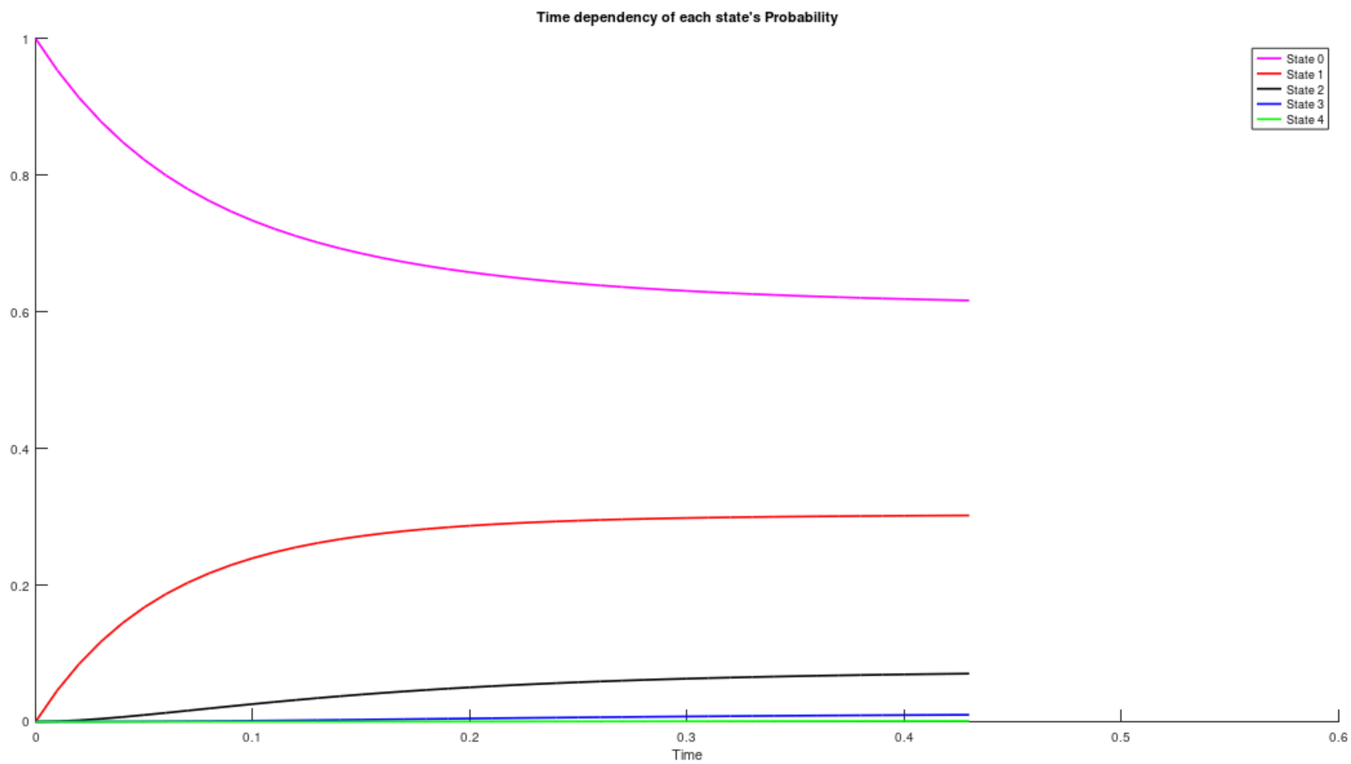
iii) Σε κατάσταση ισορροπίας, ο μέσος αριθμός πελατών που βρίσκονται στο σύστημα είναι:

```
The Mean number of clients in our system in equilibrium is: E[n(t)] = 1.49921 clients
```

iv) Η πιθανότητα απόρριψης πελάτη, όταν το σύστημα βρίσκεται σε ισορροπία προκύπτει ίση με αυτή που υπολογίσαμε προηγουμένως:

```
The blocking possibility is: P[blocking] = 0.157978 %
```

v) Για $\lambda = 5\ \pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma/sec$ και $\mu = 10\ \pi\varepsilon\lambda\acute{\alpha}\tau\varepsilon\varsigma/sec$, τα διαγράμματα πιθανοτήτων των καταστάσεων του συστήματος συναρτήσει του χρόνου είναι:



Ο κώδικας που υπολογίζει τα παραπάνω είναι:

```
% system M/M/1/4
% when there are 3 clients in the system, the capability of the server doubles.

clc;
clear all;
close all;

# Διαδικασία γεννήσεων θανάτων (birth-death process): εφαρμογή σε σύστημα M/M/1/K
# Task b

lambda = 5;
mu = 10;
states = [0, 1, 2, 3, 4]; % system with capacity 4 states
% the initial state of the system. The system is initially empty.
initial_state = [1, 0, 0, 0, 0];

% define the birth and death rates between the states of the system.
births_B = [lambda, lambda/2, lambda/3, lambda/4];
deaths_D = [mu, mu, mu, mu];

# (i)

% get the transition matrix of the birth-death process
transition_matrix = ctmcbd(births_B, deaths_D);
printf("The Transition Matrix of our M/M/1/4 system is: \n");
transition_matrix

# (ii)

% get the ergodic probabilities of the system
P = ctmc(transition_matrix);
printf("The ergodic probabilities of our system are: \n");
probabilities = 100 * P # values in %

% plot the ergodic probabilities (bar for bar chart)
figure(1);
bar(states, P, "r", 0.5);

# (iii)

E_nt = sum((1+states).*P);
printf("The Mean number of clients in our system in equilibrium is: E[n(t)] = %d clients\n", E_nt);

# (iv)

p_blocking = 100 * P(columns(P));
printf("The blocking possibility is: P[blocking] = %f %% \n", p_blocking)

# (v)



% transient probability of state 0 until convergence to ergodic probability. Convergence takes place P0 and P differ by 0.01
index = 0;
for T = 0 : 0.01 : 50
  index = index + 1;
  P0 = ctmc(transition_matrix, T, initial_state);
  Prob0(index) = P0(1);       # state 0
  Prob1(index) = P0(2);       # state 1
  Prob2(index) = P0(3);       # state 2
  Prob3(index) = P0(4);       # state 3
  Prob4(index) = P0(5);       # state 4
  if P0 - P < 0.01
    break;
  endif
endfor

T = 0 : 0.01 : T;
figure(2);
colors = "mrkbg";
hold on
title("Time dependency of each state's Probability")
plot(T, Prob0, colors(1), "linewidth", 1.3);
plot(T, Prob1, colors(2), "linewidth", 1.3);
plot(T, Prob2, colors(3), "linewidth", 1.3);
plot(T, Prob3, colors(4), "linewidth", 1.3);
plot(T, Prob4, colors(5), "linewidth", 1.3);
xlabel("Time");
ylabel("Probability");
legend("State 0", "State 1", "State 2", "State 3", "State 4");
hold off
```