<u>ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ</u>

<u>Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών</u>

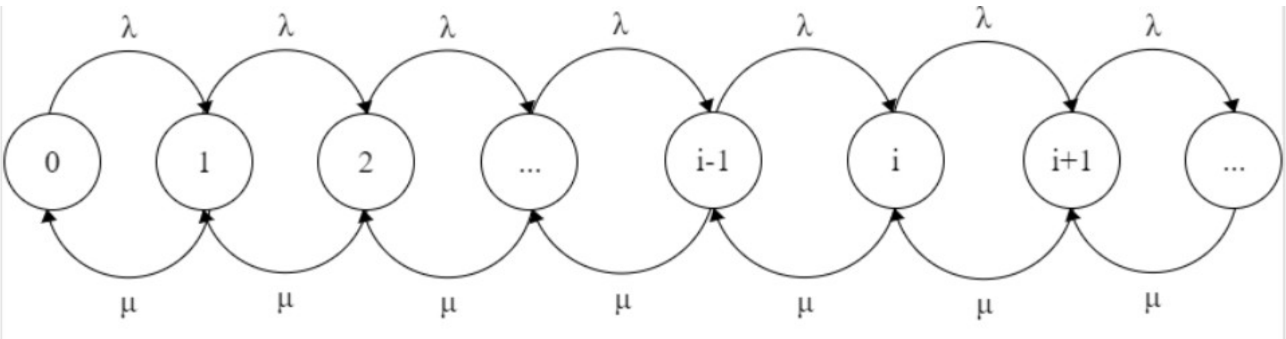
<u>Ονοματεπώνυμο</u>: Γκούμε Λαουρεντιάν
<u>Α.Μ</u>.: 031 18 014
<u>Έτος/Εξάμηνο</u>: 3ο/6ο

<u>ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ (Queuing Systems)</u>

<u>2η Σειρά Ασκήσεων</u>

<u>Θεωρητική μελέτη της ουράς M/M/1</u>

(α) Προκειμένου μια ουρά M/M/1 να είναι **εργοδική**, θα πρέπει η ένταση φορτίου ρ = (λ/μ) να είναι μικρότερη από 1 Erlang, δηλαδή **$\rho < 1 \Rightarrow \lambda < \mu$**, όπου λ ο ρυθμός άφιξης πελατών και μ ο ρυθμός εξυπηρέτησης. Για το διάγραμμα ρυθμού μεταβάσεων που παρουσιάζεται παρακάτω, οι ρυθμοί άφιξης/εξυπηρέτησης έχουν θεωρηθεί ανεξάρτητοι της παρούσας κατάστασης του συστήματος και ίσοι με λ/μ αντίστοιχα.



Εικόνα 1: Διάγραμμα ρυθμού μεταβάσεων ουράς M/M/1

Με χρήση των Εξισώσεων Λεπτομερούς Ισορροπίας (Detailed Balance Equations), υπολογίζουμε τις ζητούμενες εργοδικές πιθανότητες των καταστάσεων του συστήματος:

- $\lambda P_0 = \mu P_1 \Rightarrow P_1 = (\lambda/\mu)P_0$
- $\lambda P_1 = \mu P_2 \Rightarrow P_2 = (\lambda/\mu)P_1 \Rightarrow P_2 = (\lambda/\mu)^2 P_0$
- Επαγωγικά: $\lambda P_{i-1} = \mu P_i \Rightarrow P_i = (\lambda/\mu)P_{i-1} \Rightarrow \mathbf{P_i = (\lambda/\mu)^i P_0}$

Κανονικοποιώντας τις εργοδικές πιθανότητες, έχουμε: $P_0 + P_1 + P_2 + \ldots = 1 \Rightarrow$

$P_0 (1 + \rho + \rho^2 + \ldots) = 1 \Rightarrow P_0 \frac{1}{1-\rho} = 1 \Rightarrow$ **$P_0 = (1 - \rho)$**.
Για το παραπάνω αποτέλεσμα χρησιμοποιήθηκε η σχέση ρ = (λ/μ) και η γεωμετρική σειρά. Συνεπώς, παίρνουμε τελικά: **$P_i = \rho^i(1 - \rho)$**, για κάθε ακέραιο $i \geq 0$.

(β) Για τον μέσο χρόνο καθυστέρησης E[T], έχουμε από τον τύπο του Little ότι:

$E[T] = \frac{E[n(t)]}{\gamma}$, ωστόσο για το σύστημα M/M/1, έχουμε άπειρη ουρά οπότε P[blocking] = 0, ή ισοδύναμα γ = λ, άρα **E[T] =** $\frac{(\frac{\rho}{1-\rho})}{\lambda} = \frac{\rho}{\lambda(1-\rho)}$.

(γ) Από τις εργοδικές πιθανότητες του (α) ερωτήματος, για i = 57, έχουμε:

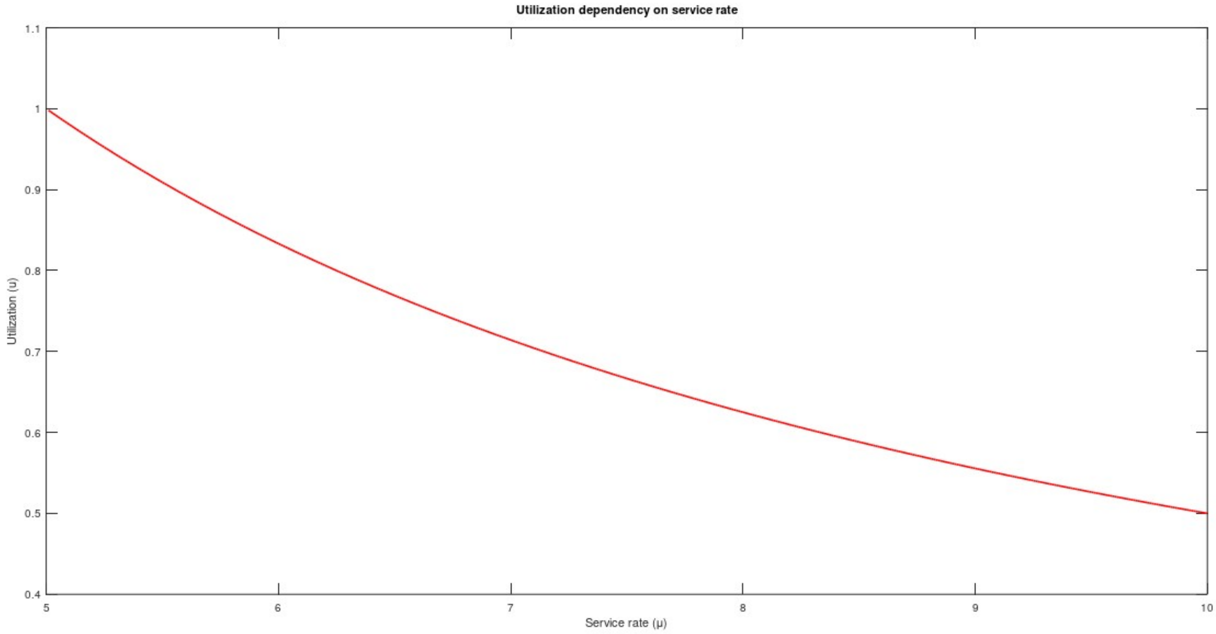
$P_{57} = \rho^{57}(1-\rho) > 0$. Επομένως, **υπάρχει πιθανότητα κάποια χρονική στιγμή το σύστημα να βρεθεί με 57 πελάτες**. Παρατηρούμε πως όσο μικρότερο γίνεται το ρ, τόσο μικρότερη γίνεται η πιθανότητα ($\frac{dP(\rho)}{d\rho} = \rho^{56}(57 - 58\rho)$ => P(ρ) γνησίως αύξουσα όσο ρ < (57/58)). Αυτό είναι αναμενόμενο, καθώς μείωση του ρ σημαίνει είτε μείωση του ρυθμού αφίξεων λ με σταθερό μ, είτε αύξηση του ρυθμού εξυπηρέτησης μ με σταθερό λ, είτε ταυτόχρονη μείωση του ρυθμού αφίξεων και αύξηση του ρυθμού εξυπηρετήσεων.

Ανάλυση ουράς M/M/1 με Octave

(α) Όπως είδαμε, εργοδικό σύστημα σημαίνει ρ < 1 => λ < μ, άρα το μ πρέπει να πάρει τιμές στο διάστημα **(5, 10] πελάτες/sec**.

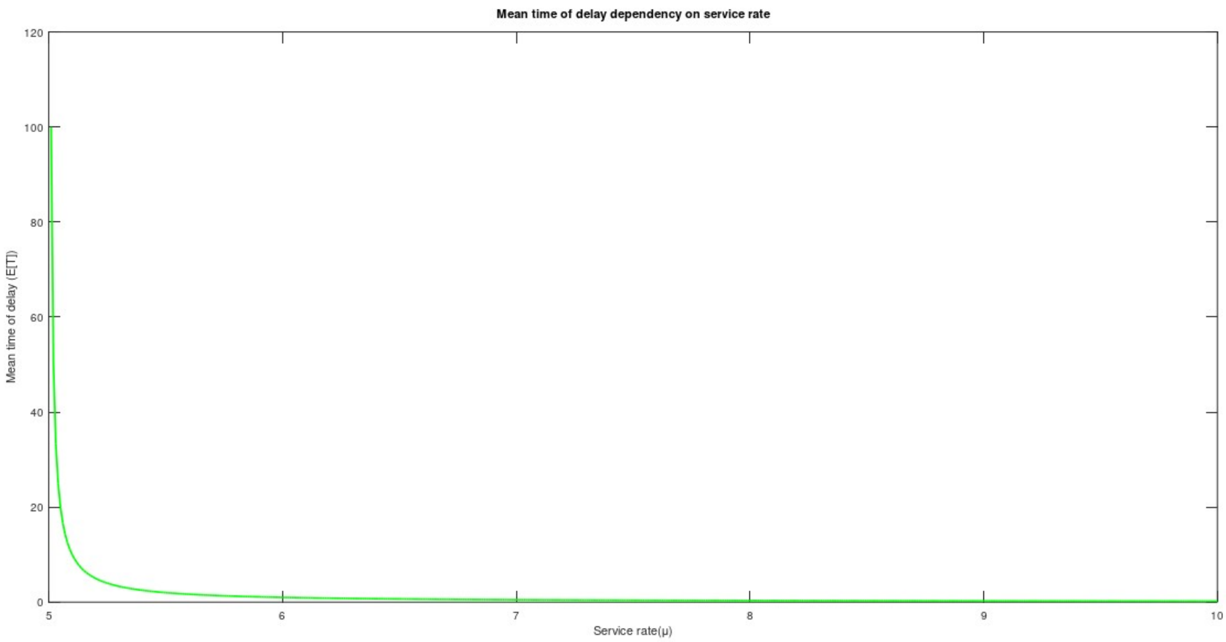
(β)

- **Βαθμός χρησιμοποίησης (u) ως προς τον ρυθμό εξυπηρέτησης (μ):** Είναι **u = (γ/μ)**, ωστόσο για το σύστημα M/M/1, έχω άπειρη ουρά, επομένως P[blocking] = 0, οπότε και γ = λ, άρα ο βαθμός χρησιμοποίησης είναι ίσος με την ένταση του φορτίου ρ = (λ/μ). Όπως βλέπουμε παρακάτω, η γραφική ξεκινάει από το 1 για λ ≈ μ και συγκλίνει στο 0.5 για μ = 10, επαληθεύοντας όσα είπαμε.



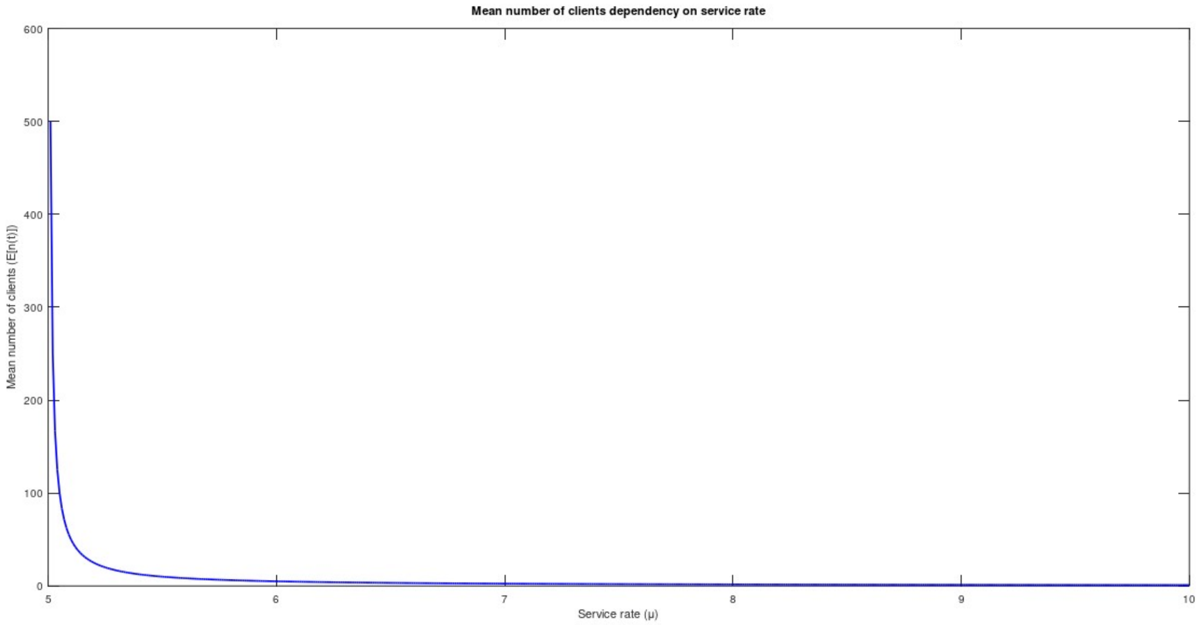
Εικόνα 2: Βαθμός χρησιμοποίησης συναρτήσει του ρυθμού εξυπηρέτησης

- **Μέσος χρόνος καθυστέρησης του συστήματος (E[T]) ως προς τον ρυθμό εξυπηρέτησης (μ):** Είναι E[T] = $\frac{1}{\mu(1-\rho)} = \frac{1}{\mu(1-(\lambda/\mu))} = \frac{1}{\mu-\lambda}$ και αφού μ > λ, αναμένουμε όταν το μ τείνει στο 5 μεγάλες τιμές του E[T], ενώ όταν τείνει στο μ = 10 = 2λ αναμένεται να γίνει ίση με (1/λ) = (1/5) seconds. Επαληθεύουμε γραφικά:



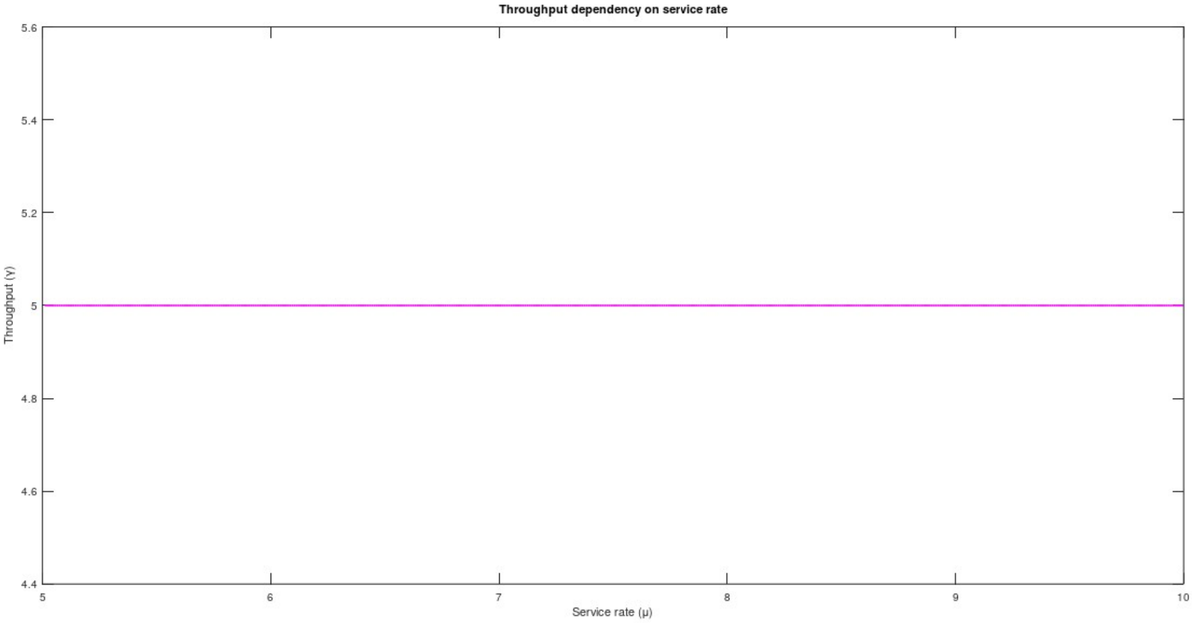
Εικόνα 3: Μέσος χρόνος καθυστέρησης του συστήματος συναρτήσει του ρυθμού εξυπηρέτησης

- **Μέσος αριθμός πελατών στο σύστημα (E[n(t)]) ως προς τον ρυθμό εξυπηρέτησης (μ):** Είναι E[n(t)] = $\frac{\rho}{1-\rho}=\frac{(\lambda/\mu)}{(1-(\lambda/\mu))}=\frac{\lambda}{\mu-\lambda}$. Αναμένουμε ίδια μορφή με του μέσου χρόνου καθυστέρησης συναρτήσει του (μ) που είδαμε προηγουμένως:



Εικόνα 4: Μέσος αριθμός πελατών συναρτήσει του ρυθμού εξυπηρέτησης

- **Ρυθμαπόδοση (γ) πελατών ως προς τον ρυθμό εξυπηρέτησης (μ):** Είναι γ = λ(1-P[blocking]), αλλά για το σύστημα M/M/1 είπαμε πως P[blocking] = 0, άρα γ = λ, ανεξάρτητο του μ. Επαληθεύουμε μέσω της προσομοίωσης:



Εικόνα 5: Ρυθμαπόδοση συναρτήσει του ρυθμού εξυπηρέτησης

Ο κώδικας που υλοποιεί τις ανωτέρω γραφικές παραστάσεις:

```
#Task2_2
clc;
clear all;
close all;

mu = 5.01 : 0.01 : 10;
lambda_rate = 5;

for i = 1 : columns(mu)
  lambda(i) = lambda_rate;          #if args of qsmm1 = 2, vectors are needed
endfor

color = "rgbm";
figure(1);
[U, R, Q, X] = qsmm1(lambda, mu);

figure(1);
plot(mu, U, color(1), "linewidth", 1.2);
title("Utilization dependency on service rate");
xlabel("Service rate (μ)");
ylabel("Utilization (u)");

figure(2);
plot(mu, R, color(2), "linewidth", 1.2);
title("Mean time of delay dependency on service rate");
xlabel("Service rate(μ)");
ylabel("Mean time of delay (E[T])");
```

```
figure(3);
plot(mu, Q, color(3), "linewidth", 1.2);
title("Mean number of clients dependency on service rate");
xlabel("Service rate (μ)");
ylabel("Mean number of clients (E[n(t)])");

figure(4);
plot(mu, X, color(4), "linewidth", 1.2);
title("Throughput dependency on service rate");
xlabel("Service rate (μ)");
ylabel("Throughput (γ)");
```

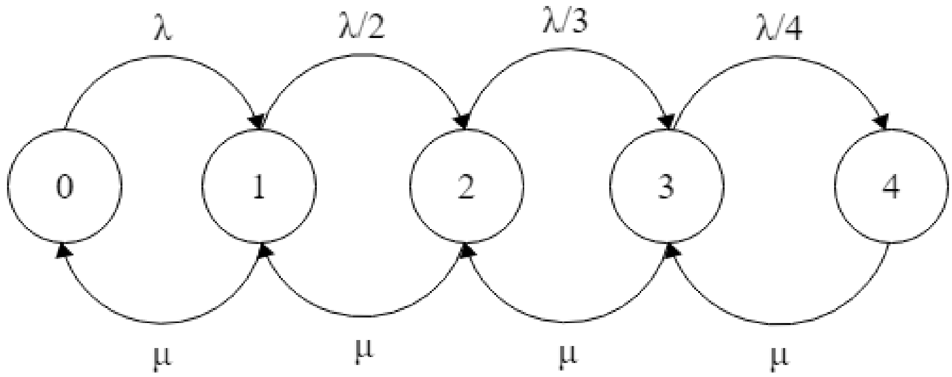
Εικόνα 6: Κώδικας σε Octave

(γ) Θα επιλέξουμε τον μέσο ρυθμό εξυπηρέτησης βασιζόμενοι στα παραπάνω γραφήματα. Για αρχή, η ρυθμαπόδοση είναι ανεξάρτητη του μ, οπότε δε μας απασχολεί εδώ. Επιθυμούμε να ελαχιστοποιήσουμε τον μέσο χρόνο καθυστέρησης των πελατών μας, διατηρώντας μια όσο το δυνατόν μεγαλύτερη τιμή για το utilization μας (μέχρι το 1 προφανώς). Εύκολα συμπεραίνουμε πως μια καλή επιλογή που πληροί τα κριτήρια αυτά, είναι να έχουμε ρυθμό εξυπηρέτησης πελατών ίσο με **μ = 6 πελάτες/sec**, καθώς από εκεί και πέρα όσο αυξάνεται προκαλεί πρακτικά αμελητέα μείωση στην καθυστέρηση, αλλά σημαντική μείωση του utilization.

(δ) Όπως είδαμε, η **ρυθμαπόδοση είναι ανεξάρτητη του ρυθμού εξυπηρέτησης για ουρές M/M/1**.

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

(α) Το διάγραμμα ρυθμού μεταβάσεων για το σύστημα αναμονής M/M/1/4 με $\lambda_i = \lambda/(i+1)$ και $\mu_i = \mu$ για i = 0, 1, 2, 3 δίδεται παρακάτω:



Εικόνα 7: Διάγραμμα ρυθμού μεταβάσεων ουράς M/M/1/4

Κάνοντας χρήση των Detailed Balance Equations, παίρνουμε:

- $\lambda P_0 = \mu P_1 \Rightarrow P_1 = (\lambda/\mu)P_0$
- $(\lambda/2)P_1 = \mu P_2 \Rightarrow P_2 = \frac{1}{2}(\lambda/\mu)P_1 \Rightarrow P_2 = \frac{1}{2!}(\lambda/\mu)^2 P_0$
- $(\lambda/3)P_2 = \mu P_3 \Rightarrow P_3 = \frac{1}{3}(\lambda/\mu)P_2 \Rightarrow P_3 = \frac{1}{3!}(\lambda/\mu)^3 P_0$
- $(\lambda/4)P_3 = \mu P_4 \Rightarrow P_4 = \frac{1}{4}(\lambda/\mu)P_3 \Rightarrow P_4 = \frac{1}{4!}(\lambda/\mu)^4 P_0$

Για τον υπολογισμό του P_0, κανονικοποιούμε ως εξής: $P_0 + P_1 + P_2 + P_3 + P_4 = 1 \Rightarrow P_0\,[1 + (\lambda/\mu) + \frac{1}{2!}(\lambda/\mu)^2 + \frac{1}{3!}(\lambda/\mu)^3 + \frac{1}{4!}(\lambda/\mu)^4] = 1 \Rightarrow P_0\,[1 + \frac{1}{2} + \frac{1}{2!}\frac{1}{4} + \frac{1}{3!}\frac{1}{8} + \frac{1}{4!}\frac{1}{16}] = 1 \Rightarrow P_0\,(211/128) = 1 \Rightarrow P_0 = (128/211) \Rightarrow P_0 \approx$

0.6066 ή $\mathbf{P_0 \approx 60.66\%}$. Με τη σειρά, βρίσκουμε τις υπόλοιπες πιθανότητες: $\mathbf{P_1 \approx 30.33\%}$, $\mathbf{P_2 \approx 7.58\%}$, $\mathbf{P_3 \approx 1.26\%}$, $\mathbf{P_4 \approx 0.16\%}$. Η **πιθανότητα απώλειας** P[blocking], είναι η εργοδική πιθανότητα $\mathbf{P_4 \approx 0.16\%}$, καθώς αν το σύστημα βρίσκεται στην κατάσταση 4 ο επόμενος πελάτης θα απορριφθεί, διότι σε διαφορετική περίπτωση θα είχαμε 5 πελάτες στο σύστημα, πράγμα που υπερβαίνει τη χωρητικότητα του.

(β)

i) Η **μήτρα του ρυθμού μεταβάσεων** είναι η εξής:

```
Transition matrix of our M/M/1/4 system:
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
```

Εικόνα 8: Μήτρα ρυθμού μεταβάσεων

ii) Οι **εργοδικές πιθανότητες** που υπολογίζονται μέσω της προσομοίωσης, επαληθεύουν τις θεωρητικά υπολογιζόμενες όπως και φαίνεται παρακάτω:

```
System's ergodic probabilities:
P_percentage =

   60.6635   30.3318    7.5829    1.2638    0.1580
```

Εικόνα 9: Εργοδικές πιθανότητες

iii) Ο **μέσος αριθμός πελατών** που βρίσκεται στο σύστημα, όταν εκείνο βρίσκεται σε κατάσταση ισορροπίας υπολογίζεται μέσω του ορισμού, και προκύπτει ίσος με:

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

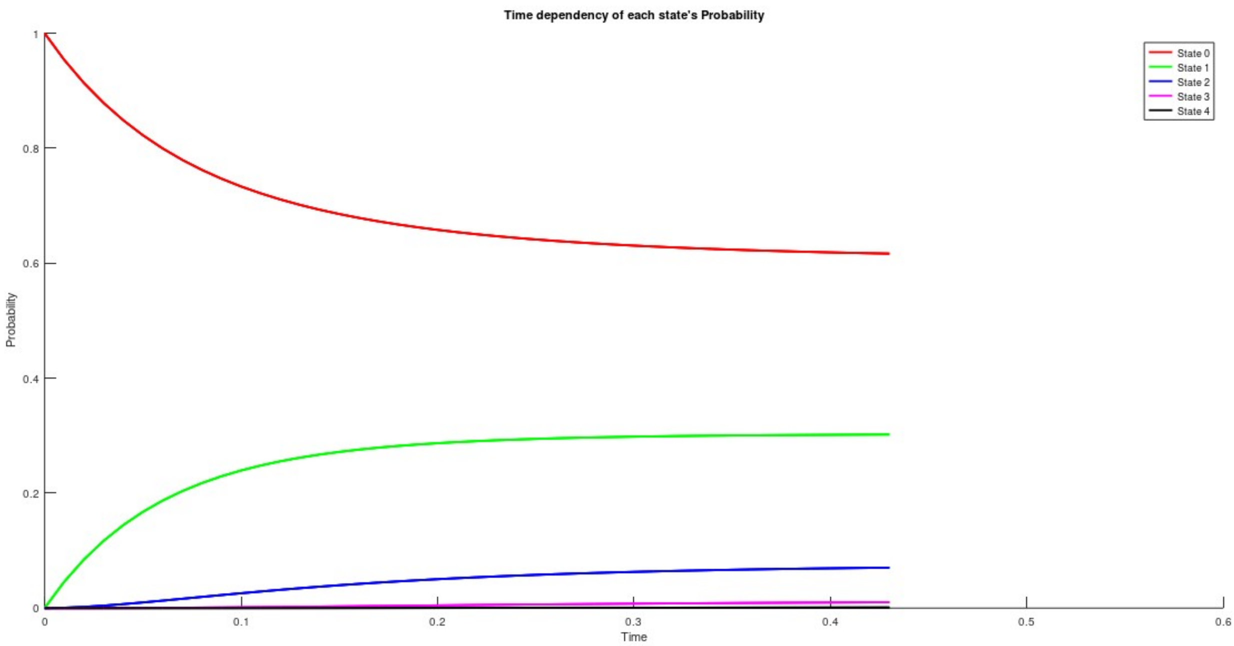
Εικόνα 10: Μέσος αριθμός πελατών στο σύστημα

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

```
The possibility of declining a client is: P_4 = 0.157978 %
```

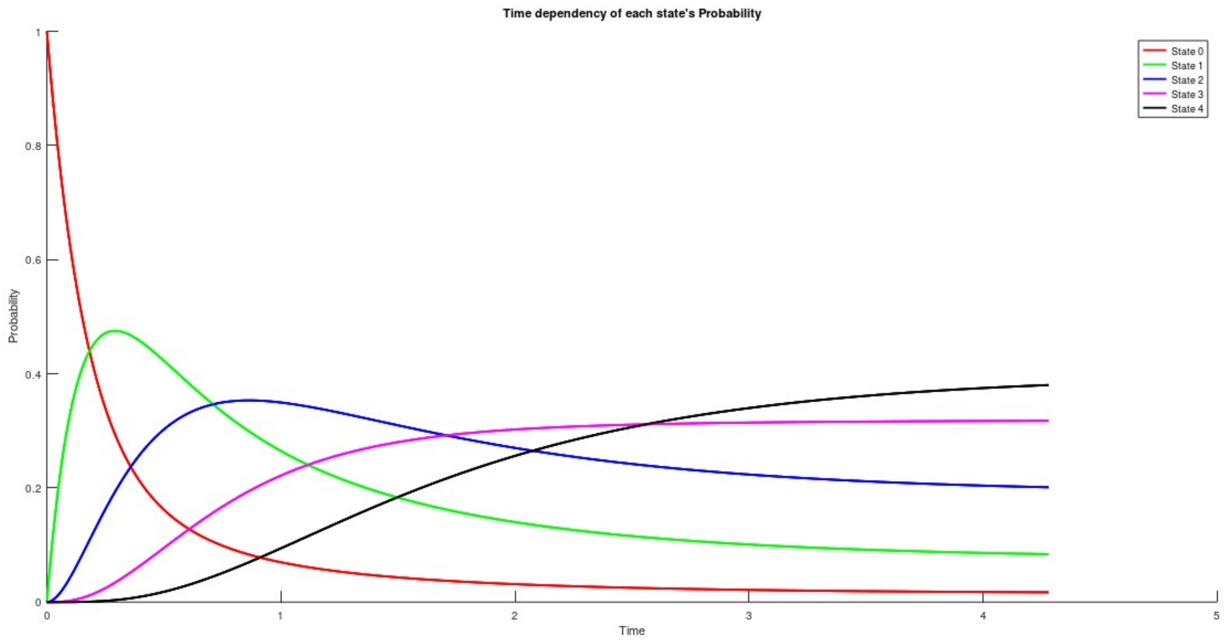
Εικόνα 11: Πιθανότητα απόρριψης πελάτη

v) Για λ = 5 πελάτες/sec και μ = 10 πελάτες/sec, τα **διαγράμματα πιθανοτήτων των καταστάσεων του συστήματος συναρτήσει του χρόνου** έχουν ως εξής:

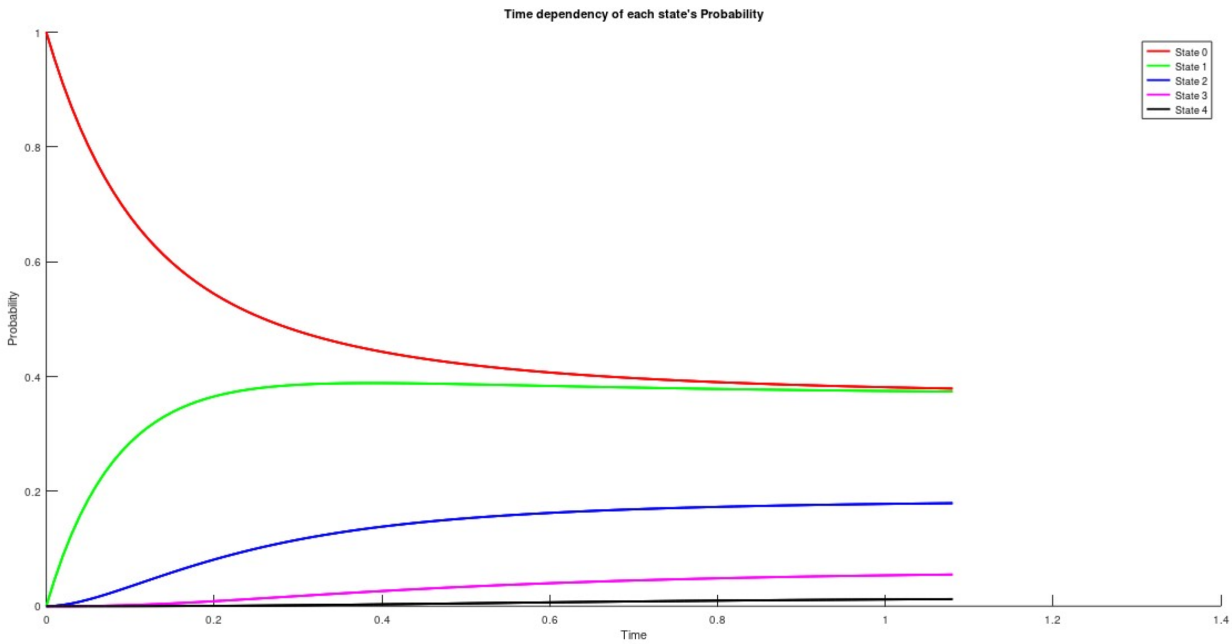


Εικόνα 12: Διάγραμμα πιθανοτήτων καταστάσεων του συστήματος (λ = 5, μ =10)

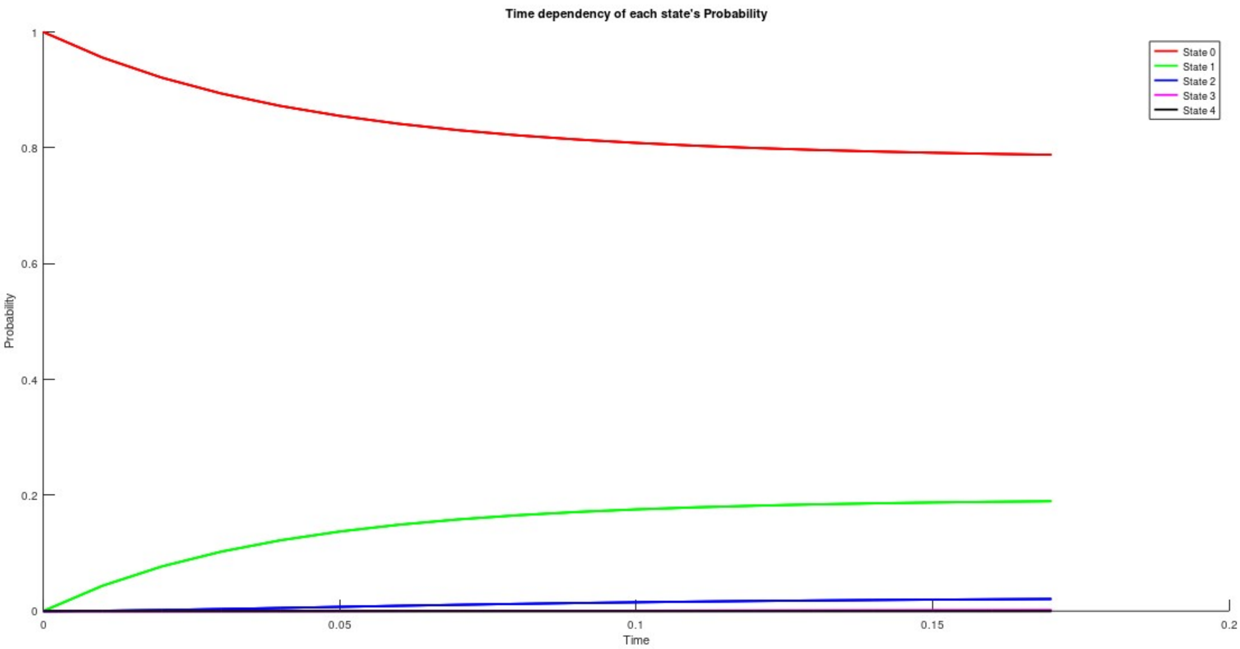
vi) Επαναλαμβάνοντας το παραπάνω ερώτημα για διαφορετικές τιμές του ρυθμού εξυπηρέτησης μ, λαμβάνουμε τα παρακάτω αποτελέσματα:



Εικόνα 13: Διάγραμμα πιθανοτήτων καταστάσεων του συστήματος (λ = 5, μ =1)



Εικόνα 14: Διάγραμμα πιθανοτήτων καταστάσεων του συστήματος (λ = 5, μ =5)



Εικόνα 15: Διάγραμμα πιθανοτήτων καταστάσεων του συστήματος (λ = 5, μ =20)

Ποιοτικά, παρατηρούμε πως όσο μικρότερο είναι το μ από το λ, τόσο ελαττώνονται οι πιθανότητες μικρότερων καταστάσεων, ενώ αυξάνονται εκείνες μεγαλύτερων καταστάσεων (όπου μικρότερη κατάσταση η 0, μεγαλύτερη η 4). Αντιστρόφως, για μ μεγαλύτερα του λ, όσο αυξάνεται η διαφορά, η μικρότερη κατάσταση έχει πολύ μεγάλη εργοδική πιθανότητα, ενώ οι πιθανότητες των υπολοίπων καταστάσεων παρουσιάζονται αρκετά μικρότερες. Όταν το μ = λ, οι πιθανότητες των καταστάσεων 0 και 1 φαίνεται να συγκλίνουν στην ίδια τιμή. Όσον αφορά τον **χρόνο σύγκλισης**, είναι προφανές πως όσο μεγαλώνει το μ, τόσο μειώνεται εκείνος.

Ο κώδικας που υλοποιεί τα ζητήματα του (β) ερωτήματος: (δίνεται για μ = 10, για την υλοποίηση του vi ερωτήματος προσαρμόζουμε το μ (mu) ανάλογα):

```
#Task2_3
pkg install -forge queueing
pkg load queueing

clc;
clear all;
close all;

lambda = 5;
mu = 10;          #changing mu, gives us results for vi
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];

#Task2_3_i
% get the transition matrix of the birth-death process
transition_matrix = ctmcbd(births_B, deaths_D);
printf("Transition matrix of our M/M/1/4 system: \n");
transition_matrix

#Task2_3_ii
% get the ergodic probabilities of the system
P = ctmc(transition_matrix);
printf("System's ergodic probabilities: \n");
P_percentage = 100 * P
```

```
#Task2_3_iii
E_n = 0;
for i = 1 : columns(states)
  E_n = E_n + (i * P(i));
endfor
printf("Mean number of clients in our system is: E[n(t)] = %d clients\n", E_n);

#Task2_3_iv
P_blocking = 100 * P(columns(P));
printf("The possibility of declining a client is: P_4 = %f %% \n", P_blocking)

#Task2_3_v
#transient probability of states until convergence to ergodic probabilities
#Convergence takes place when 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(1);
colors = "rgbmk";
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

#Task2_3_v_i
mu = 1;
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(1);
colors = "rgbmk";
hold on
```

```
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
```

Εικόνα 16: Κώδικας για τα ζητήματα του (β) ερωτήματος