

ΣΥΣΤΗΜΑΤΑ ΑΝΑΜΟΝΗΣ

Queuing Systems

Διαδικασίες Birth-Death, Ουρές Markov:

1. Διαγράμματα Μεταβάσεων Εργοδικών Καταστάσεων, Εξισώσεις Ισοροπίας
2. Προσομοιώσεις, Άσκηση Προσομοίωσης Ουράς M/M/1/10

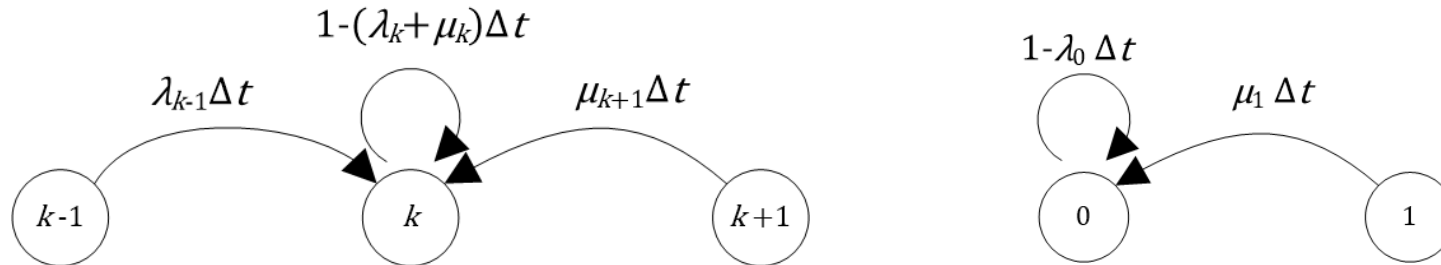
Βασίλης Μάγκλαρης

maglaris@netmode.ntua.gr

3/4/2019

ΕΞΙΣΩΣΕΙΣ ΙΣΟΡΡΟΠΙΑΣ (1/3) επανάληψη

- Birth-Death Process: Διάγραμμα **Πιθανοτήτων Μεταβάσεων** σε χρόνο $\Delta t \rightarrow 0$ προς $n(t) = k$

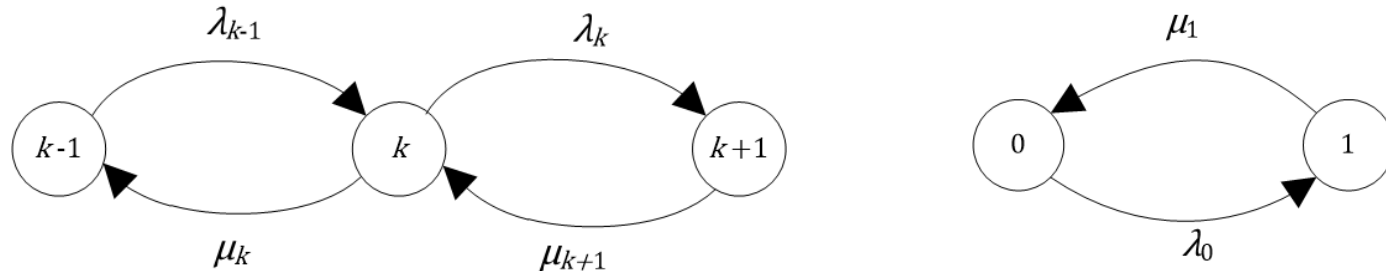


Εξισώσεις Μετάβασης (**Charpan - Kolmogorov**):

$$P_k(t) = \lambda_{k-1} \Delta t P_{k-1}(t - \Delta t) + \mu_{k+1} \Delta t P_{k+1}(t - \Delta t) + [1 - (\lambda_k + \mu_k) \Delta t] P_k(t - \Delta t), \quad k \geq 1$$

$$P_0(t) = \mu_1 \Delta t P_1(t - \Delta t) + (1 - \lambda_0 \Delta t) P_0(t - \Delta t)$$

- Birth-Death Process: Διάγραμμα **Ρυθμών Μεταβάσεων** μεταξύ **Εργοδικών** Καταστάσεων



Εξισώσεις Ισορροπίας:

$$(\lambda_k + \mu_k) P_k = \lambda_{k-1} P_{k-1} + \mu_{k+1} P_{k+1} \text{ για } k \geq 1 \text{ και } \lambda_0 P_0 = \mu_1 P_1$$

Σχετικές Πιθανότητες Μεταβάσεων $k \rightarrow (k + 1), k \rightarrow (k - 1)$:

$$P[k \rightarrow (k + 1)/\text{μετάβαση}] = \lambda_k / (\lambda_k + \mu_k), \quad P[k \rightarrow (k - 1)/\text{μετάβαση}] = \mu_k / (\lambda_k + \mu_k)$$

Dwell Time - Χρόνος Παραμονής στην $n(t) = k$ μέχρι την επόμενη μετάβαση

Εκθετική τυχαία μεταβλητή d_k με μέσο $1/(\lambda_k + \mu_k)$: Η μικρότερη δύο ανεξάρτητων εκθετικών τυχαίων μεταβλητών μέχρι (1) την επόμενη άφιξη με μέσο $1/\lambda_k$ ή (2) την ολοκλήρωση εξυπηρέτησης με μέσο $1/\mu_k$

$$d_k = \min(x, y), F_{d_k}(\tau) = P\{d_k \leq \tau\} = 1 - P\{d_k > \tau\} = 1 - e^{-(\lambda_k + \mu_k)\tau} \text{ διότι}$$

$$P\{d_k > \tau\} = P\{x > \tau, y > \tau\} = P\{x > \tau\}P\{y > \tau\} = e^{-\lambda_k \tau} e^{-\mu_k \tau} = e^{-(\lambda_k + \mu_k)\tau}$$

ΕΞΙΣΩΣΕΙΣ ΙΣΟΡΡΟΠΙΑΣ (2/3) επανάληψη

- Απείρως επισκέψιμες καταστάσεις $s = n(t)$ **positive recurrent states**: Με μη μηδενικές εργοδικές πιθανότητες $P\{n(t) = k\} = P_k(t) \rightarrow P_k > 0, k = 0,1,2, \dots$
- Σε μεγάλο χρονικό διάστημα παρατήρησης T ισορροπούν οι αριθμοί μεταβάσεων από και προς την κατάσταση s :

$$\#\{\text{ΜΕΤΑΒΑΣΕΩΝ ΠΡΟΣ ΤΗΝ ΚΑΤΑΣΤΑΣΗ } s\} = \#\{\text{ΕΚΤΟΣ ΤΗΣ } s\}$$

Σφαιρική Ισορροπία, Global Balance Equations

- Σε μεγάλο χρονικό διάστημα παρατήρησης T ισορροπούν οι αριθμοί μεταβάσεων μεταξύ δύο (όχι αναγκαστικά γειτονικών) καταστάσεων s_1 και s_2 :

$$\#\{\text{ΜΕΤΑΒΑΣΕΩΝ } s_1 \rightarrow s_2\} = \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } s_2 \rightarrow s_1\}$$

Τοπική Ισορροπία, Local Balance Equations

- Λόγω **εργοδικότητας** σε μεγάλο χρονικό διάστημα παρατήρησης T , με T_1 και T_2 τους συνολικούς χρόνους παραμονής στις s_1, s_2 :

$$(1) \quad \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } s_1 \rightarrow s_2\} = T_1 \times r_{1,2}$$

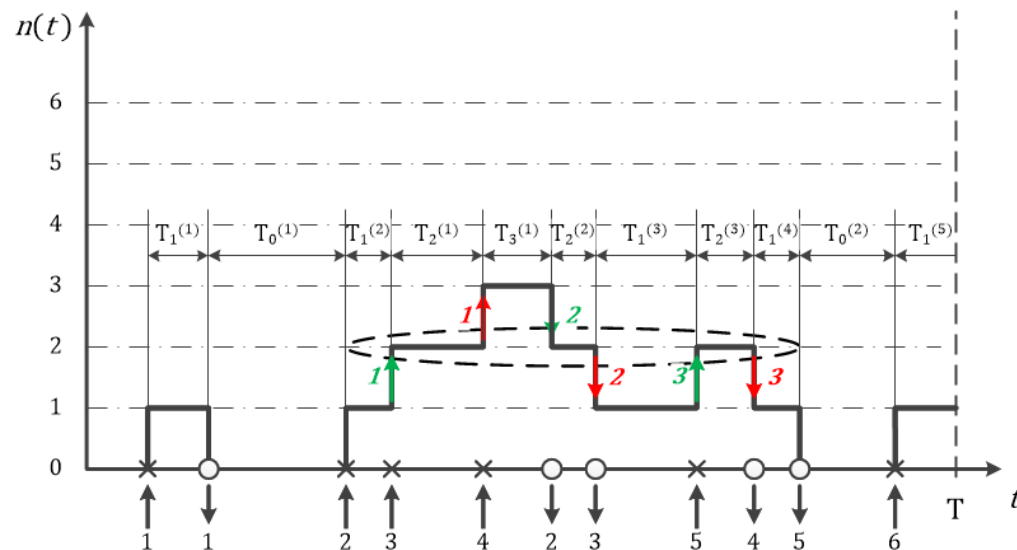
$$(2) \quad \#\{\text{ΜΕΤΑΒΑΣΕΩΝ } s_2 \rightarrow s_1\} = T_2 \times r_{2,1}$$

Όπου $r_{1,2}$ και $r_{2,1}$ οι μέσοι ρυθμοί μετάβασης από $s_1 \rightarrow s_2$ και $s_2 \rightarrow s_1$

- Λόγω **ισορροπίας**: (1) = (2) και $r_{1,2} \lim_{T \rightarrow \infty} \frac{T_1}{T} = r_{2,1} \lim_{T \rightarrow \infty} \frac{T_2}{T}$ ή

$$r_{1,2}P_1 = r_{2,1}P_2 \quad \text{Local Balance Equations}$$

ΕΞΙΣΩΣΕΙΣ ΙΣΟΡΡΟΠΙΑΣ (3/3) επανάληψη



Global Balance Equation

$$T_1 = T_1^{(1)} + T_1^{(2)} + T_1^{(3)} + T_1^{(4)} + T_1^{(5)}$$

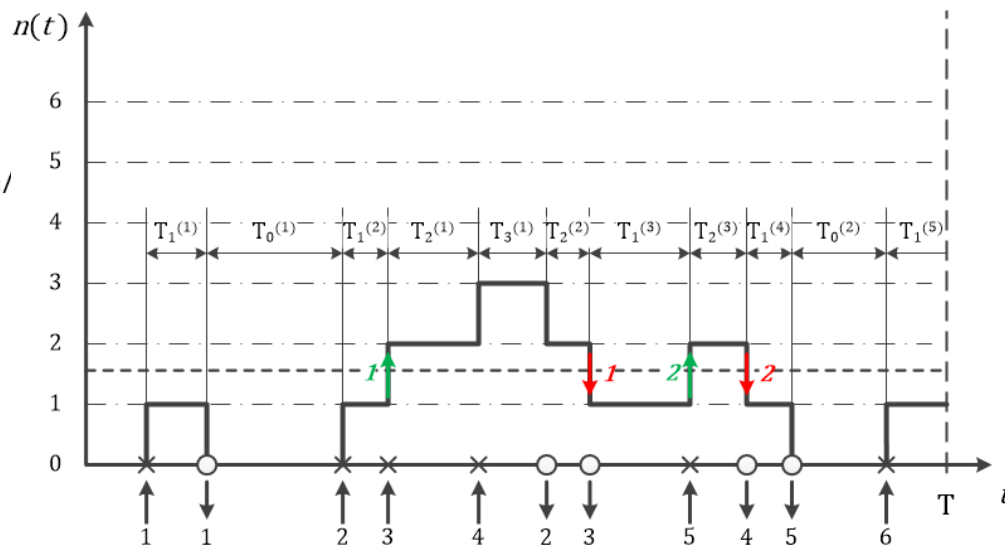
$$T_2 = T_2^{(1)} + T_2^{(2)} + T_2^{(3)}, T_3 = T_3^{(1)}$$

(State 2: **Transitions In** = **Transitions Out**)

$$\lambda T_1 + \mu T_3 \sim \lambda T_2 + \mu T_2 \Rightarrow (\lambda T_1 + \mu T_3)/T \sim (\lambda + \mu) (T_2/T)$$

$$\Rightarrow \lambda P_1 + \mu P_3 = (\lambda + \mu) P_2$$

Σχηματική Απεικόνιση
Εξισώσεις Ισορροπίας στην Εργοδική Κατάσταση
Χρόνος Παρατήρησης Δείγματος $n(t)$: T



Local Balance Equation

$$T_1 = T_1^{(1)} + T_1^{(2)} + T_1^{(3)} + T_1^{(4)} + T_1^{(5)}$$

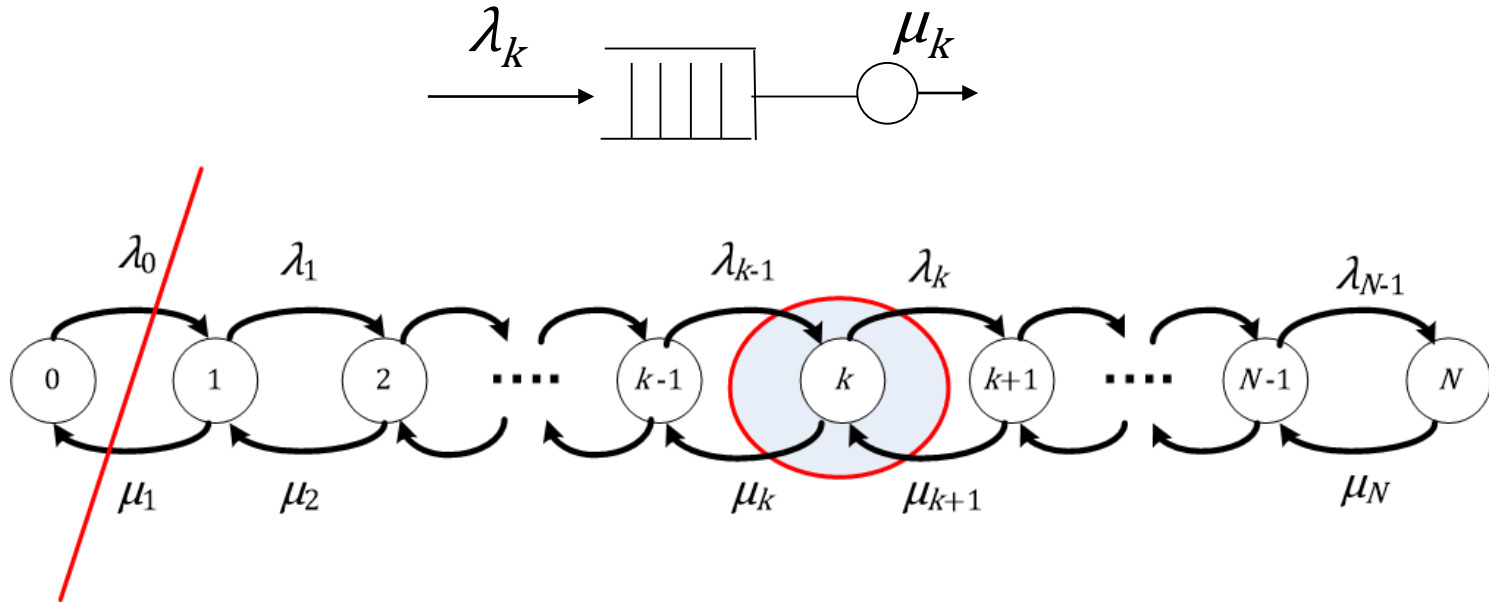
$$T_2 = T_2^{(1)} + T_2^{(2)} + T_2^{(3)}$$

(**Transitions 1->2** = **Transitions 2->1**)

$$\lambda T_1 \sim \mu T_2 \Rightarrow \lambda (T_1/T) \sim \mu (T_2/T) \Rightarrow \lambda P_1 = \mu P_2$$

ΟΥΡΑ Μ/Μ/1/Ν (1/2) επανάληψη

- Συστήματα Μ/Μ/1/Ν με ρυθμούς άφιξης και ρυθμούς εξυπηρέτησης εξαρτώμενους από τον αριθμό των πελατών στο σύστημα (από την παρούσα κατάσταση του συστήματος)
(State Dependent Μ/Μ/1/Ν Queues)



Local Balance Equation

$$\lambda_0 P_0 = \mu_1 P_1$$

$$\lambda_{k-1} P_{k-1} = \mu_k P_k \quad k = 1, 2, \dots, N$$

Global Balance Equation

$$\lambda_0 P_0 = \mu_1 P_1$$

$$(\lambda_k + \mu_k) P_k = \lambda_{k-1} P_{k-1} + \mu_{k+1} P_{k+1} \quad k = 1, \dots, N$$

Κανονικοποίηση Εργοδικών Πιθανοτήτων

$$P_0 + \dots + P_N = 1$$

ΟΥΡΑ Μ/Μ/1/Ν (2/2) επανάληψη

- Σταθεροί μέσοι ρυθμοί αφίξεων (γεννήσεων)

$$\lambda_k = \lambda, \text{ Poisson, } k = 1, 2, \dots, N$$

- Σταθεροί μέσοι ρυθμοί εξυπηρέτησης (θανάτων)

$$\mu_k = \mu, k = 1, 2, \dots, N$$

$$\text{Εκθετικοί ανεξάρτητοι χρόνοι εξυπηρέτησης } s, E(s) = 1/\mu$$

- Εργοδικές πιθανότητες καταστάσεων

$$P_k = \rho^k P_0, k = 0, 1, 2, \dots, N$$

$$P_0 + P_1 + \dots + P_{N-1} + P_N = 1$$

$$\rho = \lambda/\mu \text{ Erlangs (η } M/M/1/N \text{ είναι πάντα ευσταθής γιατί υπερβολικό φορτίο δεν προωθείται)}$$

- Αντικαθιστώντας με τον τύπο πεπερασμένου αθροίσματος γεωμετρικής προόδου:

$$P_0 = \frac{1 - \rho}{1 - \rho^{N+1}}, \quad \rho \neq 1$$

$$P_0 = \frac{1}{N + 1}, \quad \rho = 1$$

- Χρησιμοποίηση Εξυπηρετητή (Server Utilization) $U = 1 - P_0$
- Ρυθμαπόδοση (throughput) $\gamma = \lambda(1 - P_N) = \mu(1 - P_0) = \mu U$
- Πιθανότητα απώλειας $P_{blocking} = P_N$

- Στάσιμος Εργοδικός μέσος όρος πληθυσμού – κατάστασης

$$E[n(t)] \rightarrow E(k) = \sum_{k=1}^N k P_k = \rho \frac{1 - (N + 1)\rho^N + N\rho^{N+1}}{(1 - \rho)(1 - \rho^{N+1})}$$

- Νόμος του Little: $E(T) = E(k)/\gamma = E(k)/[\lambda(1 - P_N)]$

ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

BIRTH-DEATH ΟΜΟΙΟΓΕΝΩΝ ΑΦΙΞΕΩΝ (1/2)

- Σε στοχαστικό σύστημα **Birth-Death** με αφίξεις (γεννήσεις) σταθερού μέσου ρυθμού λ ανεξάρτητου του πληθυσμού $n(t) = k$ (**ομοιογενείς αφίξεις Poisson** με ρυθμό $\lambda_k = \lambda$ αφίξεις/sec) οι εργοδικές πιθανότητες (αν υπάρχουν) μπορούν να υπολογισθούν σαν λόγος αφίξεων που βρίσκουν το σύστημα στη κατάσταση $n(t) = k$ στη χρονική διάρκεια T_k , προς τον συνολικό αριθμό αφίξεων σε μεγάλο χρονικό διάστημα παρατήρησης T μιας χρονικής εξέλιξης της διαδικασίας:

$$P_n = \lim_{T \rightarrow \infty} \frac{T_k}{T} = \lim_{T \rightarrow \infty} \frac{\lambda T_k}{\lambda T} \approx \frac{\#\{\text{ΑΦΙΞΕΩΝ στη } n(t) = k \text{ σε ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ } T_k\}}{\#\{\text{ΣΥΝΟΛΟΥ ΑΦΙΞΕΩΝ σε ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ } T\}}$$

Άρα μπορούμε να προσομοιώσουμε σύστημα Birth-Death με ομοιογενείς αφίξεις καταμετρώντας τις αφίξεις στις διάφορες καταστάσεις που μεταβαίνει

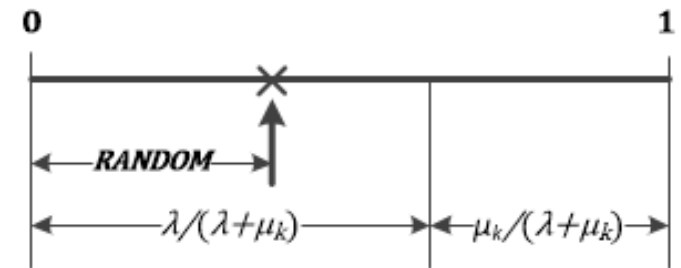
- Η εξέλιξη της κατάστασης (πληθυσμού) του συστήματος προκύπτει από τις πιθανότητες μετάβασης από την κατάσταση $n(t) = k$ στις $(k + 1)$, $(k - 1)$ με το δεδομένο ότι μια από τις δύο μεταβάσεις θα συμβεί με απόλυτη βεβαιότητα:

$$P[k \rightarrow (k + 1)/\text{μετάβαση}] = \lambda/(\lambda + \mu_k), \quad P[k \rightarrow (k - 1)/\text{μετάβαση}] = \mu_k/(\lambda + \mu_k)$$

- Η προσομοίωση ενεργοποιεί τις μεταβάσεις με κλήση τυχαίου αριθμού $RANDOM(0,1)$ ομοιόμορφα κατανεμημένου μεταξύ $(0, 1)$:

$$0 \leq RANDOM(0,1) \leq \frac{\lambda}{\lambda + \mu_k} \Rightarrow \text{ΑΦΙΞΗ, } n(t) \rightarrow k + 1$$

$$\frac{\lambda}{\lambda + \mu_k} < RANDOM(0,1) \leq 1 \Rightarrow \text{ΑΝΑΧΩΡΗΣΗ, } n(t) \rightarrow k - 1$$



- Αν το σύστημα έχει μηδενικό πληθυσμό $n(t) = 0$, η επόμενη μετάβαση είναι πάντα ΑΦΙΞΗ και $n(t) \rightarrow 1$
- Αν το σύστημα δεν επιδέχεται αύξηση πληθυσμού, η $n(t) = K$ είναι blocking state και δεν ενεργοποιείται μετάβαση κατάστασης, αλλά η διαδικασία τυχαίας δημιουργίας επόμενου γεγονότος (ΑΦΙΞΗ ή ΑΝΑΧΩΡΗΣΗ) καθώς και η μέτρηση αφίξεων συνεχίζονται κανονικά

ΠΡΟΣΟΜΟΙΩΣΗ ΣΤΟΧΑΣΤΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

BIRTH-DEATH ΟΜΟΙΟΓΕΝΩΝ ΑΦΙΞΕΩΝ (2/2)

Στατιστική Σύγκλιση Προσομοίωσης

- Η σύγκλιση της προσομοίωσης ελέγχεται ως προς την **στατιστική σύγκλιση** μεγεθών ενδιαφέροντος, π.χ. εκτίμηση μέσου πληθυσμού $E[n(t)] = \sum_{k=0}^K P_k$ που υπολογίζεται σε τακτά διαστήματα από την αρχή της προσομοίωσης π.χ. κάθε 1000 αφίξεις: (από 0 έως 1000 αφίξεις), (από 0 έως 2000 αφίξεις), (από 0 έως 3000 αφίξεις) κλπ.
- Λόγω **εργοδικότητας** δεν απαιτούνται επαναλήψεις: Η εκτέλεση του προγράμματος και οι μετρήσεις συνεχίζονται σε μία υλοποίηση που διακόπτεται προσωρινά π.χ. κάθε 1000 αφίξεις μέχρι την ικανοποίηση κριτηρίων σύγκλισης
- Η στατιστική σύγκλιση επιταχύνεται αν αγνοήσουμε στη καταμέτρηση αφίξεων στις διάφορες καταστάσεις τις πρώτες μεταβάσεις (π.χ. 1-1000 αφίξεις) που αντιστοιχούν στο **μεταβατικό φαινόμενο** προς την εργοδική κατάσταση
- **Εμπειρικός κανόνας**: Η ταχύτητα σύγκλισης είναι αντιστρόφως ανάλογη με τον βαθμό χρησιμοποίησης του συστήματος
- **Γενική Παρατήρηση**: Η σύγκλιση μιας προσομοίωσης είναι σύνηθες να διερευνάται με μαθηματικά εργαλεία στατιστικής (π.χ. τεστ χ^2) γιατί μπορεί να κρύβονται εξαρτήσεις (correlations) μεγεθών, περιοδικές συμπεριφορές που οδηγούν σε πρόωρους τερματισμούς κλπ. Μια πλήρης προσομοίωση περιλαμβάνει διαστήματα εμπιστοσύνης για τις εκτιμήσεις πιθανοτήτων και ροπών τυχαίων μεταβλητών (π.χ. μέσοι όροι, διασπορά)

ΠΡΟΣΟΜΟΙΩΣΗ ΟΥΡΑΣ Μ/Μ/1/10

<http://www.netmode.ntua.gr/courses/undergraduate/queues/simulation/>

RANDOM: Ομοιόμορφος τυχαίος αριθμός (0,1)

THRESHOLD: $\lambda / (\lambda + \mu)$

ARRIVALS: Συνολικός αριθμός αφίξεων

ARRIVAL[STATE]: Αριθμός αφίξεων στην κατάσταση STATE = 0, 1, ... , 10

COUNT: Αριθμός μεταβάσεων COUNT = 0, 1, ... , MAXIMUM

STATE: Κατάσταση ουράς (πληθυσμός συστήματος Μ/Μ/1/10) , STATE = 0, 1, ... , 10

P[STATE]: Εργοδική πιθανότητα STATE = 0, 1, ... , 10

AVERAGE: Μέσος πληθυσμός συστήματος Μ/Μ/1/10

INITIALIZE: COUNT = 0, STATE = 0, ARRIVALS = 0, ARRIVAL[0...10] = 0, P[0...10] = 0

ARRIVAL:
ARRIVALS = ARRIVALS + 1
ARRIVAL[STATE] = ARRIVAL[STATE] + 1
COUNT = COUNT + 1
IF STATE = 10 : **GO TO** LOOP
ELSE : STATE = STATE + 1
GO TO LOOP

LOOP : **IF** STATE = 0 : **GO TO** ARRIVAL
ELSE :
IF RANDOM < THRESHOLD : **GO TO** ARRIVAL
ELSE : **GO TO** DEPARTURE

DEPARTURE : COUNT = COUNT + 1 ; STATE = STATE - 1
IF COUNT < MAXIMUM : **GO TO** LOOP
ELSE : P[STATE=1...10] = ARRIVAL[STATE= 1...10] / ARRIVALS
AVERAGE = SUM { STATE ^ P[STATE] }, STATE = [1...10]

ΠΡΟΣΟΜΟΙΩΣΗ ΟΥΡΑΣ Μ/Μ/1/10

Άσκηση

Οδηγίες Προγράμματος Προσομοίωσης Ουράς Μ/Μ/1/10

- Δημιουργία προγράμματος με βάση τον ψευτο-κώδικα της προηγούμενης διαφάνειας
- Γλώσσα προγραμματισμού: Γενικής φύσεως (Octave/MATLAB, Python) και όχι εξειδικευμένη γλώσσα Προσομοίωσης
- Υπολογισμός εργοδικών πιθανοτήτων P_k , $E(k) = \sum_{k=0}^{10} P_k$ και $P_{\text{blocking}} = P_{10}$ για $\lambda = 1$ και $1/\mu = 0.5, 0.75, 0.85$ με μετρητές αφίξεων στις καταστάσεις $n(t) = k$ από την αρχή μέχρι 5.000, 10.000, 15.000 ... αφίξεις. Σύγκριση με γνωστά αναλυτικά αποτελέσματα ουράς Μ/Μ/1/10:

$$P_k = P_0 \left(\frac{\lambda}{\mu} \right)^k, \quad \sum_{k=0}^{10} P_k = 1$$

- Μέτρο σύγκλισης: Διαφορά μεταξύ διαδοχικών μετρήσεων του μέσου μήκους ουράς (από την αρχή μέχρι 5.000, 10.000, 15.000... αφίξεις) μέχρι εξαντλήσεως του διαθέσιμου χρόνου αν δεν διαπιστωθεί σύγκλιση νωρίτερα (όριο 200.000 γεγονότα, αφίξεις + εξυπηρετήσεις)
- **Υπενθύμιση:** Η στατιστική σύγκλιση επιταχύνεται αν αγνοήσουμε στη καταμέτρηση αφίξεων στις διάφορες καταστάσεις τις πρώτες μεταβάσεις (π.χ. 1-5.000 αφίξεις) που αντιστοιχούν στο **μεταβατικό φαινόμενο** προς την εργοδική κατάσταση
- **Γενική Οδηγία:** Για debugging εξάγεται αρχικά λεπτομερές **trace** των μεταβάσεων της κατάστασης του συστήματος το οποίο θα απενεργοποιηθεί μετά την επιβεβαίωση της ορθότητας του κώδικα