Ουρές Αναμονής

Σημειώσεις (πρόχειρες, υπό διαμόρφωση) 2016-2017, έκδοση 30/6/2017

Αντώνης Οικονόμου

Οι σημειώσεις αυτές αναπτύσσονται στα πλαίσια του προπτυχιακού μαθήματος ‘Ουρές Αναμονής ’ του Τμήματος Μαθηματικών του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών. Περιέχουν πολλά λάθη και ασάφειες και βρίσκονται υπό συνεχή αναθεώρηση. Η υπόδειξη λαθών και οποιαδήποτε σύσταση για την παρουσίαση ή το περιεχόμενο τους είναι ιδιαίτερα καλοδεχούμενες.

Αντώνης Οικονόμου

Περιεχόμενα

Μέρος I

Εισαγωγή

1

Περιγραφή, ονοματολογία και μέτρα απόδοσης

Στο κεφάλαιο αυτό περιγράφουμε τα βασικά χαρακτηριστικά που καθορίζουν ένα σύστημα εξυπηρέτησης. Επιπλέον, αναφερόμαστε την ονοματολογία-ταξινόμηση των συστημάτων εξυπηρέτησης του Kendall. Τέλος, αναφέρουμε τα σημαντικότερα μέτρα απόδοσης που απασχολούν τη μελέτη των συστημάτων εξυπηρέτησης.

1.1 Βασική περιγραφή και ονοματολογία Kendall

Ένα σύστημα εξυπηρέτησης ή ουρά αναμονής (queueing system, queue) είναι στην ουσία ένα σύστημα εισόδου - εξόδου διακριτών οντοτήτων (μονάδων, πελατών), στο οποίο υπεισέρχεται τυχαιότητα. Αυτός ο ορισμός είναι πολύ γενικός και περιλαμβάνει πολλές πραγματικές καταστάσεις που παρατηρούνται στην καθημερινή ζωή, καθώς και σε περίπλοκα τεχνολογικά συστήματα. Ιστορικά, η συγκεκριμένη επιστημονική περιοχή άρχισε να αναπτύσσεται στις αρχές του 20ου αιώνα, όταν ο A.K. Erlang δημοσίευσε κάποιες εργασίες του για τη μαθηματική μοντελοποίηση του συνωστισμού σε τηλεφωνικά δίκτυα. Η μεγάλη επιτυχία αυτών των μεθόδων στη μελέτη πραγματικών συστημάτων έδωσε τεράστια ώθηση στη περαιτέρω ανάπτυξη της θεωρίας των ουρών αναμονής καθώς και των εφαρμογών της και σε άλλα πεδία.

Τα βασικά χαρακτηριστικά ενός συστήματος εξυπηρέτησης είναι η διαδικασία αφίξεων (arrival process), οι χρόνοι εξυπηρέτησης (service times), ο αριθμός των (παράλληλων) παροχέων εξυπηρέτησης - υπηρετών (number of servers), η χωρητικότητα του συστήματος (system capacity) και η πειθαρχία ουράς (queue discipline). Για το λόγο αυτό ο D.G. Kendall εισήγαγε ένα σύστημα ονοματολογίας για τις πιο απλές ουρές που περιγράφει συνοπτικά αυτά τα χαρακτηριστικά. Η ονοματολογία του Kendall έχει τη μορφή $A/B/c/k(\quad)$, όπου τα A, B είναι γράμματα, τα c, k αριθμοί και μέσα στην παρένθεση γράφεται μια ακροστοιχίδα γραμμάτων. Καθεμιά από τις 5 παραμέτρους της ονοματολογίας του Kendall αναφέρεται στα 5 χαρακτηριστικά που περιγράψαμε παραπάνω.

Η διαδικασία αφίξεων περιγράφει το πώς έρχονται οι πελάτες στο σύστημα. Είναι συνήθως μια ανανεωτική διαδικασία, δηλαδή οι χρόνοι μεταξύ διαδοχικών αφίξεων πελατών είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κάποια γνωστή γενική κατανομή. Η διαδικασία αφίξεων αντιστοιχεί στο γράμμα A της ονοματολογίας Kendall. Οι πιο συχνές τιμές που παίρνει στη βιβλιογραφία είναι GI ή G (General independent), M (Memoryless, Markovian), D (Deterministic) και E_r (Erlang-r) για τις περιπτώσεις που οι ενδιάμεσοι χρόνοι μεταξύ των αφίξεων είναι γενικοί, εκθετικοί, σταθεροί και Erlang-r, αντίστοιχα. Υπάρχουν βέβαια και άλλες τιμές για το γράμμα A που αντιστοιχούν σε κατανομές που εμφανίζονται σπανιότερα στη βιβλιογραφία.

Οι χρόνοι εξυπηρέτησης θεωρούνται στα κλασικά μοντέλα επίσης ανεξάρτητοι και ισόνομοι και αντιστοιχούν στο γράμμα B της ονοματολογίας Kendall που παίρνει τις ίδιες τιμές με το γράμμα A. Οι τιμές GI και G για το A και το B σηματοδοτούν ακριβώς το ίδιο, δηλαδή ανεξάρτητους ισόνομους χρόνους με γενική κατανομή. Παρόλα αυτά συνηθίζεται εθιμικά η τιμή GI για το A και η τιμή G για το B στα περισσότερα βιβλία και σημειώσεις που κυκλοφορούν διεθνώς. Στις παρούσες σημειώσεις θα προτιμήσουμε την τιμή GI τόσο για το A και όσο και για το B.

Ο αριθμός των υπηρετών αναφέρεται στο πόσοι είναι οι παράλληλοι υπηρέτες που εξυπηρετούν τη ροή των πελατών που εισέρχεται στο σύστημα. Με την έννοια 'παράλληλοι' υπηρέτες εννοούμε ότι υπάρχει μια κοινή ουρά για όλους και οι πελάτες πηγαίνουν στον πρώτο υπηρέτη που θα αδειάσει, αν όλοι οι υπηρέτες είναι απασχολημένοι, ή διαλέγουν στην τύχη κάποιον από τους άδειους υπηρέτες, αν υπάρχουν ελεύθεροι υπηρέτες. Ο αριθμός των υπηρετών αντιστοιχεί στον αριθμό c της ονοματολογίας Kendall.

Η χωρητικότητα του συστήματος εκφράζει το μέγιστο πλήθος πελατών που μπορεί να χωρέσει το σύστημα, συμπεριλαμβανομένων τόσο αυτών που περιμένουν να εξυπηρετηθούν όσο και αυτών που βρίσκονται σε διαδικασία εξυπηρέτησης. Αν ένα σύστημα έχει φτάσει στο μέγιστο της χωρητικότητάς του και αφιχθεί ένας πελάτης τότε στο κλασικό μοντέλο των ουρών αναμονής ο πελάτης απορρίπτεται και θεωρείται χαμένος για πάντα. Φυσικά υπάρχουν και μοντέλα στα οποία οι πελάτες που αποχωρούν λόγω της περιορισμένης χωρητικότητας επανέρχονται αργότερα με την ελπίδα να υπάρχει διαθέσιμη θέση στο σύστημα. Στην περίπτωση αυτή μιλάμε για μοντέλα με επαναπροσπάθειες ή επαναδοκιμές (retrials). Σε τέτοια μοντέλα είναι απαραίτητο να αποσαφηνισθεί η διαδικασία με την οποία οι πελάτες επανέρχονται στο σύστημα και έτσι τα μοντέλα αυτά δεν περιγράφονται στο πλαίσιο της ονοματολογίας Kendall. Προς το παρόν, επομένως, μένουμε στο πλαίσιο των μοντέλων χωρίς επαναπροσπάθειες, όπου η χωρητικότητα του συστήματος αντιστοιχεί στον αριθμό k της ονοματολογίας του Kendall.

Η πειθαρχία ουράς είναι ο τρόπος με τον οποίο το σύστημα διαλέγει ποιόν πελάτη θα εξυπηρετήσει, μόλις βρεθεί κάποιος διαθέσιμος υπηρέτης. Η πλέον κλασική πειθαρχία ουράς είναι η FCFS (First-Come-First-Served) ή FIFO (First-In-First-Out), κατά την οποία οι πελάτες εξυπηρετούνται σύμφωνα με τη σειρά της άφιξής τους. Έτσι, μόλις αδειάσει ένας υπηρέτης, επιλέγεται για εξυπηρέτηση ο πελάτης που έχει αφιχθεί πρώτος από όλους που περιμένουν. Η πειθαρχία αυτή μοιάζει η πιο δίκαιη με μία πρώτη ματιά και χρησιμοποιείται περισσότερο από οποιαδήποτε άλλη στην περίπτωση που οι πελάτες είναι άνθρωποι και μάλιστα έχουν οπτική επαφή με το τί συμβαίνει στο σύστημα (και επομένως μπορούν και βλέπουν πότε φθάνουν οι άλλοι πελάτες). Σε διάφορες εφαρμογές, πάντως, χρησιμοποιούνται και άλλες πειθαρχίες ουράς, όπως η LCFS (Last-Come-First-Served) κατά την οποία οι πελάτες εξυπηρετούνται αντίστροφα από τη σειρά αφιξής τους, η SIRO (Service-In-Random-Order) όπου οι πελάτες εξυπηρετούνται τυχαία, χωρίς να λαμβάνεται υπόψιν η σειρά αφιξής τους, η SSTF (Shortest-Service-Time-First) όπου επιλέγεται προς εξυπηρέτηση ο πελάτης που έχει το μικρότερο χρόνο εξυπηρέτησης κ.α. Υπάρχουν επίσης πειθαρχίες ουράς για την περίπτωση που υπάρχουν διάφορα είδη πελατών και το σύστημα τους αντιμετωπίζει διαφορετικά. Στην περίπτωση αυτή είναι δυνατόν κάποια είδη πελατών να έχουν προτεραιότητα έναντι κάποιων άλλων, οπότε μιλάμε για πειθαρχίες ουρών με προτεραιότητες. Γενικά, αν σκεφθούμε πρακτικές εφαρμογές των ουρών αναμονής θα συνειδητοποιήσουμε ότι υπάρχει μεγάλη ποικιλία στις πειθαρχίες ουράς που χρησιμοποιούνται (σκεφτείτε τα ταμεία για πελάτες με μέχρι 10 προϊόντα στα supermarket, τα ταμεία για επιχειρηματικές συναλλαγές στις τράπεζες, τις κρατήσεις θέσεων σε εστιατόρια, τα τηλεφωνικά κέντρα που εξυπηρετούν πελάτες σε περισσότερες από μια γλώσσες κλπ.).

Η χωρητικότητα του συστήματος k και/ή πειθαρχία ουράς μπορεί να παραλείπονται στην ονοματολογία Kendall. Αυτό συμβαίνει στην περίπτωση που έχουμε απεριόριστη χωρητικότητα ($k = \infty$) ή πειθαρχία FCFS αντίστοιχα.

Παράδειγμα 1.1 Η $M/M/1$ ουρά είναι ένα σύστημα εξυπηρέτησης με Poisson διαδικασία αφίξεων (ανεξάρτητους εκθετικούς ενδιάμεσους χρόνους αφίξεων), εκθετικούς χρόνους εξυπηρέτησης και 1 υπηρέτη που έχει άπειρη χωρητικότητα και λειτουργεί υπό την πειθαρχία ουράς FCFS.

Παράδειγμα 1.2 Η $GI/E_2/1/5$ (SIRO) ουρά είναι ένα σύστημα εξυπηρέτησης με ανανεωτική διαδικασία αφίξεων (ανεξάρτητους και ισόνομους ενδιάμεσους χρόνους αφίξεων), Erlang-2 χρόνους εξυπηρέτησης και 1 υπηρέτη που έχει χωρητικότητα για 5 πελάτες και λειτουργεί υπό την πειθαρχία ουράς SIRO.

Πολλές φορές η διαδικασία αφίξεων και/η κατανομή των χρόνων εξυπηρέτησης δεν είναι ακριβώς γνωστή. Στην περίπτωση αυτή μπορεί να δίνεται κάποια αδρή πληροφορία για το πώς έρχονται και πώς εξυπηρετούνται οι πελάτες. Π.χ. μπορεί να δίνεται ο μέσος ενδιάμεσος χρόνος a μεταξύ δυο διαδοχικών αφίξεων και ο μέσος χρόνος εξυπηρέτησης b. Ισοδύναμα, μπορεί να δίνεται ο ρυθμός αφίξεων $\lambda = 1/a$ και ο ρυθμός εξυπηρέτησης $\mu = 1/b$. Τα a και b έχουν τη φυσική έννοια της (μέσης) περιόδου των διαδικασιών των αφίξεων και των εξυπηρετήσεων αντίστοιχα, ενώ τα λ και μ αντιστοιχούν στη φυσική έννοια της συχνότητας. Με τόσο ελλειπή πληροφορία, βεβαίως, τα αποτελέσματα που θα προκύψουν από τη μαθηματική ανάλυση μπορεί να είναι τελείως αναξιόπιστα. Για την εξαγωγή ασφαλών συμπερασμάτων χρειάζεται να είναι γνωστή η κατανομή των αντίστοιχων χρόνων ή τουλάχιστον κάποιες ροπές ανώτερης τάξης. Μετά τη μέση τιμή, η διασπορά των χρόνων μεταξύ των αφίξεων και/ή των χρόνων εξυπηρέτησης επηρεάζει σημαντικά την απόδοση ενός συστήματος.

1.2 Μέτρα απόδοσης συστήματος

Αφού περιγραφεί ένα σύστημα, το πρόβλημα που τίθεται είναι να προβλέψουμε πώς θα συμπεριφέρεται. Τυπικά ερωτήματα που απασχολούν τη θεωρία των ουρών αναμονής είναι:

1. Πόσοι πελάτες θα βρίσκονται στο σύστημα κατά μέσο όρο μια τυχούσα χρονική στιγμή;
2. Πόσο χρόνο θα περάσει στο σύστημα κατά μέσο όρο ένας πελάτης;

3. Ποιό ποσοστό του χρόνου του θα βρίσκεται απασχολημένος ένας υπηρέτης που δουλεύει στο συγκεκριμένο σύστημα;

Όπως βλέπουμε, υπάρχουν ερωτήματα που απασχολούν το διαχειριστή του συστήματος, που βλέπει το σύστημα συνολικά, σαν εξωτερικός παρατηρητής (ερώτημα 1), ερωτήματα που απασχολούν τους πελάτες, που επιδρούν στο σύστημα μόνο παροδικά και κατόπιν φεύγουν (ερώτημα 2) και ερωτήματα που απασχολούν τους υπηρέτες, που επιδρούν διαρκώς στο σύστημα (ερώτημα 3). Είναι σημαντικό να κατανοηθεί ότι αυτοί οι παράγοντες του συστήματος (διαχειριστής, πελάτες και υπηρέτες) έχουν διαφορετικές οπτικές και για το λόγο αυτό υπάρχουν μέτρα απόδοσης του συστήματος που σχετίζονται με την οπτική του καθενός.

Για το διαχειριστή του συστήματος η πιο σημαντική πληροφορία είναι ο αριθμός των πελατών που βρίσκονται στο σύστημα μια τυχαία χρονική στιγμή. Έτσι ορίζουμε

- $Q(t)$ τον αριθμό των πελατών στο σύστημα τη στιγμή t,
- $Q_q(t)$ τον αριθμό των πελατών στο χώρο αναμονής τη στιγμή t (ο δείκτης q μπαίνει για να θυμίζει ότι αναφερόμαστε στο πλήθος των πελατών στην ουρά ($q \to$ queue),
- $Q_s(t)$ τον αριθμό των πελατών στο χώρο εξυπηρέτησης τη στιγμή t (ο δείκτης s μπαίνει για να θυμίζει ότι αναφερόμαστε στο πλήθος των πελατών υπό εξυπηρέτηση ($s \to$ service).

Φυσικά ισχύει

$$Q(t) = Q_q(t) + Q_s(t).$$

Ενδιαφερόμαστε για το τί συμβαίνει στο σύστημα σε κατάσταση ισορροπίας, δηλαδή καθώς $t \to \infty$ και για το λόγο αυτό μας ενδιαφέρουν οι οριακές πιθανότητες

$$p_j = \lim_{t\to\infty} \Pr[Q(t) = j],\ j = 0, 1, \ldots. \tag{1.1}$$

Η p_j είναι η οριακή πιθανότητα σε συνεχή χρόνο να βρίσκονται j πελάτες στο σύστημα. Μια πρώτη της ερμηνεία επομένως είναι ότι εκφράζει την πιθανότητα να βρίσκονται j πελάτες στο σύστημα, αν το κοιτάξουμε σε κάποια χρονική στιγμή που έχει παρέλθει πολύς χρόνος από την έναρξη λειτουργίας του συστήματος και επομένως η επίδραση της αρχικής κατάστασης του συστήματος (αριθμός πελατών σε αυτό κατά την έναρξη της λειτουργίας του) έχει χαθεί. Ισχύει, επίσης, ότι

$$p_j = \lim_{t\to\infty} \int_0^t \Pr[Q(u) = j]\frac{1}{t}du,\ j = 0, 1, \ldots. \tag{1.2}$$

(διότι αν υπάρχει το σύνηθες όριο $\lim_{t\to\infty} f(t)$ μιας συνάρτησης τότε υπάρχει και το Cesaro όριο $C - \lim_{t\to\infty} f(t) = \lim_{t\to\infty} \int_0^t f(u)du/t$ και είναι και τα δυο ίσα). Η έκφραση αυτή μας οδηγεί σε μια δεύτερη ερμηνεία της p_j, ως της πιθανότητας να βρίσκονται j πελάτες στο σύστημα, αν το κοιτάξουμε σε μια τυχαία, ομοιόμορφα επιλεγμένη χρονική στιγμή σε ένα διάστημα μεγάλου μήκους. Αν το σύστημα έχει αναγεννητικό χαρακτήρα, δηλαδή η στοχαστική διαδικασία που περιγράφει την εξέλιξή του ξαναρχίζει από την αρχή, ξεχνώντας την παρελθούσα ιστορία της, σε συγκεκριμένες χρονικές στιγμές (συνήθως τις στιγμές που το σύστημα αδειάζει) τότε έχουμε και ότι με πιθανότητα 1

$$p_j = \lim_{t\to\infty} \frac{\int_0^t 1\{Q(u)=j\}du}{t}, \tag{1.3}$$

όπου με $1\{Q(u)=j\}$ συμβολίζουμε τη δείκτρια τυχαία μεταβλητή του ενδεχομένου $\{Q(u)=j\}$ που παίρνει την τιμή 1 όταν $Q(u)=j$ και την τιμή 0 διαφορετικά. Παρατηρώντας ότι ο αριθμητής του κλάσματος είναι η χρονική διάρκεια που στο σύστημα βρίσκονται j πελάτες στο χρονικό διάστημα $[0,t]$, έχουμε μια τρίτη και σημαντικότερη ερμηνεία της p_j ως το μακροπρόθεσμο ποσοστό του χρόνου που βρίσκονται j πελάτες στο σύστημα.

Πότε, όμως, είναι οι παραπάνω τρεις εκφράσεις ισοδύναμες; Η απάντηση είναι ότι ισχύουν και οι τρεις εκφράσεις κάτω από πολύ γενικές συνθήκες, που ισχύουν στη συντριπτική πλειονότητα των περιπτώσεων που εμφανίζονται στις εφαρμογές. Συγκεκριμένα, ισχύουν όταν η στοχαστική διαδικασία $\{Q(t)\}$ είναι αναγεννητική διαδικασία (regenerative process) που πρακτικά αυτό σημαίνει ότι πρέπει να υπάρχουν χρονικά σημεία στην εξέλιξη της $\{Q(t)\}$ στα οποία η διαδικασία συμπεριφέρεται σαν να ξεκινάει από την αρχή. Στις ουρές αναμονής κατά κανόνα αυτό ισχύει και μάλιστα τα σημεία αυτά είναι οι χρονικές στιγμές που στο σύστημα φθάνει ένας πελάτης που το βρίσκει κενό. Για μια αυστηρή συζήτηση πάνω στο θέμα των αναγεννητικών διαδικασιών μπορεί κανείς να ανατρέξει στα βιβλία των Φακίνου (2003) και Wolff (1989). Όλα τα συστήματα που θα μελετήσουμε είναι αναγεννητικά και κατά συνέπεια οι οριακές πιθανότητες ισούνται με τα μακροπρόθεσμα ποσοστά, όπως τα είδαμε παραπάνω. Για το λόγο αυτό σε ό,τι ακολουθεί θα χρησιμοποιούμε ελεύθερα κάποια από τις ισοδύναμες εκφράσεις, χωρίς περαιτέρω μνεία. Αυτό θα συμβαίνει και για άλλα μεγέθη που αφορούν άλλες τυχαίες μεταβλητές που έχουν ενδιαφέρον για τη μελέτη των συστημάτων. Π.χ., συμβολίζοντας με Q την οριακή τυχαία μεταβλητή που έχει την κατανομή $(p_j : j = 0,1,\ldots)$ θα γράφουμε

$$E[Q] = \lim_{t\to\infty} E[Q(t)] = \lim_{t\to\infty} \frac{1}{t}\int_0^t Q(u)du. \tag{1.4}$$

Αν έχουμε προσδιορίσει την $(p_j : n = 0,1,\ldots)$ (πράγμα που δεν είναι πάντα εύκολο) τότε μπορούμε να υπολογίσουμε την $E[Q]$ από τη σχέση $E[Q] = \sum_{j=0}^{\infty} jp_j$. Σε κάποιες περιπτώσεις η $E[Q]$ μπορεί να υπολογιστεί και απευθείας με πιθανοθεωρητικά επιχειρήματα, χρησιμοποιώντας κάποια βασικά αποτελέσματα που θα παρουσιαστούν παρακάτω. Ομοίως με την $E[Q]$ ορίζονται οι $E[Q_q]$ και $E[Q_s]$.

Για τους πελάτες το πιο σημαντικό μέτρο απόδοσης είναι ο χρόνος που παραμένουν στο σύστημα. Έτσι ορίζουμε

- S_n το χρόνο παραμονής στο σύστημα του n-οστού πελάτη,
- W_n το χρόνο αναμονής στην ουρά (μέχρι να αρχίσει η εξυπηρέτηση) του n-οστού πελάτη,
- X_n το χρόνο εξυπηρέτησης του n-οστού πελάτη.

Προφανώς έχουμε

$$S_n = W_n + X_n.$$

Ενδιαφερόμαστε και πάλι για το τί συμβαίνει σε κατάσταση ισορροπίας και γι αυτό ενδιαφερόμαστε για την οριακή κατανομή του χρόνου παραμονής. Και πάλι υπάρχουν τρεις ερμηνείες για την κατανομή αυτή που συμπίπτουν όταν η στοχαστική διαδικασία που περιγράφει την εξέλιξη του συστήματος στο χρόνο έχει αναγεννητικό χαρακτήρα. Έτσι για την οριακή κατανομή $F_S(x)$ του χρόνου παραμονής πελάτη έχουμε

$$
\begin{aligned}
F_S(x) &= \lim_{n\to\infty} \Pr[S_n \le x] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr[S_k \le x] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\{S_k \le x\},\ x \ge 0,
\end{aligned}
$$

όπου με $1\{S_k \le x\}$ συμβολίζουμε τη δείκτρια συνάρτηση του ενδεχομένου $\{S_k \le x\}$ (ο k-οστός πελάτης να παραμείνει στο σύστημα το πολύ για χρόνο x) που παίρνει την τιμή 1, όταν $S_k \le x$, και την τιμή 0, διαφορετικά.

Συμβολίζοντας με S την οριακή τυχαία μεταβλητή που έχει την κατανομή $(F_S(x) : x \ge 0)$ για τον οριακό μέσο χρόνο παραμονής έχουμε

$$
E[S] = \lim_{n\to\infty} E[S_n] = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} S_k.
$$

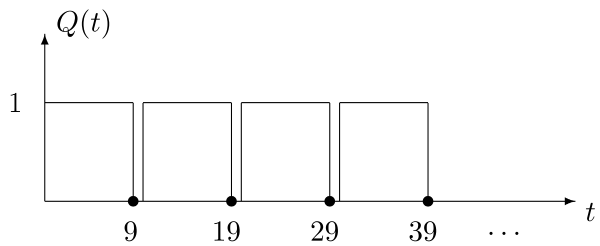
Ο υπολογισμός του $E[S]$ είναι εύκολος αν έχουμε υπολογίσει την κατανομή $(F_S(x) : x \ge 0)$ ή την αντίστοιχη συνάρτηση πυκνότητας πιθανότητας $(f_S(x) : x \ge 0)$. Πράγματι χρειάζεται να χρησιμοποιήσουμε τη σχέση $E[S] = \int_0^\infty x f_S(x) dx = \int_0^\infty (1 - F_S(x)) dx$ και επομένως πρόκειται για έναν υπολογισμό ρουτίνας (ο οποίος μπορεί πάντως να έχει αρκετές πράξεις). Σε κάποιες περιπτώσεις όμως, ο $E[S]$ μπορεί να υπολογιστεί και απευθείας ταυτόχρονα με την $E[Q]$ με πιθανοθεωρητικά επιχειρήματα. Ομοίως με το $E[S]$ ορίζονται τα $E[W]$ και $E[X]$.

Ένα άλλο σημαντικό μέγεθος που αξίζει μελέτης είναι ο κύκλος λειτουργίας (busy cycle) του συστήματος, που ορίζεται ως το διάστημα από την αναχώρηση ενός πελάτη που αφήνει το σύστημα κενό, μέχρι την επόμενη αναχώρηση πελάτη που θα αφήσει το σύστημα κενό. Κάθε τέτοιος κύκλος αρχίζει με ένα χρονικό διάστημα που το σύστημα παραμένει κενό μέχρι να εμφανιστεί ο πρώτος πελάτης. Το διάστημα αυτό αναφέρεται ως μια περίοδος αργίας (idle period) του συστήματος. Από τη στιγμή που θα αφιχθεί ο πρώτος πελάτης το σύστημα θα είναι συνεχώς απασχολημένο μέχρι να τελειώσει ο κύκλος απασχόλησης. Το διάστημα αυτό αναφέρεται ως μια περίοδος συνεχούς λειτουργίας του συστήματος (busy period). Οι διάρκειες των περιόδων αργίας I, των περιόδων συνεχούς λειτουργίας Y και των κύκλων απασχόλησης Z ενδιαφέρουν κυρίως από την οπτική σκοπιά των υπηρετών και του διαχειριστή του συστήματος αφού οι ποσότητες αυτές είναι σημαντικές για να ληφθούν αποφάσεις σχετικά με τη συντήρηση του συστήματος ή με διακοπές -διαλείματα των υπηρετών. Ισχύει ότι $Z = I + Y$. Για το λόγο αυτό ενδιαφερόμαστε για τη μελέτη των αντίστοιχων οριακών κατανομών $F_I(x)$, $F_Y(x)$ και $F_Z(x)$, ή τουλάχιστον των αντίστοιχων μέσων τιμών τους $E[I]$, $E[Y]$ και $E[Z]$.

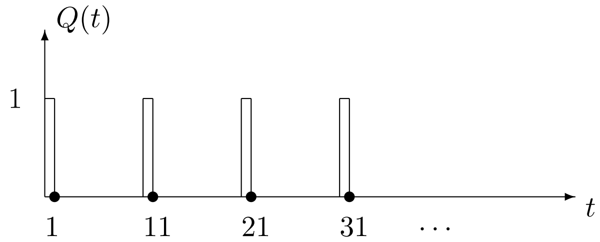
1.3 Εμφυτευμένες διαδικασίες σε στιγμές αφίξεων και αναχωρήσεων

Όπως είπαμε παραπάνω, ο διαχειριστής του συστήματος, που μπορεί να θεωρηθεί ως ένας εξωτερικός παρατηρητής του συστήματος, έχει μια διαφορετική αντίληψη από τους πελάτες του συστήματος όσον αφορά το συνωστισμό. Για το λόγο αυτό ορίσαμε και τα διάφορα μέτρα απόδοσης. Για να γίνει περισσότερο κατανοητή η διαφορά των δυο οπτικών, διαχειριστή-εξωτερικού παρατηρητή και πελατών ας φανταστούμε ότι έχουμε ένα $D/D/1$ σύστημα και ας εξετάσουμε τί αντιλαμβάνεται ο διαχειριστής και τί οι πελάτες.

Σε ένα πρώτο σενάριο, ας υποθέσουμε ότι έχουμε αφίξεις σε σταθερά χρονικά διαστήματα, κάθε 10 λεπτά και ότι οι χρόνοι εξυπητέτησης είναι επίσης σταθεροί και ίσοι με 9 λεπτά. Τότε ο διαχειριστής βλέπει το σύστημα πολύ απασχολημένο, για την ακρίβεια βλέπει ότι το 90% του χρόνου στο σύστημα υπάρχει 1 πελάτης ενώ μόνο ένα 10% του χρόνου το σύστημα είναι άδειο. Από την άλλη μεριά κάθε πελάτης βλέπει το σύστημα άδειο τη στιγμή που φθάνει σε αυτό (αφού ο προηγούμενος πελάτης έχει φύγει πριν 1 λεπτό).



Σε ένα δεύτερο σενάριο, ας υποθέσουμε ότι οι χρόνοι μεταξύ των αφίξεων είναι και πάλι σταθεροί και ίσοι με 10 λεπτά, αλλά τώρα θεωρούμε ότι οι χρόνοι εξυπηρέτησης διαρκούν 1 λεπτό για κάθε πελάτη. Τώρα ο διαχειριστής βλέπει το σύστημα πολύ λίγο απασχολημένο, αφού το 10% μόλις του χρόνου υπάρχει 1 πελάτης ενώ το 90% του χρόνου το σύστημα είναι άδειο. Από την άλλη μεριά η εντύπωση που αποκομίζει κάθε πελάτης φθάνοντας στο σύστημα δεν διαφέρει από την εντύπωση που έχουν οι πελάτες στο πρώτο σενάριο: Και πάλι κάθε πελάτης βλέπει το σύστημα άδειο τη στιγμή που εισέρχεται σε αυτό.



Το συμπέρασμα είναι ότι οι οπτικές ενός εξωτερικού παρατηρητή και ενός αφικνούμενου πελάτη μπορεί να δίνουν πολύ διαφορετικές εικόνες για το ίδιο σύστημα. Επίσης δυο συστήματα μπορεί να μοιάζουν παρόμοια υπό τη μία οπτική (όπως τα δυο σενάρια υπό την οπτική των πελατών) και να είναι πολύ διαφορετικά υπό την άλλη οπτική (όπως τα δυο σενάρια υπό την οπτική του διαχειριστή). Αν σκεφτούμε ότι οι πελάτες δεν είναι παθητικές οντότητες αλλά μπορεί να αποφασίζουν τι θα κάνουν σε σχέση με το σύστημα (π.χ. να μπουν σε αυτό ή να φύγουν) έχει μεγάλη σημασία να ποσοτικοποιήσουμε με κάποιο τρόπο το πώς αντιλαμβάνονται το συνωστισμό του συστήματος οι πελάτες κατά την άφιξή τους ή την αναχώρησή τους.

Προς το σκοπό αυτό έστω $A_1 < A_2 < A_3 < \ldots$ οι διαδοχικές στιγμές αφίξεων και $D_1 < D_2 < D_3 < \ldots$ οι διαδοχικές στιγμές αναχωρήσεων των πελατών. Ξεκινώντας από τη στοχαστική διαδικασία $\{Q(t)\}$ που περιγράφει τον αριθμό των πελατών στο σύστημα σε συνεχή χρόνο, ορίζουμε τις εμφυτευμένες διαδικασίες $\{Q_n^-\}$ και $\{Q_n^+\}$ σε στιγμές αφίξεων και αναχωρήσεων πελατών αντίστοιχα. Έτσι ορίζουμε

- $Q_n^- = Q(A_n^-)$ τον αριθμό των πελατών αμέσως πριν την n-οστή άφιξη πελάτη (δηλαδή των αριθμό των παρόντων πελατών που βλέπει ο n-οστός πελάτης καθώς εισέρχεται στο σύστημα),
- $Q_n^+ = Q(D_n^+)$ τον αριθμό των πελατών αμέσως μετά την n-οστή αναχώρηση πελάτη (δηλαδή των αριθμό των πελατών που αφήνει πίσω του κατά την έξοδο του ο n-οστός πελάτης που φεύγει από το σύστημα).

Ενδιαφερόμαστε για τις αντίστοιχες οριακές κατανομές που περιγράφουν την εντύπωση που διαμορφώνει ένας πελάτης τη στιγμή που εισέρχεται στο σύστημα και τη στιγμή που αναχωρεί από αυτό, εφόσον το σύστημα βρίσκεται σε κατάσταση ισορροπίας. Μιλάμε, αντίστοιχα για τις οριακές κατανομές του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων. Η οριακή πιθανότητα σε στιγμή άφιξης να βρίσκονται j πελάτες στο σύστημα (εξαιρουμένης της νέας άφιξης) είναι

$$
\begin{aligned}
a_j &= \lim_{n\to\infty} \Pr[Q_n^- = j] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr[Q_k^- = j] \\
&= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\{Q_k^- = j\},\ j = 0, 1, \ldots.
\end{aligned}
$$

Ομοίως η οριακή πιθανότητα σε στιγμή αναχώρησης να βρίσκονται j πελάτες στο σύστημα (εξαιρουμένης της νέας αναχώρησης) είναι

$$\begin{aligned} d_j &= \lim_{n\to\infty} \Pr[Q_n^+ = j] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \Pr[Q_k^+ = j] \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} 1\{Q_k^+ = j\},\ j = 0, 1, \ldots . \end{aligned}$$

Οι a_j και d_j είναι οι οριακές πιθανότητες να βρίσκονται j πελάτες στο σύστημα σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα, ενώ η p_j είναι η οριακή πιθανότητα σε συνεχή χρόνο. Όπως και η p_j, έτσι και οι a_j και d_j έχουν ερμηνείες ως ποσοστά. Συγκεκριμένα η a_j εκφράζει το μακροπρόθεσμο ποσοστό των αφίξεων που βρίσκουν j πελάτες στο σύστημα, ενώ η d_j εκφράζει το μακροπρόθεσμο ποσοστό των αναχωρήσεων που αφήνουν j πελάτες στο σύστημα.

Δεν υπάρχει κάποιος λόγος οι οριακές κατανομές (p_j), (a_j) και (d_j) να συμπίπτουν και γενικά αυτό δεν ισχύει. Πράγματι, παρατηρείστε ότι στα δυο σενάρια για την $D/D/1$ ουρά που περιγράψαμε παραπάνω είχαμε $a_0 = 1$ και $a_j = 0,\ j \geq 1$, ενώ στο πρώτο σενάριο είχαμε $p_0 = 0.1$, $p_1 = 0.9$, $p_j = 0,\ j \geq 2$ και στο δεύτερο είχαμε $p_0 = 0.9$, $p_1 = 0.1$, $p_j = 0,\ j \geq 2$.

2

Βασικά αποτελέσματα

Στο κεφάλαιο αυτό διατυπώνουμε τέσσερα βασικά αποτελέσματα που αφορούν τα συστήματα εξυπηρέτησης, τα οποία ισχύουν κάτω από πολύ γενικές συνθήκες. Συγκεκριμένα, χαρακτηρίζουμε την ευστάθεια μέσω του ρυθμού συνωστισμού ενός συστήματος και διατυπώνουμε την ιδιότητα των μεμονωμένων αφίξεων, την ιδιότητα PASTA (Poisson-Arrivals-See-Time-Averages και το νόμο του Little.

2.1 Ρυθμός συνωστισμού - Ευστάθεια

Το μέσο ποσό εργασίας που εισέρχεται προς διεκπεραίωση σε ένα σύστημα εξυπηρέτησης ανά χρονική μονάδα αναφέρεται ως ο ρυθμός συνωστισμού ρ και ισούται με το γινόμενο του ρυθμού αφίξεων λ επί το μέσο χρόνο εξυπηρέτησης b:

$$\rho = \lambda b.$$

Το ποσό της εργασίας που μπορεί να διεκπεραιώσει το σύστημα ανά χρονική μονάδα είναι ίσο με το πλήθος των υπηρετών c αφού κάθε υπηρέτης μπορεί να διεκπεραιώσει μια μονάδα εργασίας ανά χρονική μονάδα. Έτσι για να είναι το σύστημα ευσταθές και να να μην απειρίζεται η ουρά αναμένουμε διαισθητικά ότι θα πρέπει το μέσο ποσό εργασίας που εισέρχεται ανά χρονική μονάδα να είναι μικρότερο από τη μέγιστη δυνατότητα διεκπεραίωσης του συστήματος ανά χρονική μονάδα.

Θεώρημα 2.1 (Χαρακτηρισμός ευστάθειας) *Στο $GI/GI/c$ σύστημα με συνεχή κατανομή για τους ενδιάμεσους χρόνους μεταξύ των αφίξεων και/ή τους χρόνους εξυπηρέτησης ισχύει ένα ακριβώς από τα παρακάτω:*

1. *$\rho < c$ οπότε το σύστημα είναι ευσταθές, δηλαδή υπάρχουν οι κατανομές (p_j), (a_j) και (d_j) και είναι $p_j > 0, \ \ j \geq 0$ και $\sum_{j=0}^{\infty} p_j = 1$ (και όμοια για τις (a_j) και (d_j)).*
2. *$\rho \geq c$ οπότε το σύστημα είναι ασταθές, δηλαδή το πλήθος των πελατών απειρίζεται καθώς $t \to \infty$ και $p_j = a_j = d_j = 0, \ \ j \geq 0$.*

Η απόδειξη του αποτελέσματος αυτού είναι ιδιαίτερα περίπλοκη. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Baccelli and Bremaud (1994) όπου αποδεικνύονται θεωρήματα ευστάθειας και άλλα θεωρητικά αποτελέσματα κάτω από γενικές συνθήκες.

2.2 Ιδιότητα μεμονωμένων μεταβάσεων και ιδιότητα PASTA

Όπως είπαμε οι οριακές κατανομές (p_j), (a_j) και (d_j) γενικά δεν συμπίπτουν. Υπάρχουν όμως δυο περιπτώσεις στις οποίες κάποιες από αυτές συμπίπτουν.

Θεώρημα 2.2 (Ιδιότητα μεμονωμένων μεταβάσεων) *Σε συστήματα εξυπηρέτησης στα οποία οι πελάτες έρχονται και αναχωρούν μεμονωμένα, δηλαδή δεν υπάρχουν ομαδικές αφίξεις ούτε ομαδικές αναχωρήσεις οι οριακές κατανομές σε στιγμές αφίξεων και σε στιγμές αναχωρήσεων συμπίπτουν:* $(a_j) = (d_j)$.

Η ιδιότητα των μεμονωμένων μεταβάσεων μπορεί να αιτιολογηθεί ως εξής: Έστω $A_j(t)$ το πλήθος των αφίξεων στο $[0,t]$ που βρίσκουν j πελάτες και $A(t)$ το συνολικό πλήθος αφίξεων στο $[0,t]$. Τότε, έχουμε $a_j = \lim_{t\to\infty} \frac{A_j(t)}{A(t)}$. Ομοίως, $d_j = \lim_{t\to\infty} \frac{D_j(t)}{D(t)}$, όπου $D_j(t)$ το πλήθος των αναχωρήσεων στο $[0,t]$ που αφήνουν πίσω τους j πελάτες και $D(t)$ το συνολικό πλήθος αναχωρήσεων στο $[0,t]$. Οι μακροπρόθεσμοι ρυθμοί αφίξεων και αναχωρήσεων είναι $\lim_{t\to\infty} \frac{A(t)}{t}$ και $\lim_{t\to\infty} \frac{D(t)}{t}$.

Έχουμε ότι

$$\lim_{t\to\infty} \frac{A(t)}{t} = \lim_{t\to\infty} \frac{D(t)}{t},$$

αφού οι μακροπρόθεσμοι ρυθμοί αφίξεων και αναχωρήσεων θα είναι ίσοι υπό την υπόθεση της ευστάθειας (ο ρυθμός των αφίξεων είναι προφανώς μεγαλύτερος ή ίσος του ρυθμού των αναχωρήσεων, αλλά δεν μπορεί να είναι μεγαλύτερος αν το σύστημα είναι ευσταθές). Επίσης έχουμε ότι $|A_j(t) - D_j(t)| \leq 1$, αφού μεταξύ δυο διαδοχικών αφίξεων που βρίσκουν j πελάτες πρέπει να υπάρχει μια ακριβώς αναχώρηση που αφήνει j πελάτες στο σύστημα και μεταξύ δυο διαδοχικών αναχωρήσεων που αφήνουν j πελάτες πρέπει να υπάρχει μια ακριβώς άφιξη που βρίσκει j πελάτες. Οπότε, έχουμε ότι:

$$\lim_{t\to\infty} \frac{A_j(t)}{t} = \lim_{t\to\infty} \frac{D_j(t)}{t}.$$

Επομένως, είναι

$$a_j = \lim_{t\to\infty} \frac{A_j(t)}{A(t)} = \lim_{t\to\infty} \frac{A_j(t)/t}{A(t)/t} = \lim_{t\to\infty} \frac{D_j(t)/t}{D(t)/t} = \lim_{t\to\infty} \frac{D_j(t)/t}{D(t)/t} = d_j.$$

Θεώρημα 2.3 (Ιδιότητα PASTA (Poisson-Arrivals-See-Time-Averages)) *Σε συστήματα εξυπηρέτησης στα οποία οι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson (δηλαδή οι ενδιάμεσοι χρόνοι μεταξύ των αφίξεων είναι ανεξάρτητοι και ισόνομοι με εκθετική κατανομή) οι οριακές κατανομές σε στιγμές αφίξεων και σε συνεχή χρόνο συμπίπτουν:* $(a_j) = (p_j)$.

Η διαισθητική αιτιολόγηση της ιδιότητα PASTA είναι ότι η διαδικασία Poisson μοντελοποιεί την ιδέα των εντελώς τυχαίων αφίξεων στο χρόνο και επομένως η παρατήρηση του αριθμού των πελατών κατά τη στιγμή της άφιξη ενός πελάτη στο σύστημα ισοδυναμεί με την παρατήρηση του αριθμού των πελατών σε μια τυχαία χρονική στιγμή (σε συνεχή χρόνο). Η αιτιολόγηση αυτή μπορεί να γίνει περισσότερο κατανοητή αν σκεφθούμε ως εξής: Έστω $A(t, t+h)$ το πλήθος των αφίξεων σε ένα σύστημα εξυπηρέτησης στο διάστημα $(t, t+h]$. Τότε, έχουμε

$$\begin{aligned} a_j &= \lim_{t\to\infty} \lim_{h\to 0^+} \Pr[Q(t) = j | A(t, t+h) > 0] \\ &= \lim_{t\to\infty} \lim_{h\to 0^+} \frac{\Pr[Q(t) = j, A(t, t+h) > 0]}{\Pr[A(t, t+h) > 0]} \\ &= \lim_{t\to\infty} \lim_{h\to 0^+} \frac{\Pr[Q(t) = j] \Pr[A(t, t+h) > 0 | Q(t) = j]}{\Pr[A(t, t+h) > 0]}. \end{aligned}$$

Όμως το ενδεχόμενο $\{A(t, t+h) > 0\}$ είναι ανεξάρτητο του $\{Q(t) = j\}$ λόγω του ότι η διαδικασία αφίξεων είναι Poisson, της οποίας η (μελλοντική) εξέλιξη μετά τη χρονική στιγμή t δεν εξαρτάται από την κατάσταση του συστήματος τη χρονική στιγμή t. Επομένως $\Pr[A(t, t+h) > 0 | Q(t) = j] = \Pr[A(t, t+h) > 0]$ και άρα

$$a_j = \lim_{t\to\infty} \Pr[Q(t) = j] = p_j.$$

Η αυστηρή απόδειξη των δυο παραπάνω αποτελεσμάτων κάτω από γενικές συνθήκες είναι αρκετά απαιτητική. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στο βιβλίο των Baccelli and Bremaud (1994), όπου υπάρχουν και άλλα συναφή αποτελέσματα. Στο βιβλίο του Φακίνου (2003) υπάρχει μια περιγραφή αυτών των αποδεικτικών ιδεών.

Φυσικά μπορούμε να συνδυάσουμε τα δυο αυτά αποτελέσματα και τότε έχουμε ότι σε συστήματα που οι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson και έχουμε μεμονωμένες μεταβάσεις (αφίξεις, αναχωρήσεις) όλες οι οριακές κατανομές συμπίπτουν: $(p_j) = (a_j) = (d_j)$.

2.3 Ο νόμος του Little

Ο νόμος του Little είναι ένα πολύ γενικό αποτέλεσμα που συνδέει το μέσο πλήθος πελατών στο σύστημα $E[Q]$, τον ρυθμό αφίξεων λ και τον μέσο χρόνο παραμονής ενός πελάτη $E[S]$ σε αυτό. Συγκεκριμένα έχουμε

Θεώρημα 2.4 (Νόμος του Little) *Έστω ένα σύστημα εξυπηρέτησης με μέσο πλήθος πελατών $E[Q]$, ρυθμό αφίξεων λ και μέσο χρόνο παραμονής πελάτη $E[S]$. Τότε*

$$E[Q] = \lambda E[S].$$

Διαισθητικά, το αποτέλεσμα του Little μπορεί να γίνει κατανοητό θεωρώντας ότι κάθε πελάτης πληρώνει 1 χρηματική μονάδα ανά χρονική μονάδα παραμονής του στο σύστημα. Τότε ο διαχειριστής του συστήματος λαμβάνει $E[Q]$ χρηματικές μονάδες στη μονάδα του χρόνου, αν υποθέσουμε ότι η πληρωμή γίνεται κατά τρόπο ¨συνεχή¨. Από την άλλη μεριά η μέση είσπραξη του διαχειριστή στη μονάδα του χρόνου θα πρέπει να είναι η ίδια αν οι πελάτες πληρώνουν ¨προκαταβολικά¨, δηλαδή αν με την είσοδό τους στο σύστημα δίνουν όλο το ποσό για την παραμονή τους. Αλλά τότε θα έχουμε κατά μέσο όρο λ πελάτες ανά χρονική μονάδα και ο καθένας θα πληρώνει $E[S]$ χρηματικές μονάδες, οπότε η συνολική είσπραξη του διαχειριστή θα είναι $\lambda E[S]$

χρηματικές μονάδες στη μονάδα του χρόνου. Αφού τα δυο ποσά πρέπει να είναι ίσα (μακροπρόθεσμο μέσο έσοδο με συνεχή είσπραξη = μακροπρόθεσμο μέσο έσοδο με προκαταβολική είσπραξη) έχουμε τη σχέση $E[Q] = \lambda E[S]$.

Ένας άλλος τρόπος να αποτυπώσουμε πιό μαθηματικά την παραπάνω ιδέα είναι ο εξής: Έστω $Q(t)$ το πλήθος των πελατών τη στιγμή t, $A(t)$ το πλήθος των αφίξεων ως τη στιγμή t και $S_1, S_2, \ldots$ οι διαδοχικοί χρόνοι παραμονής των πελατών. Τότε τα μέτρα απόδοσης που εμφανίζονται στο Νόμο του Little έχουν τις εξής εκφράσεις:

$$E[Q] = \lim_{t\to\infty} \frac{\int_0^t Q(u)du}{t}, \ \lambda = \lim_{t\to\infty} \frac{A(t)}{t}, \ E[S] = \lim_{n\to\infty} \frac{\sum_{k=1}^n S_k}{n}.$$

Έστω $T_1, T_2, \ldots$ οι χρόνοι κατά τους οποίους αδειάζει ένα σύστημα εξυπηρέτησης (επομένως είναι οι στιγμές που αρχίζουν οι διαδοχικοί κύκλοι λειτουργίας του συστήματος). Τότε έχουμε ότι

$$\int_0^{T_n} Q(u)du = \sum_{k=1}^{A(T_n)} S_k.$$

Πράγματι, κάθε χρονική μονάδα παραμονής κάθε πελάτη που αφίχθει στο διάστημα $(0, T_n]$, συνεισφέρει μια μονάδα στο αριστερό μέλος, επομένως το ολοκλήρωμα του αριστερού μέλους ισούται με το συνολικό χρόνο παραμονής όλων των πελατών που αφίχθηκαν στο διάστημα $(0, T_n]$, που είναι ακριβώς το δεξιό μέλος. Συνδυάζοντας τα παραπάνω έχουμε:

$$\begin{aligned} E[Q] &= \lim_{t\to\infty} \frac{\int_0^t Q(u)du}{t} = \lim_{n\to\infty} \frac{\int_0^{T_n} Q(u)du}{T_n} \\ &= \lim_{n\to\infty} \frac{\sum_{k=1}^{A(T_n)} S_k}{T_n} = \lim_{n\to\infty} \frac{A(T_n)}{T_n} \cdot \lim_{n\to\infty} \frac{\sum_{k=1}^{A(T_n)} S_k}{A(T_n)} \\ &= \lim_{t\to\infty} \frac{A(t)}{t} \cdot \lim_{n\to\infty} \frac{\sum_{k=1}^n S_k}{n} = \lambda E[S]. \end{aligned}$$

Αυστηρές αποδείξεις του θεωρήματος Little έχουν γίνει κάτω από πολύ γενικές συνθήκες. Ο ενδιαφερόμενος αναγνώστης παραπέμπεται στις εργασίες των Little (1961) και Stidham (1974). Στο βιβλίο του Φακίνου (2003) υπάρχει ένα σκαρίφημα αυτών των αποδεικτικών ιδεών.

Το αποτέλεσμα του Little μπορεί να εφαρμοστεί και σε υποσυστήματα ενός συστήματος, δίνοντας ενδιαφέροντα αποτελέσματα. Στην περίπτωση αυτή, το $E[Q]$ θα αναφέρεται στο μέσο πλήθος πελατών στο συγκεκριμένο υποσύστημα, το λ στο ρυθμό άφιξης στο συγκεκριμένο υποσύστημα και το $E[S]$ στο χρόνο παραμονής ενός πελάτη στο συγκεκριμένο υποσύστημα.

Θεωρώντας ως υποσύστημα το χώρο αναμονής ενός συστήματος (δηλαδή την ουρά) παίρνουμε τη σχέση

$$E[Q_q] = \lambda E[W],$$

δηλαδή ο μέσος αριθμός πελατών στην ουρά ισούται με το ρυθμό αφίξεων επί το μέσο χρόνο αναμονής ενός πελάτη μέχρι να αρχίσει η εξυπηρέτησή του.

Θεωρώντας ως υποσύστημα το χώρο εξυπηρέτησης ενός συστήματος παίρνουμε τη σχέση

$$E[Q_s] = \lambda E[X] = \lambda b = \rho,$$

δηλαδή ο μέσος αριθμός πελατών στο χώρο εξυπηρέτησης, που προφανώς ταυτίζεται με τον μέσο αριθμό απασχολημένων υπηρετών, ισούται με το ρυθμό συνωστισμού του συστήματος. Έτσι έχουμε μια δεύτερη ερμηνεία του ρυθμού συνωστισμού. Όχι μόνο είναι το μέσο ποσό εργασίας που εισέρχεται στο σύστημα ανά χρονική μονάδα αλλά επιπλέον εκφράζει και το μέσο αριθμό απασχολημένων υπηρετών μια τυχούσα χρονική στιγμή. Επειδή ο μέσος αριθμός απασχολημένων υπηρετών ισούται με το πλήθος c των υπηρετών επί την πιθανότητα ένας υπηρέτης να είναι απασχολημένος συμπεραίνουμε ότι η οριακή πιθανότητα απασχολημένου υπηρέτη ή, ισοδύναμα, το ποσοστό του χρόνου απασχόλησης ενός υπηρέτη υπηρέτη είναι $\frac{\rho}{c}$.

Ειδικά για την $GI/G/1$ ουρά έχουμε

$$\rho = E[Q_s] = 0 \Pr[Q_s = 0] + 1 \Pr[Q_s = 1] = \Pr[Q_s = 1] = \Pr[Q \geq 1] = 1 - p_0,$$

επομένως στην περίπτωση αυτή έχουμε ότι η οριακή πιθανότητα κενού συστήματος είναι

$$p_0 = 1 - \rho.$$

Μέρος II

Αποτίμηση απόδοσης

3

Η ανάλυση μέσης τιμής

Η ιδιότητα PASTA σε συνδυασμό με τον νόμο του Little μπορούν να χρησιμοποιηθούν για να υπολογίσουμε με πιθανοθεωρητικούς συλλογισμούς και ελάχιστους υπολογισμούς τα μέτρα απόδοσης $E[Q]$ και $E[S]$ για αρκετά συστήματα, χωρίς να χρειαστεί να υπολογίσουμε τις αντίστοιχες κατανομές. Η ανάλυση αυτή αναφέρεται συχνά ως ανάλυση μέσης τιμής (Mean Value Analysis - MVA) και είναι ένα ιδιαίτερα ισχυρό εργαλείο. Θα την παρουσιάσουμε αναλύοντας μια σειρά από συγκεκριμένα συστήματα με τη μέθοδο αυτή.

3.1 Ανάλυση μέσης τιμής στην $M/M/1/1$ ουρά

Θεωρούμε μια $M/M/1/1$ ουρά με ρυθμό αφίξεων λ και εκθετική κατανομή χρόνων εξυπηρέτησης με παράμετρο μ. Επιθυμούμε να υπολογίσουμε τις $E[Q]$ και $E[S]$.

Από το νόμο του Little έχουμε

$$E[Q] = \lambda E[S].$$

Για το μέσο χρόνο παραμονής ενός πελάτη δεσμεύοντας στον αριθμό των πελατών Q^- που βλέπει ένας πελάτης κατά την άφιξή του, έχουμε:

$$E[S] = a_0 \cdot \frac{1}{\mu} + a_1 \cdot 0 = \frac{a_0}{\mu} = \frac{p_0}{\mu},$$

όπου η τελευταία ισότητα ισχύει λόγω της PASTA. Συνδυάζοντας αυτές τις σχέσεις και λαμβάνοντας υπόψη ότι $E[Q] = p_1$, έχουμε:

$$p_1 = \lambda \frac{p_0}{\mu} = \rho p_0,$$

όπου $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού. Επίσης το σύστημα είναι ευσταθές αφού έχει πεπερασμένη χωρητικότητα, και άρα

$$p_0 + p_1 = 1$$

. Λύνοντας το σύστημα παίρνουμε

$$p_0 = \frac{1}{1+\rho},\ p_1 = \frac{\rho}{1+\rho},$$

οπότε

$$E[Q] = \frac{\rho}{1+\rho},\ E[S] = \frac{1}{\mu(1+\rho)}.$$

3.2 Ανάλυση μέσης τιμής στην $M/M/1$ ουρά

Θεωρούμε μια $M/M/1$ ουρά με ρυθμό αφίξεων λ και εκθετική κατανομή χρόνων εξυπηρέτησης με παράμετρο μ. Επιθυμούμε να υπολογίσουμε τις $E[Q]$ και $E[S]$.

Από το νόμο του Little έχουμε:

$$E[Q] = \lambda E[S].$$

Για το μέσο χρόνο παραμονής ενός πελάτη, δεσμεύοντας στον αριθμό των πελατών Q^- που βλέπει ένας πελάτης κατά την άφιξή του, έχουμε:

$$E[S] = \frac{E[Q^-]+1}{\mu} = \frac{E[Q]+1}{\mu}.$$

Πράγματι, ένας πελάτης που βρίσκει κατά την άφιξή του j πελάτες στο σύστημα θα παραμείνει σε αυτό για $j+1$ εκθετικούς χρόνους εξυπηρέτησης με παράμετρο μ. Ο πρώτος χρόνος αντιστοιχεί στον υπολειπόμενο χρόνο εξυπηρέτησης του πελάτη που εξυπηρετείται. Λόγω της αμνήμονης ιδιότητας της εκθετικής κατανομής είναι και αυτός εκθετικός με παράμετρο μ. Οι άλλοι χρόνοι είναι οι χρόνοι εξυπηρέτησης των πελατών που βρίσκονται στο χώρο αναμονής και του αφικνούμενου πελάτη. Οπότε προκύπτει η πρώτη ισότητα. Η δεύτερη ισότητα είναι άμεση από την ιδιότητα PASTA.

Λύνοντας το σύστημα των $E[S]$ και $E[Q]$ προκύπτει ότι:

$$E[Q] = \frac{\rho}{1-\rho},\ E[S] = \frac{1}{\mu(1-\rho)},$$

όπου $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού.

3.3 Ασκήσεις

1. Με χρήση του νόμου του Little και της ιδιότητας PASTA, να βρείτε τα μακροπρόθεσμα ποσοστά του χρόνου κενού και απασχολημένου υπηρέτη, δηλαδή τις πιθανότητες p_0 και p_1, σε μια $M/G/1/1$ ουρά με ρυθμό αφίξεων λ και μέσο χρόνο εξυπηρέτησης b.

2. Βρείτε τον οριακό μέσο αριθμό πελατών στο σύστημα, $E[Q]$ στην $GI/G/1$ ουρά με μέσο ενδιάμεσο χρόνο αφίξεων a και μέσο χρόνο εξυπηρέτησης b.

3. Θεωρήστε μια $M/M/c$ ουρά με ρυθμό αφίξεων 5 πελάτες την ώρα και μέσο χρόνο εξυπηρέτησης ανά πελάτη 78 λεπτά.

   1. Ποιός είναι ο ελάχιστος αριθμός υπηρετών c που χρειάζεται για να είναι το σύστημα ευσταθές (δηλαδή να μην απειρίζεται η ουρά);
   2. Ποιός είναι ο ελάχιστος αριθμός υπηρετών που χρειάζεται αν η εργατική νομοθεσία επιβάλλει κάθε υπηρέτης να είναι απασχολημένος το πολύ το 80% του χρόνου του;

4. Θεωρούμε ένα σύστημα εξυπηρέτησης με έναν υπηρέτη, στο οποίο καταφθάνουν πελάτες δύο τύπων 1 και 2, σύμφωνα με δυο ανεξάρτητες διαδικασίες Poisson με ρυθμούς λ_1 και λ_2, αντίστοιχα. Κάθε πελάτης, ανεξαρτήτως τύπου έχει $Exp(\mu)$ χρόνο εξυπηρέτησης. Οι πελάτες τύπου 1 έχουν απόλυτη προτεραιότητα έναντι των πελατών τύπου 2, δηλαδή όταν υπάρχουν πελάτες τύπου 1 στο σύστημα ο υπηρέτης εξυπηρετεί αυτούς και αρχίζει να εξυπηρετεί πελάτες τύπου 2 μόνο όταν δεν υπάρχουν πελάτες τύπου 1. Επιπλέον, αν ένας πελάτης τύπου 2 εξυπηρετείται και αφιχθεί πελάτης τύπου 1, ο υπηρέτης διακόπτει την εξυπηρέτηση και πηγαίνει να εξυπηρετήσει τον νεοαφιχθέντα πελάτη τύπου 1. Να βρεθούν οι μέσοι οριακοί αριθμοί πελατών τύπων 1 και 2, $E[Q_1]$ και $E[Q_2]$, αντίστοιχα.

5. Να βρείτε τις οριακές κατανομές (p_n), (a_n) και (d_n) των αριθμών των πελατών σε συνεχή χρόνο, σε στιγμές αφίξεων και σε στιγμές αναχωρήσεων, αντίστοιχα, σε μια ευσταθή $M/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ, χρησιμοποιώντας το νόμο του Little και την ιδιότητα PASTA. Για το σκοπό αυτό θεωρήστε ως 'σύστημα' την i θέση του συστήματος εξυπηρέτησης για $i = 1, 2, \ldots$.

4

Μαρκοβιανές αλυσίδες συνεχούς χρόνου

Οι Μαρκοβιανές αλυσίδες συνεχούς χρόνου είναι μια κλάση στοχαστικών διαδικασιών με μεγάλες δυνατότητες προτυποποίησης στα πλαίσια της θεωρίας των ουρών αναμονής. Πραγματικά τα πλέον κλασικά μοντέλα ουρών προτυποποιούνται από τέτοιες διαδικασίες. Σε αυτό το κεφάλαιο γίνεται μια σύντομη ανασκόπηση της βασικής θεωρίας τους.

4.1 Βασικοί ορισμοί

Μια στοχαστική διαδικασία $\{X(t) : t \geq 0\}$ με διακριτό χώρο καταστάσεων $\mathcal{S}$ (το πολύ αριθμήσιμο) λέγεται Μαρκοβιανή αλυσίδα συνεχούς χρόνου (Μ.α.ς.χ.) αν έχει τη λεγόμενη Μαρκοβιανή ιδιότητα, δηλαδή αν για κάθε $s, t > 0$ και $i, j \in \mathcal{S}$ ισχύει

$$\Pr[X(t+s) = j | X(u), 0 \leq u < s, X(s) = i] = \Pr[X(t+s) = j | X(s) = i],$$

δηλαδή αν δεδομένης της ακριβούς πληροφορίας για την παρούσα κατάσταση το παρελθόν και το μέλλον της διαδικασίας είναι ανεξάρτητα.

Η πιθανότητα $p_{ij}(s, s+t) = \Pr[X(t+s) = j | X(s) = i]$ αναφέρεται ως πιθανότητα μετάβασης από την i στη j σε χρόνο t τη στιγμή s. Αν δεν εξαρτάται από το s η Μ.α.ς.χ. αναφέρεται ως (χρονικά) ομογενής. Τότε συμβολίζουμε την $p_{ij}(s, s+t)$ με $p_{ij}(t)$. Από εδώ και στο εξής θα ασχοληθούμε με ομογενείς Μ.α.ς.χ. Επίσης θα θεωρούμε ότι $\mathcal{S} = \mathbb{N}_0$ ή κάποιο γνήσιο υποσύνολό του. Ο πίνακας πιθανοτήτων μετάβασης μιας Μ.α.ς.χ. είναι ο πίνακας

$$\mathbb{P}(t) = (p_{ij}(t)) = \begin{pmatrix} p_{00}(t) & p_{01}(t) & p_{02}(t) & \cdots \\ p_{10}(t) & p_{11}(t) & p_{12}(t) & \cdots \\ p_{20}(t) & p_{21}(t) & p_{22}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Η $\{X(t)\}$ είναι πλήρως ορισμένη, αν δίνονται η αρχική κατανομή της $(p_i(0) : i \in \mathcal{S})$ με

$$p_i(0) = \Pr[X(0) = i],\ i \in \mathcal{S}$$

, και ο πίνακας πιθανοτήτων μετάβασης. Ειδικότερα μπορούμε να υπολογίσουμε τότε την από κοινού πιθανότητα μιας συγκεκριμένης διαδοχής καταστάσεων σε συγκεκριμένες χρονικές στιγμές $0 < t_1 < t_2 < \cdots < t_n$:

$$\Pr[X(0) = i_0, X(t_1) = i_1, X(t_2) = i_2, \ldots, X(t_n) = i_n]$$
$$= p_{i_0}(0) p_{i_0 i_1}(t_1) p_{i_1 i_2}(t_2 - t_1) \cdots p_{i_{n-1} i_n}(t_n).$$

4.2 Ρυθμοί μετάβασης

Για κάθε Μαρκοβιανή αλυσίδα, αποδεικνύεται ότι υπάρχουν τα όρια

$$q_{ij} = \lim_{h \to 0^+} \frac{p_{ij}(h)}{h} = p'_{ij}(0+), i, j \in \mathcal{S} \text{ με } i \neq j,$$
$$q_i = \lim_{h \to 0^+} \frac{1 - p_{ii}(h)}{h} = -p'_{ii}(0+), i, j \in \mathcal{S}$$

που αναφέρονται αντίστοιχα ως ο ρυθμός μετάβασης από την κατάσταση i στην κατάσταση j και ως ο ρυθμός εξόδου από την κατάσταση i. Προφανώς ισχύει

$$q_i = \sum_{j \neq i} q_{ij}, \ i \in \mathcal{S}.$$

Εναλλακτικά μπορούμε να σκεφτόμαστε ότι

$$p_{ij}(h) = q_{ij} h + o(h), \ j \neq i$$
$$p_{ii}(h) = 1 - q_i h + o(h),$$

όπου $o(h)$ τέτοια ώστε $o(h)/h \to 0$, καθώς $h \to 0^+$. Επομένως οι ρυθμοί q_{ij} καθορίζουν την 'τοπική συμπεριφορά' της $\{X(t)\}$, από την οποία μπορούν να συναχθούν οι $p_{ij}(t)$ για οποιοδήποτε t. Επομένως, μια Μ.α.ς.χ. μπορεί να περιγραφεί από την αρχική κατανομή της και τον πίνακα ρυθμών μετάβασης

$$\mathbb{Q} = (q_{ij}) = \begin{pmatrix} -q_0 & q_{01} & q_{02} & \cdots \\ q_{10} & -q_1 & q_{12} & \cdots \\ q_{20} & q_{21} & -q_2 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Ο πίνακας αυτός αναφέρεται και ως απειροστικός γεννήτορας της Μ.α.ς.χ. Τα εκτός διαγωνίου στοιχεία του είναι μη αρνητικά, ενώ τα διαγώνια είναι μη-θετικά (ένα διαγώνιο στοιχείο q_{ii} είναι 0, μόνο αν η κατάσταση i είναι απορροφητική). Επιπλέον τα αθροίσματα των γραμμών του πίνακα είναι 0.

4.3 Χρόνοι παραμονής σε καταστάσεις, πιθανότητες μετάβασης

Όσον αφορά το χρόνο παραμονής T_i σε μια κατάσταση i πριν από κάποια μετάβαση σε κατάσταση $j \neq i$ έχουμε ότι είναι εκθετικός με παράμετρο q_i. Πράγματι έχουμε

$$\Pr[T_i > t + h | T_i > t] = \Pr[X(t+h) = i | X(t) = i] = p_{ii}(h) = 1 - q_i h + o(h),$$

ανεξάρτητος του t, οπότε η T_i έχει την αμνήμονη ιδιότητα και άρα έχει εκθετική κατανομή. Επιπλέον είναι εύκολο να δούμε ότι η παράμετρος της εκθετικής είναι η q_i.

Όταν η $\{X(t)\}$ φύγει από κάποια κατάσταση i, η επόμενη κατάσταση θα είναι η j με πιθανότητα $p_{ij} = \frac{q_{ij}}{q_i}$.

Επομένως, ένας εναλλακτικός τρόπος για να σκεφτόμαστε την εξέλιξη μιας Μ.α.ς.χ. με βάση ρυθμούς είναι ότι όντας σε μια κατάσταση i μένει σε αυτήν για εκθετικό χρονο q_i και κατόπιν πηδάει σε κάποια κατάσταση $j \neq i$ με πιθανότητα $\frac{q_{ij}}{q_i}$.

Γνωρίζουμε ότι αν $T_1, T_2, \ldots, T_k$ είναι ανεξάρτητες εκθετικές τ.μ. με αντίστοιχες παραμέτρους $\lambda_1, \lambda_2, \ldots, \lambda_k$, τότε η $\min(T_1, T_2, \ldots, T_k)$ είναι επίσης εκθετική με παράμετρο $\sum_{i=1}^k \lambda_i$. Επιπλέον, το ενδεχόμενο $\{T_j = \min(T_1, T_2, \ldots, T_k)\}$ είναι ανεξάρτητο της τ.μ. $\min(T_1, T_2, \ldots, T_k)$ και $\Pr[T_j = \min(T_1, T_2, \ldots, T_k)] = \lambda_j / \sum_{i=1}^k \lambda_i$. Με βάση την ιδιότητα αυτή μπορούμε να συναγάγουμε το ακόλουθο κριτήριο για το πότε μια στοχαστική διαδικασία είναι Μ.α.ς.χ.

Θεώρημα 4.1 (Κριτήριο Μαρκοβιανής αλυσίδας συνεχούς χρόνου) *Έστω $\{X(t)\}$ μια στοχαστική διαδικασία με διακριτό χώρο καταστάσεων $\mathcal{S}$ και δοθέντος ότι $X(t) = i$ ισχύει ότι:*

(i) *υπάρχουν χρόνοι $T_{ik} \sim Exp(q_{ik})$, $k \in \mathcal{S} \setminus \{i\}$,*
(ii) *ο χρόνος που θα γίνει η επόμενη μετάβαση είναι $\min_k T_{ik}$ και η κατάσταση j στην οποία πηγαίνει η $\{X(t)\}$ είναι αυτή για την οποία $T_{ij} = \min_k T_{ik}$.*

Τότε η $\{X(t)\}$ είναι Μ.α.ς.χ. με ρυθμούς q_{ij}

Μια Μ.α.ς.χ. περιγράφεται είτε από το διάγραμμα των ρυθμών μετάβασής της, είτε από τον πίνακα των ρυθμών μετάβασής της, $\mathbb{Q} = (q_{ij})$.

Παράδειγμα 4.2 (Η στοχαστική διαδικασία Poisson) Έστω ότι πελάτες έρχονται σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και $X(t)$ είναι το πλήθος των αφίξεων στο $[0, t]$. Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$ και σχετικά με τις μεταβάσεις έχουμε

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$n \geq 0$	$n+1$	$T_{n,n+1} \sim Exp(\lambda)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μ.α.ς.χ.

Παράδειγμα 4.3 (Η $M/M/1/1$ ουρά) Έστω η $M/M/1/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρετήσεων μ, και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t. Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \{0, 1\}$ και σχετικά με τις μεταβάσεις έχουμε

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim Exp(\lambda)$
1	0	$T_{10} \sim Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μ.α.ς.χ.

Παράδειγμα 4.4 (Η $M/M/1$ ουρά) Έστω η $M/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρετήσεων μ, και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t. Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$ και σχετικά με τις μεταβάσεις έχουμε

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim Exp(\lambda)$
$n \geq 1$	$n+1$	$T_{n,n+1} \sim Exp(\lambda)$
	$n-1$	$T_{n,n-1} \sim Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{X(t)\}$ είναι Μ.α.ς.χ.

Παράδειγμα 4.5 (Η $D/M/1$ ουρά) Έστω η $D/M/1$ ουρά με σταθερούς ενδιάμεσους χρόνους αφίξεων a και ρυθμό εξυπηρετήσεων μ, και έστω $X(t)$ το πλήθος των πελατών τη στιγμή t. Τότε για τη στοχαστική διαδικασία $\{X(t)\}$ έχουμε ότι ο χώρος καταστάσεων είναι ο $\mathcal{S} = \mathbb{N}_0$ και σχετικά με τις μεταβάσεις έχουμε

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \not\sim Exp$
$n \geq 1$	$n+1$	$T_{n,n+1} \not\sim Exp$
	$n-1$	$T_{n,n-1} \sim Exp(\mu)$

Αφού έστω κι ένας χρόνος δεν είναι εκθετικός έχουμε ότι η $\{X(t)\}$ δεν είναι Μ.α.ς.χ.

4.4 Εξισώσεις Chapman-Kolmogorov

Για τις πιθανότητες μετάβασης έχουμε

$$\begin{aligned} p_{ij}(s+t) &= \Pr[X(t+s) = j | X(0) = i] \\ &= \sum_k \Pr[X(s) = k | X(0) = i] \Pr[X(t+s) = j | X(s) = k] \\ &= p_{ik}(s) p_{kj}(t), \end{aligned}$$

δηλαδή η $p_{ij}(s+t)$ είναι το εσωτερικό γινόμενο της i γραμμής του πίνακα $\mathbb{P}(s)$ και της j στήλης του πίνακα $\mathbb{P}(t)$. Σε επίπεδο πινάκων έχουμε, επομένως, τη σχέση

$$\mathbb{P}(s+t) = \mathbb{P}(s)\mathbb{P}(t).$$

Επίσης

$$p_{ij}(0) = \begin{cases} 1 & \text{αν } i = j, \\ 0 & \text{αν } i \neq j, \end{cases}$$

που σε επίπεδο πινάκων γράφεται

$$\mathbb{P}(0) = I.$$

Οι εξισώσεις αυτές αναφέρονται ως εξισώσεις Chapman-Kolmogorov για τις πιθανότητες μετάβασης. Εφαρμόζοντάς τες στο διάστημα $[0, t+h]$, για $h \to 0^+$, έχουμε

$$\begin{aligned} p_{ij}(t+h) &= \sum_k p_{ik}(t)p_{kj}(h) \\ &= p_{ij}(1-q_j h) + \sum_{k \neq j} p_{ik}(t)q_{kj}h + o(h),\ h \to 0^+ \end{aligned}$$

οπότε

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = -p_{ij}(t)q_j + \sum_{k \neq j} p_{ik}(t)q_{kj} + \frac{o(h)}{h},\ h \to 0^+.$$

Παίρνοντας $h \to 0^+$ συνάγουμε

$$p'_{ij}(t) = -p_{ij}(t)q_j + \sum_{k \neq j} p_{ik}(t)q_{kj}.$$

Οι εξισώσεις αυτές γράφονται σε επίπεδο πινάκων ως

$$\mathbb{P}'(t) = \mathbb{P}(t)\mathbb{Q}$$

και αναφέρονται ως (προδρομικές διαφορικές) εξισώσεις Chapman-Kolmogorov με αρχική συνθήκη την

$$\mathbb{P}(0) = I.$$

Αποδεικνύεται ότι έχουν μοναδική λύση, κάτω από πολύ γενικές συνθήκες που ισχύουν στις εφαρμογές. Συγκεκριμένα, έχουμε

$$\mathbb{P}(t) = e^{\mathbb{Q}t} = \sum_{n=0}^{\infty} \mathbf{Q}^n \frac{t^n}{n!}.$$

4.5 Η μεταβατική κατανομή

Έστω $\{X(t)\}$ μια Μ.α.ς.χ. και $\mathbf{p}(t) = (p_j(t) : j \in \mathcal{S})$ η μεταβατική κατανομή της (διάνυσμα-γραμμή), όπου

$$p_j(t) = \Pr[X(t) = j],\ j \in \mathcal{S}.$$

Εφαρμόζοντας το θεώρημα ολικής πιθανότητας έχουμε

$$p_j(t) = \sum_i p_i(0)p_{ij}(t),$$

δηλαδή σε μορφή διανυσμάτων-πινάκων

$$\mathbf{p}(t) = \mathbf{p}(0)\mathbb{P}(t).$$

Οπότε από τις προδρομικές διαφορικές εξισώσεις Chapman-Kolmogorov για τις πιθανότητες μετάβασης, παίρνουμε τις αντίστοιχες εξισώσεις Chapman-Kolmogorov για τη μεταβατική κατανομή:

$$\mathbf{p}'(t) = \mathbf{p}(0)\mathbb{P}'(t) = \mathbf{p}(0)\mathbb{P}(t)\mathbb{Q} = \mathbf{p}(t)\mathbb{Q}.$$

Αναλυτικά, οι εξισώσεις αυτές γράφονται ως

$$p_j'(t) = -p_j(t)q_j + \sum_{i \neq j} p_i(t)q_{ij},\ j \in \mathcal{S}.$$

Παράδειγμα 4.6 (Στοχαστική διαδικασία Poisson) Έστω $X(t)$ το πλήθος των αφίξεων πελατών στο $[0,t]$ που συμβαίνουν σύμφωνα με μια στοχαστική διαδικασία Poisson με ρυθμό λ. Οι εξισώσεις Chapman-Kolmogorov για τη μεταβατική κατανομή είναι

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t), \\ p_n'(t) &= -\lambda p_n(t) + \lambda p_{n-1}(t),\ n \geq 1. \end{aligned}$$

Παράδειγμα 4.7 (Η $M/M/1$ ουρά) Έστω $X(t)$ το πλήθος των πελατών τη στιγμή t σε μια $M/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρετήσεων μ. Οι εξισώσεις Chapman-Kolmogorov για τη μεταβατική κατανομή είναι

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) + \mu p_1(t), \\ p_n'(t) &= -(\lambda+\mu)p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t),\ n \geq 1. \end{aligned}$$

4.6 Η κατανομή ισορροπίας

Για $t \to \infty$ αποδεικνύεται ότι $p_j(t) \to p_j$, για κάθε $j \in \mathcal{S}$. Οι ποσότητες p_j ικανοποιούν τις εξισώσεις που προκύπτουν παίρνοντας $t \to \infty$ στις προδρομικές εξισώσεις Chapman-Kolmogorov, υποθέτοντας ότι $\lim_{t\to\infty} p_j'(t) = 0$. Οι εξισώσεις που προκύπτουν είναι οι

$$0 = -p_j q_j + \sum_{i \neq j} p_i q_{ij},\ j \in \mathcal{S},$$

που αναφέρονται και ως εξισώσεις ισορροπίας της Μ.α.ς.χ. Πιο συγκεκριμένα έχουμε το ακόλουθο αποτέλεσμα.

Θεώρημα 4.8 (Θεμελιώδες εργοδικό θεώρημα Μ.α.ς.χ.) *Έστω μια αδιαχώριστη Μ.α.ς.χ. $\{X(t)\}$ (δηλαδή όλες οι καταστάσεις της επικοινωνούν - υπάρχει μονοπάτι θετικών ρυθμών μετάβασης που να συνδέει οποιεσδήποτε δυο από αυτές). Τότε, ξεκινώντας από οποιαδήποτε κατάσταση της Μ.α.ς.χ. θα επιστρέψουμε σε αυτή με βεβαιότητα και ο αντίστοιχος μέσος χρόνος επανόδου είναι πεπερασμένος αν και μόνο αν το σύστημα των εξισώσεων ισορροπίας*

$$p_j \sum_{i \neq j} q_{ji} = \sum_{i \neq j} p_i q_{ij},\ j \in \mathcal{S}$$

και της εξίσωσης κανονικοποίησης

$$\sum_j p_j = 1,$$

έχει λύση. Στην περίπτωση αυτή η λύση $\mathbf{p} = (p_j : j \in \mathcal{S})$ είναι μοναδική, όλες οι συντεταγμένες της είναι θετικές και ισχύουν:

(i) *$p_j = \lim_{t\to\infty} p_j(t)$, δηλαδή η p_j είναι η οριακή πιθανότητα η $\{X(t)\}$ να βρίσκεται στην κατάσταση j,*

(ii) *$p_j = \lim_{t\to\infty} \frac{\int_0^t 1\{X(u)=j\}du}{t}$ με πιθανότητα 1, δηλαδή η p_j είναι το μακροπρόθεσμο ποσοστό του χρόνου που η $\{X(t)\}$ περνά στην κατάσταση j.*

(iii) *$p_j = \lim_{t\to\infty} \frac{\int_0^t \Pr[X(u)=j]du}{t} = \lim_{t\to\infty} \frac{E[\int_0^t 1\{X(u)=j\}du]}{t}$, δηλαδή η p_j είναι η $C-$οριακή πιθανότητα η $\{X(t)\}$ να βρίσκεται στην κατάσταση j, ή ισοδυναμα το μέσο μακροπρόθεσμο ποσοστό του χρόνου που η $\{X(t)\}$ περνά στην κατάσταση j.*

(iv) *$p_j = \frac{1/q_j}{m_j}$, όπου m_j είναι ο μέσος χρόνος μεταξύ δυο διαδοχικών εισόδων στην κατάσταση j.*

(v) *Αν η αρχική κατανομή $\mathbf{p}(0)$ της $\{X(t)\}$ είναι η $\mathbf{p}$ τότε κάθε μεταβατική κατανομή $\mathbf{p}(t)$ της $\{X(t)\}$ ισούται με $\mathbf{p}$.*

Η $\mathbf{p}$ αναφέρεται ως οριακή, στάσιμη ή κατανομή ισορροπίας της $\{X(t)\}$. Όταν το σύστημα των εξισώσεων ισορροπίας και της εξίσωσης κανονικοποίησης δεν έχει λύση, τότε οι ποσότητες στα (i)-(iv) παραπάνω είναι όλες 0.

Υπό το φως του εργοδικού θεωρήματος και λαμβάνοντας υπόψη την ερμηνεία των ρυθμών μετάβασης έχουμε ότι για δεδομένες καταστάσεις i και j μιας Μ.α.ς.χ. η ποσότητα $p_j q_{ji}$ εκφράζει το μακροπρόθεσμο ρυθμό μεταβάσεων $j \to i$ ανά χρονική μονάδα. Υπό αυτή την έννοια οι εξισώσεις ισορροπίας είναι εξισώσεις διατήρησης και απαιτούν ο μακροπρόθεσμος ρυθμός μεταβάσεων που εξέρχονται από μια κατάσταση j (δηλ. ο $p_j q_j = p_j \sum_{i\neq j} q_{ji}$ να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων που εισέρχονται στην ίδια κατάσταση (δηλ. με τον $\sum_{i\neq j} p_i q_{ij}$). Οι εξισώσεις ισορροπίας αναφέρονται συχνά και ως εξισώσεις πλήρους ισορροπίας.

Η στάσιμη κατανομή $\mathbf{p}$ ικανοποιεί επίσης και τις λεγόμενες εξισώσεις γενικευμένης ισορροπίας που απαιτούν ο μακροπρόθεσμος ρυθμός μεταβάσεων που εξέρχονται από ένα σύνολο καταστάσεων A να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων που εισέρχονται στο ίδιο σύνολο, δηλαδή

$$\sum_{j\in A}\sum_{i\in A^c} p_j q_{ji} = \sum_{i\in A^c}\sum_{j\in A} p_i q_{ij}, \; A \subseteq \mathcal{S}.$$

Αν ο χώρος καταστάσεων μιας Μ.α.ς.χ. είναι πεπερασμένος τότε υπάρχει πάντοτε στάσιμη κατανομή.

4.7 Ασκήσεις

1. Θεωρούμε την τροποποίηση της $M/M/c/c$ ουράς με Poisson διαδικασία αφίξεων ρυθμού λ και $Exp(\mu)$ χρόνους εξυπηρέτησης, όπου ορισμένοι πελάτες αποχωρούν από το σύστημα αμέσως μόλις αφιχθούν σε αυτό χωρίς να εξυπηρετηθούν (αποθαρρυνόμενοι πελάτες). Συγκεκριμένα, κάθε πελάτης που βρίσκει n άλλα άτομα στο σύστημα κατά την άφιξή του αναχωρεί με πιθανότητα $\frac{n}{c}$, $n = 0, 1, 2, \ldots, c$. Να δικαιολογηθεί ότι η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. και να γίνει το διάγραμμα ρυθμών μετάβασής της .

2. Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και εκθετικούς χρόνους εξυπηρέτησης $Exp(\mu)$, όπου κάθε πελάτης που συμπληρώνει την εξυπηρέτησή του δεν μένει ικανοποιημένος από αυτή με πιθανότητα q $(0 < q < 1)$ και την επαναλαμβάνει ευθύς αμέσως κ.ο.κ. Διαδοχικές επαναλήψεις έχουν ανεξάρτητες χρονικές διάρκειες. Να δικαιολογηθεί ότι η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. και να γίνει το διάγραμμα ρυθμών μετάβασής της.
3. Θεωρούμε μια μηχανή η οποία εξυπηρετεί εργασίες σύμφωνα με τη σειρά άφιξής τους (FCFS). Λόγω υψηλών λειτουργικών εξόδων, η μηχανή απενεργοποιείται μόλις το σύστημα αδειάσει. Όταν μια νέα εργασία αφίκνυται η μηχανή τίθεται εκ νέου σε λειτουργία, αλλά χρειάζεται κάποιο χρόνο προθέρμανσης (setup time). Υποθέτουμε ότι οι εργασίες φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και οι χρόνοι εξυπηρέτησης είναι $Exp(\mu)$. Οι χρόνοι προθέρμανσης της μηχανής είναι $Exp(\theta)$. Θα αναφέρουμε το παραπάνω σύστημα ως η $M/M/1/k$ ουρά με χρόνους εκκίνησης, όταν η χωρητικότητα του συστήματος είναι k. Έστω η στοχαστική διαδικασία $\{Q(t), I(t)\}$, όπου $Q(t)$ είναι το πλήθος των πελατών τη στιγμή t και $I(t)$ η κατάσταση της μηχανής (1 αν εξυπηρετεί και 0 αν είναι απενεργοποιημένη). Να δικαιολογηθεί ότι η $\{Q(t), I(t)\}$ είναι Μ.α.ς.χ. και να γίνει το διάγραμμα των ρυθμών μετάβασής της για $k = 1$ και $k = \infty$ ($M/M/1/1$ και $M/M/1$ ουρές με χρόνους εκκίνησης).

5

Απλές Μαρκοβιανές ουρές

Στο κεφάλαιο αυτό μελετάμε τις πλέον απλές ουρές, στις οποίες ο αριθμός των πελατών στο σύστημα είναι μια Μ.α.ς.χ. τύπου γέννησης θανάτου, δηλαδή σε κάθε χρονική στιγμή το σύστημα όντας σε μια κατάσταση n μπορεί να μεταβεί μόνο στην $n+1$ (λόγω άφιξης πελάτη) ή στην $n-1$ λόγω αναχώρησης. Παρόλο που η μελέτη αυτών των συστημάτων είναι πολύ απλή, η θέση τους στη Θεωρία Ουρών είναι πολύ σημαντική δεδομένης της μεγάλης εφαρμοσιμότητάς τους.

5.1 Ορισμός και στάσιμη κατανομή

Ένα σύστημα εξυπηρέτησης αναφέρεται ως απλή Μαρκοβιανή ουρά αν η στοχαστική διαδικασία $\{Q(t)\}$ που καταγράφει τον αριθμό των πελατών στο σύστημα είναι Μ.α.ς.χ. τύπου γέννησης θανάτου με ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda_i & \text{αν } i \geq 0,\ j = i+1, \\ \mu_i & \text{αν } i \geq 1,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Σε ένα τέτοιο σύστημα οι αφίξεις και οι αναχωρήσεις γίνονται μεμονωμένα. Οι ρυθμοί αφίξεων και αναχωρήσεων είναι λ_j και μ_j αντίστοιχα, όταν υπάρχουν j πελάτες στο σύστημα.

Θεωρώντας τα σύνολα $A_n = \{0, 1, 2, \ldots, n-1\}$, $n \geq 1$ έχουμε τις εξισώσεις γενικευμένης ισορροπίας

$$\lambda_{n-1} p_{n-1} = \mu_n p_n,\ n \geq 1,$$

απ΄ όπου παίρνουμε

$$p_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} p_0,\ n \geq 1.$$

Από την εξίσωση κανονικοποίησης, $\sum_{n=0}^{\infty} p_n = 1$, συνάγουμε ότι το σύστημα είναι ευσταθές, δηλαδή υπάρχει στάσιμη κατανομή αν και μόνο αν

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

Τότε, έχουμε

$$p_n = \begin{cases} B & \text{αν } n = 0, \\ B\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} & \text{αν } n \geq 1. \end{cases}$$

5.2 Ο ρυθμός διαπέρασης και οι εμφυτευμένες κατανομές

Ως ρυθμό διαπέρασης (throughput) μ^* ενός συστήματος εξυπηρέτησης εννοούμε τον μακροπρόθεσμο μέσο ρυθμό εξυπηρετούμενων πελατών ανά χρονική μονάδα. Ο ρυθμός αυτός, στην περίπτωση που εξυπηρετούνται όλοι οι πελάτες που εισέρχονται (δηλαδή δεν συμβαίνουν υπαναχωρήσεις πελατών), ισούται με τον μακροπρόθεσμο ρυθμό εισερχομένων πελατών ανά χρονική μονάδα λ^*. Έστω $A(t,t{+}h)$ και $D(t,t{+}h)$ τα πλήθη αφίξεων και αναχωρήσεων αντίστοιχα σε μια απλή Μαρκοβιανή ουρά στο διάστημα $(t, t+h]$. Τότε έχουμε

$$\begin{aligned}
\lambda^* &= \lim_{t\to\infty}\lim_{h\to 0^+}\frac{E[A(t,t+h)]}{h} = \frac{\Pr[A(t,t+h)=1]+o(h)}{h} \\
&= \lim_{t\to\infty}\lim_{h\to 0^+}\frac{\sum_{n=0}^{\infty}\Pr[Q(t)=n]\Pr[A(t,t+h)=1|Q(t)=n]}{h} \\
&= \sum_{n=0}^{\infty} p_n\lambda_n.
\end{aligned}$$

Ομοίως,

$$\mu^* = \lim_{t\to\infty}\lim_{h\to 0^+}\frac{E[D(t,t+h)]}{h} = \sum_{n=1}^{\infty} p_n\mu_n.$$

Ισχύει, όπως είπαμε $\lambda^* = \mu^*$, που μπορεί να αποδειχθεί και αλγεβρικά αθροίζοντας τις εξισώσεις ισορροπίας για όλες τις καταστάσεις. Για τις οριακές κατανομές σε στιγμές αφίξεων κα αναχωρήσεων έχουμε:

$$\begin{aligned}
a_j &= \lim_{t\to\infty}\lim_{h\to 0^+}\Pr[Q(t)=j|A(t,t+h)=1] \\
&= \lim_{t\to\infty}\lim_{h\to 0^+}\frac{\Pr[Q(t)=j, A(t,t+h)=1]}{\Pr[A(t,t+h)=1]} \\
&= \lim_{t\to\infty}\lim_{h\to 0^+}\frac{\Pr[Q(t)=j]\Pr[A(t,t+h)=1|Q(t)=j]}{\Pr[A(t,t+h)=1]} \\
&= \lim_{t\to\infty}\lim_{h\to 0^+}\frac{\Pr[Q(t)=j]\Pr[A(t,t+h)=1|Q(t)=j]/h}{\Pr[A(t,t+h)=1]/h} \\
&= \frac{p_j\lambda_j}{\lambda^*},
\end{aligned}$$

και ομοίως

$$\begin{aligned} d_j &= \lim_{t\to\infty}\lim_{h\to 0^+} \Pr[Q(t+h)=j|D(t,t+h)=1] \\ &= \lim_{t\to\infty}\lim_{h\to 0^+} \frac{\Pr[Q(t+h)=j, D(t,t+h)=1]}{\Pr[D(t,t+h)=1]} \\ &= \lim_{t\to\infty}\lim_{h\to 0^+} \frac{\Pr[Q(t)=j+1]\Pr[D(t,t+h)=1|Q(t)=j+1]}{\Pr[D(t,t+h)=1]} \\ &= \lim_{t\to\infty}\lim_{h\to 0^+} \frac{\Pr[Q(t)=j+1]\Pr[D(t,t+h)=1|Q(t)=j+1]/h}{\Pr[D(t,t+h)=1]/h} \\ &= \frac{p_{j+1}\mu_{j+1}}{\mu^*}. \end{aligned}$$

Βλέπουμε, λοιπόν ότι στην περίπτωση των απλών Μαρκοβιανών ουρών η ιδιότητα των μεμονωμένων αφίξεων και η ιδιότητα PASTA προκύπτουν άμεσα. Πράγματι, οι εξισώσεις ισορροπίας δίνουν $\lambda_j p_j = \mu_{j+1}p_{j+1}$ για κάθε $j \geq 0$, οπότε $\lambda^* = \mu^*$ και $a_j = d_j$, για κάθε $j \geq 0$. Επίσης, όταν η διαδικασία αφίξεων είναι Poisson έχουμε $\lambda_j = \lambda$, για κάθε $j \geq 0$, οπότε έχουμε $\lambda^* = \lambda$ και $a_j = \frac{p_j\lambda_j}{\lambda^*} = \frac{p_j\lambda}{\lambda} = p_j$, για $j \geq 0$.

5.3 Η $M/M/1/1$ ουρά

Η $M/M/1/1$ ουρά χρησιμοποιήθηκε ως το μοντέλο μιας τηλεφωνικής γραμμής που μπορεί να είναι ελεύθερη ή κατειλημμένη (θεωρώντας ότι δεν υπάρχει δυνατότητα κράτησης μιας κλήσης σε αναμονή). Αν ο ρυθμός αφίξεων είναι λ, ο ρυθμός εξυπηρέτησης είναι μ και $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού τότε έχουμε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. τύπου γέννησης -θανάτου με χώρο καταστάσεων $\mathcal{S} = \{0,1\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i=0,\ j=1, \\ \mu & \text{αν } i=1,\ j=0, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ \mu p_1 &= \lambda p_0, \end{aligned}$$

ενώ η εξίσωση κανονικοποίησης δίνει

$$p_0 + p_1 = 1.$$

Παρατηρήστε ότι η μια εξίσωση ισορροπίας είναι περιττή και μπορεί να παραληφθεί. Αυτό ισχύει πάντα, μία από τις εξισώσεις ισορροπίας μπορεί πάντα να παραληφθεί αφού προκύπτει από τις υπόλοιπες (με άθροισή τους). Λύνοντας το σύστημα παίρνουμε

$$p_0 = \frac{1}{1+\rho},$$
$$p_1 = \frac{\rho}{1+\rho}.$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της ιδιότητας PASTA έχουμε

$$d_0 = a_0 = p_0 = \frac{1}{1+\rho},\ d_1 = a_1 = p_1 = \frac{\rho}{1+\rho}.$$

Το ποσοστό των χαμένων πελατών, δηλαδή των πελατών που τελικά δεν εξυπηρετούνται, είναι το ποσοστό των πελατών που βρίσκουν έναν πελάτη κατά την άφιξή τους, δηλαδή δίνεται από την πιθανότητα a_1.

Ο μέσος αριθμός πελατών στο σύστημα είναι

$$E[Q] = 0p_0 + 1p_1 = p_1 = \frac{\rho}{1+\rho}.$$

Εφαρμόζοντας το νόμο του Little έχουμε ότι ο μέσος χρόνος παραμονής ενός αφικνούμενου πελάτη στο σύστημα είναι

$$E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu(1+\rho)}.$$

Προσέξτε ότι αυτός ο μέσος χρόνος αναφέρεται σε όλους τους πελάτες που φθάνουν στο σύστημα. Ο υπολογισμός του θα μπορούσε να γίνει και ως εξής:

$$E[S] = a_0\frac{1}{\mu} + a_1 0 = \frac{1}{\mu(1+\rho)},$$

δηλαδή δεσμεύοντας στον αριθμό των πελατών που βρίσκει ένας αφικνούμενος πελάτης. Αν μας ενδιέφερε ο μέσος χρόνος παραμονής ενός εισερχόμενου πελάτη στο σύστημα, που προφανώς είναι ο μέσος χρόνος εξυπηρέτησής του, ο νόμος του Little δίνει και πάλι το σωστό αποτέλεσμα, αρκεί να χρησιμοποιήσουμε όχι το συνολικό ρυθμό αφίξεων λ, αλλά το ρυθμό εισερχομένων πελατών $\lambda^* = \lambda a_0$. Πράγματι θα είχαμε

$$E[S_{entered}] = \frac{E[Q]}{\lambda^*} = \frac{1}{\mu}.$$

Όσον αφορά τη μελέτη του κύκλου απασχόλησης του συστήματος, έχουμε ότι η μέση περίοδος αργίας είναι $E[I] = \frac{1}{\lambda}$, αφού $I \sim Exp(\lambda)$ ενώ η μέση περίοδος συνεχούς λειτουργίας είναι $E[Y] = \frac{1}{\mu}$ αφού $Y \sim Exp(\mu)$. Ο μέσος κύκλος απασχόλησης είναι $E[Z] = E[I] + E[Y] = \frac{1}{\lambda} + \frac{1}{\mu}$.

5.4 Η $M/M/1$ ουρά

Η $M/M/1$ ουρά είναι το απλούστερο μοντέλο συστήματος εξυπηρέτησης με άπειρο χώρο αναμονής. Αν ο ρυθμός αφίξεων είναι λ, ο ρυθμός εξυπηρέτησης είναι μ και $\rho = \frac{\lambda}{\mu}$ είναι ο ρυθμός συνωστισμού τότε συνάγουμε ότι η στοχαστική διαδικασία

$\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. τύπου γέννησης - θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0,\ j = i+1, \\ \mu & \text{αν } i \geq 1,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Χρησιμοποιώντας τη γενική θεωρία, έχουμε

$$\begin{aligned} B^{-1} &= 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \rho^n \\ &= \begin{cases} \frac{1}{1-\rho} & \text{αν } \rho < 1 \text{ (ευστάθεια)} \\ \infty & \text{αν } \rho \geq 1 \text{ (αστάθεια).} \end{cases} \end{aligned}$$

Οπότε, για την οριακή κατανομή του αριθμού των πελατών, έχουμε

$$\begin{aligned} p_n &= \begin{cases} B & \text{αν } n = 0 \\ B \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} & \text{αν } n \geq 1 \end{cases} \\ &= (1-\rho)\rho^n,\ n \geq 0. \end{aligned}$$

Με άλλα λόγια $(p_n) \sim Geom(\rho)$ στο $\mathbb{N}_0$. Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο. Για το μέσο αριθμό πελατών στο σύστημα έχουμε

$$E[Q] = \sum_{n=0}^{\infty} n p_n = (1-\rho) \sum_{n=0}^{\infty} n \rho^n = \frac{\rho}{1-\rho}.$$

Από το νόμο του Little έχουμε για το χρόνο παραμονής ότι

$$E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu(1-\rho)}.$$

Όσον αφορά τη μελέτη του κύκλου απασχόλησης του συστήματος, έχουμε ότι η περίοδος αργίας $I \sim Exp(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$. Επιπλέον λόγω του αναγεννητικού χαρακτήρα του συστήματος, έχουμε ότι η πιθανότητα κενού συστήματος που ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που το σύστημα είναι κενό, ισούται επίσης με το ποσοστό του χρόνου σε έναν αναγεννητικό κύκλο που το σύστημα είναι κενό. Στην περίπτωση του συστήματος αυτού, ένας αναγεννητικός κύκλος είναι ένας κύκλος απασχόλησης, και στη διάρκειά του το σύστημα είναι κενό μόνο κατά την αντίστοιχη περίοδο αργίας. Επομένως,

$$p_0 = \frac{E[I]}{E[Z]},$$

όπου $E[Z]$ η μέση διάρκεια ενός κύκλου απασχόλησης. Οπότε,

$$E[Z] = \frac{E[I]}{p_0} = \frac{1}{\lambda(1-\rho)},$$

και η μέση περίοδος συνεχούς λειτουργίας θα είναι

$$E[Y] = E[Z] - E[I] = \frac{1}{\lambda(1-\rho)} - \frac{1}{\lambda} = \frac{1}{\mu(1-\rho)}.$$

Όσον αφορά την κατανομή $F_S(x)$ του χρόνου παραμονής ενός πελάτη στο σύστημα, υπολογίζουμε τον αντίστοιχο μετασχηματισμό Laplace-Stieltjes, δεσμεύοντας στο πλήθος των πελατών που βλέπει ο πελάτης κατά την άφιξή του σε αυτό. Έχουμε ότι αν ο πελάτης βρει n πελάτες στο σύστημα κατά την άφιξή του τότε θα περιμένει συνολικά όσο το άθροισμα $n+1$ χρόνων εξυπηρέτησης, οπότε ο αντίστοιχος δεσμευμένος μετασχηματισμός Laplace-Stietjes είναι αυτός της $Erlang(n+1,\mu)$ (δηλαδή $\left(\frac{\mu}{\mu+s}\right)^{n+1}$). Οπότε έχουμε

$$\begin{aligned}
F_S^*(s) = E[e^{-sS}] &= \sum_{n=0}^{\infty} \Pr[Q^- = n] E[e^{-sS} | Q^- = n] \\
&= \sum_{n=0}^{\infty} (1-\rho)\rho^n \left(\frac{\mu}{\mu+s}\right)^{n+1} \\
&= \frac{(1-\rho)\mu}{\mu+s} \left(1 - \frac{\rho\mu}{\mu+s}\right)^{-1} \\
&= \frac{(1-\rho)\mu}{(1-\rho)\mu+s}.
\end{aligned}$$

Ο τελευταίος είναι ο μετασχηματισμός Laplace-Stieltjes της $Exp((1-\rho)\mu)$, οπότε συνάγουμε ότι $S \sim Exp((1-\rho)\mu)$.

5.5 Τροποποιήσεις της $M/M/1$ ουράς

Εδώ θα εξετάσουμε τρεις ενδιαφέρουσες τροποποιήσεις της $M/M/1$ ουράς που εμφανίζονται στις εφαρμογές.

5.5.1 Η $M/M/1/k$ ουρά

Θεωρούμε την $M/M/1/k$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρέτησης μ, ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$ και χωρητικότητα k. Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι και εδώ Μ.α.ς.χ. τύπου γέννησης θανάτου με χώρο καταστάσεων $\mathcal{S} = \{0, 1, \ldots, k\}$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } 0 \le i \le k-1,\ j = i+1, \\ \mu & \text{αν } 1 \le i \le k,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Επειδή ο χώρος καταστάσεων είναι πεπερασμένος, το σύστημα είναι πάντα ευσταθές και έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{k} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{k} \rho^n$$
$$= \begin{cases} \frac{1-\rho^{k+1}}{1-\rho} & \text{αν } \rho \neq 1 \\ k+1 & \text{αν } \rho = 1. \end{cases}$$

Οπότε, για την οριακή κατανομή του αριθμού των πελατών, έχουμε

$$p_n = \begin{cases} \frac{(1-\rho)\rho^n}{1-\rho^{k+1}} & \text{αν } 0 \leq n \leq k, \ \rho \neq 1 \\ \frac{1}{k+1} & \text{αν } 0 \leq n \leq k, \ \rho = 1. \end{cases}$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο. Ο ρυθμός διαπέρασης του συστήματος, που αναφέρεται στους τελικά εισερχόμενους πελάτες στο σύστημα, είναι

$$\lambda^* = \sum_{n=0}^{k} \lambda_n p_n = \sum_{n=0}^{k-1} \lambda p_n = \lambda(1 - p_k).$$

Οπότε αν ενδιαφερόμαστε για την οριακή κατανομή του αριθμού των πελατών σε στιγμές 'πραγματικών αφίξεων', δηλαδή σε στιγμές εισόδων πελατών έχουμε

$$a_n^{enter} = \frac{\lambda_n p_n}{\lambda^*}$$
$$= \begin{cases} \frac{p_n}{1-p_k} & \text{αν } 0 \leq n \leq k-1, \\ 0 & \text{αν } n = k, \end{cases}$$

δηλαδή η PASTA δεν είναι εφαρμόσιμη. Για το μέσο αριθμό πελατών στο σύστημα έχουμε

$$E[Q] = \sum_{n=0}^{k} n p_n$$
$$= \begin{cases} \frac{\rho}{1-\rho} \cdot \frac{k\rho^{k+1} - (k+1)\rho^k + 1}{1-\rho^{k+1}} & \text{αν } \rho \neq 1, \\ \frac{k}{2} & \text{αν } \rho = 1. \end{cases}$$

Από το νόμο του Little έχουμε για το μέσο χρόνο παραμονής με αναφορά σε όλους τους αφικνούμενους πελάτες (είτε εισέρχονται στο σύστημα είτε όχι):

$$E[S] = \frac{E[Q]}{\lambda}.$$

Όμως, αν μας ενδιαφέρει ο μέσος χρόνος παραμονής μόνο των πελατών που εισέρχονται, έχουμε

$$E[S^{enter}] = \frac{E[Q]}{\lambda^*}.$$

Το ποσοστό των χαμένων πελατών βρίσκεται ως

$$1 - \frac{\lambda^*}{\lambda} = 1 - \frac{\lambda(1-p_k)}{\lambda} = p_k.$$

Εναλλακτικά, το ποσοστό των χαμένων πελατών ισούται με την πιθανότητα ένας πελάτης να βρει το σύστημα γεμάτο και να αναγκαστεί να φύγει, είναι δηλαδή ίσο με a_k. Οπότε, λόγω της PASTA συνάγουμε ότι είναι ίσο με p_k.

Όσον αφορά τη μελέτη του κύκλου απασχόλησης του συστήματος, έχουμε ότι η περίοδος αργίας $I \sim Exp(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$. Επιπλέον λόγω του αναγεννητικού χαρακτήρα του συστήματος, έχουμε ότι η πιθανότητα κενού συστήματος που ισούται με το μακροπρόθεσμο ποσοστό του χρόνου που το σύστημα είναι κενό, ισούται επίσης με το ποσοστό του χρόνου σε έναν αναγεννητικό κύκλο που το σύστημα είναι κενό. Ένας αναγεννητικός κύκλος στην περίπτωση του συστήματος αυτού είναι ένας κύκλος απασχόλησης και στη διάρκειά του το σύστημα είναι κενό μόνο κατά την αντίστοιχη περίοδο αργίας. Επομένως

$$p_0 = \frac{E[I]}{E[Z]},$$

όπου $E[Z]$ η μέση διάρκεια ενός κύκλου απασχόλησης. Οπότε ισχύει ακριβώς η ίδια λογική προσδιορισμού που περιγράφηκε για την $M/M/1$ ουρά. Λύνουμε την παραπάνω σχέση ως προς $E[Z]$, αφού τα $E[I]$ και p_0 έχουν προσδιοριστεί και κατόπιν προσδιορίζω και τη μέση περίοδο συνεχούς λειτουργίας από τη σχέση $E[Y] = E[Z] - E[I]$.

5.5.2 Η $M/M/1$ με αποθαρρυνόμενους πελάτες

Θεωρούμε τώρα την $M/M/1$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρέτησης μ, ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$, όπου υποθέτουμε ότι οι πελάτες μπορεί να αποθαρρυνθούν να μπουν στο σύστημα, αφού παρατηρήσουν τον υπάρχοντα συνωστισμό σε αυτό. Συγκεκριμένα, υποθέτουμε ότι ένας αφικνούμενος πελάτης που βρίσκει n πελάτες φεύγει (balks) με πιθανότητα q_n. Ισοδύναμα, το ποσοστό των πελατών που αναχωρούν άμεσα επί αυτών που βρίσκουν το σύστημα με n πελάτες είναι q_n. Και σε αυτή την περίπτωση μπορούμε εύκολα να δούμε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. τύπου γέννησης -θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda(1-q_i) & \text{αν } i \geq 0,\ j = i+1, \\ \mu & \text{αν } i \geq 1,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Στα πραγματικά συστήματα είναι λογικό να υποθέσουμε ότι η q_i είναι αύξουσα ως προς i, και ότι τείνει στο 1, καθώς το i τείνει στο άπειρο. Το προηγούμενο σύστημα της $M/M/1/k$ ουράς μπορεί να θεωρηθεί μια ειδική περίπτωση της $M/M/1$ ουράς με αποθαρρυνόμενους πελάτες που εμπίπτει σε αυτή την κατηγορία (εκεί είναι $q_i = 0$ για $0 \leq i \leq k-1$ και $q_i = 1$ για $i \geq k$). Εδώ θα εξετάσουμε την ειδική περίπτωση όπου $q_i = \frac{i}{i+1}$ που δίνει κομψά αποτελέσματα. Έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0\lambda_1 \cdots \lambda_{n-1}}{\mu_1\mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = e^{\rho} < \infty,$$

δηλαδή το συγκεκριμένο σύστημα είναι πάντα ευσταθές. Οπότε, για την οριακή κατανομή του αριθμού των πελατών, έχουμε

$$p_n = e^{-\rho}\frac{\rho^n}{n!},\ n \geq 0,$$

δηλαδή $(p_n) \sim Poisson(\rho)$. Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο. Αυτό βέβαια ισχύει όταν αναφερόμαστε σε όλους τους αφικνούμενους πελάτες, είτε μπαίνουν στο σύστημα είτε όχι. Ο ρυθμός διαπέρασης του συστήματος που αναφέρεται στους τελικά εισερχόμενους πελάτες στο σύστημα είναι

$$\begin{aligned} \lambda^* &= \sum_{n=0}^{\infty} \lambda_n p_n = \sum_{n=0}^{\infty} \frac{\lambda}{n+1} e^{-\rho}\frac{\rho^n}{n!} = \frac{\lambda e^{-\rho}}{\rho} \sum_{n=0}^{\infty} \frac{\rho^{n+1}}{(n+1)!} \\ &= \frac{\lambda e^{-\rho}}{\rho}(e^{\rho} - 1) = \mu(1 - e^{-\rho}). \end{aligned}$$

Το ίδιο αποτέλεσμα θα μπορούσε να προκύψει αν σκεφτόμαστε με εξυπηρετήσεις:

$$\mu^* = \sum_{n=1}^{\infty} \mu_n p_n = \sum_{n=1}^{\infty} \mu p_n = \mu(1 - p_0) = \mu(1 - e^{-\rho}).$$

Αν ενδιαφερόμαστε για την οριακή κατανομή του αριθμού των πελατών σε στιγμές 'πραγματικών αφίξεων', δηλαδή σε στιγμές εισόδων πελατών έχουμε

$$a_n^{enter} = \frac{\lambda_n p_n}{\lambda^*} = \frac{e^{-\rho}}{1 - e^{-\rho}} \cdot \frac{\rho^{n+1}}{(n+1)!},\ n \geq 0,$$

δηλαδή η PASTA δεν είναι εφαρμόσιμη. Ο μέσος αριθμός πελατών στο σύστημα είναι $E[Q] = \rho$. Από το νόμο του Little έχουμε για το μέσο χρόνο παραμονής με αναφορά σε όλους τους αφικνούμενους πελάτες (είτε εισέρχονται στο σύστημα είτε όχι) ότι $E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu}$. Αυτό μοιάζει κάπως παράδοξο, αφού ο μέσος χρόνος εξυπηρέτησης είναι $\frac{1}{\mu}$. Όμως, δεν είναι παράδοξο αφού η μέση τιμή του χρόνου παραμονής αναφέρεται σε όλους τους αφικνούμενους πελάτες, οπότε κάποιοι από αυτούς (αυτοί που φεύγουν αμέσως) έχουν μηδενικό χρόνο παραμονής. Αν μας ενδιαφέρει ο μέσος χρόνος παραμονής μόνο των πελατών που εισέρχονται, έχουμε

$$E[S^{enter}] = \frac{E[Q]}{\lambda^*} = \frac{\rho}{\mu(1 - e^{-\rho})}.$$

Το ποσοστό των χαμένων πελατών βρίσκεται ως

$$1 - \frac{\lambda^*}{\lambda} = 1 - \frac{\mu(1 - e^{-\rho})}{\lambda} = 1 - \frac{1 - e^{-\rho}}{\rho}.$$

Εναλλακτικά, το ποσοστό των χαμένων πελατών μπορεί να υπολογιστεί ως η πιθανότητα ένας πελάτης να αποχωρήσει από το σύστημα μόλις αφιχθεί σε αυτό. Οπότε, δεσμεύοντας στον αριθμό των πελατών που βρίσκει φθάνοντας στο σύστημα, έχουμε ότι είναι

$$\sum_{n=0}^{\infty} a_n q_n = \sum_{n=0}^{\infty} e^{-\rho} \frac{\rho^n}{n!} \cdot \frac{n}{n+1} = \cdots = 1 - \frac{1-e^{-\rho}}{\rho}.$$

Όσον αφορά τη μελέτη του κύκλου απασχόλησης του συστήματος, αυτή μπορεί να γίνει όπως και σε προηγούμενα παραδείγματα, δηλαδή να ξεκινήσουμε με το ότι η περίοδος αργίας $I \sim Exp(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$ και κατόπιν να λύσουμε ως προς το μέσο κύκλο απασχόλησης τη σχέση $p_0 = \frac{E[I]}{E[Z]}$, αφού το p_0 έχει ήδη υπολογιστεί. Τέλος, η μέση περίοδος συνεχούς λειτουργίας βρίσκεται από τη σχέση $E[Y] = E[Z] - E[I]$.

5.5.3 Η $M/M/1$ ουρά με μεταβλητή ταχύτητα εξυπηρέτησης

Θεωρούμε τώρα την $M/M/1$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρέτησης μ, ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$, όπου υποθέτουμε ότι ο υπηρέτης δουλεύει με ταχύτητα v_n όταν υπάρχουν n πελάτες στο σύστημα. Αυτό σημαίνει ότι όταν έναν εκθετικός χρόνος εξυπηρέτησης X διανύεται όταν στο σύστημα υπάρχουν n πελάτες, τότε η αντίστοιχη διάρκεια είναι X/v_n, δηλαδή, στην περίπτωση αυτή, η πραγματική διάρκεια εξυπηρέτης είναι $Exp(\mu v_n)$, οπότε ο ρυθμός εξυπηρέτησης είναι μv_n. Σε αυτή την περίπτωση η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. τύπου γέννησης -θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \geq 0,\ j = i+1, \\ \mu v_i & \text{αν } i \geq 1,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Στα πραγματικά συστήματα είναι λογικό να υποθέσουμε ότι η v_i είναι αύξουσα ως προς i. Εδώ θα εξετάσουμε την ειδική περίπτωση όπου $v_i = i$ που δίνει κομψά αποτελέσματα. Έχουμε

$$B^{-1} = 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = e^{\rho} < \infty,$$

δηλαδή το συγκεκριμένο σύστημα είναι πάντα ευσταθές. Οπότε για την οριακή κατανομή του αριθμού των πελατών έχουμε

$$p_n = e^{-\rho} \frac{\rho^n}{n!},\ n \geq 0,$$

δηλαδή $(p_n) \sim Poisson(\rho)$. Είναι ενδιαφέρον να παρατηρήσουμε ότι η οριακή κατανομή του αριθμού των πελατών συμπίπτει με την αντίστοιχη κατανομή για το σύστημα με αποθαρρυνόμενους πελάτες που μελετήσαμε πρωτύτερα, αν και τα δυο συστήματα είναι πολύ διαφορετικά. Η PASTA είναι εδώ εφαρμόσιμη, όπως και η ιδιότητα των μεμονωμένων μεταβάσεων οπότε οι οριακές κατανομές του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή του αριθμού των πελατών σε συνεχή χρόνο: $p_n = a_n = d_n$, $n \geq 0$. Ο μέσος αριθμός πελατών στο σύστημα είναι $E[Q] = \rho$. Από το νόμο του Little έχουμε ότι ο μέσος χρόνος παραμονής στο σύστημα $E[S] = \frac{E[Q]}{\lambda} = \frac{1}{\mu}$. Τέλος, η μελέτη του κύκλου απασχόλησης του συστήματος μπορεί να γίνει όπως και στα προηγούμενα παραδείγματα.

5.6 Η $M/M/c$ ουρά

Θεωρούμε την $M/M/c$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρέτησης μ, ρυθμό συνωστισμού $\rho = \frac{\lambda}{\mu}$, c υπηρέτες και απεριόριστη χωρητικότητα. Η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι και εδώ Μ.α.ς.χ. τύπου γέννησης θανάτου με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } 0 \leq i \leq k-1,\ j = i+1, \\ \min(i,c)\mu & \text{αν } 1 \leq i \leq k,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Εδώ έχουμε

$$\begin{aligned} B^{-1} &= 1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} = \sum_{n=0}^{c} \frac{\rho^n}{n!} + \sum_{n=c+1}^{\infty} \frac{\rho^n}{c! c^{n-c}} \\ &= \begin{cases} \sum_{n=0}^{c} \frac{\rho^n}{n!} + \frac{\rho^c}{c!} \frac{\rho}{c-\rho} & \text{αν } \rho < c \text{ (ευστάθεια)} \\ \infty & \text{αν } \rho \geq c \text{ (αστάθεια).} \end{cases} \end{aligned}$$

Οπότε για την οριακή κατανομή του αριθμού των πελατών, όταν το σύστημα είναι ευσταθές, έχουμε

$$p_n = \begin{cases} B \frac{\rho^n}{n!} & \text{αν } 0 \leq n \leq c, \\ B \frac{\rho^n}{c! c^{n-c}} & \text{αν } n \geq c+1. \end{cases}$$

Λόγω της ιδιότητας των μεμονωμένων μεταβάσεων και της PASTA, έχουμε ότι και οι οριακές κατανομές σε στιγμές αφίξεων και αναχωρήσεων συμπίπτουν με την οριακή κατανομή σε συνεχή χρόνο.

Η κατανομή του αριθμού των πελατών στο χώρο αναμονής, δεδομένου ότι όλοι οι πελάτες είναι απασχολημένοι βρίσκεται ως εξής:

$$\begin{aligned} \Pr[Q_q = n | Q \geq c] = \Pr[Q = n + c | Q \geq c] &= \frac{B \frac{\rho^{c+n}}{c! c^n}}{\sum_{m=c}^{\infty} B \frac{\rho^m}{c! c^{m-c}}} \\ &= \left(1 - \frac{\rho}{c}\right) \left(\frac{\rho}{c}\right)^n,\ n \geq 0, \end{aligned}$$

δηλαδή $(Q_q | Q \geq c) \sim Geom(\rho/c)$ στο $\mathbb{N}_0$. Οπότε εχόυμε και ότι $E[Q_q | Q \geq c] = \frac{\rho/c}{1-\rho/c}$. Η μέση τιμή του αριθμού των πελατών στο χώρο αναμονής μπορεί τώρα να βρεθεί εύκολα, ως εξής:

$$\begin{aligned}E[Q_q] &= \Pr[Q < c]E[Q_q|Q < c] + \Pr[Q \geq c]E[Q_q|Q \geq c] \\ &= \Pr[Q < c] \cdot 0 + \Pr[Q \geq c] \cdot \frac{\rho/c}{1-\rho/c} \\ &= B\frac{\rho^{c+1}}{(c-1)!(c-\rho)^2}.\end{aligned}$$

Εφαρμόζοντας το νόμο του Little στο χώρο αναμονής παίρνουμε το μέσο χρόνο αναμονής ενός πελάτη που είναι

$$E[W] = \frac{E[Q_q]}{\lambda} = B\frac{\rho^c}{(c-1)!(c-\rho)^2}\frac{1}{\mu}.$$

Ο μέσος χρόνος παραμονής είναι $E[S] = E[W] + E[X] = E[W] + \frac{1}{\mu}$, οπότε εφαρμόζοντας και πάλι το νόμο του Little παίρνουμε το μέσο αριθμό πελατών στο σύστημα

$$E[Q] = \lambda E[S] = B\frac{\rho^{c+1}}{(c-1)!(c-\rho)^2} + \rho.$$

Όσον αφορά τον προσδιορισμό του μέσου κύκλου απασχόλησης του συστήματος $E[Z]$, έχουμε καταρχήν ότι η περίοδος αργίας $I \sim Exp(\lambda)$, οπότε $E[I] = \frac{1}{\lambda}$. Επίσης, έχουμε $p_0 = \frac{E[I]}{E[Z]}$, οπότε $E[Z] = \frac{E[I]}{p_0} = \frac{1}{\lambda B}$. Τέλος η μέση περίοδος συνεχούς λειτουργίας είναι $E[Y] = E[Z] - E[I] = \frac{1}{\lambda B} - \frac{1}{\lambda}$.

Ο μετασχηματισμός Laplace-Stieltjes $F_W^*(s)$ του χρόνου αναμονής ενός πελάτη στο σύστημα βρίσκεται εύκολα, δεσμεύοντας στο πλήθος των πελατών που βρίσκει ο πελάτης κατά την άφιξή του σε αυτό. Αν ο πελάτης βρει λιγότερους από c πελάτες στο σύστημα, τότε ο χρόνος αναμονής του είναι μηδενικός. Αν βρεί $n \geq c$ πελάτες, τότε θα περιμένει να γίνουν $n+1-c$ αναχωρήσεις για να αρχίσει να εξυπηρετείται από κάποιον ελεύθερο υπηρέτη. Οι χρόνοι μεταξύ αυτών των εξυπηρετήσεων είναι $Exp(c\mu)$, αφού καθένας τους αντιστοιχεί στον ελάχιστο από τους c ανεξάρτητους $Exp(\mu)$ χρόνους εξυπηρέτησης που τρέχουν παράλληλα. Επομένως έχουμε

$$\begin{aligned}F_W^*(s) &= E[e^{-sW}] = \sum_{n=0}^{\infty}\Pr[Q^- = n]E[e^{-sW}|Q^- = n] \\ &= \sum_{n=0}^{c-1}\Pr[Q^- = n]\cdot 1 + \sum_{n=c}^{\infty}\Pr[Q^- = n]\cdot\left(\frac{c\mu}{s+c\mu}\right)^{n+1-c} \\ &= \sum_{n=0}^{c-1}B\frac{\rho^n}{n!} + \sum_{n=c}^{\infty}B\frac{\rho^n}{c!c^{n-c}}\left(\frac{c\mu}{s+c\mu}\right)^{n+1-c} \\ &= \sum_{n=0}^{c-1}B\frac{\rho^n}{n!} + B\frac{\rho^c}{c!}\cdot\frac{c\mu}{c\mu-\lambda}\cdot\frac{c\mu-\lambda}{s+c\mu-\lambda} \\ &= \Pr[Q < c]\cdot 1 + \Pr[Q \geq c]\frac{c\mu-\lambda}{s+c\mu-\lambda}.\end{aligned}$$

Ο τελευταίος μετασχηματισμός Laplace-Stieltjes δείχνει ότι η τ.μ. W είναι 0 με πιθανότητα $\Pr[Q < c]$ και ακολουθεί την $Exp(c\mu - \lambda)$ με πιθανότητα $\Pr[Q \geq c]$.

5.7 Ασκήσεις

1. Θεωρήστε την $M/M/c$ ουρά με ρυθμό αφίξεων λ και $Exp(\mu)$ χρόνους εξυπηρέτησης. Έστω ότι ένας πελάτης βρίσκει $n > c$ πελάτες στο σύστημα κατά την άφιξή του. Να προσδιοριστούν
   1. η κατανομή του χρόνου μέχρι την επόμενη αναχώρηση πελάτη από το σύστημα.
   2. η κατανομή του χρόνου παραμονής του πελάτη στο σύστημα.

2. Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και $Exp(\mu)$ χρόνους εξυπηρέτησης, όπου ο χρόνος υπομονής κάθε πελάτη που περιμένει στο χώρο αναμονής έχει την εκθετική κατανομή με παράμετρο ν και μόλις αυτός συμπληρωθεί ο πελάτης αναχωρεί. Αιτιολογήστε ότι η στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ. και δώστε το διάγραμμα των ρυθμών μετάβασής της. Για την ειδική περίπτωση $\nu = \mu$, βρείτε τη στάσιμη κατανομή του αριθμού των πελατών σε συνεχή χρόνο.

3. Θεωρούμε την τροποποίηση της $M/M/1$ ουράς με ρυθμό αφίξεων λ και $Exp(\mu)$ χρόνους εξυπηρέτησης, με αποθαρρυνόμενους πελάτες, όπου κάθε πελάτης που βρίσκει n πελάτες κατά την άφιξή του αναχωρεί με πιθανότητα q_n, με $q_0 = \frac{1}{4}$ και $q_n = \frac{3}{4}$ για $n \geq 1$. Να βρεθούν:
   1. η συνθήκη ευστάθειας (στασιμότητας) για το σύστημα,
   2. οι κατανομές (p_n), (a_n) και (d_n) του αριθμού των πελατών σε συνεχή χρόνο, σε στιγμές αφίξεων και αναχωρήσεων, αντίστοιχα,
   3. το ποσοστό των χαμένων πελατών.

4. Θεωρούμε μια απλή Μαρκοβιανή ουρά με ρυθμούς αφίξεων και εξυπηρετήσεων $\lambda_n = \alpha^n \lambda$ $(n \geq 0)$ και $\mu_n = n\alpha^n \mu$ $(n \geq 1)$ αντίστοιχα, όπου $\alpha \in (0,1)$, $\lambda, \mu > 0$ γνωστές παράμετροι.
   1. Πότε η ουρά είναι ευσταθής; Να βρεθεί η στάσιμη κατανομή (p_n) του αριθμού των πελατών στο σύστημα. Τί είδους κατανομή είναι;
   2. Να βρεθούν οι οριακές κατανομές (a_n) και (d_n) του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα. Τί κατανομές είναι;
   3. Να βρεθεί ο μέσος χρόνος παραμονής $E[S]$ ενός πελάτη στο σύστημα, καθώς και οι μέσοι χρόνοι συνεχούς λειτουργίας, αργίας και κύκλου απασχόλησης $E[Y]$, $E[I]$ και $E[Z]$, αντίστοιχα.

5. Θεωρούμε την τροποποίηση της $M/M/c/c$ ουράς με Poisson διαδικασία αφίξεων ρυθμού λ και $Exp(\mu)$ χρόνους εξυπηρέτησης, όπου ορισμένοι πελάτες αναχωρούν από το σύστημα αμέσως μόλις αφιχθούν, χωρίς να εξυπηρετηθούν (αποθαρρυνόμενοι πελάτες). Συγκεκριμένα, κάθε πελάτης που βρίσκει n άλλα άτομα στο σύστημα κατά την άφιξή του αναχωρεί με πιθανότητα $\frac{n}{c}$, $0 \leq n \leq c$.

1. Να βρεθεί η στάσιμη κατανομή (p_n) του αριθμού των πελατών στο σύστημα $\{Q(t)\}$. Τί είδους κατανομή είναι;
2. Να βρεθεί το μακροπρόθεσμο ποσοστό των άμεσα αποχωρούντων πελατών (χαμένων πελατών).
3. Να βρεθούν οι οριακές κατανομές (a_n) και (d_n) του αριθμού των πελατών σε στιγμές αφίξεων και αναχωρήσεων αντίστοιχα, που αναφέρονται στους πραγματικά εισερχόμενους πελάτες του συστήματος.
4. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη στο σύστημα, $E[S]$, λαμβάνοντας υπόψη όλους τους αφιχθέντες πελάτες. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτης στο σύστημα, $E[S']$, λαμβάνοντας υπόψη μόνο τους τελικά εισερχόμενους πελάτες.
5. Να βρεθεί ο μέσος κύκλος απασχόλησης του συστήματος, $E[Z]$.

6

Γενικές Μαρκοβιανές ουρές

Στο κεφάλαιο αυτό μελετάμε ουρές που μπορούν να παρασταθούν από μια Μ.α.ς.χ., που όμως μπορεί να μην είναι του τύπου γέννησης θανάτου. Στην πιό απλή περίπτωση η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα είναι Μ.α.ς.χ., αλλά όχι τύπου γέννησης θανάτου, καθώς από μια κατάσταση i είναι δυνατή η άμεση μετάβαση σε κάποια κατάσταση j με $j \neq i-1, i+1$. Αυτό συμβαίνει όταν οι χρόνοι μεταξύ των αφίξεων είναι εκθετικοί και οι χρόνοι εξυπηρέτησης είναι επίσης εκθετικοί, αλλά οι πελάτες έρχονται και/ή εξυπηρετούνται σε ομάδες. Σε πιο περίπλοκες περιπτώσεις, η διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα δεν είναι Μ.α.ς.χ. αλλά μπορεί να εμφυτευτεί σε μια Μ.α.ς.χ., δηλαδή υπάρχει κάποια άλλη στοχαστική διαδικασία $\{I(t)\}$ έτσι ώστε η $\{(Q(t), I(t)\}$ να είναι Μ.α.ς.χ. Το βασικό εργαλείο για τη μελέτη της οριακής κατανομής γενικών Μαρκοβιανών ουρών είναι οι πιθανογεννήτριες. Για τη μελέτη των χρόνων παραμονής χρησιμοποιούνται μετασχηματισμοί Laplace-Stieltjes όπως και στις απλές Μαρκοβιανές ουρές.

6.1 Η $M/M/1$ ουρά με ομαδικές αφίξεις

Θεωρούμε την $M/M/1$ ουρά με ομαδικές αφίξεις. Το μοντέλο αυτό είναι παρόμοιο με την $M/M/1$ ουρά, με μόνη διαφορά ότι οι αφίξεις συμβαίνουν κατά ομάδες και τα μεγέθη των διαδοχικά αφικνούμενων ομάδων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Πιο συγκεκριμένα, μια αφικνούμενη ομάδα είναι μεγέθους j με πιθανότητα g_j. Έστω, λοιπόν μια τέτοια ουρά, με ρυθμό αφίξεων λ, συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων $(g_j : j \geq 1)$, $Exp(\mu)$ χρόνους εξυπηρέτησης, 1 υπηρέτη και απεριόριστη χωρητικότητα. Αν εξετάσουμε τη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα, τότε βλέπουμε ότι για τις μεταβάσεις της έχουμε τον παρακάτω πίνακα:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim Exp(\lambda g_k)$
$n \geq 1$	$n+k,\ k \geq 1$ $n-1$	$T_{n,n+k} \sim Exp(\lambda g_k)$ $T_{n,n-1} \sim Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μ.α.ς.χ. με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0,\ j = i + k, \\ \mu & \text{αν } i \geq 1,\ j = i - 1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Έστω $(p_n : n \geq 0)$ η οριακή κατανομή της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda + \mu) p_n &= \sum_{i=0}^{n-1} \lambda g_{n-i} p_i + \mu p_{n+1},\ n \geq 1. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, θα ακολουθήσουμε τη λεγόμενη μέθοδο των πιθανογεννητριών. Για την εφαρμογή αυτής της μεθόδου εισάγουμε καταρχήν τις πιθανογεννήτριες

$$\begin{aligned} P(z) &= \sum_{n=0}^{\infty} p_n z^n,\ |z| \leq 1, \\ G(z) &= \sum_{n=0}^{\infty} g_n z^n,\ |z| \leq 1. \end{aligned}$$

Κατόπιν μετασχηματίζουμε τις εξισώσεις ισορροπίας ώστε να προκύψει κάποια εξίσωση για την άγνωστη πιθανογεννήτρια $P(z)$, την οποία και επιλύουμε. Στο τέλος ἀντιστρέφουμε΄ την πιθανογεννήτρια, δηλαδή την αναπτύσσσουμε σε δυνάμεις του z^n, ώστε να προκύψουν οι οριακές πιθανότητες p_n.

Για το μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\begin{aligned} & \lambda p_0 + \sum_{n=1}^{\infty} (\lambda + \mu) p_n z^n = \mu p_1 + \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n + \sum_{n=1}^{\infty} \mu p_{n+1} z^n \\ \Leftrightarrow\ & (\lambda + \mu) P(z) - \mu p_0 = \lambda \sum_{i=0}^{\infty} p_i z^i \sum_{n=i+1}^{\infty} g_{n-i} z^{n-i} + \mu \sum_{n=0}^{\infty} p_{n+1} z^n \\ \Leftrightarrow\ & (\lambda + \mu) P(z) = \mu p_0 + \lambda P(z) G(z) + \frac{\mu}{z} (P(z) - p_0) \\ \Leftrightarrow\ & \left(\lambda + \mu - \lambda G(z) - \frac{\mu}{z} \right) P(z) = \mu p_0 \left(1 - \frac{1}{z} \right) \end{aligned}$$

Λύνοντας για την $P(z)$, παίρνουμε

$$P(z) = \frac{\mu p_0 (1 - z)}{\mu - (\lambda + \mu) z + \lambda z G(z)}.$$

Ακολούθως, από την εξίσωση κανονικοποίησης έχουμε $\sum_{n=0}^{\infty} p_n = P(1) = 1$. Χρησιμοποιώντας αυτή τη σχέση στην έκφραση για την $P(z)$, παίρνουμε μια απροσδιόριστη μορφή $\frac{0}{0}$, και εφαρμόζοντας τον κανόνα L'Hospital έχουμε

$$1 = P(1) = \frac{-\mu p_0}{-(\lambda+\mu) + \lambda G(1) + \lambda G'(1)} = \frac{\mu p_0}{\lambda + \mu - \lambda - \lambda m},$$

όπου $m = G'(1) = \sum_{i=1}^{\infty} i g_i$ είναι το μέσο μέγεθος αφικνούμενων ομάδων. Επομένως,

$$p_0 = 1 - \frac{\lambda m}{\mu}.$$

Ασφαλώς, το σύστημα είναι ευσταθές αν και μόνο αν $p_0 > 0$, δηλαδή αν

$$\frac{\lambda m}{\mu} < 1.$$

Η συνθήκη αυτή είναι φυσιολογική αφού απαιτεί ο ρυθμός αφίξεων πελατών λm να είναι μικρότερος από το μέγιστο ρυθμό εξυπηρέτησης μ που μπορεί να παρέχει το σύστημα. Τελικά, όταν το σύστημα ειναι ευσταθές, έχουμε

$$P(z) = \frac{\mu\left(1 - \frac{\lambda m}{\mu}\right)(1-z)}{\mu - (\lambda+\mu)z + \lambda z G(z)}.$$

Από την πιθανογεννήτρια μπορούμε να ανακτήσουμε την οριακή κατανομή p_n, χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!},\ n \geq 0.$$

Επίσης μπορούμε να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2)\cdots(Q-n+1)] = P^{(n)}(1),\ n \geq 1.$$

Οι σχέσεις αυτές βέβαια έχουν θεωρητική κυρίως αξία. Σε ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων μπορούμε να ανακτήσουμε την οριακή κατανομή σε κλειστή μορφή.

6.1.1 Η περίπτωση των μεμονωμένων αφίξεων

Στην ειδική περίπτωση των μεμονωμένων αφίξεων, δηλαδή $g_1 = 1$ και $g_j = 0$ για $j \geq 2$ έχουμε ότι $G(z) = z$, οπότε

$$P(z) = \frac{\mu\left(1 - \frac{\lambda}{\mu}\right)(1-z)}{\mu - (\lambda+\mu)z + \lambda z^2} = \frac{1 - \frac{\lambda}{\mu}}{1 - \frac{\lambda}{\mu} z},$$

και αναπτύσσοντας σε δυνάμεις του z παίρνουμε τη γνωστή γεωμετρική κατανομή του αριθμού των πελατών στην $M/M/1$ ουρά.

6.1.2 Η περίπτωση των ομάδων γεωμετρικού μεγέθους

Στην ειδική περίπτωση που η συνάρτηση πιθανότητας αφικνούμενων ομάδων $(g_i : i = 1, 2, \ldots) \sim Geom(\alpha)$ στο $\mathbb{N}$, δηλαδή $g_i = (1-\alpha)\alpha^{i-1}$, $i \geq 1$ έχουμε

$$G(z) = \frac{(1-\alpha)z}{1-\alpha z}$$

και $m = G'(1) = \frac{1}{1-\alpha}$, οπότε

$$P(z) = \frac{\mu\left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)(1-z)}{\mu - (\lambda+\mu)z + \lambda z \frac{(1-\alpha)z}{1-\alpha z}} = \frac{\left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)(1-\alpha z)}{1 - \frac{\lambda+\alpha\mu}{mu} z}.$$

Χρησιμοποιώντας τη γεωμετρική σειρά για να αναπτύξουμε το $\left(1 - \frac{\lambda+\alpha\mu}{mu} z\right)^{-1}$ σε δυνάμεις του z, παίρνουμε

$$\begin{aligned} P(z) &= \left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)(1-\alpha z)\sum_{k=0}^{\infty}\left(\frac{\lambda+\alpha\mu}{\mu}\right)^k z^k \\ &= \sum_{n=0}^{\infty}\left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)\left(\frac{\lambda+\alpha\mu}{\mu}\right)^n z^n \\ &\quad - \sum_{n=1}^{\infty}\left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)\alpha\left(\frac{\lambda+\alpha\mu}{\mu}\right)^{n-1} z^n, \end{aligned}$$

οπότε

$$\begin{aligned} p_0 &= 1 - \frac{\lambda}{(1-\alpha)\mu}, \\ p_n &= \left(1 - \frac{\lambda}{(1-\alpha)\mu}\right)\left(\frac{\lambda+\alpha\mu}{\mu}\right)^{n-1}\left(\frac{\lambda+\alpha\mu}{\mu} - \alpha\right),\ n \geq 1. \end{aligned}$$

6.1.3 Οριακές κατανομές πλήθους πελατών σε στιγμές αφίξεων

Αν ενδιαφερόμαστε για τις οριακές πιθανότητες του αριθμού των πελατών σε στιγμές αφίξεων στην $M/M/1$ ουρά με ομαδικές αφίξεις, τότε είναι σημαντικό να συνειδητοποιήσουμε ότι υπάρχουν δυο ειδών τέτοιες πιθανότητες. Συμβολίζουμε με a_n^{group} την πιθανότητα μια αφικνούμενη ομάδα να βρει n πελάτες στο σύστημα κατά την άφιξή της και με a_n την πιθανότητα ένας αφικνούμενος πελάτης να βρει n πελάτες στο σύστημα κατά την άφιξή του. Λόγω της ιδιότητας PASTA έχουμε ότι $a_n^{group} = p_n$, όπου p_n είναι η οριακή πιθανότητα n πελατών στο σύστημα σε συνεχή χρόνο, την οποία έχουμε ήδη υπολογίσει. Για να βρούμε την a_n, πρέπει να διευκρινήσουμε τον τρόπο με τον οποίο οι πελάτες μιας αφικνούμενης ομάδας εισέρχονται στο σύστημα. Θεωρούμε ότι οι πελάτες μπαίνουν ακολουθιακά, αλλά ακαριαία. Οπότε αν μια ομάδα μεγέθους k φθάσει στο σύστημα όταν υπάρχουν n πελάτες σε αυτό, τότε ο πρώτος πελάτης της θα δει n άτομα και θα μπει στην $n+1$ θέση, ο δεύτερος θα δει $n+1$ άτομα και θα μπει στην $n+2$ θέση, ο τρίτος θα δει $n+2$ άτομα και θα μπει στην

$n+3$ θέση κ.ο.κ. Επομένως έχουμε

$$a_n = \sum_{k=0}^{n} a_k^{group} \tilde{g}_{n-k+1} = \sum_{k=0}^{n} p_k \tilde{g}_{n-k+1},\ n \geq 0,$$

όπου $\tilde{g}_n$ είναι η πιθανότητα ένας πελάτης να είναι ο n-οστός στην ομάδα του. Επομένως, χρειάζεται να βρούμε τις πιθανότητες $\tilde{g}_n$, $n \geq 1$, όταν έχουμε τη συνάρτηση πιθανότητας του μεγέθους αφικνούμενων ομάδων $(g_n : n \geq 1)$.

Η πιθανότητα $\tilde{g}_n$ ένας πελάτης να είναι ο n-οστός της ομάδας του μπορεί να ερμηνευθεί και ως το μακροπρόθεσμο ποσοστό των πελατών που είναι n-οστοί στις ομάδες τους. Επομένως, συμβολίζοντας με Y_i, $i \geq 1$, τα μεγέθη των διαδοχικά αφικνούμενων ομάδων στο σύστημα έχουμε

$$\begin{aligned} \tilde{g}_n &= \lim_{N\to\infty} \frac{\text{πλήθος πελατών που είναι } n\text{-οστοί στις ομάδες } 1,2,\ldots,N}{Y_1+Y_2+\cdots+Y_N} \\ &= \lim_{N\to\infty} \frac{\text{πλήθος ομάδων με μέγεθος } \geq n \text{ στις ομάδες } 1,2,\ldots,N}{Y_1+Y_2+\cdots+Y_N}, \end{aligned}$$

διότι μια ομάδα έχει έναν n-οστό πελάτη αν και μόνο αν το μέγεθός της είναι τουλάχιστον n. Επομένως, θέτοντας $X_i = 1_{\{Y_i \geq n\}}$, $i \geq 1$, έχουμε

$$\tilde{g}_n = \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i} = \frac{\sum_{i=1}^{N} X_i/N}{\sum_{i=1}^{N} Y_i/N} = \frac{E[X_i]}{E[Y_i]},$$

όπου η τελευταία ισότητα προκύπτει από το νόμο των μεγάλων αριθμών (και ισχύει με πιθανότητα 1). Όμως $E[Y_i] = \sum_{k=1}^{\infty} k g_k$ και $E[X_i] = \Pr[Y_i \geq n] = \sum_{k=n}^{\infty} g_k$, οπότε

$$\tilde{g}_n = \frac{\sum_{k=n}^{\infty} g_k}{\sum_{k=1}^{\infty} k g_k},\ k \geq 1.$$

6.2 Το Μαρκοβιανό σύστημα ολικών εξυπηρετήσεων με ομαδικές αφίξεις

Στο Μαρκοβιανό σύστημα ολικών εξυπηρετήσεων με ομαδικές αφίξεις, οι πελάτες φθάνουν κατά ομάδες σύμφωνα με μια διαδικασία Poisson ρυθμού λ με συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων $(g_j : j \geq 1)$. Υπάρχει ένα υπηρέτης που εξυπηρετεί όλους τους παρόντες πελάτες μαζί, οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι $Exp(\mu)$ και ο χώρος αναμονής είναι απεριόριστός. Ένα τέτοιο σύστημα είναι κατάλληλο για τη μοντελοποίηση της εξυπηρέτησης σε συστήματα μεταφορών. Π.χ. στην περίπτωση αυτή το σύστημα είναι μια στάση μεταφορικού μέσου και ο υπηρέτης είναι το μεταφορικό μέσο. Ο χρόνος εξυπηρέτησης αντιστοιχεί σε έναν ενδιάμεσο χρόνο μεταξύ δυο διαδοχικών επισκέψεων του μέσου από τη στάση και κάθε φορά που γίνεται επίσκεψη του μέσου η στάση αδειάζει. Για τις μεταβάσεις της στοχαστικής διαδικασίας $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα έχουμε τον παρακάτω πίνακα:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim Exp(\lambda g_k)$
$n \geq 1$	$n+k,\ k \geq 1$ 0	$T_{n,n+k} \sim Exp(\lambda g_k)$ $T_{n0} \sim Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μ.α.ς.χ. με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0,\ j = i+k, \\ \mu & \text{αν } i \geq 1,\ j = 0, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Έστω $(p_n : n \geq 0)$ η οριακή κατανομή της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\lambda p_0 = \sum_{i=1}^{\infty} \mu p_i,$$
$$(\lambda+\mu)p_n = \sum_{i=0}^{n-1} \lambda g_{n-i} p_i,\ n \geq 1.$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, θα ακολουθήσουμε πάλι τη μέθοδο των πιθανογεννητριών. Έστω οι πιθανογεννήτριες

$$P(z) = \sum_{n=0}^{\infty} p_n z^n,\ |z| \leq 1,$$
$$G(z) = \sum_{n=0}^{\infty} g_n z^n,\ |z| \leq 1.$$

Για το μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\lambda p_0 + \sum_{n=1}^{\infty} (\lambda+\mu) p_n z^n = \mu \sum_{i=1}^{\infty} p_i + \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n$$
$$\Leftrightarrow (\lambda+\mu)P(z) - \mu p_0 = \mu(1-p_0) + \lambda \sum_{i=0}^{\infty} p_i z^i \sum_{n=i+1}^{\infty} g_{n-i} z^{n-i}$$
$$\Leftrightarrow (\lambda+\mu)P(z) = \mu + \lambda P(z) G(z).$$

Λύνοντας για την $P(z)$, παίρνουμε

$$P(z) = \frac{\mu}{\mu + \lambda - \lambda G(z)}.$$

Εδώ το σύστημα είναι πάντα ευσταθές και από την πιθανογεννήτρια μπορούμε να ανακτήσουμε την οριακή κατανομή p_n, χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!},\ n \geq 0.$$

Επίσης μπορούμε να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2)\cdots(Q-n+1)] = P^{(n)}(1),\ n \geq 1.$$

Οι σχέσεις αυτές βέβαια έχουν θεωρητική κυρίως αξία. Σε ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων μπορούμε να ανακτήσουμε την οριακή κατανομή σε κλειστή μορφή.

6.2.1 Η περίπτωση των μεμονωμένων αφίξεων

Στην ειδική περίπτωση που έχουμε μεμονωμένες αφίξεις, δηλαδή $g_1 = 1$ και $g_j = 0$ για $j \geq 2$ έχουμε ότι $G(z) = z$, οπότε

$$P(z) = \frac{\mu}{\mu + \lambda - \lambda z}.$$

Αναπτύσοντας σε δυνάμεις του z χρησιμοποιώντας τη γεωμετρική σειρά έχουμε

$$P(z) = \frac{\mu}{\mu+\lambda} \cdot \frac{1}{1 - \frac{\lambda}{\lambda+\mu} z} = \sum_{n=0}^{\infty} \frac{\mu}{\mu+\lambda} \left(\frac{\lambda}{\mu+\lambda}\right)^n z^n,$$

απ᾽ όπου

$$p_n = \frac{\mu}{\mu+\lambda} \left(\frac{\lambda}{\mu+\lambda}\right)^n,\ n \geq 0,$$

δηλαδή η $(p_n : n \geq 0) \sim Geom(\frac{\lambda}{\lambda+\mu})$ στο $\mathbb{N}_0$.

6.3 Η $M/M/1$ ουρά με ομαδικές εξυπηρετήσεις

Θεωρούμε την $M/M/1$ ουρά με ομαδικές εξυπηρετήσεις σταθερού μεγέθους. Πιο συγκεκριμένα, οι πελάτες φθάνουν στο σύστημα σύμφωνα με μια διαδικασία Poisson με ρυθμό λ, οι χρόνοι εξυπηρέτησης είναι ανεξάρτητοι $Exp(\mu)$, υπάρχει 1 υπηρέτης που εξυπηρετεί ταυτόχρονα r πελάτες ως μια ομάδα και η χωρητικότητα του συστήματος είναι απεριόριστη. Όταν στο σύστημα υπάρχουν λιγότεροι από r πελάτες, ο υπηρέτης δεν παρέχει εξυπηρέτηση, αλλά περιμένει πρώτα να συγκεντρωθούν r. Για τις μεταβάσεις της στοχαστικής διαδικασίας $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα έχουμε τον παρακάτω πίνακα

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$n,\ 0 \le n \le r-1$	$n+1$	$T_{n,n+1} \sim Exp(\lambda)$
$n \ge r$	$n+1$ $n-r$	$T_{n,n+1} \sim Exp(\lambda)$ $T_{n,n-r} \sim Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{Q(t)\}$ είναι Μ.α.ς.χ. με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$ και ρυθμούς μετάβασης

$$q_{ij} = \begin{cases} \lambda & \text{αν } i \ge 0,\ j = i+1, \\ \mu & \text{αν } i \ge r,\ j = i-r, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Έστω $(p_n : n \ge 0)$ η οριακή κατανομή της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_r, \\ \lambda p_n &= \lambda p_{n-1} + \mu p_{n+r},\ 1 \le n \le r-1 \\ (\lambda+\mu) p_n &= \lambda p_{n-1} + \mu p_{n+r},\ n \ge r. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, θα ακολουθήσουμε τη μέθοδο των πιθανογεννητριών.

Έστω $P(z) = \sum_{n=0}^{\infty} p_n z^n$, $|z| \le 1$, η πιθανογεννήτρια της (p_n). Πολλαπλασιάζοντας την n-οστή εξίσωση ισορροπίας με z^n και αθροίζοντας παίρνουμε:

$$\begin{aligned} &\lambda p_0 + \sum_{n=1}^{r-1} \lambda p_n z^n + \sum_{n=r}^{\infty} (\lambda+\mu) p_n z^n = \mu p_r + \sum_{n=1}^{\infty} \lambda p_{n-1} z^n + \sum_{n=1}^{\infty} \mu p_{n+r} z^n \\ \Leftrightarrow\ &(\lambda+\mu) P(z) - \mu \sum_{n=0}^{r-1} p_n z^n = \lambda z \sum_{n=1}^{\infty} p_{n-1} z^{n-1} + \frac{\mu}{z^r} \sum_{n=0}^{\infty} p_{n+r} z^{n+r} \\ \Leftrightarrow\ &(\lambda+\mu) P(z) - \mu \sum_{n=0}^{r-1} p_n z^n = \lambda z \sum_{k=0}^{\infty} p_k z^k + \frac{\mu}{z^r} \sum_{k=r}^{\infty} p_k z^k \\ \Leftrightarrow\ &(\lambda+\mu) P(z) - \mu \sum_{n=0}^{r-1} p_n z^n = \lambda z P(z) + \frac{\mu}{z^r} \left(P(z) - \sum_{k=0}^{r-1} p_k z^k \right). \end{aligned}$$

Λύνοντας για την $P(z)$, έχουμε

$$\begin{aligned} P(z) &= \frac{\mu \sum_{n=0}^{r-1} p_n z^n \left(1 - \frac{1}{z^r}\right)}{\lambda + \mu - \lambda z - \frac{\mu}{z^r}} \\ &= \frac{\mu \sum_{n=0}^{r-1} p_n z^n (z^r - 1)}{(\lambda+\mu) z^r - \lambda z^{r+1} - \mu} \\ &= \frac{\sum_{n=0}^{r-1} p_n z^n (z^r - 1)}{(\rho+1) z^r - \rho z^{r+1} - 1}, \end{aligned}$$

όπου $\rho = \lambda/\mu$.

Έστω $N(z)$ και $D(z)$ ο αριθμητής και ο παρονομαστής του παραπάνω κλάσματος. Γνωρίζουμε ότι η παραπάνω πιθανογεννήτρια συγκλίνει για τιμές του z στον κλειστό μοναδιαίο δίσκο, όταν δηλαδή $|z| \leq 1$. Παρατηρούμε ότι ο βαθμός του παρονομαστή $D(z) = (1+\rho)z^r - \rho z^{r+1} - 1$ είναι $r+1$, και συνεπώς θα έχει $r+1$ ρίζες, έστω τις $z_0, z_1, ..., z_r$. Ας υποθέσουμε ότι έχει μια ρίζα στο εσωτερικό του μοναδιαίου δίσκου, και έστω, χωρίς βλάβη της γενικότητας, ότι είναι η ρίζα z_0, με $|z_0| < 1$. Τότε η $P(z)$ θα έχει πόλο στο z_0, αν $N(z_0) \neq 0$, το οποίο είναι άτοπο, εφόσον η πιθανογεννήτρια $P(z)$ συγκλίνει $\forall \ \ z \in \{z \ : |z| \leq 1\}$.

Συμπεραίνουμε λοιπόν ότι η ρίζα z_0 θα πρέπει να είναι και ρίζα του αριθμητή $N(z) = (z^r - 1)\sum_{n=0}^{r-1} p_n z^n$. Κατ΄ επέκταση κάθε ρίζα του παρονομαστή $D(z) = (1+r)z^r - rz^{r+1} - 1$ που βρίσκεται στον μοναδιαίο δίσκο, δηλαδή κάθε $z_i \in \{z \ : |z| \leq 1\}$, θα πρέπει να είναι και ρίζα του αριθμητή $N(z) = (z^r - 1)\sum_{n=0}^{r-1} p_n z^n$.

Για να προχωρήσουμε, λοιπόν στην απλοποίηση της μορφής της $P(z)$ πρέπει να βρούμε τις ρίζες του παρονομαστή $D(z)$ που βρίσκονται στον μοναδιαίο δίσκο, δηλαδή πρέπει να προσδιορίσουμε το πλήθος των $z_i \in \{z \ : \ |z| \leq 1\}$. Τη λύση στο πρόβλημα μας δίνει το παρακάτω θεώρημα που προκύπτει ως εφαρμογή του Θεωρήματος Rouché από τη Μιγαδική Ανάλυση:

Θεώρημα 6.1 (Πόρισμα Θεωρήματος Rouché) *Έστω μια ακολουθία $\{a_n\}_{n\in\mathbb{N}_0}$ με $a_n \geq 0$ για την οποία ισχύουν*

$$\sum_{n=0}^{\infty} a_n \leq 1, \quad \sum_{n=0}^{\infty} n a_n < \infty \quad \text{και} \quad \sum_{n=0}^{\infty} n^2 a_n < \infty.$$

Έστω N θετικός ακέραιος. Θέτουμε

$$a(z) = \sum_{n=0}^{\infty} a_n z^n, \quad A(z) = z^N - a(z) \quad \text{και}$$

$$k = \text{ΜΚΔ}\{j - N \mid \text{για τους δείκτες } j \text{ που } a_j \neq 0\}.$$

Τότε, σχετικά με τα πλήθη των ριζών της $A(z)$ στα σύνολα $\{z : |z| < 1\}$, $\{z : |z| = 1\}$ και $\{z : |z| > 1\}$ έχουμε:

Συνθήκες	*ρίζες* $\in \{\lvert z\rvert < 1\}$	*ρίζες* $\in \{\lvert z\rvert = 1\}$	*ρίζες* $\in \{\lvert z\rvert \leq 1\}$
$a(1) < 1$	N	0	N
$a(1) = 1,\ A'(1) > 0$	$N-k$	k *(απλές, τις k-οστές ρίζες της 1)*	N
$a(1) = 1,\ A'(1) = 0$	$N-k$	k *(διπλές, τις k-οστές ρίζες της 1)*	$N+k$
$a(1) = 1,\ A'(1) < 0$	N	k *(απλές, τις k-οστές ρίζες της 1)*	$N+k$

Σκοπός μας είναι τώρα να απλοποιήσουμε την πιθανογεννήτρια $P(z)$, χρησιμοποιώντας το θεώρημα αυτό, για να μελετήσουμε το πλήθος των ριζών του παρονομαστή $D(z)$ στον κλειστό μοναδιαίο δίσκο. Βλέπουμε ότι ο παρονομαστής $D(z)$ γράφεται στη μορφή

$$D(z) = (1+\rho)\left(z^r - \frac{1}{1+\rho} - \frac{\rho}{1+\rho}z^{r+1}\right).$$

Για να εφαρμόσουμε το θεώρημα παίρνουμε $N = r$ και ορίζουμε την ακολουθία $(a_n)_{n\in\mathbb{N}_0}$ με $a_0 = \frac{1}{1+\rho}$, $a_1 = a_2 = \cdots = a_r = 0$, $a_{r+1} = \frac{\rho}{1+\rho}$ και $a_n = 0$ για $n \geq r+2$. Τότε για την ακολουθία που ορίσαμε ισχύουν οι υποθέσεις του θεωρήματος:

$$\sum_{n=0}^{\infty} a_n = 1, \quad \sum_{n=0}^{\infty} n a_n < \infty \quad \text{και} \quad \sum_{n=0}^{\infty} n^2 a_n < \infty,$$

και θέτουμε

$$a(z) = \sum_{n=0}^{\infty} a_n z^n = \frac{1}{1+\rho} + \frac{\rho}{1+\rho}z^{r+1}, \quad A(z) = z^r - a(z).$$

Εδώ,

$$\begin{aligned} k &= \text{ΜΚΔ}\{j - r \mid \text{για τους δείκτες } j \text{ που } a_j \neq 0\} \\ &= \text{ΜΚΔ}\{j - r \mid \text{για } j = 0 \text{ και } j = r+1\} = \text{ΜΚΔ}\{-r, 1\} = 1. \end{aligned}$$

Για να δούμε σε ποιά από τις τέσσερεις περιπτώσεις του θεωρήματος εμπίπτει η συγκεκριμένη κατάσταση, υπολογίζουμε την $a(1)$ και $A'(1)$. Είναι

$$a(1) = \frac{1}{1+r} + \frac{r}{1+r} = 1$$

και

$$A'(1) = r - (r+1)\frac{\rho}{1+\rho} = \frac{r-\rho}{1+\rho}.$$

Επομένως διακρίνουμε τρεις περιπτώσεις

(i) $A'(1) > 0 \;\Leftrightarrow \rho < r$. Στην περίπτωση αυτή εφαρμόζοντας το θεώρημα έπεται ότι το πολυώνυμο $D(z)$ έχει $r-1$ ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις $z_1, z_2, ..., z_{r-1}$, δηλαδή $z_i \in \{z : |z| < 1\}$ για $i = 1, \ldots, r-1$, έχει 1 ρίζα τη μονάδα, έστω $z_r = 1$, και άρα θα έχει μια ρίζα, έστω z_0, με $|z_0| > 1$, άρα $D(z) = c_1(z - z_0)(z - z_1)\cdots(z - z_{r-1})(z - 1)$, όπου c_1 κάποια σταθερά. Επειδή η $P(z)$ συγκλίνει στον κλειστό μοναδιαίο δίσκο, ως πιθανογεννήτρια, συνάγουμε ότι ο αριθμητής $N(z)$ έχει ρίζες τις $z_1, z_2, ..., z_{r-1}$, καθώς και την 1. Αλλά οι ρίζες του $z^r - 1$ είναι οι r-οστές ρίζες της μονάδας που έχουν όλες μέτρο 1, οπότε οι ρίζες $z_1, z_2, ..., z_{r-1}$ είναι αναγκαστικά ρίζες του άλλου παράγοντα του αριθμητή $N(z)$, δηλαδή του $\sum_{n=0}^{r-1} p_n z^n$ που έχει βαθμό $r-1$, όσο δηλαδή και το πλήθος των $z_1, z_2, ..., z_{r-1}$. Επομένως έχουμε αναγκαστικά ότι $\sum_{n=0}^{r-1} p_n z^n = c_2(z-z_1)(z-z_2)\cdots(z-z_{r-1})$, όπου c_2 κάποια σταθερά, οπότε $N(z) = c_2(z^r-1)(z-z_1)(z-z_2)...(z-z_{r-1})$. Τελικά:

$$\begin{aligned} P(z) = \frac{N(z)}{D(z)} &= \frac{c_2(z^r-1)(z-z_1)(z-z_2)\cdots(z-z_{r-1})}{c_1(z-z_0)(z-z_1)(z-z_2)\cdots(z-z_{r-1})(z-1)} \\ &= c\frac{(z^r-1)}{(z-z_0)(z-1)}, \end{aligned}$$

όπου $c = c_2/c_1$ μια σταθερά που θα προσδιοριστεί από την εξίσωση κανονικοποίησης $P(1) = 1$. Πράγματι, χρησιμοποιώντας τη σχέση αυτή και εφαρμόζοντας τον κανόνα του $L'Hospital$ παίρνουμε $c = \frac{1-z_0}{r}$. Επομένως η $P(z)$ παίρνει την απλή μορφή

$$P(z) = \frac{(z_0-1)(z^r-1)}{r(z-z_0)(z-1)},$$

όπου η z_0 είναι η μοναδική ρίζα του $D(z)$ με μέτρο μεγαλύτερο της μονάδας. Για τον υπολογισμό της z_0, παρατηρείστε ότι $D(z_0) = 0$ σημαίνει ότι

$$z_0^r - \frac{1}{1+\rho} - \frac{\rho}{1+\rho} z_0^{r+1} = 0$$

ή ισοδύναμα (διαιρώντας με z_0^{r+1})

$$\frac{1}{z_0} - \frac{1}{1+\rho}\left(\frac{1}{z_0}\right)^{r+1} - \frac{\rho}{1+\rho} = 0.$$

Επομένως, η z_0 μπορεί να χαρακτηριστεί ως η αντίστροφη της μοναδικής ρίζας του πολυωνύμου $f(z) = z - \frac{1}{1+\rho}z^{r+1} - \frac{\rho}{1+\rho}$ με μέτρο μικρότερο της μονάδας. Όμως, έχουμε $f(0) < 0$, $f(1) = 0$ και $f'(1) = -\frac{r}{\rho+1} < 0$ οπότε το $f(z)$ έχει πραγματική ρίζα στο $(0,1)$ που μπορεί να βρεθεί με τη μέθοδο της διχοτόμησης.

Έχοντας προσδιορίσει την z_0 με αυτό τον τρόπο, μπορούμε να αντιστρέψουμε εύκολα την πιθανογεννήτρια $P(z)$ και να προσδιορίσουμε τη στάσιμη κατανομή (p_n). Έχουμε:

$$\begin{aligned} P(z) &= \frac{z_0-1}{z-z_0} \cdot \frac{z^r-1}{r(z-1)} \\ &= \frac{1-z_0^{-1}}{1-z_0^{-1}z} \cdot \frac{z^r-1}{r(z-1)}. \end{aligned}$$

Αναγνωρίζοντας ότι ο πρώτος παράγοντας είναι η πιθανογεννήτρια της $Geom(z_0^{-1})$ με συνάρτηση πιθανότητας $a_n = (1-z_0^{-1})(z_0^{-1})^n$, $n \geq 0$ και ο δεύτερος παράγοντας είναι η πιθανογεννήτρια της $Uniform(\{0,1,\ldots,r-1\})$ με συνάρτηση πιθανότητας $b_n = 1/r$, για $0 \leq n \leq r-1$ και $b_n = 0$, για $n \geq r$, έχουμε ότι:

$$\begin{aligned} p_n &= \sum_{k=0}^{n} b_k a_{n-k} \\ &= \sum_{k=0}^{min(n,r-1)} \frac{1}{r}(1 - z_0^{-1})(z_0^{-1})^{n-k}. \end{aligned}$$

(ii) $A'(1) = 0 \Leftrightarrow \rho = r$. Στην περίπτωση αυτή, εφαρμόζοντας το θεώρημα, έπεται ότι το πολυώνυμο $D(z)$ έχει $r-1$ ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις $z_1, z_2, ..., z_{r-1}$, με $z_i \in \{z : |z| < 1\}$ για $i = 1, \dots, r-1$, και έχει μία διπλή ρίζα τη μονάδα. Δεδομένου ότι το $D(z)$ είναι βαθμού $r+1$ παραγοντοποιείται αναγκαστικά ως $D(z) = c_1(z - z_1)...(z - z_{r-1})(z-1)^2$. Για να συγκλίνει η $P(z)$ στον κλειστό μοναδιαίο δίσκο θα πρέπει ο αριθμητής $N(z)$ να έχει ρίζες τις $z_1, z_2, ..., z_{r-1}$, καθώς και διπλή ρίζα τη μονάδα. Οι ρίζες του $z^r - 1$ είναι οι r-οστές ρίζες της μονάδας που έχουν όλες μέτρο 1, οπότε οι ρίζες $z_1, z_2, ..., z_{r-1}$ είναι αναγκαστικά ρίζες του άλλου παράγοντα του αριθμητή $N(z)$, δηλαδή του $\sum_{n=0}^{r-1} p_n z^n$ που έχει βαθμό $r-1$, όσο δηλαδή και το πλήθος των $z_1, z_2, ..., z_{r-1}$. Επομένως έχουμε αναγκαστικά ότι $\sum_{n=0}^{r-1} p_n z^n = c_2(z-z_1)(z-z_2)\cdots(z-z_{r-1})$, όπου c_2 κάποια σταθερά, οπότε $N(z) = (z^r-1)\sum_{n=0}^{r-1} p_n z^n = c_2(z^r-1)(z-z_1)...(z-z_{r-1})$. Όμως το 1 δεν μπορεί να είναι διπλή ρίζα του $N(z)$, οπότε στην περίπτωση αυτή το μόνο που είναι δυνατό είναι το $N(z)$ να είναι ταυτοτικά 0, οπότε και $P(z) = 0$. Δηλαδή στην περίπτωση αυτή το σύστημα δεν είναι ευσταθές.

(iii) $A'(1) < 0 \Leftrightarrow \rho > r$. Στην περίπτωση αυτή, εφαρμόζοντας το θεώρημα, έπεται ότι το πολυώνυμο $D(z)$ έχει r ρίζες στο εσωτερικό του μοναδιαίου δίσκου, έστω τις $z_1, z_2, ..., z_r$ και έχει μία απλή ρίζα τη μονάδα, άρα $D(z) = c_1(z - z_1)...(z - z_r)(z-1)$. Από τον προηγούμενο ισχυρισμό απαιτούμε ο αριθμητής $N(z)$ να έχει ρίζες τις $z_1, z_2, ..., z_r$, καθώς και το 1. Θα έπρεπε (ανάλογα με την προηγούμενη περίπτωση) το πολυώνυμο $r-1$-οστού βαθμού $\sum_{n=0}^{r-1} p_n z^n = (z^r - 1)c_1(z - z_1)...(z - z_r)$ να έχει r ρίζες, δηλαδή να είναι ταυτοτικά 0. Οπότε στην περίπτωση αυτή έχουμε $P(z) = 0$ και το σύστημα δεν είναι ευσταθές.

Βλέπουμε, λοιπόν, ότι το πόρισμα του Θεωρήματος του Rouché μας επιτρέπει να χαρακτηρίσουμε πότε το σύστημα είναι ευσταθές και επιπλέον να βρούμε την αντίστοιχη οριακή κατανομή. Η προκύπτουσα συνθήκη ευστάθειας είναι φυσιολογική: Για να είναι το σύστημα ευσταθές θα πρέπει ο ρυθμός αφίξεων να είναι μικρότερος από τον μέγιστο ρυθμό εξυπηρέτησης του συστήματος, δηλαδή θα πρέπει $\lambda < r\mu$ ή ισοδύναμα $\rho < r$.

6.4 Η $M/M/\infty$ ουρά με ομαδικές αφίξεις

Θεωρούμε την $M/M/\infty$ ουρά με ομαδικές αφίξεις και υποθέτουμε ότι τα μεγέθη των διαδοχικά αφικνούμενων ομάδων είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές. Πιο συγκεκριμένα μια αφικνούμενη ομάδα είναι μεγέθους j με πιθανότητα

g_j. Έστω, λοιπόν μια τέτοια ουρά, με ρυθμό αφίξεων λ, συνάρτηση πιθανότητας μεγέθους αφικνούμενων ομάδων $(g_j : j \geq 1)$, $Exp(\mu)$ χρόνους εξυπηρέτηση και άπειρους υπηρέτες (επομένως και η χωρητικότητα είναι απεριόριστη). Για τη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα βλέπουμε ότι για τις μεταβάσεις της έχουμε τον παρακάτω πίνακα:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$k \geq 1$	$T_{0k} \sim Exp(\lambda g_k)$
$n \geq 1$	$n+k,\ k \geq 1$ $n-1$	$T_{n,n+k} \sim Exp(\lambda g_k)$ $T_{n,n-1} \sim Exp(n\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί έχουμε ότι η $\{Q(t)\}$ είναι Μ.α.ς.χ. με χώρο καταστάσεων $\mathcal{S} = \mathbb{N}_0$. Οι ρυθμοί μετάβασης είναι

$$q_{ij} = \begin{cases} \lambda g_k & \text{αν } i \geq 0,\ j = i+k, \\ i\mu & \text{αν } i \geq 1,\ j = i-1, \\ 0 & \text{διαφορετικά.} \end{cases}$$

Έστω $(p_n : n \geq 0)$ η οριακή κατανομή της $\{Q(t)\}$. Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned} \lambda p_0 &= \mu p_1, \\ (\lambda + n\mu)p_n &= \sum_{i=0}^{n-1} \lambda g_{n-i} p_i + (n+1)\mu p_{n+1},\ n \geq 1. \end{aligned}$$

Για την επίλυση αυτών των εξισώσεων και τον προσδιορισμό της οριακής κατανομής, ακολουθούμε και πάλι τη μέθοδο των πιθανογεννητριών. Έστω οι πιθανογεννήτριες

$$\begin{aligned} P(z) &= \sum_{n=0}^{\infty} p_n z^n,\ |z| \leq 1, \\ G(z) &= \sum_{n=0}^{\infty} g_n z^n,\ |z| \leq 1. \end{aligned}$$

Για το μετασχηματισμό των εξισώσεων ισορροπίας, πολλαπλασιάζουμε την εξίσωση που αναφέρεται στην κατάσταση n με z^n και κατόπιν αθροίζουμε για όλα τα $n \geq 0$. Έτσι έχουμε:

$$\lambda p_0 + \sum_{n=1}^{\infty} (\lambda + n\mu)p_n z^n = \mu p_1 + \sum_{n=1}^{\infty}\sum_{i=0}^{n-1} \lambda g_{n-i} p_i z^n + \sum_{n=1}^{\infty} (n+1)\mu p_{n+1} z^n.$$

Όπως στην ανάλυση της $M/M/1$ ουράς με ομαδικές αφίξεις, έχουμε

$$\sum_{n=1}^{\infty}\sum_{i=0}^{n-1} g_{n-i} p_i z^n = P(z)G(z).$$

Για να προχωρήσουμε, βλέπουμε ότι πρέπει να εκφράσουμε με όρους πιθανογεννητριών τις σειρές $\sum_{n=1}^{\infty} n p_n z^n$ και $\sum_{n=0}^{\infty} (n+1)p_{n+1} z^n$. Παραγωγίζοντας την $P(z)$, παίρνουμε

$$P'(z) = \sum_{n=1}^{\infty} n p_n z^{n-1},$$

οπότε

$$\sum_{n=1}^{\infty} n p_n z^n = z P'(z) \text{ και } \sum_{n=0}^{\infty} (n+1) p_{n+1} z^n = P'(z).$$

Επομένως, η εξίσωση που προέκυψε από το μετασχηματισμό των εξισώσεων ισορροπίας παίρνει τη μορφή:

$$\begin{aligned} &\lambda P(z) + \mu z P'(z) = \lambda P(z) G(z) + \mu P'(z) \\ \Leftrightarrow\ &\lambda(1 - G(z)) P(z) = \mu (1 - z) P'(z), \end{aligned}$$

δηλαδή αυτή τη φορά η μέθοδος των πιθανογεννητριών έδωσε μια διαφορική εξίσωση για την $P(z)$. Αυτό συμβαίνει γενικά όταν υπάρχουν ρυθμοί $q_{i,i-1}$ ανάλογοι του i, όπως εδώ που έχουμε άπειρους υπηρέτες ή παρόμοιες καταστάσεις. Η προκύπτουσα διαφορική εξίσωση είναι ομογενής γραμμική πρώτης τάξης και λύνεται εύκολα:

$$\begin{aligned} &\frac{P'(z)}{P(z)} = \frac{\lambda}{\mu} \cdot \frac{1 - G(z)}{1 - z} \\ \Leftrightarrow\ &\frac{d \log P(z)}{dz} = \frac{\lambda}{\mu} \cdot \frac{1 - G(z)}{1 - z} \\ \Leftrightarrow\ &\log P(1) - \log P(z) = \frac{\lambda}{\mu} \int_z^1 \frac{1 - G(u)}{1 - u} du. \end{aligned}$$

Λύνοντας για την $P(z)$, έχουμε

$$P(z) = \exp\left(-\frac{\lambda}{\mu} \int_z^1 \frac{1 - G(u)}{1 - u} du \right).$$

Από την πιθανογεννήτρια μπορούμε θεωρητικά να ανακτήσουμε την οριακή κατανομή p_n, χρησιμοποιώντας τη σχέση

$$p_n = \frac{P^{(n)}(0)}{n!},\ n \geq 0,$$

και να πάρουμε τις παραγοντικές ροπές του οριακού αριθμού πελατών στο σύστημα:

$$E[Q_{(n)}] = E[Q(Q-1)(Q-2)\cdots(Q-n+1)] = P^{(n)}(1),\ n \geq 1.$$

Βέβαια οι σχέσεις αυτές στην πράξη μπορούν να χρησιμοποιηθούν μόνο για μικρά n, π.χ. όταν μας ενδιαφέρει η πιθανότητα κενού συστήματος ή η μέση τιμή και η διασπορά του αριθμού των πελατών στο σύστημα. Σε λίγες ειδικές περιπτώσεις για την κατανομή του μεγέθους αφικνούμενων ομάδων μπορούμε να ανακτήσουμε την οριακή κατανομή σε κλειστή μορφή.

6.5 Ασκήσεις

1. Έστω μια $M/M/1$ ουρά με ομαδικές εξυπηρετήσεις σταθερού μεγέθους, όπου οι πελάτες φθάνουν σύμφωνα με μια διαδικασία Poisson με ρυθμό λ, οι χρόνοι εξυπηρέτησης είναι $Exp(\mu)$, υπάρχει 1 υπηρέτης που εξυπηρετεί ταυτόχρονα r πελάτες ως μια ομάδα και η χωρητικότητα του συστήματος είναι απεριόριστη. Θεωρούμε την τροποποίησή της με τις ίδιες παραμέτρους, όπου ο υπηρέτης δεν περιμένει για τη συμπλήρωση r πελατών, αλλά εξυπηρετεί όταν υπάρχει τουλάχιστον ένας πελάτης στο σύστημα. Αν ένας χρόνος εξυπηρέτησης τελειώσει και υπάρχουν λιγότεροι από r πελάτες στο σύστημα, τότε αναχωρούν όλοι, ενώ αν υπάρχουν περισσότεροι από r, τότε αναχωρούν ακριβώς r.

    1. Να γραφούν οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή (p_n) του αριθμού των πελατών στο σύστημα σε συνεχή χρόνο.
    2. Αποδείξτε ότι όταν $\lambda < r\mu$, τότε το σύστημα είναι ευσταθές και βρείτε τη στάσιμη κατανομή του. Τι είδους κατανομή είναι;

2. Θεωρούμε την $M/M/1$ ουρά με ομαδικές αφίξεις και μεμονωμένες εξυπηρετήσεις, όπου ομάδες πελατών μεγέθους 2 φθάνουν σύμφωνα με μια στοχαστική διαδικασία Poisson με ρυθμό λ, οι χρόνοι εξυπηρέτησης είναι $Exp(\mu)$, υπάρχει ένας υπηρέτης και απεριόριστη χωρητικότητα. Έστω $\{Q(t)\}$ η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα.

    1. Να αιτιολογήσετε γιατί η $\{Q(t)\}$ είναι Μ.α.ς.χ. και να κάνετε το διάγραμμα των ρυθμών μετάβασής της.
    2. Να υπολογίσετε την πιθανογεννήτρια $P(z)$ της στάσιμης κατανομής (p_n) της $\{Q(t)\}$, συναρτήσει των λ, μ και p_0.
    3. Να βρεθεί η συνθήκη ευστάθειας του συστήματος και να υπολογιστεί η p_0.
    4. Να βρεθεί συναρτήσει των πιθανοτήτων p_n, η πιθανότητας ένας πελάτης αμέσως μετά την άφιξή του να καταλάβει την n-οστή θέση στο σύστημα (δηλαδή να έχει μπροστά του $n-1$ πελάτες).
    5. Για $\lambda = 1$ και $\mu = 6$ να βρείτε έναν γενικό τύπο για την p_n.

3. Θεωρούμε την $M/M/1$ ουρά με ομαδικές αφίξεις και μεμονωμένες εξυπηρετήσεις. Οι ομάδες φθάνουν στο σύστημα με ρυθμό λ και το μέγεθός τους έχει συνάρτηση πιθανότητας $(g_j : j \geq 1)$. Μια ομάδα που βρίσκει κατά την άφιξή της το σύστημα κενό εισέρχεται με πιθανότητα 1 σε αυτό, ενώ αν βρει έστω κι έναν πελάτη εισέρχεται με πιθανότητα p. Ο ρυθμός εξυπηρέτης είναι μ. Έστω $\{Q(t)\}$ η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα.

    1. Να αιτιολογήσετε γιατί η $\{Q(t)\}$ είναι Μ.α.ς.χ. και να κάνετε το διάγραμμα των ρυθμών μετάβασής της. Γράψτε τις εξισώσεις ισορροπίας για τη στάσιμη κατανομή της, (p_n).
    2. Να υπολογίσετε την πιθανογεννήτρια $P(z)$ της στάσιμης κατανομής και τη συνθήκη ευστάθειας του συστήματος.
    3. Να βρεθεί συναρτήσει των (p_n) και (g_n), η πιθανότητα ένας πελάτης αμέσως

μετά την άφιξή του στο σύστημα να καταλάβει την n-οστή θέση (να έχει μπροστά του $n-1$ πελάτες).

4. Στην περίπτωση των μεμονωμένων αφίξεων (δηλαδή όταν $g_1 = 1$ και $g_j = 0$ για $j \geq 2$) να βρεθεί η στάσιμη κατανομή (p_n).

4. Θεωρούμε την $M/M/1/2$ ουρά με ομαδικές αφίξεις, όπου ο ρυθμός αφίξεων ομάδων είναι λ, ο ρυθμός εξυπηρέτησης είναι μ και κάθε αφικνούμενη ομάδα έχει μέγεθος 1 με πιθανότητα $\frac{1}{2}$ και μέγεθος 2 με πιθανότητα $\frac{1}{2}$. Σε περίπτωση που μια αφικνούμενη ομάδα δεν χωράει να εισέλθει ολόκληρη στο σύστημα, τότε απορρίπτεται ολόκληρη (κανένας πελάτης - μέλος της δεν εισέρχεται).
   1. Να γραφούν οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή (p_n) του αριθμού των πελάτων στο σύστημα σε συνεχή χρονο.
   2. Να υπολογιστεί η (p_n).
   3. Να βρεθεί το ποσοστό των χαμένων πελατών.
   4. Να απαντηθούν τα ίδια ερωτήματα για το τροποποιημένο μοντέλο, όπου μια αφικνούμενη ομάδα που δεν χωράει να εισέλθει ολόκληρη στο σύστημα δεν απορρίπτεται, αλλά εισέρχονται όσοι πελάτες της μπορούν να χωρέσουν σε αυτό.

7

Αντιστρέψιμες Μαρκοβιανές ουρές

Η αντιστρεψιμότητα είναι μια ιδιότητα που μπορεί να έχει μια Μαρκοβιανή αλυσίδα, η οποία διευκολύνει πολύ τους υπολογισμούς που σχετίζονται με αυτήν, όταν υπάρχει. Στο κεφάλαιο αυτό θα δούμε τις έννοιες της αντίστροφης μιας Μ.α.ς.χ. και της αντιστρέψιμης Μ.α.ς.χ., καθώς και πως οι έννοιες αυτές μπορούν να χρησιμεύσουν για τη μελέτη κάποιων συστημάτων εξυπηρέτησης.

7.1 Αντίστροφη Μαρκοβιανής αλύσίδας

Έστω $\{X(t)\}$ μια Μ.α.ς.χ. με ρυθμούς q_{ij} και στάσιμη κατανομή (p_i). Η αντίστροφη $\{\hat{X}_\tau(t)\}$ της $\{X(t)\}$ ως προς τ, ορίζεται ως

$$\hat{X}_\tau(t) = X(\tau - t),\ t \in \mathbb{R}.$$

Διαισθητικά, η $\{\hat{X}_\tau(t)\}$ προκύπτει αν πάρω την $\{X(t)\}$, θεωρήσω ως αρχή του χρόνου τη στιγμή τ και αντιστρέψω τη φορά του χρόνου στην $\{X(t)\}$. Για $\tau = 0$, η $\hat{X}_0(t)$ λέγεται τυπική αντίστροφη της $\{X(t)\}$ και συμβολίζεται απλά με $\hat{X}(t)$. Το ακόλουθο αποτέλεσμα συνοψίζει κάποιες σημαντικές ιδιότητες της αντίστροφης μιας στάσιμης αδιαχώριστης Μ.α.ς.χ.

Θεώρημα 7.1 (Ιδιότητες αντίστροφης Μ.α.ς.χ.) *Η αντίστροφη $\hat{X}(t)$ μιας στάσιμης αδιαχώριστης Μ.α.ς.χ. $X(t)$ με χώρο καταστάσεων $\mathcal{S}$, ρυθμούς μετάβασης q_{ij} και στάσιμη κατανομή (p_i) είναι επίσης Μ.α.ς.χ. με ρυθμούς $\hat{q}_{ij} = \frac{p_j q_{ji}}{p_i}$ και στάσιμη κατανομή $(\hat{p}_i)$ με $\hat{p}_i = p_i$, για κάθε i, j.*

Μπορούμε να δικαιολογήσουμε εύκολα αυτές τις σχέσεις. Πράγματι, όσον αφορά τις οριακές πιθανότητες της αντίστροφης έχουμε

$$\hat{p}_i = \lim_{t\to\infty} \Pr[\hat{X}(t) = i] = \lim_{t\to\infty} \Pr[X(-t) = i] = p_i,\ i \in \mathcal{S},$$

ενώ όσον αφορά τους ρυθμούς μετάβασής της παίρνουμε

$$\begin{aligned}\hat{q}_{ij} &= \lim_{h\to 0^+} \frac{\Pr[\hat{X}(t+h) = j|\hat{X}(t) = i]}{h} \\ &= \lim_{h\to 0^+} \frac{\Pr[X(-t-h) = j|X(-t) = i]}{h},\ i, j \in \mathcal{S}, i \neq j.\end{aligned}$$

Χρησιμοποιώντας τον κανόνα του Bayes έχουμε

$$\begin{aligned}\hat{q}_{ij} &= \lim_{h\to 0^+} \frac{\Pr[X(-t-h)=j]\Pr[X(-t)=i|X(-t-h)=j]}{\Pr[X(-t)=i]h} \\ &= \frac{p_j q_{ji}}{p_i},\ i,j\in\mathcal{S}, i\neq j.\end{aligned}$$

Τα αποτελέσματα είναι εύλογα και από διαισθητική σκοπιά. Πράγματι, η πιθανότητα $\hat{p}_i$ είναι το μακροπρόθεσμο ποσοστό του χρόνου που η Μαρκοβιανή αλυσίδα $\{\hat{X}(t)\}$ περνάει στην κατάσταση i, ενώ η πιθανότητα p_i είναι το μακροπρόθεσμο ποσοστό του χρόνου που η $\{X(t)\}$ περνάει στην κατάσταση i αν κοιτάμε το χρόνο με αντίστροφη φορά. Αυτά τα δυο ποσοστά είναι ίσα, οπότε έπεται ότι $\hat{p}_i = p_i$, $i\in\mathcal{S}$.

Επίσης, η ποσότητα $p_j q_{ji}$ δίνει το μακροπρόθεσμο ρυθμό μεταβάσεων της $\{X(t)\}$ από την j στην i ανά χρονική μονάδα. Κάθε τέτοια μετάβαση αντιστοιχεί σε μια μετάβαση της $\{\hat{X}(t)\}$ από την i στην j, αφού η $\{\hat{X}(t)\}$ προκύπτει κοιτάζοντας την $\{X(t)\}$ αντίστροφα στο χρόνο. Άρα θα πρέπει ο μακροπρόθεσμος ρυθμός μεταβάσεων της $\{X(t)\}$ από την j στην i ανά χρονική μονάδα να ισούται με τον μακροπρόθεσμο ρυθμό μεταβάσεων της $\{\hat{X}(t)\}$ από την i στην j ανά χρονική μονάδα, δηλαδή $p_j q_{ji} = \hat{p}_i\hat{q}_{ij}$. Λύνοντας ως προς την $\hat{q}_{ij}$, και χρησιμοποιώντας ότι $\hat{p}_i = p_i$, παίρνουμε $\hat{q}_{ij} = \frac{p_j q_{ji}}{p_i}$.

7.2 Αντιστρέψιμες Μαρκοβιανές αλυσίδες

Μια Μ.α.ς.χ. λέγεται αντιστρέψιμη αν η αντίστροφη της είναι όμοια με αυτήν όσον αφορά την πιθανοθεωρητική συμπεριφορά της. Πιο συγκεκριμένα, μια Μ.α.ς.χ. $\{X(t)\}$ λέγεται αντιστρέψιμη αν είναι στοχαστικά ισοδύναμη με την αντίστροφή της, $\{\hat{X}(t)\}$, δηλαδή για οποιαδήποτε πεπερασμένη ακολουθία χρονικών στιγμών $t_1 < t_2 < \cdots < t_n$ οι $(X(t_1), X(t_2), \ldots, X(t_n))$ και $(\hat{X}(t_1), \hat{X}(t_2), \ldots, \hat{X}(t_n))$ έχουν την ίδια κατανομή.

Ενώ κάθε Μ.α.ς.χ. έχει αντίστροφη, μόνο λίγες Μ.α.ς.χ. είναι αντιστρέψιμες. Πράγματι, όπως έχουμε δει, μια στάσιμη Μ.α.ς.χ. $\{X(t)\}$ χαρακτηρίζεται από τους ρυθμούς μετάβασής της, q_{ij}. Επομένως, η $\{X(t)\}$ είναι αντιστρέψιμη αν και μόνο αν $q_{ij} = \hat{q}_{ij}$, για κάθε $i,j\in\mathcal{S}$, δηλαδή αν

$$p_i q_{ij} = p_j q_{ji},\ i,j\in\mathcal{S},\ i\neq j.$$

Οι εξισώσεις αυτές αναφέρονται ως εξισώσεις λεπτομερούς ισορροπίας και ισχύουν μόνο για αντιστρέψιμες Μ.α.ς.χ., σε αντίθεση με τις εξισώσεις πλήρους ισορροπίας και τις εξισώσεις γενικευμένης ισορροπίας που ικανοποιεί η στάσιμη κατανομή κάθε Μ.α.ς.χ. Οι εξισώσεις λεπτομερούς ισορροπίας επιτρέπουν να συνδέει κανείς άμεσα δυο οποιεσδήποτε στάσιμες πιθανότητες και γι αυτό διευκολύνουν τον υπολογισμό της στάσιμης κατανομής μιας αντιστρέψιμης Μ.α.ς.χ. Είναι σημαντικό, επομένως, να μπορούμε να ελέγξουμε πότε μια Μ.α.ς.χ. είναι αντιστρέψιμη, χωρίς να χρειάζεται να υπολογίσουμε τη στάσιμη κατανομή της. Αυτό γίνεται με το ακόλουθο κριτήριο αντιστρεψιμότητας.

Θεώρημα 7.2 (Κριτήριο αντιστρεψιμότητας του Kolmogorov) *Μια Μ.α.ς.χ. είναι αντιστρέψιμη αν και μόνο αν για κάθε κύκλο καταστάσεων $(i_0, i_1, i_2, \ldots, i_n, i_0)$, το γινόμενο των ρυθμών μετάβασης είναι το ίδιο ως προς τις δυο φορές διαγραφής του κύκλου, δηλαδή αν*

$$q_{i_0 i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n i_0} = q_{i_0 i_n} q_{i_n i_{n-1}} q_{i_{n-1} i_{n-2}} \cdots q_{i_2 i_1} q_{i_1 i_0}.$$

Είναι εύκολο να δούμε ότι για να είναι μια Μ.α.ς.χ. αντιστρέψιμη, αρκεί να περιοριστούμε στον έλεγχο της ισότητας των γινομένων των ρυθμών μετάβασης, μόνο σε απλούς κύκλους καταστάσεων (δηλ. χωρίς επαναλήψεις καταστάσεων) με τουλάχιστον τρεις καταστάσεις. Ως άμεση εφαρμογή του κριτηρίου αντιστρεψιμότητας του Kolmogorov έχουμε ότι κάθε Μ.α.ς.χ. της οποίας το διάγραμμα ρυθμών μετάβασης είναι αμφίδρομο δέντρο (δηλαδή δεν υπάρχουν κύκλοι, και αν υπάρχει θετικός ρυθμός μετάβασης από μια κατάσταση i σε μια κατάσταση j υπάρχει και θετικός ρυθμός μετάβασης από την j στην i) είναι αντιστρέψιμη. Ειδικότερα κάθε Μ.α.ς.χ. γέννησης - θανάτου είναι αντιστρέψιμη.

Ο τρόπος με τον οποίο χρησιμοποιείται η έννοια της αντιστρεψιμότητας για τον υπολογισμό της στάσιμης κατανομής (p_n) μιας Μ.α.ς.χ. $\{X(t)\}$ είναι ο εξής:

Βήμα 1ο: Διαπιστώνεται ότι η $\{X(t)\}$ είναι αντιστρέψιμη με το κριτήριο Kolmogorov.

Βήμα 2ο: Ορίζουμε μια κατάσταση αναφοράς i_0 και για κάθε κατάσταση j βρίσκουμε ένα μονοπάτι που τη συνδέει με την i_0, έστω το $i_0 \to i_1 \to i_2 \to \cdots \to i_n \to j$. Τότε οι εξισώσεις λεπτομερούς ισορροπίας δίνουν

$$p_j = p_{i_0} \frac{q_{i_0 i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} q_{i_n j}}{q_{i_1 i_0} q_{i_2 i_1} q_{i_3 i_2} \cdots q_{i_n i_{n-1}} q_{j i_n}}.$$

Με άλλα λόγια, η p_j υπολογίζεται συναρτήσει της p_{i_0} σαν να είχαμε αλυσίδα γέννησης - θανάτου με χώρο καταστάσεων $\{i_0, i_1, \ldots, i_n, j\}$.

Βήμα 3ο: Η p_{i_0} υπολογίζεται από την εξίσωση κανονικοποίησης.

7.3 Η $M/M/c$ με ετερογενείς υπηρέτες

Θεωρούμε την $M/M/c$ ουρά με ετερογενείς υπηρέτης. Το μοντέλο αυτό είναι παρόμοιο με την $M/M/c$ ουρά, με μόνη διαφορά ότι κάθε υπηρέτης έχει έναν διαφορετικό ρυθμό εξυπηρέτησης. Πιο συγκεκριμένα, οι αφίξεις συμβαίνουν σύμφωνα με μια διαδικασία Poisson, υπάρχουν c υπηρέτες, οι χρόνοι εξυπηρέτησης των πελατών που εξυπηρετούνται από τον υπηρέτη i είναι $Exp(\mu_i)$, $i = 1, 2, \ldots, c$, και, η χωρητικότητα είναι απεριόριστη. Αν εξετάσουμε τη στοχαστική διαδικασία $\{Q(t)\}$ του αριθμού των πελατών στο σύστημα τότε βλέπουμε ότι για τις μεταβάσεις της έχουμε τον πίνακα

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	1	$T_{01} \sim Exp(\lambda)$
$n \geq c$	$n+1$ $n-1$	$T_{n,n+1} \sim Exp(\lambda)$ $T_{n,n-1} \sim Exp(\mu_1 + \mu_2 + \cdots + \mu_c)$
$1 \leq n < c$	$n+1$ $n-1$	$T_{n,n+1} \sim Exp(\lambda)$;;;

Πράγματι, όταν το πλήθος των πελατών στο σύστημα είναι μικρότερο από c, δηλαδή όταν δεν είναι όλοι οι υπηρέτες απασχολημένοι, χρειάζεται να ξέρουμε ποιοί υπηρέτες είναι απασχολημένοι για να προσδιορίσουμε το χρόνο ως την επόμενη αναχώρηση πελάτη. Επομένως, για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος, χρειάζεται εκτός από τον αριθμό των πελατών $Q(t)$ στο σύστημα να καταγράφουμε και το σύνολο των υπηρετών που είναι απασχολημένοι κάθε στιγμή t (φυσικά για $Q(t) = 0$ το σύνολο των απασχολημένων υπηρετών είναι κενό, ενώ για $Q(t) \geq c$ είναι το σύνολο όλων των υπηρετών, και επομένως, η καταγραφή του συνόλου των απασχολημένων υπηρετών είναι αναγκαία μόνο όταν $1 \leq Q(t) < c$).

Για να κάνουμε πιο συγκεκριμένη τη μελέτη μας, θα περιοριστούμε στην περίπτωση των δύο ετερογενών υπηρετών.

7.3.1 Η $M/M/2$ ουρά με ετερογενείς υπηρέτες

Θεωρούμε την $M/M/2$ ουρά με ετερογενείς υπηρέτες και υποθέτουμε ότι ένας πελάτης που βρίσκει το σύστημα κενό διαλέγει τον υπηρέτη 1 ή τον υπηρέτη 2, τυχαία, με πιθανότητες p και $q = 1 - p$, αντίστοιχα. Έστω $\{Q(t)\}$ η στοχαστική διαδικασία που καταγράφει τον αριθμό των πελατών. Για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος, σύμφωνα με τα παραπάνω, πρέπει να καταγράφουμε ποιός υπηρέτης είναι απασχολημένος στην περίπτωση που $Q(t) = 1$. Έστω, λοιπόν $C(t)$ ο υπηρέτης που είναι απασχολημένος τη στιγμή t. Η $\{(Q(t), C(t)\}$ είναι μια διαδικασία με χώρο καταστάσεων $\mathcal{S} = \{0, (1,1), (1,2), 2, 3, \ldots\}$ και για τις μεταβάσεις της έχουμε τον παρακάτω πίνακα

Κατάσταση	Επόμενη κατάσταση	Χρόνος
0	$(1,1)$	$T_{0,(1,1)} \sim Exp(\lambda p)$
	$(1,2)$	$T_{0,(1,2)} \sim Exp(\lambda q)$
$(1,1)$	2	$T_{(1,1),2} \sim Exp(\lambda)$
	0	$T_{(1,1),0} \sim Exp(\mu_1)$
$(1,2)$	2	$T_{(1,2),2} \sim Exp(\lambda)$
	0	$T_{(1,2),0} \sim Exp(\mu_2)$
2	3	$T_{23} \sim Exp(\lambda)$
	$(1,1)$	$T_{2,(1,1)} \sim Exp(\mu_2)$
	$(1,2)$	$T_{2,(1,2)} \sim Exp(\mu_1)$
$n \geq 3$	$n+1$	$T_{n,n+1} \sim Exp(\lambda)$
	$n-1$	$T_{n,n-1} \sim Exp(\mu_1+\mu_2)$

Η Μ.α.ς.χ. είναι μορφής γέννησης -θανάτου για τις καταστάσεις $2, 3, \ldots$. Άρα ο μοναδικός απλός κύκλος καταστάσεων με μήκος μεγαλύτερο του 2 είναι ο κύκλος καταστάσεων $0 \to (1,1) \to 2 \to (1,2) \to 0$. Από το κριτήριο του Kolmogorov, η Μ.α.ς.χ. θα είναι αντιστρέψιμη αν και μόνο αν το γινόμενο των ρυθμών μετάβασης είναι το ίδιο ως προς τις δυο φορές διαγραφής του κύκλου, δηλαδή αν

$$q_{0,(1,1)}q_{(1,1),2}q_{2,(1,2)}q_{(1,2),0} = q_{0,(1,2)}q_{(1,2),2}q_{2,(1,1)}q_{(1,1),0}.$$

Αυτή η συνθήκη ικανοποιείται μόνο αν $p = q$, δηλαδή αν οι πελάτες που βρίσκουν το σύστημα κενό επιλέγουν έναν υπηρέτη ισοπίθανα για να εξυπηρετηθούν. Στην περίπτωση αυτή, θεωρούμε την κατάσταση 0 ως την κατάσταση αναφοράς, και για κάθε κατάσταση βρίσκουμε ένα μονοπάτι που τη συνδέει με την 0. Έτσι, παίρνουμε εύκολα τη στάσιμη κατανομή ως

$$\begin{aligned} p_{(1,1)} &= \frac{\lambda}{2\mu_1}p_0, \\ p_{(1,2)} &= \frac{\lambda}{2\mu_2}p_0, \\ p_n &= \frac{\lambda^2}{2\mu_1\mu_2}\left(\frac{\lambda}{\mu_1+\mu_2}\right)^{n-2}p_0, \ n \geq 3. \end{aligned}$$

Η πιθανότητα p_0 υπολογίζεται από την εξίσωση κανονικοποίησης

$$\begin{aligned} & p_0 + p_{(1,1)} + p_{(1,2)} + \sum_{n=2}^{\infty} p_n = 1 \\ \Leftrightarrow\ & p_0\left(1 + \frac{\lambda}{2\mu_1} + \frac{\lambda}{2\mu_2} + \sum_{n=2}^{\infty}\frac{\lambda^2}{2\mu_1\mu_2}\left(\frac{\lambda}{\mu_1+\mu_2}\right)^{n-2}\right) = 1. \end{aligned}$$

Για να είναι το σύστημα ευσταθές θα πρέπει η άπειρη σειρά να συγκλίνει δηλαδή $\lambda < \mu_1 + \mu_2$, που εκφράζει και πάλι τη φυσική συνθήκη ο ρυθμός αφίξεων να είναι μικρότερος του μέγιστου ρυθμού εξυπηρέτησης που μπορεί να παρέχει το σύστημα. Υπολογίζοντας τη γεωμετρική σειρά, παίρνουμε

$$p_0 = \left(1 + \frac{\lambda}{2\mu_1} + \frac{\lambda}{2\mu_2} + \frac{\lambda^2}{2\mu_1\mu_2} \cdot \frac{\mu_1 + \mu_2}{\mu_1 + \mu_2 - \lambda}\right)^{-1}.$$

7.4 Η διαδικασία αναχωρήσεων σε απλές Μαρκοβιανές ουρές με Poisson διαδικασία αφίξεων

Η διαπίστωση ότι μια Μαρκοβιανή ουρά είναι αντιστρέψιμη επιτρέπει τον εύκολο προσδιορισμό της στάσιμης κατανομής της, αφού αυτή ικανοποιεί το ισχυρότερο σύστημα των εξισώσεων λεπτομερούς ισορροπίας. Πέραν, όμως, της υπολογιστικής ευκολίας που συνεπάγεται η αντιστρεψιμότητα, οδηγεί και σε ποιοτικά συμπεράσματα σχετικά με τη συμπεριφορά μιας ουράς. Το σημαντικότερο από αυτό είναι το Θεώρημα του Burke.

Θεώρημα 7.3 (Θεώρημα του Burke) *Η διαδικασία των αναχωρήσεων σε μια στάσιμη απλή Μαρκοβιανή ουρά με Poisson διαδικασία αφίξεων με ρυθμό λ, είναι επίσης διαδικασία Poisson με τον ίδιο ρυθμό. Επιπλέον για κάθε χρονική στιγμή t το πλήθος των πελατών τη στιγμή t είναι ανεξάρτητο από τη διαδικασία των αναχωρήσεων στο χρονικό διάστημα $(-\infty, t)$.*

Η ισχύς του θεωρήματος αυτού φαίνεται άμεσα αρκεί κανείς να συγκρίνει τη διαδικασία του αριθμού των πελατών $\{Q(t)\}$ και την αντίστροφή της $\{\hat{Q}(t)\}$. Πράγματι η διαδικασία αφίξεων της $\{Q(t)\}$ είναι Poisson και επιπλέον η εξέλιξή της στο $(-t, \infty)$ είναι ανεξάρτητη της $Q(-t)$ για κάθε t. Επομένως, αφού η $\{Q(t)\}$ είναι αντιστρέψιμη, τα ίδια ισχύουν και για την $\{\hat{Q}(t)\}$, δηλαδή η διαδικασία των αφίξεων της $\{\hat{Q}(t)\}$ είναι Poisson και επιπλέον η εξέλιξή της στο $(-t, \infty)$ είναι ανεξάρτητη της $\hat{Q}(-t)$ για κάθε t. Μπορούμε τώρα να μεταφράσουμε αυτές τις δυο ιδιότητες με όρους της αρχικής διαδικασίας $\{Q(t)\}$ και να συναγάγουμε το συμπέρασμα. Πράγματι κάθε άφιξη στην $\{\hat{Q}(t)\}$ αντιστοιχεί σε αναχώρηση στην $\{Q(t)\}$. Επομένως η διαδικασία των αφίξεων της $\{\hat{Q}(t)\}$ είναι στοχαστικά ισοδύναμη με τη διαδικασία των αναχωρήσεων της $\{Q(t)\}$. Επομένως, η διαδικασία των αναχωρήσεων της $\{Q(t)\}$ είναι Poisson με τον ρυθμό λ. Επίσης, αφού η εξέλιξη της $\{\hat{Q}(t)\}$ στο $(-t, \infty)$ είναι ανεξάρτητη της $\hat{Q}(-t)$ για κάποιο t, έχουμε ότι η εξέλιξη της $\{Q(t)\}$ στο $(-\infty, t)$ είναι ανεξάρτητη της $Q(t)$.

7.5 Ασκήσεις

1. Βρείτε τη στάσιμη κατανομή του αριθμού των πελατών στο σύστημα και το ποσοστό των χαμένων πελατών σε μια $M/M/2/2$ ουρά με ετερογενείς υπηρέτες, οι οποίοι επιλέονται ισοπίθανα όταν είναι και οι δυο ελεύθεροι. Δίνονται ο ρυθμός αφίξεων λ και οι ρυθμοί εξυπηρέτησης μ_1 και μ_2 των δυο υπηρετών.

2. Θεωρούμε μια $M/M/1$ ουρά με ρυθμό αφίξεων λ και $Exp(\mu)$ χρόνους εξυπηρέτησης, στην οποία ο υπηρέτης υπόκειται σε διαδοχικές βλάβες κι επισκευές και επομένως εναλλάσσεται μεταξύ περιόδων λειτουργίας και περιόδων αργίας (επισκευής). Οι περίοδοι λειτουργίας του υπηρέτη έχουν την $Exp(\theta)$ κατανομή, ενώ οι περίοδοι αργίας (επισκευής) την $Exp(\xi)$ κατανομή. Όταν ο υπηρέτης λειτουργεί οι αφίξεις και οι εξυπηρετήσεις γίνονται κανονικά, όπως σε μια συνήθη $M/M/1$ ουρά, ενώ όταν δεν λειτουργεί σταματούν τόσο οι αφίξεις όσο και οι εξυπηρετήσεις.

    1. Να περιγραφεί το σύστημα ως Μ.α.ς.χ. και να γίνει το διάγραμμα των ρυθμών μετάβασής της.
    2. Να βρεθεί η συνθήκη ευστάθειας (στασιμότητας) του συστήματος.
    3. Να υπολογιστεί η οριακή κατανομή του αριθμού των πελατών στο σύστημα.
    4. Να βρεθεί ο μέσος χρόνος παραμονής ενός πελάτη που βρίσκει κατά την άφιξή του στο σύστημα n πελάτες.

3. Έστω μια αδιαχώριστη Μαρκοβιανή αλυσίδα $\{X(t)\}$ με χώρο καταστάσεων $\mathcal{S}$ και ρυθμούς μετάβασης q_{ij}. Έστω, επίσης, ένα γνήσιο μη-κενό υποσύνολο $\mathcal{A}$ του $\mathcal{S}$. Θεωρούμε τη Μαρκοβιανή αλυσίδα $\{X_A(t)\}$ με χώρο καταστάσεων $\mathcal{A}$ και ρυθμούς μετάβασης

$$q_{Aij} = \begin{cases} q_{ij} & \text{αν } i,j \in \mathcal{A} \\ 0 & \text{διαφορετικά,} \end{cases}$$

    που αναφέρεται ως ο περιορισμός της $\{X(t)\}$ στο $\mathcal{A}$.

    1. Δείξτε ότι, αν η $\{X(t)\}$ είναι αντιστρέψιμη τότε και η $\{X_A(t)\}$ είναι αντιστρέψιμη και η στάσιμη κατανομή (p_{Aj}) της $\{X_A(t)\}$ βρίσκεται κανονικοποιώντας τη στάσιμη κατανομή (p_j) της $\{X(t)\}$ στο $\mathcal{A}$, δηλαδή

$$p_{Aj} = \frac{p_j}{\sum_{i\in\mathcal{A}} p_i},\ j \in \mathcal{A}.$$

    2. Δώστε κάποιο αντιπαράδειγμα για να δείξετε ότι, αν η $\{X(t)\}$ δεν είναι αντιστρέψιμη τότε δεν ισχύει το παραπάνω αποτέλεσμα.
    3. Χρησιμοποιώντας το παραπάνω αποτέλεσμα, βρείτε τη στάσιμη κατανομή της $M/M/1/k$ ουράς αν είναι γνωστό ότι η στάσιμη κατανομή της $M/M/1$ ουράς είναι $p_j = (1-\rho)\rho^j$, $j \geq 0$, όπου ρ ο ρυθμός συνωστισμού.
    4. Χρησιμοποιώντας το παραπάνω αποτέλεσμα, βρείτε τη στάσιμη κατανομή της $M/M/k/k$ ουράς αν είναι γνωστό ότι η στάσιμη κατανομή της $M/M/\infty$ ουράς είναι $p_j = e^{-\rho}\frac{\rho^j}{j!}$, $j \geq 0$, όπου ρ ο ρυθμός συνωστισμού.
    5. Θεωρούμε δυο ανεξάρτητες (παράλληλες) $M/M/1$ ουρές με ρυθμούς αφίξεων λ_1, λ_2 αντίστοιχα και ρυθμούς εξυπηρέτησης μ_1, μ_2 αντίστοιχα, με ρυθμούς συνωστισμού μικρότερους του 1. Υποθέτουμε ότι οι δυο αυτές ουρές έχουν έναν κοινό χώρο αναμονής πελατών (στον οποίο παραμένουν όσοι δεν εξυπηρετούνται) χωρητικότητας k. Βρείτε τη στάσιμη κατανομή της στοχαστικής

διαδικασίας $\{(Q_1(t), Q_2(t))\}$ του αριθμού των πελατών στις δυο ουρές, χρησιμοποιώντας το παραπάνω αποτέλεσμα.

4. Έστω $\{X(t)\}$ αδιαχώριστη στάσιμη Μαρκοβιανή αλυσίδα με χώρο καταστάσεων $\mathcal{S}$ και ρυθμούς μετάβασης q_{ij}. Αν υπάρχει συνάρτηση πιθανότητας $(p_j : j \in \mathcal{S})$ και σύνολο μη-αρνητικών αριθμών $\{\hat{q}_{ij} : i, j \in \mathcal{S}, i \neq j\}$ με τις ιδιότητες
$$\hat{q}_i = \sum_{j \neq i} \hat{q}_{ij} = \sum_{j \neq i} q_{ij} = q_i, \; i \in \mathcal{S},$$
$$p_i \hat{q}_{ij} = p_j q_{ji}, \; i, j \in \mathcal{S} \text{ με } i \neq j,$$
να αποδείξετε ότι η $\{X(t)\}$ έχει στάσιμη κατανομή την $(p_j : j \in \mathcal{S})$ και η αντίστροφή της έχει ρυθμούς μετάβασης $\hat{q}_{ij}$.

5. Έστω ένα σύστημα εξυπηρέτησης με Poisson διαδικασία αφίξεων ρυθμού λ και δυο υπηρέτες 1 και 2 τοποθετημένους σε σειρά με εκθετικούς χρόνους εξυπηρέτησης με ρυθμούς μ_1 και μ_2, αντίστοιχα. Κάθε πελάτης που φθάνει στο σύστημα πηγαίνει να εξυπηρετηθεί πρώτα στον υπηρέτη 1 και όταν τελειώσει πηγαίνει στον υπηρέτη 2. Κάθε υπηρέτης έχει το δικό του απεριόριστο χώρο αναμονής και ακολουθεί την FCFS πειθαρχία ουράς. Να υπολογιστεί η στάσιμη κατανομή της στοχαστικής διαδικασίας $\{(Q_1(t), Q_2(t))\}$ του αριθμού των πελατών στους δυο υπηρέτες (συμπεριλαμβανομένων των πελατών που εξυπηρετούνται και των πελατών σε αναμονή).

8

Διδιάστατες Μαρκοβιανές ουρές

Στο κεφάλαιο αυτό εξετάζουμε μοντέλα στα οποία η στοχαστική διαδικασία του αριθμού των πελατών στο σύστημα, $\{Q(t)\}$, δεν είναι Μ.α.ς.χ., αλλά μπορεί να εμφυτευτεί σε μια Μ.α.ς.χ. υπό την έννοια ότι η $\{(Q(t), I(t))\}$ είναι Μ.α.ς.χ., με την $I(t)$ να καταγράφει κάποια επιπλέον πληροφορία για την κατάσταση του συστήματος τη χρονική στιγμή t που μπορεί να αναφέρεται στην διαδικασία των αφίξεων ή των εξυπηρετήσεων, στην κατάσταση του υπηρέτη κ.τ.λ. Θα δούμε διάφορα παραδείγματα μοντέλων που εμπίπτουν σε αυτή την κατηγορία, καθώς και μεθόδους επίλυσής τους.

8.1 Η μέθοδος των φάσεων

Η κατανομή $Erlang(k, \alpha)$ ορίζεται ως η κατανομή του αθροίσματος k ανεξάρτητων $Exp(\alpha)$. Υπό αυτή την έννοια ένας χρόνος S (εξυπηρέτησης, ενδιάμεσος μεταξύ διαδοχικών αφίξεων, κ.τ.λ.) που ακολουθεί την κατανομή $Erlang(k, \alpha)$, μπορεί να θεωρηθεί ότι αποτελείται από k εκθετικά κατανεμημένες 'φάσεις ': $S = X_1+X_2+\cdots+X_k$, με $X_i \sim Exp(\alpha)$, $i = 1, 2, \ldots, k$. Βεβαίως έχουμε $E[S] = \frac{k}{\alpha}$ και $Var[S] = \frac{k}{\alpha^2}$. Για μια συγκεκριμένη μέση τιμή $\frac{1}{\nu}$ υπάρχουν πολλές κατανομές Erlang που αντιστοιχούν σε αυτή. Συγκεκριμένα, είναι όλες οι $Erlang(k, k\nu)$. Πράγματι, αν $S_k \sim Erlang(k, k\nu)$, τότε $E[S_k] = \frac{k}{k\nu} = \frac{1}{\nu}$ και καθώς το k αυξάνει η διασπορά της S_k γίνεται μικρότερη, αφού $Var[S_k] = \frac{k}{k^2\nu^2} = \frac{1}{k\nu^2}$. Έτσι μπορούμε να χρησιμοποιούμε κάποια S_k, ώστε να μοντελοποιήσουμε χρόνους με μέση τιμή $\frac{1}{\nu}$ και μικρές διασπορές. Αυτό είναι σημαντικό σε συστήματα εξυπηρέτησης που είτε οι χρόνοι εξυπηρέτησης είτε οι ενδιάμεσοι χρόνοι αφίξεων έχουν μικρή διασπορά, πράγμα που συμβαίνει συχνά στις εφαρμογές.

Όταν έχω μια τέτοια κατάσταση όπου οι ενδιάμεσοι χρόνοι αφίξεων ή οι χρόνοι εξυπηρέτησης είναι Erlang, για να διατηρήσω τη Μαρκοβιανή ιδιότητα, θεωρώ ότι οι αντίστοιχοι χρόνοι αναλύονται σε φάσεις και κάθε χρονική στιγμή θα πρέπει να γνωρίζουμε σε ποιά φάση βρίσκεται ο αντίστοιχος χρόνος. Π.χ. αν οι αφίξεις γίνονται με ρυθμό λ και οι ενδιάμεσοι χρόνοι αφίξεων είναι $Erlang(k, k\lambda)$, πρέπει να σκεφτόμαστε ότι κάθε φορά που αφίκνυται ένας πελάτης στο σύστημα, ένας άλλος ξεκινά να έρχεται και θα φθάσει στο σύστημα αφού περάσει k το πλήθος φάσεις διάρκειας $Exp(k\lambda)$. Ομοίως, αν οι εξυπηρετήσεις γίνονται με ρυθμό μ και οι χρόνοι εξυπηρέτησης είναι $Erlang(s, s\mu)$, κάθε πελάτης περνάει από s $Exp(s\mu)$ φάσεις κατά την εξυπηρέτησή του.

Για να γίνει κατανοητη η μέθοδος θα δούμε ορισμένα παραδείγματα στα οποία είτε οι ενδιάμεσοι χρόνοι των αφίξεων είτε οι χρόνοι εξυπηρέτησης είναι Erlang.

8.2 Η $E_2/M/1$ ουρά

Θεωρούμε την $E_2/M/1$ ουρά, όπου οι χρόνοι μεταξύ διαδοχικών αφίξεων είναι ανεξάρτητοι $Erlang(2, 2\lambda)$ (επομένως ο ρυθμός των αφίξεων είναι λ), οι χρόνοι εξυπηρέτησης είναι $Exp(\mu)$, υπάρχει 1 υπηρέτης και απεριόριστη χωρητικότητα. Έστω $Q(t)$ το πλήθος των πελατών στο σύστημα της στιγμή t. Η $\{Q(t)\}$ δεν είναι Μ.α.ς.χ. αφού αν $Q(t) = 0$, ο χρόνος μέχρι την επόμενη μετάβαση που είναι αναγκαστικά ως προς την κατάσταση 1 είναι ο υπολειπόμενος χρόνος μιας $Erlang(2, 2\lambda)$ που δεν είναι εκθετικός.

Για να πάρουμε μια Μαρκοβιανή περιγραφή του συστήματος, θεωρούμε ότι κάθε πελάτης περνάει από δυο $Exp(2\lambda)$ φάσεις για να φθάσει στο σύστημα και θέτουμε $A(t)$ να είναι η φάση που βρίσκεται ο πελάτης σε διαδικασία άφιξης τη στιγμή t. Τότε η διαδικασία $\{(Q(t), A(t))\}$ έχει χώρο καταστάσεων $\{(n, i) : n \geq 0,\ i = 1, 2\}$ και ο σχετικός πίνακας μεταβάσεων είναι

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(0,1)$	$(0,2)$	$T_{(0,1),(0,2)} \sim Exp(2\lambda)$
$(0,2)$	$(1,1)$	$T_{(0,2),(1,1)} \sim Exp(2\lambda)$
$(1,1)$	2 0	$T_{(1,1),2} \sim Exp(\lambda)$ $T_{(1,1),0} \sim Exp(\mu_1)$
$(n,1),\ n \geq 1$	$(n,2)$ $(n-1,1)$	$T_{(n,1),(n,2)} \sim Exp(2\lambda)$ $T_{(n,1),(n-1,1)} \sim Exp(\mu)$
$(n,2),\ n \geq 1$	$(n+1,1)$ $(n-1,2)$	$T_{(n,2),(n+1,1)} \sim Exp(2\lambda)$ $T_{(n,2),(n-1,2)} \sim Exp(\mu)$

Οι εξισώσεις ισορροπίας είναι

$$\begin{aligned}
2\lambda p_{(0,1)} &= \mu p_{(1,1)} \\
2\lambda p_{(0,2)} &= 2\lambda p_{(0,1)} + \mu p_{(1,2)} \\
(2\lambda + \mu) p_{(n,1)} &= 2\lambda p_{(n-1,2)} + \mu p_{(n+1,1)},\ n \geq 1 \\
(2\lambda + \mu) p_{(n,2)} &= 2\lambda p_{(n,1)} + \mu p_{(n+1,2)},\ n \geq 1
\end{aligned}$$

8.3 Η $E_k/E_s/1$ ουρά

Η $E_k/E_s/1$ ουρά είναι το σύστημα με $Erlang - k$ ανεξάρτητους ενδιάμεσους χρόνους αφίξεων, με $Erlang - s$ ανεξάρτητους εκθετικούς χρόνους εξυπηρέτησης, 1 υπηρέτη και απεριόριστη χωρητικότητα. Υποθέτουμε ότι ο ρυθμός αφίξεων είναι λ και ο ρυθμός εξυπηρέτησης μ, οπότε οι ενδιάμεσοι χρόνοι των αφίξεων έχουν την $Erlang(k, k\lambda)$ κατανομή, ενώ οι χρόνοι εξυπηρέτησης έχουν την $Erlang(s, s\mu)$ κατανομή. Για να έχουμε μια Μαρκοβιανή περιγραφή του συστήματος θεωρούμε τη διαδικασία $\{(Q(t), A(t), S(t)\}$, όπου η $Q(t)$ καταγράφει τον αριθμό των πελατών τη στιγμή t, η $A(t)$ τη φάση στην οποία βρίσκεται ο πελάτης που ειναι σε διαδικασία άφιξης της στιγμή t, και η $S(t)$ τη φάση στην οποία βρίσκεται ο πελάτης που είναι σε διαδικασία εξυπηρετησης τη στιγμή t. Βεβαίως, η $S(t)$ έχει νόημα μόνο όταν $Q(t) \geq 1$.

Ο χώρος καταστάσεων είναι

$$\mathcal{S} = \{(n,i,j) : n \geq 1, 1 \leq i \leq k, 1 \leq j \leq s\} \cup \{(0,i) : 1 \leq i \leq k\}.$$

Όντας σε μια κατάσταση (n,i,j), $n \geq 1$, για τις μεταβάσεις έχουμε:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(n,i,j), i < k, j < s$	$(n,i+1,j)$	$Exp(k\lambda)$
	$(n,i,j+1)$	$Exp(s\mu)$
$(n,k,j), j < s$	$(n+1,1,j)$	$Exp(k\lambda)$
	$(n,k,j+1)$	$Exp(s\mu)$
$(n,i,s), i < k$	$(n,i+1,s)$	$Exp(k\lambda)$
	$(n-1,i,1)$	$Exp(s\mu)$
(n,k,s)	$(n+1,1,s)$	$Exp(k\lambda)$
	$(n-1,k,1)$	$Exp(s\mu)$

Ανάλογος είναι και ο πίνακας δυνατών μεταβάσεων από καταστάσεις με $n = 0$.

8.4 Η $M/M/1/1$ ουρά με επαναπροσπάθειες

Θεωρούμε την $M/M/1/1$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρετήσεων μ, 1 υπηρέτη και χωρητικότητα για 1 πελάτη, όπου οι πελάτες που βρίσκουν τον υπηρέτη απασχολημένο κατά την άφιξή τους μπαίνουν σε 'τροχιά επαναπροσπάθειας ', δηλαδή ξαναδοκιμάζουν να αποκτήσουν πρόσβαση σε αυτόν μέχρι να εξυπηρετηθούν, ανεξάρτητα ο ένας από τον άλλο. Οι χρόνοι μεταξύ διαδοχικών προσπαθειών ενός πελάτη θεωρούνται ανεξάρτητοι $Exp(\nu)$. Ο ν αναφέρεται ως ρυθμός επαναπροσπαθειών ανά πελάτη. Έστω $Q(t)$ ο αριθμός των πελατών τη στιγμή t στο σύστημα. Είναι φανερό ότι η $\{Q(t)\}$ δεν είναι Μ.α.ς.χ., αφού όταν $Q(t) = 0$, ο χρόνος μέχρι την επόμενη μετάβαση στην κατάσταση 1 είναι $Exp(\lambda + n\nu)$, όπου n είναι το πλήθος των πελατών σε τροχιά επαναπροσπάθειας (διότι η επόμενη νέα άφιξη στο σύστημα θα γίνει σε χρόνο $Exp(\lambda)$, ενώ οι n πελάτες σε τροχιά επαναπροσπάθειας επανέρχονται σε $Exp(\nu)$ χρόνους ο καθένας). Επομένως, για μια Μαρκοβιανή περιγραφή του συστηματος χρειάζεται επίσης να καταγράφουμε σε κάθε χρονική στιγμή t τον αριθμό των πελατών σε τροχιά επαναπροσπάθειας, $R(t)$. Για τις μεταβάσεις της $\{(Q(t), R(t))\}$ έχουμε τότε τον πίνακα:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(0,0)$	$(1,0)$	$Exp(\lambda)$
$(0,n),\ n \geq 1$	$(1,n)$	$Exp(\lambda)$
	$(1,n-1)$	$Exp(n\nu)$
$(1,n),\ n \geq 0$	$(1,n+1)$	$Exp(\lambda)$
	$(0,n)$	$Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), R(t))\}$ είναι Μ.α.ς.χ. Έστω $(p_{(i,n)})$ η στάσιμη κατανομή της. Οι εξισώσεις ισορροπίας της είναι

$$\begin{aligned}
(\lambda + n\nu)p_{(0,n)} &= \mu p_{(1,n)},\ n \geq 0,\\
(\lambda + \mu)p_{(1,0)} &= \lambda p_{(0,0)} + \nu p_{(0,1)},\\
(\lambda + \mu)p_{(1,n)} &= \lambda p_{(1,n-1)} + \lambda p_{(0,n)} + (n+1)\nu p_{(0,n+1)},\ n \geq 1.
\end{aligned}$$

Για να προσδιορίσουμε τη στάσιμη κατανομή, εισάγουμε τις μερικές πιθανογεννήτριες

$$\begin{aligned}
P_0(z) &= \sum_{n=0}^{\infty} p_{(0,n)} z^n,\ |z| \leq 1,\\
P_1(z) &= \sum_{n=0}^{\infty} p_{(1,n)} z^n,\ |z| \leq 1.
\end{aligned}$$

Πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(0,n)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα n παίρνουμε

$$\begin{aligned}
&\sum_{n=0}^{\infty}(\lambda + n\nu)p_{(0,n)} z^n = \sum_{n=0}^{\infty} \mu p_{(1,n)} z^n\\
\Leftrightarrow\ &\lambda P_0(z) + \nu z P_0(z) = \mu P_1(z).
\end{aligned}$$

Ομοίως, πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(1,n)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα n παίρνουμε

$$\begin{aligned}
&\sum_{n=0}^{\infty}(\lambda + \mu)p_{(1,n)} z^n = \sum_{n=1}^{\infty} \lambda p_{(1,n-1)} z^n + \sum_{n=0}^{\infty} \lambda p_{(0,n)} z^n + \sum_{n=0}^{\infty} \nu(n+1) p_{(0,n+1)} z^n\\
\Leftrightarrow\ &(\lambda + \mu)P_1(z) = \lambda z P_1(z) + \lambda P_0(z) + \nu P_0'(z).
\end{aligned}$$

Διαιρώντας με μ και αντικαθιστώντας το λ/μ με ρ, έχουμε το παρακάτω σύστημα 2 εξισώσεων και 2 αγνώστων για τις $P_0(z)$, $P_1(z)$:

$$\begin{aligned}
\rho P_0(z) + \frac{\nu}{\mu} z P_0'(z) &= P_1(z),\\
(1 + \rho - \rho z)P_1(z) &= \rho P_0(z) + \frac{\nu}{\mu} P_0'(z).
\end{aligned}$$

Αντικαθιστώντας την 1^η εξίσωση στη 2^η, έχουμε

$$\begin{aligned}&(1+\rho-\rho z)\rho P_0(z)+(1+\rho-\rho z)\frac{\nu}{\mu}zP_0'(z)=\rho P_0(z)+\frac{\nu}{\mu}P_0'(z)\\ \Leftrightarrow\ &\rho^2(1-z)P_0(z)=\frac{\nu}{\mu}(1-z)(1-\rho z)P_0'(z),\end{aligned}$$

οπότε

$$(\log P_0(z))'=\frac{P_0'(z)}{P_0(z)}=\frac{\rho^2\mu}{\nu(1-\rho z)}=\frac{\lambda\rho}{\nu(1-\rho z)}.$$

Ολοκληρώνοντας κατά μέλη, έχουμε:

$$\log P_0(z)-\log P_0(0)=\frac{\lambda\rho}{\nu}\int_0^z\frac{1}{1-\rho u}du=-\frac{\lambda}{\nu}\log(1-\rho z).$$

Επομένως,

$$P_0(z)=(1-\rho z)^{-\frac{\lambda}{\nu}}p_{(0,0)}.$$

Αντικαθιστώντας στην 1^η εξίσωση του συστήματος των πιθανογεννητριών και κάνοντας μερικές απλοποιήσεις, παίρνουμε:

$$P_1(z)=\rho(1-\rho z)^{-\left(\frac{\lambda}{\nu}+1\right)}p_{(0,0)}.$$

Το $p_{(0,0)}$ βρίσκεται από την εξίσωση κανονικοποίησης $P_0(1)+P_1(1)=1$ και είναι $p_{(0,0)}=(1-\rho)^{\frac{\lambda}{\nu}+1}$. Για να βρούμε τις στάσιμες πιθανότητες $p_{(0,n)}$, $p_{(1,n)}$, $n\geq 0$ χρησιμοποιούμε το ανάπτυγμα

$$(1+t)^x=\sum_{n=0}^{\infty}\binom{x}{n}t^n,\ |t|<1.$$

Παίρνουμε

$$\begin{aligned}p_{(0,n)}&=(1-\rho)^{\frac{\lambda}{\nu}+1}\binom{-\lambda/nu}{n}(-1)^n\rho^n\\ &=(1-\rho)^{\frac{\lambda}{\nu}+1}\frac{\lambda(\lambda+\nu)(\lambda+2\nu)\cdots(\lambda+(n-1)\nu)\rho^n}{n!\nu^n},\ n\geq 0.\end{aligned}$$

Κατόπιν βρίσκουμε τα $p_{(1,n)}$:

$$\begin{aligned}p_{(1,n)}&=\frac{\lambda+n\nu}{\mu}p_{(0,n)}\\ &=(1-\rho)^{\frac{\lambda}{\nu}+1}\frac{(\lambda+\nu)(\lambda+2\nu)\cdots(\lambda+n\nu)\rho^{n+1}}{n!\nu^n},\ n\geq 0.\end{aligned}$$

8.5 Η $M/M/1$ ουρά με χρόνους προθέρμανσης

Θεωρούμε την $M/M/1$ ουρά με ρυθμό αφίξεων λ, ρυθμό εξυπηρετήσεων μ, 1 υπηρέτη και απεριόριστη χωρητικότητα υπό την πειθαρχία ουράς FCFS. Λόγω υψηλών λειτουργικών εξόδων, υποθέτουμε ότι ο υπηρέτης απενεργοποιείται ακαριαία μόλις το σύστημα αδειάσει. Όταν ένας νέος πελάτης αφιχθεί, ο υπηρέτης ξεκινά έναν χρόνο προθέρμανσης (setup time) με το πέρας του οποίου ο υπηρέτης αρχίζει να παρέχει εξυπηρέτηση μέχρι το σύστημα να μείνει ξανά κενό. Στη διάρκεια ενός χρόνου προθέρμανσης το σύστημα δέχεται νέες αφίξεις, αλλά δεν παρέχει εξυπηρέτηση. Οι χρόνοι προθέρμανσης του υπηρέτη είναι $Exp(\theta)$. Το σύστημα αυτό, αναφέρεται ως η $M/M/1$ ουρά με χρόνους προθέρμανσης (όταν το σύστημα μείνει κενό). Έστω $Q(t)$ ο αριθμός των πελατών τη στιγμή t στο σύστημα. Είναι φανερό ότι η $\{Q(t)\}$ δεν είναι Μ.α.ς.χ., αφού όταν $Q(t) = n \geq 1$, ο χρόνος μέχρι την επόμενη αναχώρηση μπορεί να είναι είτε $Exp(\mu)$, είτε το άθροισμα δυο ανεξάρτητων $Exp(\theta)$ και $Exp(\mu)$, ανάλογα με το αν ο υπηρέτης είναι ενεργός ή όχι, αντίστοιχα. Επομένως για μια Μαρκοβιανή περιγραφή του συστηματος χρειάζεται επίσης να καταγράφουμε σε κάθε χρονική στιγμή t την κατάσταση του υπηρέτη $I(t)$ ($I(t) = 1$ αν ο υπηρέτης είναι ενεργός, ενώ $I(t) = 0$ αν είναι ανενεργός). Για τις μεταβάσεις της $\{(Q(t), I(t))\}$ έχουμε τότε τον παρακάτω πίνακα:

Κατάσταση	Επόμενη κατάσταση	Χρόνος
$(0,0)$	$(1,0)$	$Exp(\lambda)$
$(n,0),\ n \geq 1$	$(n+1,0)$	$Exp(\lambda)$
	$(n,1)$	$Exp(\theta)$
$(1,1)$	$(2,1)$	$Exp(\lambda)$
	$(0,0)$	$Exp(\mu)$
$(n,1),\ n \geq 2$	$(n+1,1)$	$Exp(\lambda)$
	$(n-1,1)$	$Exp(\mu)$

Αφού όλοι οι χρόνοι είναι εκθετικοί, έχουμε ότι η $\{(Q(t), I(t))\}$ είναι Μ.α.ς.χ. Έστω $(p_{(i,n)})$ η στάσιμη κατανομή της. Οι εξισώσεις ισορροπίας της είναι

$$\begin{aligned}
\lambda p_{(0,0)} &= \mu p_{(1,1)}, \\
(\lambda+\theta) p_{(n,0)} &= \lambda p_{(n-1,0)},\ n \geq 1, \\
(\lambda+\mu) p_{(1,1)} &= \theta p_{(1,0)} + \mu p_{(2,1)}, \\
(\lambda+\mu) p_{(n,1)} &= \theta p_{(n,0)} + \mu p_{(n+1,1)} + \lambda p_{(n-1,1)},\ n \geq 2.
\end{aligned}$$

Για να προσδιορίσουμε τη στάσιμη κατανομή, εισάγουμε όπως προηγουμένως τις μερικές πιθανογεννήτριες

$$\begin{aligned}
P_0(z) &= \sum_{n=0}^{\infty} p_{(0,n)} z^n,\ |z| \leq 1, \\
P_1(z) &= \sum_{n=1}^{\infty} p_{(1,n)} z^n,\ |z| \leq 1.
\end{aligned}$$

Πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n,0)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα $n \geq 0$, παίρνουμε

$$\begin{aligned} &\lambda p_{(0,0)} + (\lambda+\theta)\sum_{n=1}^{\infty} p_{(n,0)} z^n = \mu p_{(1,1)} + \lambda \sum_{n=1}^{\infty} p_{(n-1,0)} z^n \\ \Leftrightarrow\ &(\lambda+\theta) P_0(z) - \theta p_{(0,0)} = \lambda p_{(0,0)} + \lambda z P_0(z). \end{aligned}$$

Λύνοντας για την $P_0(z)$, παίρνουμε

$$P_0(z) = \frac{(\lambda+\theta)p_{(0,0)}}{\lambda+\theta-\lambda z}.$$

Ομοίως, πολλαπλασιάζοντας κάθε εξίσωση ισορροπίας που αναφέρεται σε κατάσταση $(n,1)$ με το αντίστοιχο z^n και αθροίζοντας για όλα τα $n \geq 1$, παίρνουμε

$$\begin{aligned} &(\lambda+\mu)\sum_{n=1}^{\infty} p_{(n,1)} z^n = \theta \sum_{n=1}^{\infty} p_{(n,0)} z^n + \mu \sum_{n=1}^{\infty} p_{(n+1,1)} z^n + \lambda \sum_{n=2}^{\infty} p_{(n-1,1)} z^n \\ \Leftrightarrow\ &(\lambda+\mu) P_1(z) = \theta(P_0(z) - p_{(0,0)}) + \frac{\mu}{z}(P_1(z) - p_{(1,1)} z) + \lambda z P_1(z) \\ \Leftrightarrow\ &\left(\lambda+\mu-\lambda z - \frac{\mu}{z}\right) P_1(z) = \theta P_0(z) - \theta p_{(0,0)} - \mu p_{(1,1)}. \end{aligned}$$

Αντικαθιστώντας το $\mu p_{(1,1)}$ από το $\lambda p_{(0,0)}$, χρησιμοποιώντας την εξίσωση ισορροπίας για την $(0,0)$ και πολλαπλασιάζοντας με z, έχουμε

$$\left(\lambda z + \mu z - \lambda z^2 - \mu\right) P_1(z) = \theta z P_0(z) - (\lambda+\theta) z p_{(0,0)}.$$

Παραγοντοποιώντας το συντελεστή του $P_1(z)$, αντικαθιστώντας την $P_0(z)$ και λύνοντας ως προς $P_1(z)$, παίρνουμε

$$P_1(z) = \frac{\lambda(\lambda+\theta) z p_{(0,0)}}{(\lambda+\theta-\lambda z)(\mu-\lambda z)}.$$

Το $p_{(0,0)}$ βρίσκεται από την εξίσωση κανονικοποίησης $P_0(1) + P_1(1) = 1$, και είναι

$$p_{(0,0)} = \frac{\theta(\mu-\lambda)}{(\lambda+\theta)\mu}.$$

Αντικαθιστώντας το στους τύπους για τις $P_0(z)$ και $P_1(z)$ παίρνουμε

$$\begin{aligned} P_0(z) &= \frac{\theta(\mu-\lambda)}{\mu(\lambda+\theta-\lambda z)}, \\ P_1(z) &= \frac{\lambda\theta(\mu-\lambda) z}{\mu(\lambda+\theta-\lambda z)(\mu-\lambda z)}. \end{aligned}$$

Για να βρούμε τις στάσιμες πιθανότητες $p_{(0,n)}$, $n \geq 0$, χρησιμοποιούμε το γεωμετρικό ανάπτυγμα $(1-t)^{-1} = \sum_{n=0}^{\infty} t^n$, $|t| < 1$ και έχουμε

$$P_0(z) = \frac{\theta(\mu-\lambda)}{\mu(\lambda+\theta)} \cdot \frac{1}{1 - \frac{\lambda}{\lambda+\theta} z} = \sum_{n=0}^{\infty} \frac{\theta(\mu-\lambda)}{\mu(\lambda+\theta)} \left(\frac{\lambda}{\lambda+\theta}\right)^n z^n,$$

απ' όπου και συμπεραίνουμε ότι

$$p_{(n,0)} = \frac{\theta(\mu-\lambda)}{\mu(\lambda+\theta)}\left(\frac{\lambda}{\lambda+\theta}\right)^n,\ n \geq 0.$$

Για να βρούμε τις στάσιμες πιθανότητες , $p_{(n,1)}$, $n \geq 1$, διακρίνουμε δυο περιπτώσεις:

Περίπτωση 1: $\lambda + \theta \neq \mu$.

Στην περίπτωση αυτή η $P_1(z)$ αναλύεται σε απλά κλάσματα ως

$$P_1(z) = \frac{A}{\lambda+\theta-\lambda z} + \frac{B}{\mu-\lambda z}.$$

Οι σταθερές A και B κατά τα γνωστά θα προκύπτουν ως:

$$A = \lim_{z\to\frac{\lambda+\theta}{\lambda}}(\lambda+\theta-\lambda z)P_1(z) = \frac{\theta(\mu-\lambda)(\lambda+\theta)}{\mu(\mu-\lambda-\theta)},$$

$$B = \lim_{z\to\frac{\mu}{\lambda}}(\mu-\lambda z)P_1(z) = \frac{\theta(\mu-\lambda)}{\lambda+\theta-\mu}.$$

Επομένως, έχουμε ότι

$$P_1(z) = \frac{A}{\lambda+\theta}\cdot\frac{1}{1-\frac{\lambda}{\lambda+\theta}z} + \frac{B}{\mu}\cdot\frac{1}{1-\frac{\lambda}{\mu}z}$$

και χρησιμοποιώντας το γεωμετρικό ανάπτυγμα $(1-t)^{-1} = \sum_{n=0}^{\infty} t^n$, $|t| < 1$ συμπεραίνουμε ότι

$$p_{(n,1)} = \frac{\theta(\mu-\lambda)}{\mu(\mu-\lambda-\theta)}\left(\frac{\lambda}{\lambda+\theta}\right)^n - \frac{\theta(\mu-\lambda)}{\mu(\mu-\lambda-\theta)}\left(\frac{\lambda}{\mu}\right)^n,\ n \geq 0.$$

Περίπτωση 2: $\lambda + \theta = \mu$.

Στην περίπτωση αυτή η $P_1(z)$ γράφεται ως

$$P_1(z) = \frac{\lambda\theta^2 z}{(\lambda+\theta)(\lambda+\theta-\lambda z)^2} = \frac{\lambda\theta^2 z}{(\lambda+\theta)^3}\cdot\frac{1}{\left(1-\frac{\lambda}{\lambda+\theta}z\right)^2}.$$

Χρησιμοποιώντας το ανάπτυγμα

$$(1+t)^x = \sum_{n=0}^{\infty}\binom{x}{n}t^n,\ |t| < 1$$

παίρνουμε

$$\begin{aligned} p_{(n,1)} &= \frac{\lambda\theta^2}{(\lambda+\theta)^3}\binom{-2}{n-1}(-1)^{n-1}\left(\frac{\lambda}{\lambda+\theta}\right)^{n-1} \\ &= \frac{n\theta^2\lambda^n}{(\lambda+\theta)^{n+2}},\ n \geq 1. \end{aligned}$$

8.6 Ασκήσεις

1. Βρείτε τη στάσιμη κατανομή του αριθμού των πελατών στην $M/E_s/1/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ.

2. Βρείτε τη στάσιμη κατανομή του αριθμού των πελατών στην $M/E_2/2/2$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ.

3. Βρείτε τη στάσιμη κατανομή του αριθμού των πελατών στην $M/E_s/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ.

4. Βρείτε τη στάσιμη κατανομή του αριθμού των πελατών στην $E_k/M/1$ ουρά με ρυθμό αφίξεων λ και ρυθμό εξυπηρέτησης μ.

5. Θεωρούμε δυο συστήματα εξυπηρέτησης O_1 και O_2, συνδεδεμένα σε σειρά, με έναν υπηρέτη το καθένα, χωρίς χώρο αναμονής. Οι χρόνοι εξυπηρέτησης στα δυο συστήματα είναι εκθετικά κατανεμημένοι με παραμέτρους μ_1 και μ_2, αντίστοιχα. Οι αφίξεις στο πρώτο σύστημα O_1 γίνονται σύμφωνα με μια διαδικασία Poisson με ρυθμό λ και κάθε πελάτης που βρίσκει το σύστημα αυτό κενό αναχωρεί άμεσα. Κάθε πελάτης που αναχωρεί από το σύστημα O_1 πηγαίνει στο O_2 αν αυτό είναι κενό. Διαφορετικά παραμένει στο σύστημα O_1 μέχρι να αδειάσει το σύστημα O_2. Στην περίπτωση αυτή, το σύστημα O_1 δεν παρέχει εξυπηρέτηση και αδρανοποιείται. Μόλις όμως, το σύστημα O_2 αδειάσει, ο πελάτης που περιμένει στο O_1 μεταφέρεται ακαριαία σε αυτό. Να ορίσετε μια Μ.α.ς.χ. που να περιγράφει την εξέλιξη του συστήματος στο χρόνο και να παραθέσετε το διάγραμμα των ρυθμών μετάβασής της.

9

Απλά Μαρκοβιανά δίκτυα ουρών

Ένα δίκτυο συστημάτων εξυπηρέτησης αποτελείται από πολλά συστήματα εξυπηρέτησης τα οποία συνδέονται μεταξύ τους, υπό την έννοια ότι οι πελάτες που αναχωρούν από ένα σύστημα μπορεί να πηγαίνουν σε κάποιο άλλο για να συνεχίσουν την εξυπηρέτησή τους και από εκει σε ένα άλλο κ.ο.κ. Καθώς τα περισσότερα πραγματικά συστήματα εξυπηρέτησης περιλαμβάνουν πολλά στάδια εξυπηρέτησης, η μαθηματική μελέτη των δικτύων συστημάτων εξυπηρέτησης είναι πολύτιμη για την μελέτη του συνωστισμού σε πραγματικά συστήματα. Στο πλαίσιο αυτών των σημειώσεων, θα γίνει μόνο μια σύντομη εισαγωγή στη θεωρία των δικτύων συστημάτων εξυπηρέτησης.

9.1 Μαρκοβιανά δίκτυα

Για να περιγραφεί η κατάσταση ενός δικτύου συστημάτων εξυπηρέτησης με N σταθμούς πρέπει κατ᾽ ελάχιστον να καταγράφεται ο αριθμός των πελατών στους διάφορους σταθμούς του δικτύου. Έστω, επομένως, ένα δίκτυο με N σταθμούς εξυπηρέτησης. Θα συμβολίζουμε με $Q_i(t)$ το πλήθος των παρόντων πελατών στο σταθμό i του δικτύου τη χρονική στιγμή t. Θέτουμε $\mathbf{Q}(t) = (Q_1(t), Q_2(t), \ldots, Q_N(t))$, όπου η $\mathbf{Q}(t)$ είναι η βασική στοχαστική διαδικασία του πλήθους πελατών στους διάφορους σταθμούς του δικτύου. Ένα δίκτυο συστημάτων εξυπηρέτησης λέγεται Μαρκοβιανό, αν η διαδικασία αυτή είναι Μ.α.ς.χ. Αν επιπλέον οι πελάτες κινούνται μεμονωμένα, τότε έχουμε ένα απλό Μαρκοβιανό δίκτυο.

Έστω ότι $\mathbf{Q}(t) = \mathbf{n} = (n_1, n_2, \ldots, n_N)$ μια χρονική στιγμή t και έστω επίσης $\mathbf{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$, $i = 1, 2, \ldots, N$ τα μοναδιαία διανύσματα. Τότε, σε ένα απλό Μαρκοβιανό δίκτυο η επόμενη μετάβαση μπορεί να είναι:

- $\mathbf{n} \to \mathbf{n} + \mathbf{e}_i$: εξωτερική άφιξη στο σταθμό i,
- $\mathbf{n} \to \mathbf{n} - \mathbf{e}_j$: αναχώρηση από το σταθμό i προς το εξωτερικό του δικτύου (δυνατή μόνο όταν $n_j \geq 1$),
- $\mathbf{n} \to \mathbf{n} - \mathbf{e}_j + \mathbf{e}_i$: μετάβαση πελάτη από το σταθμό j στον i (δυνατή μόνο όταν $n_j \geq 1$).

Οι αντίστοιχοι ρυθμοί είναι:

- $\lambda_i(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} + \mathbf{e}_i)$: εξωτερικός ρυθμός αφίξεων στην i,
- $\mu_{j0}(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} - \mathbf{e}_j)$: εξωτερικός ρυθμός αναχωρήσεων από την j,
- $\mu_{ji}(\mathbf{n}) = q(\mathbf{n}, \mathbf{n} - \mathbf{e}_j + \mathbf{e}_i)$: ρυθμός μεταβάσεων από την j στην i.

Επίσης, θέτουμε

- $\mu_j(\mathbf{n}) = \sum_{i=0}^{N} \mu_{ji}(\mathbf{n})$: συνολικός ρυθμός εξυπηρέτησης στην j.

9.2 Δίκτυα Jackson

Ένα δίκτυο συστημάτων εξυπηρέτησης λέγεται δίκτυο Jackson αν είναι απλό Μαρκοβιανό δίκτυο με τις επιπλέον ιδιότητες

(i) $\lambda_i(\mathbf{n}) = \lambda_i$ (δηλαδή η διαδικασία εξωτερικών αφίξεων σε κάθε σταθμό είναι Poisson)

(ii) $\mu_j(\mathbf{n}) = \mu_j(n_j)$ (δηλαδή ο ρυθμός εξυπηρέτησης σε κάθε σταθμό εξαρτάται μόνο από το πλήθος των πελατών σε αυτόν)

(iii) $\mu_{ji}(\mathbf{n}) = \mu_j(n_j)p_{ji}$ (δηλαδή οι πελάτες επιλέγουν τους σταθμούς που θα επισκεφθούν σύμφωνα με μια Μαρκοβιανή αλυσίδα διακριτού χρόνου (Μαρκοβιανές διαδρομές πελατών».

Ένα δίκτυο Jackson λέγεται ανοικτό αν υπάρχουν σταθμοί i, j τέτοιοι ώστε $\lambda_i > 0$ και $\mu_{j0}(n_j) > 0$ για κάποιο $n_j \geq 1$. Σε αντίθετη περίπτωση, δηλαδή όταν $\lambda_i = 0$ και $\mu_{j0}(n_j) = 0$ για όλους τους σταθμούς i και j και όλα τα $n_j \geq 1$, τότε το δίκτυο λέγεται κλειστό.

Ο χώρος καταστάσεων $\mathcal{S}$ της διαδικασίας $\{\mathbf{Q}(t)\}$ που καταγράφει τον αριθμό των πελατών στους διάφορορους σταθμούς του δικτύου είναι υποσύνολο του $\mathbb{N}_0^N$. Πιο συγκεκριμένα, για ανοικτά δίκτυα Jackson, ο χώρος καταστάσεων είναι

$$\mathcal{S}_o = \{(n_1, n_2, \ldots, n_N) : n_i \geq 0, i = 1, 2, \ldots, N\} = \mathbb{N}_0^N.$$

Για κλειστά δίκτυα Jackson, ο χώρος καταστάσεων είναι

$$\mathcal{S}_c = \{(n_1, n_2, \ldots, n_N) : n_i \geq 0, i = 1, 2, \ldots, N \text{ και } \sum_{i=1}^{N} n_i = M\}.$$

Για ένα ανοικτό δίκτυο Jackson ο πίνακας Μαρκοβιανών διαδρομών ενός πελάτη ορίζεται ως

$$\mathbb{P} = (p_{ij}) = \begin{pmatrix} 0 & \frac{\lambda_1}{\lambda} & \frac{\lambda_2}{\lambda} & \cdots & \frac{\lambda_{N-1}}{\lambda} & \frac{\lambda_N}{\lambda} \\ p_{10} & p_{11} & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{20} & p_{21} & p_{22} & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,0} & p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} & p_{N-1,N} \\ p_{N0} & p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} \end{pmatrix},$$

όπου $\lambda = \sum_{i=1}^{N} \lambda_i$ είναι ο συνολικός ρυθμός αφίξεων στο δίκτυο. Ο πίνακας αυτός αντιστοιχεί στον πίνακα πιθανοτήτων μετάβασης από σταθμό σε σταθμό για κάθε συγκεκριμένο πελάτη που παρακολουθούμε εσαεί να εισέρχεται στο και να εξέρχεται από το δίκτυο. Για ένα κλειστό δίκτυο Jackson, ο αντίστοιχος πίνακας Μαρκοβιανών διαδρομών ενός πελάτη είναι

$$\mathbb{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1,N-1} & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2,N-1} & p_{2N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p_{N-1,1} & p_{N-1,2} & \cdots & p_{N-1,N-1} & p_{N-1,N} \\ p_{N1} & p_{N2} & \cdots & p_{N,N-1} & p_{NN} \end{pmatrix}.$$

Ο πίνακας Μαρκοβιανών διαδρομών και στις δυο περιπτώσεις αντιστοιχεί στον πίνακα πιθανοτήτων μετάβασης μιας αδιαχώριστης Μαρκοβιανής αλυσίδας διακριτού χρόνου.

9.3 Ρυθμοί διαπέρασης δικτύων Jackson

Ο ρυθμός διαπέρασης (throughput) Λ_i ενός σταθμού i σε ένα δίκτυο Jackson ορίζεται ως ο συνολικός ρυθμός περατώσεων εξυπηρετήσεων στο σταθμό i, $i = 1, 2, \ldots, N$. Λόγω του ότι όλοι οι πελάτες που φθάνουν στο δίκτυο εξυπηρετούνται και δεν αναχωρούν πρόωρα (δεν υπάρχουν δηλαδή αποθαρρυνόμενοι, ούτε ανυπόμονοι πελάτες, ούτε κάποια διαδικασία άλλης αποχώρησης πελατών), ο ρυθμός διαπέρασης Λ_i ισούται επίσης με το συνολικό ρυθμό αφίξεων και το συνολικό ρυθμό αναχωρήσεων στο σταθμό i.

Θεώρημα 9.1 *Οι ρυθμοί διαπέρασης Λ_i, $i = 1, 2, \ldots, N$, σε ένα ανοικτό δίκτυο Jackson είναι η μοναδική θετική λύση του συστήματος των εξισώσεων κίνησης*

$$\Lambda_i = \lambda_i + \sum_{j=1}^{N} \Lambda_j p_{ji},\ i = 1, 2, \ldots, N.$$

Επιπλέον οι ρυθμοί ικανοποιούν τη σχέση

$$\sum_{j=1}^{N} \Lambda_j p_{j0} = \sum_{j=1}^{N} \lambda_j.$$

Το ότι οι ρυθμοί διαπέρασης ικανοποιούν τις εξισώσεις κίνησης είναι άμεσο, αρκεί να ερμηνεύσει κανείς κατάλληλα τα δυο μέλη κάθε εξίσωσης κίνησης. Πράγματι το αριστερό μέλος της εξίσωσης κίνησης για το σταθμό i είναι ο συνολικός ρυθμός αφίξεων Λ_i στο σταθμό αυτό. Από την άλλη μεριά, ο πρώτος όρος του δεξιού μέλους είναι ο ρυθμός λ_i των εξωτερικών αφίξεων στο σταθμό i, ενώ κάθε προσθετέος του δεύτερου όρου του δεξιού μέλους, δηλαδή η ποσότητα $\Lambda_j p_{ji}$ αντιστοιχεί στον ρυθμό μεταβάσεων πελατών από το σταθμό j στο σταθμό i. Προσθέτοντας κατά μέλη όλες τις εξισώσεις κίνησης παίρνουμε

$$\sum_{i=1}^{N} \Lambda_i = \sum_{i=1}^{N} \lambda_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \Lambda_j p_{ji} = \sum_{i=1}^{N} \lambda_i + \sum_{j=1}^{N} \Lambda_j (1 - p_{j0}).$$

Απλοποιώντας το $\sum_{i=1}^{N} \Lambda_i$, παίρνουμε τη σχέση $\sum_{j=1}^{N} \Lambda_j p_{j0} = \sum_{j=1}^{N} \lambda_j$ που εξισώνει το ρυθμό αναχωρήσεων από το δίκτυο με το ρυθμό των αφίξεων σε αυτό. Θα μπορούσαμε να σκεφτόμαστε την εξίσωση αυτή ως εξίσωση κίνησης για τον εικονικό σταθμό 0 που παριστάνει το εξωτερικό του δικτύου. Επομένως αποδείξαμε ότι οι ρυθμοί διαπέρασης ικανοποιούν τις εξισώσεις κίνησεις.

Για να δούμε ότι οι εξισώσεις κίνησης έχουν μοναδική λύση, ορίζουμε $\Lambda_0 = \sum_{j=1} \lambda_j = \lambda$. Τότε το σύστημα των εξισώσεων κίνησης για τους σταθμούς $1, 2, \ldots, N$ και 0 γράφεται στη μορφή

$$\Lambda_i = \Lambda_0 \frac{\lambda_i}{\lambda} + \sum_{j=1}^{N} \Lambda_j p_{ji}, \; i = 1, 2, \ldots, N,$$
$$\Lambda_0 = \sum_{j=1}^{N} \Lambda_j p_{j0}.$$

Το σύστημα αυτό ταυτίζεται με το σύστημα των εξισώσεων ισορροπίας της Μαρκοβιανής αλυσίδας διακριτού χρόνου με πίνακα πιθανοτήτων τον πίνακα Μαρκοβιανών διαδρομών $\mathbb{P}$. Επομένως, το διάνυσμα $(\Lambda_0, \Lambda_1, \cdots, \Lambda_N)$ θα είναι πολλαπλάσιο του στάσιμου διανύσματος, δηλαδή $(\Lambda_0, \Lambda_1, \cdots, \Lambda_N) = c(\pi_0, \pi_1, \ldots, \pi_N)$, όπου $(\pi_0, \pi_1, \ldots, \pi_N)$ η μοναδική στάσιμη κατανομή του πίνακα των Μαρκοβιανών διαδρομών (η μοναδικότητα προκύπτει λόγω της υπόθεσης της αδιαχωρισιμότητας του πίνακα $\mathbb{P}$). Όμως $\Lambda_0 = \lambda$, οπότε $c = \frac{\lambda}{\pi_0}$, οπότε η λύση είναι μοναδική.

9.4 Στάσιμη κατανομή δικτύων Jackson

Οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή της διαδικασίας $\{\mathbf{Q}(t)\}$ του αριθμού των πελατών στους διάφορους σταθμούς ενός Μαρκοβιανού δικτύου είναι ιδιαίτερα πολύπλοκες, ακόμη και στην απλούστερη περίπτωση των δικτύων Jackson. Στη γενική περίπτωση οι εξισώσεις αυτές δεν επιδέχονται λύση σε κλειστή μορφή και για το λόγο αυτό επιστρατεύονται αριθμητικές και προσεγγιστικές μέθοδοι. Στην περίπτωση όμως των δικτύων Jackson η στάσιμη κατανομή βρίσκεται σε μια εξαιρετικά απλή κλειστή μορφή, όπως αποδεικνύεται στο παρακάτω θεώρημα.

Θεώρημα 9.2 (Θεώρημα Jackson) *Έστω δίκτυο Jackson N σταθμών, με ρυθμούς εξωτερικών αφίξεων λ_i, $i = 1, 2, \ldots, N$, ρυθμούς εξυπηρετήσεων $\mu_j(n_j)$, $j = 1, 2, \ldots, N$ και πιθανότητες μετάβασης πελατών p_{ji}, $j = 1, 2, \ldots, N$ και $i = 0, 1, \ldots, N$. Έστω Λ_i, $i = 1, 2, \ldots, N$ οι ρυθμοί διαπέρασης των σταθμών του δικτύου. Τότε:*

(i) *Το δίκτυο είναι ευσταθές αν κάθε σταθμός $i = 1, 2, \ldots, N$ μεμονωμένα είναι ευσταθής όταν τροφοδοτείται από μια διαδικασία Poisson με ρυθμό Λ_i και λειτουργεί σύμφωνα με τους ρυθμούς εξυπηρέτησης $\mu_i(n_i)$. Δηλαδή, το δίκτυο είναι ευσταθές, αν*

$$B_i^{-1} = 1 + \sum_{n_i=1}^{\infty} \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)} < \infty, \; i = 1, 2, \ldots, N.$$

(ii) *Όταν το δίκτυο είναι ευσταθές, τότε η στάσιμη κατανομή του είναι*

$$p(\mathbf{n}) = p(n_1, n_2, \ldots, n_N) = \prod_{i=1}^{N} p_i(n_i),\ \mathbf{n} = (n_1, n_2, \ldots, n_N),\ n_i \geq 0,$$

όπου η $(p_i(n_i) : n_i \geq 0)$ είναι η στάσιμη κατανομή του σταθμού i, όταν θεωρηθεί μεμονωμένα τροφοδοτούμενος από μια διαδικασία Poisson με ρυθμό Λ_i και λειτουργώντας με τους ρυθμούς $\mu_i(n_i)$. Δηλαδή η $p_i(n_i)$ δίνεται από τον τύπο

$$p_i(n_i) = B_i \frac{\Lambda_i^{n_i}}{\mu_i(1)\mu_i(2)\cdots\mu_i(n_i)},\ n_i \geq 0.$$

Η απόδειξη του θεωρήματος αυτού γίνεται δείχνοντας ότι η στάσιμη κατανομή που δίνει το θεώρημα ικανοποιεί τις εξισώσεις ισορροπίας της $\{\mathbf{Q}(t)\}$. Οι εξισώσεις ισορροπίας για τη στάσιμη κατανομή είναι:

$$\begin{aligned} &p(\mathbf{n}) \left(\sum_{i=1}^{N} \lambda_i + \sum_{j=1}^{N} \mu_j(n_j) 1\{n_j > 0\} \right) \\ &= \sum_{j=1}^{N} p(\mathbf{n} - \mathbf{e}_j)\lambda_j 1\{n_j > 0\} + \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i)\mu_i(n_i + 1)p_{i0} \\ &\quad + \sum_{j=1}^{N} \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)\mu_i(n_i + 1)p_{ij} 1\{n_j > 0\},\ \mathbf{n} \in \mathcal{S}_o \end{aligned}$$

που γράφονται ισοδύναμα ως

$$\begin{aligned} &p(\mathbf{n}) \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i)\mu_i(n_i + 1)p_{i0} \\ &\quad + \sum_{j=1}^{N} \left(p(\mathbf{n})\mu_j(n_j) - p(\mathbf{n} - \mathbf{e}_j)\lambda_j - \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)\mu_i(n_i + 1)p_{ij} \right) 1\{n_j > 0\} \\ &= 0,\ \mathbf{n} \in \mathcal{S}_o. \end{aligned}$$

Ορίζοντας

$$\Sigma_0 = p(\mathbf{n}) \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i)\mu_i(n_i + 1)p_{i0}$$

$$\Sigma_j = p(\mathbf{n})\mu_j(n_j) - p(\mathbf{n} - \mathbf{e}_j)\lambda_j - \sum_{i=1}^{N} p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)\mu_i(n_i + 1)p_{ij},\ 1 \leq j \leq N,$$

η τελευταία σχέση γράφεται $\Sigma_0 + \sum_{j=1}^{N} \Sigma_j 1\{n_j > 0\} = 0$. Θα αποδείξουμε ότι $\Sigma_j = 0$ για $j = 0, 1, \ldots, N$, αντικαθιστώντας την $p(\mathbf{n})$ από τον τύπο που δίνει το θεώρημα. Παρατηρούμε ότι

$$\frac{p(\mathbf{n} + \mathbf{e}_i)}{p(\mathbf{n})} = \frac{\Lambda_i}{\mu_i(n_i + 1)},\quad \frac{p(\mathbf{n})}{p(\mathbf{n} - \mathbf{e}_j)} = \frac{\Lambda_j}{\mu_j(n_j)},\quad \frac{p(\mathbf{n} + \mathbf{e}_i - \mathbf{e}_j)}{p(\mathbf{n} - \mathbf{e}_j)} = \frac{\Lambda_i}{\mu_i(n_i + 1)}.$$

Χρησιμοποιώντας τις σχέσεις αυτές και διαιρώντας με $p(\mathbf{n})$ τη σχέση $\Sigma_0 = 0$ έχουμε

$$\begin{aligned}\Sigma_0 = 0 &\Leftrightarrow \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \frac{p(\mathbf{n}+\mathbf{e}_i)}{p(\mathbf{n})}\mu_i(n_i+1)p_{i0} = 0 \\ &\Leftrightarrow \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \frac{\Lambda_i}{\mu_i(n_i+1)}\mu_i(n_i+1)p_{i0} = 0 \\ &\Leftrightarrow \sum_{i=1}^{N} \lambda_i - \sum_{i=1}^{N} \Lambda_i p_{i0} = 0,\end{aligned}$$

που ισχύει από την εξίσωση κίνησης για το σταθμό 0. Ομοίως, διαιρώντας με $p(\mathbf{n}-\mathbf{e}_j)$ τη σχέση $\Sigma_j = 0$, για $j = 1, 2, \ldots, N$ έχουμε

$$\begin{aligned}\Sigma_j = 0 &\Leftrightarrow \frac{p(\mathbf{n})}{p(\mathbf{n}-\mathbf{e}_j)}\mu_j(n_j) - \lambda_j - \sum_{i=1}^{N} \frac{p(\mathbf{n}+\mathbf{e}_i-\mathbf{e}_j)}{p(\mathbf{n}-\mathbf{e}_j)}\mu_i(n_i+1)p_{ij} = 0 \\ &\Leftrightarrow \frac{\Lambda_j}{\mu_j(n_j)}\mu_j(n_j) - \lambda_j - \sum_{i=1}^{N} \frac{\Lambda_i}{\mu_i(n_i+1)}\mu_i(n_i+1)p_{ij} = 0 \\ &\Leftrightarrow \Lambda_j - \lambda_j - \sum_{i=1}^{N} \Lambda_i p_{ij} = 0,\end{aligned}$$

που ισχύει από την εξίσωση κίνησης για το σταθμό $j = 1, 2, \ldots, N$.

9.5 Ασκήσεις

1. Θεωρούμε δυο συστήματα εξυπηρέτησης O_1 και O_2. Η ουρά O_i έχει Poisson εξωτερική διαδικασία αφίξεων με ρυθμό λ_i, έναν υπηρέτη και εκθετικούς χρόνους εξυπηρέτησης με παράμετρο μ_i. Συνδέουμε τις O_1 και O_2 έτσι ώστε κάθε πελάτης που φεύγει από την O_i να πηγαίνει στην άλλη ουρά με πιθανότητα p_i ή να αναχωρεί οριστικά από το σύστημα με πιθανότητα $q_i = 1 - p_i$. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του δικτύου και να υπολογιστεί η αντίστοιχη στάσιμη κατανομή των αριθμών των πελατών στους σταθμούς του δικτύου.

2. Θεωρούμε ένα δίκτυο πέντε συστημάτων εξυπηρέτησης $O_1, O_2, \ldots, O_5$. Για κάθε $i = 1, 2, 3, 4$ το σύστημα O_i έχει Poisson εξωτερική διαδικασία αφίξεων με ρυθμό $i\lambda$ και έναν υπηρέτη με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ_i. Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i πηγαίνει στο σύστημα O_{i+1} με πιθανότητα $\frac{1}{i}$ ή επαναλαμβάνει την εξυπηρέτησή του με πιθανότητα $\frac{i-1}{i}$ για $i = 1, 2, 3, 4$. Το σύστημα O_5 έχει Poisson διαδικασία αφίξεων και άπειρους υπηρέτες με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ_5 και κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του σε αυτό αναχωρεί από το δίκτυο.
   1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.

2. Να βρεθεί η στάσιμη κατανομή των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
3. Να υπολογιστεί ο μέσος αριθμός πελατών στο δίκτυο.
4. Να υπολογιστεί ο μέσος χρόνος παραμονής ενός πελάτη στο δίκτυο.
5. Να βρεθεί η κατανομή του αριθμού των επισκέψεων ενός πελάτη στο σύστημα O_i, για $i = 1, 2, 3, 4$.

3. Θεωρούμε ένα δίκτυο τριών συστημάτων O_1, O_2 και O_3. Για κάθε $i = 1, 2, 3$, το σύστημα O_i έχει εξωτερική διαδικασία αφίξεων Poisson με ρυθμό $\lambda_i = i$ και i υπηρέτες με εκθετικούς χρόνους εξυπηρέτησης μ_i. Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i αναχωρεί από το δίκτυο με πιθανότητα $\frac{i}{i+1}$, ενώ διαφορετικά πηγαίνει στην O_{i+1}, όπου $O_4 = O_1$.
   1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.
   2. Να βρεθεί η στάσιμη κατανομή των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
   3. Να υπολογιστεί η κατανομή του αριθμού των επισκέψεων ενός πελάτη στο σταθμό O_2, δεδομένου ότι ο πελάτης εισήλθε στο δίκτυο από τον σταθμό O_2, καθώς και το μέσο πλήθος επισκέψεων σε αυτόν.

4. Θεωρούμε ένα δίκτυο τριών συστημάτων O_1, O_2 και O_3. Το σύστημα O_1 έχει Poisson εξωτερική διαδικασία αφίξεων ρυθμού λ και έναν υπηρέτη με εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ. Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_1 πηγαίνει στο O_2 ή στο O_3 με πιθανότητες $\frac{1}{3}$ και $\frac{2}{3}$, αντίστοιχα. Τα συστήματα O_2 και O_3 έχουν άπειρους υπηρέτες και εκθετικούς χρόνους εξυπηρέτησης με ρυθμό μ. Κάθε πελάτης που ολοκληρώνει την εξυπηρέτησή του στο σύστημα O_i, $i = 2, 3$ αναχωρεί με πιθανότητα $\frac{1}{2}$ από το δίκτυο ή επιστρέφει με πιθανότητα $\frac{1}{2}$ στο ίδιο σύστημα O_i.
   1. Να βρεθεί ικανή και αναγκαία συνθήκη για την ευστάθεια (στασιμότητα) του παραπάνω Μαρκοβιανού δικτύου.
   2. Να βρεθεί η στάσιμη κατανομή των αριθμών των πελατών στους σταθμούς του δικτύου, όταν αυτό είναι ευσταθές.
   3. Να υπολογιστεί ο μέσος συνολικός αριθμός πελατών στο δίκτυο και ο μέσος αριθμός πελατών στο σύστημα O_2.
   4. Να υπολογιστεί ο μέσος αριθμός εξυπηρετήσεων που θα λάβει ένας πελάτης μέχρι να αναχωρήσει από το δίκτυο.
   5. Να υπολογιστεί ο μέσος χρόνος παραμονής ενός πελάτη στο δίκτυο.