ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ

ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ Η/Υ

Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών
<<Μαθηματικά των Υπολογιστών και των Αποφάσεων>>

Κατσαβίδα Ευτυχία

Α.Μ. 255

ΜΟΝΤΕΛΑ ΘΕΩΡΙΑΣ ΑΝΑΜΟΝΗΣ

Διπλωματική Εργασία

Επιβλέπων: Αναπληρωτής Καθηγητής Τσάντας Νικόλαος

Πάτρα Φεβρουάριος 2014

Εισαγωγή.

Στη σύγχρονη κοινωνία, η αναμονή σε κάποια ουρά είναι ένα δυσάρεστο φαινόμενο το οποίο συναντάμε καθημερινά. Ουρές αναμονής δημιουργούνται στα διόδια, στα super markets, στα εστιατόρια fast food, στις τράπεζες, τα ταχυδρομεία κλπ. Προφανώς η αναμονή είναι μια δυσάρεστη και μη επιθυμητή κατάσταση και από πλευράς πελατών (που αναμένουν), αλλά και από πλευράς διευθυντών, διότι η ακραία αναμονή μπορεί να βλάψει την επιχείρηση.

Μια ουρά αναμονής δημιουργείται όταν η τρέχουσα ζήτηση για μια εξυπηρέτηση είναι μεγαλύτερη από την τρέχουσα ικανότητα εξυπηρέτησης του συστήματος. Αυτό μπορεί να συμβαίνει για διάφορους λόγους, για παράδειγμα μπορεί να μην υπάρχουν αρκετοί σταθμοί εξυπηρέτησης, ή μπορεί να μην συμφέρει οικονομικά την επιχείρηση να παρέχει την απαραίτητη υπηρεσία προκειμένου να μην υπάρχει αναμονή, ή ακόμα μπορεί να υπάρχει όριο χώρου στον αριθμό των εξυπηρετήσεων που μπορούν να παρασχεθούν.

Για να συμπεράνει κάποιος αν αξίζει ή όχι να προχωρήσει σε μια επένδυση με σκοπό να μειώσει τον χρόνο αναμονής πρέπει να εξετάσει κατά πόσο αυτή η επένδυση επηρεάζει τον χρόνο αναμονής. Ο αντικειμενικός σκοπός του προβλήματος της ουράς είναι να βρεθεί μια οικονομική ισορροπία μεταξύ του κόστους εξυπηρέτησης και του κόστους αναμονής στην ουρά. Η θεωρία ουρών δίνει την πληροφόρηση που χρειάζεται για μια τέτοια απόφαση, με το να προσδιορίζει τα διάφορα χαρακτηριστικά του συστήματος και παρέχει ένα μεγάλο αριθμό μαθηματικών προτύπων για την περιγραφή των καταστάσεων της ουράς αναμονής.

Πολλές φορές για να μελετήσουμε θεωρητικά ένα πραγματικό σύστημα, χρησιμοποιούμε ένα κατάλληλο μαθηματικό μοντέλο, το οποίο είναι μια προσομοίωση του πραγματικού μοντέλου και στο οποίο οι σημαντικές σχέσεις μεταξύ των στοιχείων του έχουν αντικατασταθεί από αντίστοιχες μαθηματικές, ενώ τυχόν μη σημαντικές έχουν αγνοηθεί. Τα περισσότερα από τα πραγματικά συστήματα είναι στοχαστικά, δηλαδή η λειτουργία τους επηρεάζεται σημαντικά από τον λεγόμενο παράγοντα τύχη ή αλλιώς η μελλοντική συμπεριφορά τους δεν μπορεί να προβλεφθεί επακριβώς, αλλά μόνο πιθανοθεωρητικά. Οι στοχαστικές ανελίξεις είναι τα κατάλληλα μαθηματικά μοντέλα (στοχαστικά μοντέλα) για να περιγράψουμε και να μελετήσουμε στοχαστικά συστήματα. Μια σημαντική κατηγορία στοχαστικών συστημάτων είναι αυτή των λεγόμενων ουρών αναμονής.

Η παρούσα διπλωματική εργασία σκοπό έχει να παρουσιάσει κάποια μοντέλα ουρών ξεκινώντας από το πιο απλό όπως η ουρά M/M/1 στην οποία οι αφίξεις των πελατών γίνονται μεμονωμένα σύμφωνα με μια διαδικασία Poisson, οι χρόνοι εξυπηρέτησης είναι εκθετικοί, υπάρχει ένας σταθμός εξυπηρέτησης, δεν υπάρχει κανένας περιορισμός στο σχηματισμό της ουράς και οι πελάτες εξυπηρετούνται με τη σειρά την οποία

καταφθάνουν. Λόγω του ότι οι χρόνοι εξυπηρέτησης είναι εκθετικοί, το μήκος της ουράς μπορεί να περιγραφεί με την βοήθεια των αλυσίδων Markov, στοχαστική διαδικασία η οποία είναι αρκετή στο να περιγράψει το εν λόγω σύστημα όταν αυτό βρίσκεται σε κατάσταση στατιστικής ισορροπίας (μετά την παρέλευση δηλαδή ενός μεγάλου χρονικού διαστήματος).
Στη συνέχεια παρουσιάζονται αναλυτικά οι γενικεύσεις του παραπάνω μοντέλου οι ουρές M/G/1 και G/M1. Στην M/G/1 οι αφίξεις γίνονται μεμονωμένα σύμφωνα με μια διαδικασία Poisson, οι χρόνοι εξυπηρέτησης ακολουθούν μια γενική κατανομή και υπάρχει ένα σημείο εξυπηρέτησης. Στην G/M/1 οι αφίξεις γίνονται σύμφωνα με μια γενική διαδικασία, οι χρόνοι εξυπηρέτησης είναι εκθετικοί και υπάρχει ένα σημείο εξυπηρέτησης. Τα συστήματα αυτά δεν είναι δυνατόν να περιγραφούν με την βοήθεια των αλυσίδων Markov, παρά μόνο όταν τα μελετάμε κατά τη διάρκεια ορισμένων χρονικών στιγμών.
Τέλος στο τέταρτο και τελευταίο κεφάλαιο της εργασίας ασχολούμαστε με την μελέτη ενός συστήματος με τη χρήση της προσομοίωσης

Περιεχόμενα

Κεφάλαιο 1: ΕΙΣΑΓΩΣΗ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΟΥΡΩΝ

1.1 Βασικό Σύστημα Ουράς……....4
1.2 Χαρακτηριστικά Συστημάτων Ουράς……...5
1.3 Συμβολική Παράσταση……...8
1.4 Γενικά Συμπεράσματα……....9
1.5 Θεώρημα του Little……10

Κεφάλαιο 2: ΜΟΝΤΕΛΑ ΟΥΡΩΝ

2.1 Ανέλιξη Γέννησης-Θανάτου……..12
2.2 Ουρά M/M/1……14
2.3 Ουρά M/M/c……18
2.4 Ουρά M/M/c/K……..24
2.5 Φόρμουλα Erlang M/M/c/c……28
2.6 Ουρές με Απεριόριστη Εξυπηρέτηση M/M/∞……..29
2.7 Ουρές με Πεπερασμένη Πηγή……..30
2.8 Μοντέλα με Ανταλλακτικά……..32
2.9 Εξυπηρέτηση Εξαρτώμενη από τον Αριθμό Πελατών……..35
2.10 Ουρές με Ανυπόμονους Πελάτες……37

Κεφάλαιο 3: ΟΥΡΕΣ ΜΕ ΓΕΝΙΚΗ ΚΑΤΑΝΟΜΗ

3.1 Ουρά M/G/1……..40
3.2 Γενική Εξυπηρέτηση Πολλοί Εξυπηρετητές……57
3.3 Η Ουρά G/M/1……..61

Κεφάλαιο 4: ΠΡΟΣΟΜΟΙΩΣΗ

4.1 Discrete-event Stochastic Simulation……..…67

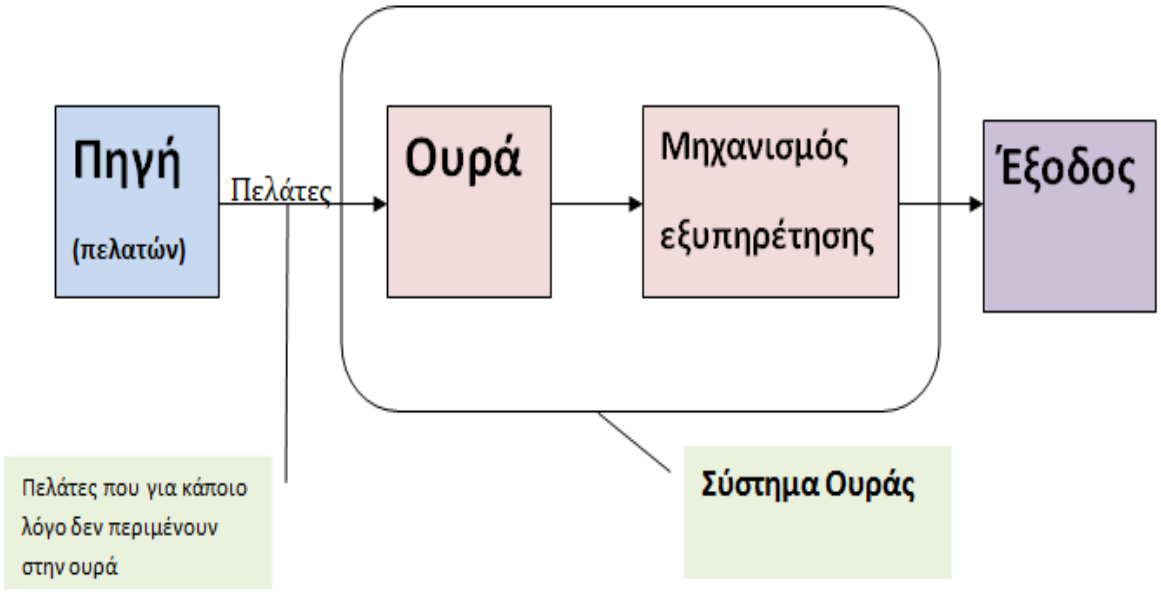
Βιβλιογραφία……81

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ ΣΤΑ ΣΥΣΤΗΜΑΤΑ ΟΥΡΩΝ

1.1 Βασικό Σύστημα Ουράς

Έχουμε μια ροή (αφίξεις) πελατών, οι οποίοι φτάνουν σ ένα σημείο εξυπηρέτησης, ενώ ταυτόχρονα υπάρχει ένα φράγμα στο ρυθμό εξυπηρέτησής τους, με αποτέλεσμα το σχηματισμό μιας ή περισσοτέρων ουρών αναμονής. Οι πελάτες αποχωρούν από το σύστημα αφού εξυπηρετηθούν. Ο όρος <<πελάτης>> είναι γενικός και δεν αναφέρεται απαραίτητα σ έναν άνθρωπο πελάτη. Για παράδειγμα πελάτης μπορεί να είναι ένα αεροπλάνο το οποίο περιμένει στην ουρά για ν απογειωθεί, ή ένα υπολογιστικό πρόγραμμα που περιμένει να τρέξει. Ένα τέτοιο σύστημα περιγράφεται γραφικά στην παρακάτω εικόνα.

***Εικόνα 1**: Γραφική αναπαράσταση ενός συστήματος ουράς.*



Η θεωρία ουρών αναπτύχθηκε για να παρέχει μοντέλα που προβλέπουν την συμπεριφορά των συστημάτων που προσπαθούν να παρέχουν εξυπηρέτηση για τυχαία εμφανιζόμενες απαιτήσεις, όχι αφύσικα, έτσι τα πρώτα προβλήματα που μελετήθηκαν ήταν εκείνα της τηλεφωνικής κυκλοφοριακής συμφόρησης.

Ο πρώτος ερευνητής ήταν ο Δανός μαθηματικός A.K. Erlang το 1909, ο οποίος στη πορεία παρατήρησε ότι ένα τηλεφωνικό σύστημα χαρακτηρίζεται από ένα από τα δυο:

- Poisson αφίξεις, εκθετικούς χρόνους εξυπηρέτησης και πολλούς εξυπηρετητές, ή
- Poisson άφιξη , σταθερό χρόνο εξυπηρέτησης και έναν εξυπηρετητή

Στον Erlang οφείλεται επίσης η έννοια της στατικής ισορροπίας

1.2 Χαρακτηριστικά Συστημάτων Ουράς

Στις περισσότερες περιπτώσεις, 6 βασικά χαρακτηριστικά μπορούν να μας παρέχουν επαρκή περιγραφή του συστήματος της ουράς, τα εξής:

- **1.2.1 Η διαδικασία άφιξης των πελατών.**
  Αναφέρεται στον τρόπο με τον οποίο φτάνουν οι πελάτες στο σύστημα και καθορίζεται συνήθως από τον μέσο ρυθμό άφιξης των πελατών και την κατανομή των αφίξεων ή τον μέσο χρόνο αναμονής μεταξύ δυο διαδοχικών αφίξεων και την εξάρτηση των χρόνων αυτών. Μπορούμε να έχουμε είτε *κανονικές* είτε *τυχαίες* αφίξεις. Έχουμε κανονικές αφίξεις όταν οι πελάτες φτάνουν ένας-ένας σε ίσα χρονικά διαστήματα. Πολλές φορές οι αφίξεις των πελατών δεν γίνονται σε ίσα χρονικά διαστήματα αλλά ακολουθούν κάποια συγκεκριμένη κατανομή, ή ενώ οι πελάτες θα έπρεπε να φτάνουν σε ίσα χρονικά διαστήματα φτάνουν με καθυστερήσεις, ή δεν φτάνουν ένας-ένας αλλά κατά ομάδες, ή ακόμα η διαδικασία άφιξης των πελατών δεν είναι στάσιμη καθ' όλη τη διάρκεια της εξυπηρέτησης αλλά μεταβάλλεται από χρονική στιγμή σε χρονική στιγμή, ή μπορεί να συμβεί οι αφίξεις να εξαρτώνται από διάφορα χαρακτηριστικά του συστήματος ή να έχουμε συνεχή ροή πελατών. Για όλες τις παραπάνω περιπτώσεις λέμε ότι έχουμε τυχαίες αφίξεις, κάτι που είναι πιο συνηθισμένο στην πράξη.
  Στις συνήθεις περιπτώσεις ουρών, η διαδικασία αφίξεων των πελατών είναι στοχαστική και είναι αναγκαίο να γνωρίζουμε τη συνάρτηση κατανομής των χρόνων μεταξύ των επιτυχημένων αφίξεων (των πελατών). Είναι επίσης απαραίτητο να γνωρίζουμε εάν οι πελάτες φτάνουν συγχρόνως (batch ή bulk αφίξεις) και αν ναι την συνάρτηση κατανομής που περιγράφει το μέγεθος του batch. Επιπλέον θα πρέπει να γνωρίζουμε την αντίδραση ενός πελάτη κατά την είσοδο του στο σύστημα. Ένας πελάτης μπορεί να αποφασίσει να περιμένει χωρίς να τον ενδιαφέρει το μέγεθος της ουράς, από την άλλη αν η ουρά είναι πολύ μεγάλη ο πελάτης μπορεί να αποφασίσει να μην εισέλθει στο σύστημα. Ένας πελάτης μπορεί να εισέλθει στην ουρά, αλλά ύστερα από λίγο να χάσει την υπομονή του και να αποχωρήσει. Στην περίπτωση που υπάρχουν δύο ή περισσότερες παράλληλες ουρές αναμονής, ο πελάτης μπορεί να αλλάξει από την μία στην άλλη. Ένας τελικός παράγοντας που πρέπει να ληφθεί υπόψη σχετικά με τη διαδικασία άφιξης είναι ο τρόπος με τον οποίο μεταβάλλεται η

διαδικασία σε σχέση με το χρόνο. Όταν η διαδικασία μεταβάλλεται από χρονική στιγμή σε χρονική στιγμή τότε λέμε ότι η διαδικασία άφιξης είναι μη στάσιμη, ενώ στην περίπτωση που μεταβάλλεται λέμε ότι είναι στάσιμη.

1.2.2 Η διαδικασία εξυπηρέτησης των πελατών.

Τρία είναι τα βασικά στοιχεία της διαδικασίας εξυπηρέτησης, ο χρόνος, η δυνατότητα και η διαθεσιμότητα εξυπηρέτησης. Ο χρόνος εξυπηρέτησης είναι ο χρόνος που χρειάζεται για την εξυπηρέτηση του πελάτη. Υποθέτουμε ότι οι τ.μ. που εκφράζουν τους χρόνους εξυπηρέτησης των πελατών είναι ανεξάρτητες μεταξύ τους και ακολουθούν την ίδια κατανομή που ονομάζεται κατανομή του χρόνου εξυπηρέτησης. Ο χρόνος αυτός μπορεί να υπολογιστεί επίσης αν είναι γνωστή η διαδικασία με την οποία οι πελάτες που έχουν ήδη εξυπηρετηθεί εγκαταλείπουν το σύστημα. Η δυνατότητα εξυπηρέτησης αναφέρεται στον μέγιστο αριθμό πελατών που μπορεί να εξυπηρετήσει το σύστημα σε μια δεδομένη χρονική στιγμή. Συνήθως ο αριθμός των πελατών που εξυπηρετούνται είναι μικρότερος από την δυνατότητα εξυπηρέτησης του συστήματος. Επίσης είναι δυνατόν να υπάρχουν συστήματα με απεριόριστη δυνατότητα εξυπηρέτησης. Η διαθεσιμότητα αναφέρεται στο χρονικό διάστημα κατά το οποίο είναι δυνατή η εξυπηρέτηση.

Η εξυπηρέτηση μπορεί να γίνεται κατ' άτομο ή κατ' ομάδα. Γενικά στο μυαλό μας υπάρχει η εξυπηρέτηση ενός ατόμου κάθε φορά από έναν εξυπηρετητή, αλλά υπάρχουν περιπτώσεις που πολλοί πελάτες εξυπηρετούνται ταυτόχρονα από τον ίδιο εξυπηρετητή, όπως ένας υπολογιστής που κάνει παράλληλη επεξεργασία ή τουρίστες οι οποίοι παρακολουθούν μια οργανωμένη ξενάγηση, ή άνθρωποι οι οποίοι επιβιβάζονται σ' ένα τραίνο. Όπως αναφέρθηκε, η διαδικασία εξυπηρέτησης μπορεί να εξαρτηθεί από τον αριθμό τον πελατών που περιμένουν να εξυπηρετηθούν. Ένας εξυπηρετητής μπορεί να δουλεύει γρηγορότερα αν η ουρά μεγαλώνει, ή αντίθετα μπορεί να ταραχθεί και να γίνει λιγότερο αποδοτικός.

1.2.3 Η πειθαρχία της ουράς.

Είναι ο τρόπος με τον οποίο επιλέγονται οι πελάτες από το σύστημα για να εξυπηρετηθούν. Η πιο κοινή πειθαρχία ουράς που παρατηρείται καθημερινώς είναι η FCFS (First Come First Served), που όμως δεν είναι η μοναδική πειθαρχία ουράς. Συχνά συναντάμε την LCFS (Last Come First Served), η οποία έχει εφαρμογή πχ σε συστήματα απογραφής όπου δεν γίνεται παλαίωση των αποθηκευμένων μονάδων δεδομένου ότι είναι ευκολότερη η αναζήτηση των κοντινότερων αντικειμένων, τα οποία εισήχθησαν τελευταία. Επίσης υπάρχει η επιλογή για εξυπηρέτηση ανεξάρτητα από την ώρα άφιξης στην ουρά RSS, και

μια ποικιλία συστημάτων προτεραιότητας, όπου δίνεται προτεραιότητα στον πελάτη κατά την είσοδο του στο σύστημα, έτσι ο πελάτης με την υψηλότερη προτεραιότητα θα επιλεγεί για εξυπηρέτηση, έναντι κάποιου άλλου με την χαμηλότερη προτεραιότητα, άσχετα με την ώρα άφιξης στο σύστημα. Υπάρχουν δύο γενικές περιπτώσεις στην πειθαρχεία ουράς με βάση τη προτεραιότητα. Στη πρώτη περίπτωση ο πελάτης με την υψηλότερη προτεραιότητα επιτρέπεται να μπει στην φάση της εξυπηρέτησης άμεσα, ακόμα και εκείνη τη στιγμή ένας πελάτης χαμηλότερης προτεραιότητας εξυπηρετείται όταν αυτός εισέλθει στο σύστημα, τότε η εξυπηρέτηση του πελάτη χαμηλής προτεραιότητας σταματάει με σκοπό να συνεχιστεί αφού εξυπηρετηθεί ο πελάτης υψηλής προτεραιότητας. Στην δεύτερη περίπτωση ο υψηλής προτεραιότητας πελάτης μπαίνει πρώτος στην ουρά, αλλά δεν μπορεί να μπει στην διαδικασία εξυπηρέτησης μέχρι να εξυπηρετηθεί ο πελάτης που βρίσκεται στην διαδικασία εξυπηρέτησης, ακόμα και αν αυτός είναι χαμηλότερης προτεραιότητας.

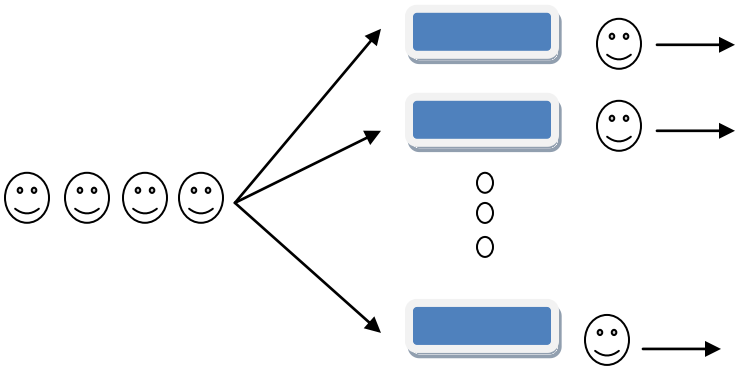
1.2.4 Η χωρητικότητα του συστήματος.

Σε κάποια συστήματα ουρών υπάρχει περιορισμός στον χώρο αναμονής, έτσι όταν η ουρά φτάσει σε κάποιο συγκεκριμένο μήκος δεν επιτρέπεται να εισέλθει άλλος πελάτης, μέχρι να δημιουργηθεί χώρος από την εξυπηρέτηση κάποιου.

1.2.5 Ο αριθμός των σημείων εξυπηρέτησης.

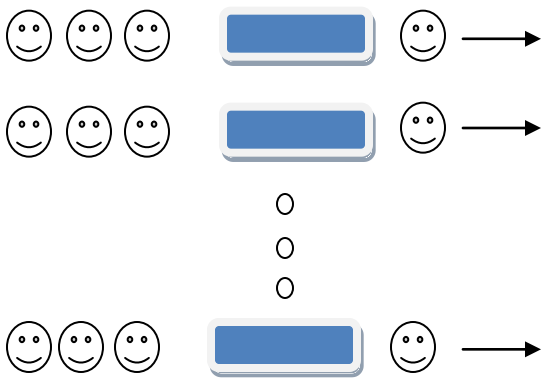
Ο αριθμός των σημείων εξυπηρέτησης αναφέρεται στον αριθμό των παράλληλων σταθμών εξυπηρέτησης που εξυπηρετούν τους πελάτες συγχρόνως. Υπάρχουν δύο περιπτώσεις συστημάτων με πολλούς σταθμούς εξυπηρέτησης:

1. Όταν δημιουργείται μια ουρά για όλους τους εξυπηρετητές:



Εικόνα 2: Ένα τέτοιο παράδειγμα είναι ένα κομμωτήριο με πολλές καρέκλες

2. Όταν κάθε εξυπηρετητής έχει δική του ουρά:



Εικόνα 3: Ένα τέτοιο παράδειγμα είναι ένα supermarket ή ένα fast food εστιατόριο.

- **1.2.6 Στάδια της εξυπηρέτησης.**
  Ένα σύστημα ουράς μπορεί να έχει ένα μόνο στάδιο εξυπηρέτησης, όπως ένα κομμωτήριο, ή μπορεί να έχει περισσότερα στάδια εξυπηρέτησης. Ένα τέτοιο σύστημα είναι οι διαδικασία γενικών εξετάσεων υγείας, όπου κάθε ασθενής πρέπει να περάσει διάφορα στάδια, όπως εξέταση αίματος, ηλεκτροκαρδιογράφημα, οφθαλμολογική εξέταση κλπ

1.3 Συμβολική Παράσταση

Ένα σύστημα ουράς περιγράφεται με τη βοήθεια μιας σειράς από 5 σύμβολα όπως φαίνεται παρακάτω. Τα 3 πρώτα σύμβολα προτάθηκαν από τον Kendal (1953) (γνωστή ως κωδικοποίηση Kendal), ενώ τα δύο τελευταία από τον A. Lee (1966)

a/b/s/k/N

a: εκφράζει τη διαδικασία αφίξεων, πχ

- **M: Poisson**
- **D: κανονικές αφίξεις**
- **E_k: κατανομή Erlang παραμέτρου κ**
- **GI: γενική ανεξάρτητη**

b: εκφράζει την κατανομή του χρόνου εξυπηρέτησης, πχ

- **M: Εκθετική**
- **D: Κανονική**
- **$E_κ$: κατανομή Erlang παραμέτρου κ**
- **G: Γενική**

s: ο αριθμός των παράλληλων σημείων εξυπηρέτησης (**1,2,…..∞)**

k: είναι η χωρητικότητα του συστήματος εξυπηρέτησης (**1,2,…..∞)**

N: πλήθος πελατών στην πηγή, όταν αυτό είναι πεπερασμένο.

1.4 Γενικά συμπεράσματα

Σ' αυτή την ενότητα θα παρουσιαστούν κάποια γενικά αποτελέσματα και σχέσεις για τις ουρές G/G/1 και G/G/c. Έστω λ ο μέσος ρυθμός άφιξης των πελατών στο σύστημα και μ ο μέσος ρυθμός εξυπηρέτησης, ένα μέτρο κυκλοφοριακής συμφόρησης για ένα σύστημα με c εξυπηρετητές είναι το $\rho \equiv \frac{\lambda}{c\ m}$ που συχνά καλείται traffic intensity.

Όταν $\rho > 1\ δηλαδή\ \lambda > c\mu$ ο μέσος όρος των αφίξεων στο σύστημα υπερβαίνει το μέγιστο ρυθμό εξυπηρέτησης του συστήματος και είναι αναμενόμενο όσο περνά η ώρα η ουρά να μεγαλώνει μέχρι το σημείο που οι πελάτες δεν θα μπορούν πλέον να εισέλθουν στο σύστημα. Επομένως αν επιθυμούμε συνθήκες ισορροπίας (δηλαδή της κατάστασης του συστήματος αφού έχει λειτουργήσει για μεγάλο διάστημα) παίρνουμε αυστηρά το $\rho < 1$.

Επομένως αν γνωρίζουμε τον μέσο ρυθμό άφιξης και τον μέσο ρυθμό εξυπηρέτησης, ο ελάχιστος αριθμός παράλληλων εξυπηρετητών που απαιτούνται μπορεί να υπολογιστεί βρίσκοντας το μικρότερο c έτσι ώστε $\frac{\lambda}{c\ \mu} < 1$.

Έστω N(t) ο συνολικός αριθμός πελατών στο σύστημα την χρονική στιγμή t, $N_q(t)$ ο αριθμός πελατών που περιμένουν στην ουρά και $N_s(t)$ ο αριθμός των πελατών του εξυπηρετούνται, τότε προφανώς: $N(t) = N_q(t) + N_s(t)$.

Αν $p_n(t) = P\{N(t) = n\} \kappa\alpha\iota\ p_n = P\{N = n\}$, τότε ο μέσος αριθμός πελατών στο σύστημα είναι:

$L = E(N) = \sum_{n=0}^{\infty} np_n$

Και ο αναμενόμενος αριθμός πελατών στην ουρά είναι:

$L_q = E(N_q) = \sum_{n=c+1}^{\infty}(n-c)p_n$.

1.5 Θεώρημα του Little

Μία από τις ισχυρότερες σχέσεις στην θεωρία ουρών αναπτύχθηκε από τον John D.C. Little στις αρχές του 1960. Το θεώρημα αυτό διαισθητικά αναφέρει ότι εάν σ' ένα σύστημα ουράς το (μέσο) μήκος της ουράς είναι μεγάλο, τότε ο (μέσος) χρόνος αναμονής σ' αυτήν αναμένεται να είναι μεγάλος. Υποθέτουμε ότι στο σύστημα έχουμε κατάσταση στατιστικής ισορροπίας, γεγονός που σημαίνει ότι η πιθανότητα να έχουμε έναν συγκεκριμένο αριθμό πελατών είναι ανεξάρτητο του χρόνου, τότε εάν:

T_q : ο χρόνος αναμονής ενός πελάτη στην ουρά

T: ο συνολικός χρόνος παραμονής του πελάτη στο σύστημα

S: ο χρόνος εξυπηρέτησης του πελάτη

Τότε προφανώς: $T = T_q + S,$ όπου T, T_q, S είναι τυχαίες μεταβλητές.

Εάν $W_q = E(T_q)$ είναι ο μέσος χρόνος αναμονής ενός πελάτης την ουρά και W=E(T) είναι ο μέσος χρόνος παραμονής στο σύστημα, τότε σύμφωνα με το θεώρημα του Little:

$\boldsymbol{L = \lambda W}$ και $\boldsymbol{L_q = \lambda W_q}$.

Και αφού $E(T) = E(T_q) + E(S)$ τότε έχουμε: $W = W_q + \frac{1}{\mu}$, όπου μ ο μέσος ρυθμός εξυπηρέτησης.

Ένα πολύ σημαντικό συμπέρασμα που προκύπτει από το θεώρημα του Little και της σχέσης μεταξύ των W και W_q είναι το εξής:

$$L - L_q = \lambda(W - W_q) = \lambda \frac{1}{\mu} = \frac{\lambda}{\mu}$$

Αλλά:

$$L - L_q = E(N) - E(N_q) = E(N - N_q) = E(N_s),$$

Έτσι ο αναμενόμενος αριθμός πελατών που εξυπηρετούνται σε κατάσταση στατιστικής ισορροπίας είναι $\frac{\lambda}{\mu} = r$.

Στην περίπτωση συστήματος με έναν σταθμό εξυπηρέτηση, τότε r=ρ και

$$L - L_q = \sum_{n-1}^{\infty} np_n - \sum_{n=1}^{\infty} (n-1)p_n = \sum_{n=1}^{\infty} p_n = 1 - p_0$$

Από το παραπάνω μπορούμε εύκολα να βρούμε την πιθανότητα p_b ενός σταθμού εξυπηρέτησης να είναι απασχολημένος σ' ένα σύστημα με πολλούς σταθμούς εξυπηρέτησης σε στατιστική ισορροπία. Έχουμε δείξει ότι ο αναμενόμενος αριθμός που βρίσκονται στην εξυπηρέτηση είναι r, τότε αν έχουμε c σταθμούς εξυπηρέτησης ο αναμενόμενος αριθμός πελατών σ' ένα σταθμό είναι r/c. Τότε εύκολα μπορούμε να δείξουμε ότι p_b=ρ από τη στιγμή που: r/c=ρ= 0 (1-p_b) + 1 p_b.

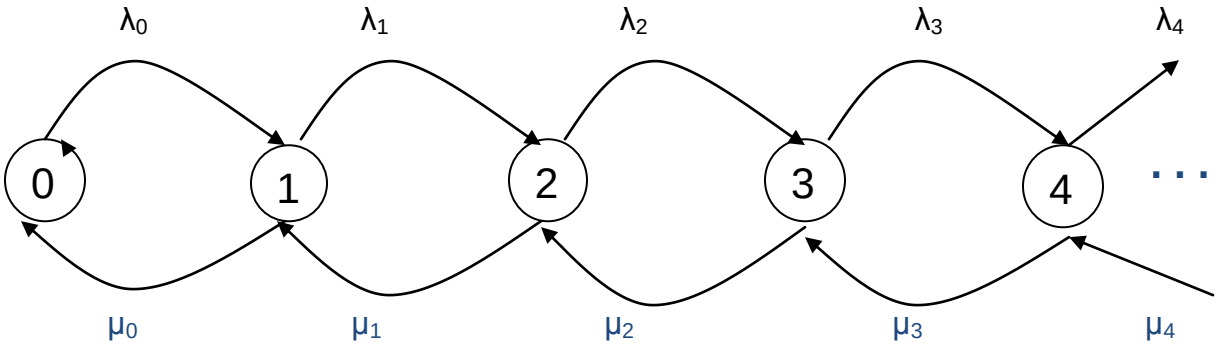
Στην περίπτωση που έχουμε ένα σταθμό εξυπηρέτησης (G/G/1) η πιθανότητα το σύστημα να είναι άεργο (N=0) είναι ίδια με την πιθανότητα ο σταθμός εξυπηρέτησης να είναι άεργος. Δηλαδή $p_0 = 1 - p_b$ και σ' αυτή την περίπτωση $p_0 = 1 - \rho = 1 - r = 1 - \frac{\lambda}{\mu}$.

Κατά μέσο όρο ένας πελάτης χρειάζεται 1/μ μονάδες εξυπηρέτησης και ο μέσος αριθμός πελατών που φτάνουν ανά μονάδα χρόνου είναι λ, έτσι η ποσότητα λ (1/μ) είναι το ποσό της δουλείας που φτάνει στο σύστημα ανά μονάδα χρόνου. Αν διαιρέσουμε με τον αριθμό των σταθμών εξυπηρέτησης δηλαδή με c θα έχουμε το μέσο όρο δουλείας που φτάνει σε κάθε σταθμό εξυπηρέτησης ανά μονάδα χρόνου.

ΚΕΦΑΛΑΙΟ 2: ΜΟΝΤΕΛΑ ΟΥΡΩΝ

2.1 Ανέλιξη Γέννησης – Θανάτου

Μια ανέλιξη γέννησης-θανάτου είναι μια συνεχής αλυσίδα Markov. Αποτελείται από ένα σύνολο καταστάσεων {0,1,2,….} που δείχνουν τον πληθυσμό κάποιου συστήματος. Η μετάβαση από τη μία κατάσταση στην άλλη εμφανίζονται σαν ένα άλμα προς τα πάνω ή κάτω από την τρέχουσα κατάσταση. Ειδικότερα όταν το σύστημα είναι στην κατάσταση $n \geq 0$, ο χρόνος μέχρι την επόμενη άφιξη (γέννηση) είναι μια τυχαία μεταβλητή που ακολουθεί εκθετική κατανομή με ρυθμό λ_n. Με την άφιξη το σύστημα περνά από την κατάσταση n στην κατάσταση n+1. Όταν το σύστημα είναι στην κατάσταση $n \geq 1$, ο χρόνος μέχρι την επόμενη αποχώρηση (θάνατο) είναι τυχαία μεταβλητή που ακολουθεί την εκθετική κατανομή με ρυθμό μ_n. Κατά την αναχώρηση το σύστημα περνά από την κατάσταση n στην κατάσταση n-1. Αυτή είναι μια συνεχούς χρόνου Μαρκοβιανή αλυσίδα, με το παρακάτω διάγραμμα μετάβασης: *Εικόνα 4:*



Στη θεωρία ουρών οι καταστάσεις δείχνουν τον αριθμό των πελατών στο σύστημα. Οι γεννήσεις αντιστοιχούν στις αφίξεις των πελατών και οι θάνατοι στις αναχωρήσεις. Για παράδειγμα η ουρά M/M/1 είναι μια ανέλιξη γέννησης-θανάτου με $\lambda_n=\lambda$ και $\mu_n=\mu$. Αν p_n είναι το μεγαλύτερο μέρος του χρόνου που το σύστημα είναι στην κατάσταση n, τότε μια υπάρχει μία λύση για την $\{p_n\}$ και μπορεί να καθοριστεί από τις παρακάτω εξισώσεις:

$$0 = -(\lambda_n + \mu_n)p_n + \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1}, \qquad n \geq 1$$

$$0 = -\lambda_0 p_0 + \mu_1 p_1$$

Ή

$$(\lambda_n + \mu_n)p_n = \lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1}, \quad n \geq 1 \qquad (1.1)$$

$$\lambda_0 p_0 = \mu_1 p_1$$

Σε κατάσταση ισορροπίας ο ρυθμός των μεταβάσεων από μια δεδομένη κατάσταση πρέπει να είναι ίδιος με τον ρυθμό των μεταβάσεων στην ίδια κατάσταση. Το αριστερό μέλος της σχέσεις (1.1) μας δείχνει των ρυθμό των μεταβάσεων από την κατάσταση n, ενώ το δεξί μέλος της μας δείχνει τον αριθμό των μεταβάσεων στην κατάσταση n. Αυτό εξηγείται ως εξής:

Όταν το σύστημα είναι στην κατάσταση n, ο μέσος ρυθμός αφίξεων (γεννήσεων) είναι λ_n αφίξεις ανά μονάδα χρόνου. Όταν το σύστημα είναι στην κατάσταση n ένα κλάσμα του χρόνου p_n, τότε $\lambda_n p_n$ είναι ο μακροπρόθεσμος ρυθμός των μεταβάσεων από την n στην n+1. Ομοίως στην κατάσταση n ο μέσος ρυθμός αναχωρήσεων (θανάτων) είναι μ_n αναχωρήσεις ανά μονάδα χρόνου. Έτσι το $\mu_n p_n$ είναι ο μακροπρόθεσμος ρυθμός των μεταβάσεων από την n στην n-1. Αφού η μετάβαση από την κατάσταση n μπορεί να είναι είτε προς τα επάνω είτε προς τα κάτω, το $(\lambda_n + \mu_n)p_n$ είναι ο μακροπρόθεσμος ρυθμός μεταβάσεων από την κατάσταση n. Ομοίως από τη στιγμή που οι μεταβάσεις στην κατάσταση n μπορούν να μετρηθούν από την από κάτω κατάσταση (n-1) είτε από την από πάνω (n+1), ο μακροπρόθεσμος ρυθμός των μεταβάσεων στην κατάσταση n είναι $\lambda_{n-1}p_{n-1} + \mu_{n+1}p_{n+1}$.

Η δεύτερη εξίσωση $\lambda_0 p_0 = \mu_1 p_1$ αφορά την κατάσταση 0, αυτή η κατάσταση είναι διαφορετική από τις υπόλοιπες μιας και καμία αναχώρηση δεν μπορεί να πραγματοποιηθεί όταν είναι 0 στο σύστημα και καμία άφιξη μπορεί να πραγματοποιηθεί με αποτέλεσμα 0 στο σύστημα.

$$p_{n+1} = \frac{\lambda_n + \mu_n}{\mu_{n+1}} p_n - \frac{\lambda_{n-1}}{\mu_{n+1}} p_{n-1}, \qquad n \geq 1$$

$$p_1 = \frac{\lambda_0}{\mu_1} p_0$$

Από τις παραπάνω προκύπτει:

$$p_2 = \frac{\lambda_1 + \mu_1}{\mu_2} p_1 - \frac{\lambda_0}{\mu_2} p_0 = \frac{\lambda_1 + \mu_1}{\mu_2} \frac{\lambda_0}{\mu_1} p_0 - \frac{\lambda_0}{\mu_2} p_0 = \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0$$

$$p_3 = \frac{\lambda_2 + \mu_2}{\mu_3} p_2 - \frac{\lambda_1}{\mu_3} p_1 = \frac{\lambda_2 + \mu_2}{\mu_3} \frac{\lambda_1 \lambda_0}{\mu_2 \mu_1} p_0 - \frac{\lambda_1 \lambda_0}{\mu_3 \mu_1} p_0 = \frac{\lambda_2 \lambda_1 \lambda_0}{\mu_3 \mu_2 \mu_1} p_0$$

και καταλήγουμε:

$$p_n = \frac{\lambda_{n-1}\lambda_{n-2} \dots \lambda_0}{\mu_n \mu_{n-1} \dots \mu_0} p_0, \quad n \geq 1 \quad = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}$$

Η παραπάνω σχέση ισχύει για n=0 και δείξαμε παραπάνω ότι ισχύει για n=1,2,3. Έστω ότι ισχύει για n=k, τότε θα αποδείξουμε ότι ισχύει για n=k+1:

$$p_{\kappa+1} = \frac{\lambda_\kappa + \mu_\kappa}{\mu_{\kappa+1}} p_\kappa - \frac{\lambda_{\kappa-1}}{\mu_{\kappa+1}} p_{\kappa-1}$$

$$= \frac{\lambda_\kappa + \mu_\kappa}{\mu_{\kappa+1}} p_0 \prod_{i=1}^{\kappa} \frac{\lambda_{i-1}}{\mu_\iota} - \frac{\lambda_{\kappa-1}}{\mu_{\kappa+1}} p_0 \prod_{i=1}^{\kappa-1} \frac{\lambda_{i-1}}{\mu_\iota}$$

$$= \frac{p_0 \lambda_\kappa}{\mu_{\kappa+1}} \prod_{i=1}^{\kappa} \frac{\lambda_{i-1}}{\mu_\iota} + \frac{p_0 \mu_\kappa}{\mu_{\kappa+1}} \prod_{i=1}^{\kappa} \frac{\lambda_{i-1}}{\mu_\iota} - \frac{p_0 \mu_\kappa}{\mu_{\kappa+1}} \prod_{i=1}^{\kappa} \frac{\lambda_{i-1}}{\mu_\iota} = p_0 \prod_{i=1}^{\kappa+1} \frac{\lambda_{i-1}}{\mu_\iota}$$

Και επειδή το άθροισμα των πιθανοτήτων πρέπει να είναι 1, έχουμε ότι:

$$p_0 = (1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i})^{-1}$$

2.2 Η ουρά M/M/1

Έστω ότι οι ενδιάμεσοι χρόνοι αφίξεων και οι χρόνοι εξυπηρέτησης ακολουθούν εκθετική κατανομή με συναρτήσεις πυκνοτήτων:

$$\alpha(t) = \lambda e^{-\lambda t}$$

$$b(t) = \mu e^{-\mu t}$$

Έστω n ο αριθμός των πελατών στο σύστημα. Οι αφίξεις μπορούν να θεωρηθούν ως γεννήσεις στο σύστημα, ενώ οι αναχωρήσεις ως θάνατοι. Ο ρυθμός λ των αφίξεων και ο ρυθμός της εξυπηρέτησης μ είναι καθορισμένοι και ανεξάρτητοι από των αριθμό των πελατών στο σύστημα (εφόσον είναι τουλάχιστον ένας πελάτης στο σύστημα).

Η M/M/1 ουρά είναι μια διαδικασία γέννησης-θανάτου με $\lambda_n=\lambda$ ($n\geq 0$), $\mu_n=\mu$ ($n\geq 1$). Επομένως έχουμε τις εξισώσεις:

$$(\lambda + \mu) p_n = \mu p_{n+1} + \lambda p_{n-1}\,, \quad n \geq 1$$

$$\lambda p_0 = \mu p_1$$

Και $$p_n = (1-\rho)\rho^n,\ \text{όπου}\ \rho = \frac{\lambda}{\mu} < 1$$

2.2.1 Μέτρα Αποδοτικότητας

Έστω N η τυχαία μεταβλητή που αντιπροσωπεύει τον αριθμό των πελατών στο σύστημα σε κατάσταση ισορροπίας και L η αναμενόμενη τιμή της, τότε:

$$L = E(N) = \sum_{n=0}^{\infty} np_n = (1-\rho)\sum_{n=0}^{\infty} n\rho^n =$$
$$(1-\rho)\rho\sum_{n=1}^{\infty} n\rho^{n-1} = (1-\rho)\rho\frac{1}{(1-\rho)^2} = \frac{\rho}{1-\rho}$$

Επομένως: $L = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda}$

Έστω N_q η τυχαία μεταβλητή που αντιπροσωπεύει των αριθμό των πελατών στην ουρά και L_q η αναμενόμενη τιμή της, τότε:

$$L_q = \sum_{n=1}^{\infty}(n-1)p_n = \sum_{n=1}^{\infty} np_n - \sum_{n=1}^{\infty} p_n = L - (1-p_0) =$$
$$\frac{\rho}{1-\rho} - \rho = \frac{\rho^2}{1-\rho}$$

Επομένως: $L_q = \frac{\rho^2}{1-\rho} = \frac{\lambda^2}{\mu(\mu-\lambda)}$

Έστω L'_q το αναμενόμενο μέγεθος της ουράς στην περίπτωση μη κενών ουρών, όταν θέλουμε δηλαδή να αγνοήσουμε τις περιπτώσεις που η ουρά είναι κενή, τότε:

$$L'_q = E\left[N_q | N_q \neq 0\right] = \sum_{n=1}^{\infty}(n-1)p'_n = \sum_{n=2}^{\infty}(n-1)p'_n$$

Όπου:

$$p'_n = P(n\ στο\ σύστημα\ /n \geq 2) = \frac{P(n\ στο\ σύστημα\ \ και\ n \geq 2)}{P(n \geq 2)} =$$
$$\frac{p_n}{\sum_{n=2}^{\infty} p_n}\ όπου\ n \geq 2 = \frac{p_n}{1-(1-\rho)-(1-\rho)\rho} = \frac{p_n}{\rho^2}$$

Επομένως:

$$L'_q = \sum_{n=2}^{\infty}(n-1)\frac{p_n}{\rho^2} = \frac{L-p_1-(1-p_0-p_1)}{\rho^2}$$ και καταλήγουμε:

$$L'_q = \frac{1}{1-\rho} = \frac{\mu}{\mu-\lambda}$$

Αποδεικνύεται ότι:

$$P(N \geq n) = \rho^n$$

Έχουμε ότι: $P(N \geq n) = \sum_{k=n}^{\infty}(1-\rho)\rho^k = (1-\rho)\rho^n \sum_{k=n}^{\infty} \rho^{k-n} = \frac{(1-\rho)\rho^n}{1-\rho} = \rho^n$

Από το Θεώρημα του Little έχουμε:

$$L = \lambda W \ \ \kappa\alpha\iota \ \ L_q = \lambda W_q$$

Επομένως:

$$W = \frac{L}{\lambda} = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda}$$

$$W_q = \frac{L_q}{\lambda} = \frac{\rho}{\mu(1-\rho)} = \frac{\rho}{\mu - \lambda}$$

2.2.2 Κατανομή του χρόνου αναμονής:

Έστω T_q η τυχαία μεταβλητή του χρόνου αναμονής στην ουρά (σε κατάσταση ισορροπίας) και $W_q(t)$ αντιπροσωπεύει την αθροιστική κατανομή πιθανότητας. Έστω ότι η πειθαρχία ουράς είναι FCFS (First Come First Served), όποιος δηλαδή εισέλθει πρώτος εξυπηρετείται. Η τυχαία μεταβλητή του χρόνου αναμονής έχει την εξής ενδιαφέρουσα ιδιότητα είναι κατά ένα μέρος διακριτή και κατά ένα άλλο συνεχής. Η αναμονή στην ουρά είναι κατά το μεγαλύτερο μέρος μια συνεχής τυχαία μεταβλητή, εκτός όταν υπάρχει μια μη μηδενική πιθανότητα όταν η αναμονή είναι μηδέν. Αυτό συμβαίνει όταν το σύστημα είναι άδειο και ένας πελάτης που φτάνει αρχίζει την εξυπηρέτηση άμεσα κατά την άφιξη. Έστω q_n η πιθανότητα (σε κατάσταση ισορροπίας) ένας πελάτης που φτάνει στο σύστημα να βρει n (ακριβώς πριν την άφιξη), τότε:

$$W_q = P\left(T_q \leq 0\right) = P\left(T_q = 0\right) =$$
$$P(\tau o\ \sigma\acute{\upsilon}\sigma\tau\eta\mu\alpha\ \nu\alpha\ \epsilon\acute{\iota}\nu\alpha\iota\ \acute{\alpha}\delta\epsilon\iota o\ \sigma\epsilon\ \mu\iota\alpha\ \acute{\alpha}\varphi\iota\xi\eta) = q_0$$

Στην περίπτωση της κατανομής Poisson $q_n=p_n$, έτσι:

$$W_q(0) = p_0 = 1 - \rho$$

Έτσι απομένει να βρούμε το $W_q(t)\ \ \acute{o}\tau\alpha\nu\ t > 0$.

Θεωρούμε $W_q(t)$ την πιθανότητα ένας πελάτης να περιμένει για εξυπηρέτηση χρόνο λιγότερο ή ίσο με t. Αν υπάρχουν n πελάτες στο σύστημα για εξυπηρέτηση κατά την άφιξη, τότε αν θέλουμε ο πελάτης να εξυπηρετηθεί σε χρόνο μεταξύ 0 και t τότε όλοι οι προηγούμενοι n πελάτες πρέπει να έχουν εξυπηρετηθεί μέχρι το χρόνο t. Η κατανομή του χρόνου που απαιτείται για n ολοκληρώσεις είναι ανεξάρτητη είναι ανεξάρτητη του χρόνου της τρέχουσας άφιξης και είναι συνέλιξη n εκθετικών τυχαίων μεταβλητών. Αυτή είναι μια Erlang κατανομή. Έτσι:

$$W_q(t) = P\big(T_q \le t\big) =$$

$$= W_q(0) + \sum_{n=1}^{\infty} P(n\ ολοκληρώσεις\ σε\ χρόνο \le t\ \ /\ \ άφιξη\ βρίσκει\ n\ στο\ συστημα)\ p_n =$$

$$= 1 - \rho + (1-\rho)\sum_{n=1}^{\infty} \rho^n \int_0^t \frac{\mu(\mu\chi)^{n-1}}{(n-1)!} e^{-\mu\chi}\, dx$$

$$= 1 - \rho + \rho \int_0^t \mu(1-\rho) e^{-\mu\chi} \sum_{n=1}^{\infty} \frac{(\mu\chi\rho)^{n-1}}{(n-1)!} dx =$$

$$= 1 - \rho + \rho \int_0^t \mu(1-\rho) e^{-\mu(1-\rho)\chi}\, dx$$

Στην τελευταία γραμμή φαίνεται ότι το $W_q(t)$ αντιπροσωπεύει τη μίξη μιας διακριτής τυχαίας μεταβλητής και μιας συνεχούς. Δεδομένου ότι το δεξί μέρος είναι μια αθροιστική συνάρτηση πιθανότητας μιας εκθετικής, μια απλοποιημένη μορφή της παραπάνω σχέσης είναι η εξής:

$$W_q(t) = 1 - \rho e^{-(\mu-\lambda)t}, \qquad t \ge 0$$

Η μέση τιμή της κατανομής είναι: $W_q = \frac{\rho}{\mu-\lambda}$

$$W_q = \int_0^{\infty} \big[1 - W_q(t)\big] dt = \int_0^{\infty} \rho e^{-\mu(1-\rho)t} dt = \frac{\rho}{\mu-\lambda}$$

Έστω T ο συνολικός χρόνος που ένας πελάτης που φτάνει περνάει στο σύστημα. Η T είναι μια τυχαία μεταβλητή που ακολουθεί εκθετική κατανομή με μέση τιμή 1/(μ-λ). Έχουμε:

$$W(t) = 1 - e^{-(\mu-\lambda)t}, \qquad t \ge 0$$

$$w(t) = (\mu-\lambda) e^{-(\mu-\lambda)t}, \qquad t > 0$$

2.3 Η ουρά M/M/c

Η ουρά M/M/c είναι μια ουρά στην οποία οι αφίξεις γίνονται σύμφωνα με μια ανέλιξη Poisson με ρυθμό λ και οι χρόνοι εξυπηρέτησης είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές με κατανομή την εκθετική με παράμετρο μ. Υπάρχουν c εξυπηρετητές στο σύστημα της ουράς που εργάζονται ανεξάρτητα και παράλληλα ο ένας με τον άλλο. Αυτή η ουρά μοντελοποιείται σαν μια ανέλιξη γέννησης-θανάτου, με ρυθμό γέννησης $\lambda_n=\lambda$ για όλα τα n, ανεξάρτητα από τον αριθμό των πελατών στο σύστημα, αντίθετα ο ρυθμός των ολοκληρωμένων εξυπηρετήσεων (θανάτων) εξαρτάται από τον αριθμό των πελατών στο σύστημα. Κάθε εξυπηρετητής εξυπηρετεί με ρυθμό μ, οπότε ο ρυθμός εξυπηρέτησης για το σύστημα είναι cμ. Όταν υπάρχουν λιγότεροι από c πελάτες στο σύστημα n<c, μόνο n από τους c εξυπηρετητές είναι απασχολημένοι και ο ρυθμός εξυπηρέτησης του συστήματος είναι nμ. Ως εκ τούτου το μ_n μπορεί να γραφεί:

$$\mu_n = \begin{cases} n\mu & 1 \le n < c \\ c\mu & n \ge c \end{cases}$$

Αντικαθιστούμε τα λ_n και μ_n στον τύπο υπολογισμού των p_n σε μια ανέλιξη γέννησης θανάτου και έχουμε:

$$p_n = \begin{cases} \frac{\lambda^n}{n!\,\mu^n} p_{0,} & 0 \le n < c \\ \frac{\lambda^n}{c^{n-c} c!\,\mu^n} p_{0,} & n \ge c \end{cases}$$

Παρατηρούμε ότι η p_n έχει τη μορφή μιας Poisson τυχαίας μεταβλητής για $0 \le n < c$ και τη μορφή μιας γεωμετρικής τυχαίας μεταβλητής για $n \ge c$.

Όπου:

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\mu^n} + \sum_{n=c}^{\infty} \frac{\lambda^n}{c^{n-c} c!\mu^n} \right)^{-1}$$

Αντικαθιστούμε r=λ/μ και ρ=r/c=λ/cμ και έχουμε:

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \sum_{n=c}^{\infty} \frac{r^n}{c^{n-c}c!}\right)^{-1}$$

$$\sum_{n=c}^{\infty} \frac{r^n}{c^{n-c}c!} = \frac{r^c}{c!}\sum_{n-c}^{\infty}\left(\frac{r}{c}\right)^{n-c} = \frac{r^c}{c!}\sum_{m=0}^{\infty}\left(\frac{r}{c}\right)^{m} = \frac{r^c}{c!}\frac{1}{1-\frac{r}{c}}, \quad \frac{r}{c} = \rho < 1$$

Επομένως:

$$p_0 = \left(\frac{r^c}{c!}\frac{1}{1-\frac{r}{c}} + \sum_{n=0}^{c-1}\frac{r^n}{n!}\right)^{-1}, \frac{r}{c} = \rho < 1$$

Ο όρος για την ύπαρξη στατιστικής ισορροπίας είναι: λ/cμ<1, αυτό σημαίνει ότι ο μέσος ρυθμός αφίξεων είναι μικρότερος από τον μέσο δυνατό ρυθμό εξυπηρέτησης του συστήματος, γεγονός που αναμένεται διαισθητικά.

2.3.1 Μέτρα Λειτουργικότητας της M/M/c Ουράς

- **2.3.1.1 Μέσο Μήκος Ουράς:**

$$L_q = \sum_{n=c+1}^{\infty}(n-c)\,p_n = \sum_{n=c+1}^{\infty}(n-c)\frac{r^n}{c^{n-c}c!}p_0 \quad \begin{matrix}\rho = \frac{r}{c} \\ = \end{matrix} \quad \frac{r^c p_0}{c!}\sum_{n=c+1}^{\infty}(n-c)\,\rho^{n-c}$$

$$= \frac{r^c p_0}{c!}\sum_{m=1}^{\infty} m\rho^m = \frac{r^c \rho p_0}{c!}\sum_{m=1}^{\infty} m\rho^{m-1} = \frac{r^c \rho p_0}{c!}\sum_{m=1}^{\infty}\frac{d}{d\rho}\{\rho^m\}$$

$$= \frac{r^c \rho p_0}{c!}\frac{d}{d\rho}\{\sum_{m=1}^{\infty}\rho^m\} = \frac{r^c \rho p_0}{c!}\frac{d}{d\rho}\left\{\frac{1}{1-\rho} - 1\right\} = \frac{r^c \rho p_0}{c!\,(1-\rho)^2}$$

2.3.1.2 Μέσος Χρόνος Αναμονής Ενός Πελάτη Στην Ουρά

Από το Θεώρημα του Little:

$$L_q = \lambda W_q => W_q = \frac{L_q}{\lambda} = \left(\frac{r^c}{c!\,(c\mu)(1-\rho)^2}\right)p_0$$

2.3.1.3 Μέσος Χρόνος Παραμονής Ενός Πελάτη Στο Σύστημα:

$$W = W_q + \frac{1}{\mu} = \frac{1}{\mu} + \left(\frac{r^c}{c!\,(c\mu)(1-\rho)^2}\right)p_0$$

2.3.1.4 Μέσος Αριθμός Πελατών Στο Σύστημα:

$$L = \lambda W = r + L_q = r + \frac{r^c \rho p_0}{c!\,(1-\rho)^2}$$

Έστω $W_q(0)$ η πιθανότητα ένας πελάτης να έχει 0 καθυστέρηση στην ουρά πριν εξυπηρετηθεί. Έτσι 1- $W_q(0)$ είναι η πιθανότητα ένας πελάτης να έχει μη μηδενική καθυστέρηση στην ουρά πριν ξεκινήσει η εξυπηρέτηση.

Έστω T_q η τυχαία μεταβλητή του χρόνου που ξοδεύει ένας πελάτης στην ουρά (σε κατάσταση ισορροπίας). Τότε:

$$W_q(0) = P(T_q = 0) = P(\lambda\iota\gamma\acute{o}\tau\varepsilon\rho o\iota\ \acute{\eta}\ \acute{\iota}\sigma o\iota\ \mu\varepsilon\ c-1\ \sigma\tau o\ \sigma\acute{\upsilon}\sigma\tau\eta\mu\alpha) = \sum_{n=0}^{c-1} p_n = p_0 \sum_{n=0}^{c-1} \frac{r^n}{n!}$$

Όμως: $\sum_{n=0}^{c-1} \frac{r^n}{n!} = \frac{1}{p_0} - \frac{r^c}{c!(1-\rho)}$, έτσι:

$$W_q(0) = p_0\left(\frac{1}{p_0} - \frac{r^c}{c!\,(1-\rho)}\right)$$

Συνεπώς η πιθανότητα ένας πελάτης που φτάνει να έχει μη μηδενική αναμονή στην ουρά είναι:

$$C(c,r) = 1 - W_q(0) = \frac{\frac{r^c}{c!\,(1-\rho)}}{\left(\frac{r^c}{c!\,(1-\rho)} + \sum_{n=0}^{c-1}\frac{r^n}{n!}\right)}$$

Το C(c,r) καλείται φόρμουλα Erlang-C, και δίνει την πιθανότητα ένας πελάτης που φτάνει να καθυστερήσει στην ουρά σαν συνάρτηση των παραμέτρων c και r.

Για $T_q>0$ και θεωρώντας πειθαρχία ουράς την FCFS έχουμε:

$$\begin{aligned} W_q(t) = P\big(T_q \le t\big) \\ &= W_q(0) \\ &+ \sum_{n=c}^{\infty} P(n-c+1\ \varepsilon\xi\upsilon\pi\eta\rho\varepsilon\tau\acute{\eta}\sigma\varepsilon\iota\varsigma\ \sigma\varepsilon\ \chi\rho\acute{o}\nu o \\ &\le t\ \ /\ \ \ \mu\varepsilon\ \tau\eta\nu\ \acute{a}\varphi\iota\xi\eta\ \beta\rho\acute{\iota}\sigma\kappa\varepsilon\iota\ n\ \sigma\tau o\ \sigma\acute{\upsilon}\sigma\tau\eta\mu\alpha)\ p_n \end{aligned}$$

Όταν n≥c οι έξοδοι από το σύστημα ακολουθούν την Poisson με μέσο ρυθμό cμ, έτσι οι χρόνοι μεταξύ των ολοκληρωμένων εξυπηρετήσεων ακολουθούν την εκθετική με μέση τιμή 1/cμ, και η κατανομή του χρόνου για n-c+1 ολοκληρωμένες εξυπηρετήσεις είναι η Erlang τυπου n-c+1. Έτσι έχουμε:

$$\begin{aligned} \mathrm{W_q(t)} = \mathrm{W_q(0)} + \mathrm{p_0} \sum_{\mathrm{n=c}}^{\infty} \frac{\mathrm{r^n}}{\mathrm{c^{n-c}c!}} \int_0^{\mathrm{t}} \frac{\mathrm{c\mu(c\mu\chi)^{n-c}}}{\mathrm{(n-c)!}} \mathrm{e^{-c\mu x}dx} \\ &= \ W_q(0) + \frac{r^c p_0}{(c-1)!} \int_0^t \mu e^{-c\mu\chi} \sum_{n=c}^{\infty} \frac{(\mu r x)^{n-c}}{(n-c)!} dx \ \ = W_q(0) \\ &+ \frac{r^c p_0}{(c-1)!} \int_0^t \mu e^{-\mu\chi\,(c-r)} dx = \ \ W_q(0) \\ &+ \frac{r^c p_0}{(c-1)!\,(c-r)} \int_0^t \mu(c-r) e^{-\mu(c-r)x} dx \\ &= W_q(0) + \frac{r^c p_0}{c!\,(1-\rho)} (1 - e^{-(c\mu-\lambda)t}) \end{aligned}$$

Και αντικαθιστώντας το $W_q(0)$ έχουμε:

$$W_q(t) = 1 - \frac{r^c p_0}{c!\,(1-\rho)} e^{-(c\mu-\lambda)t}$$

Παρατηρούμε ότι:

$$P\big(T_q > t\big) = 1 - W_q(t) = \frac{r^c p_0}{c!\,(1-\rho)} e^{-(c\mu-\lambda)t}$$

$$P(T_q > t\ /\ \ T_q > 0) = \ e^{-(c\mu-\lambda)t}$$

Έτσι:

$$W_q = E(T_q) = \int_0^\infty (1 - W_q(t))dt = \frac{r^c}{c!\, c\mu(1-\rho)^2} p_0$$

Για να βρούμε τη συνάρτηση κατανομής του χρόνου αναμονής στο σύστημα, χωρίζουμε την κατάσταση σε δυο ξεχωριστές πιθανότητες, τους πελάτες που δεν περιμένουν στην ουρά (αυτό συμβαίνει με πιθανότητα $W_q(0)$) και αυτούς που έχουν μια θετική αναμονή στην ουρά (αυτό συμβαίνει με πιθανότητα $1\text{-}W_q(0)$). Ο χρόνος στο σύστημα για τους πελάτες της πρώτης περίπτωσης είναι απλά ο χρόνος εξυπηρέτησης, μιας και δεν υπάρχει αναμονή στην ουρά. Γι' αυτούς τους πελάτες η συνάρτηση κατανομής του χρόνου στο σύστημα είναι η εκθετική με μέση τιμή 1/μ.

Για τους πελάτες της δεύτερης περίπτωσης, ο χρόνος στο σύστημα είναι το άθροισμα του χρόνου στην ουρά και του χρόνου εξυπηρέτησης. Επομένως η συνάρτηση κατανομής του χρόνου στο σύστημα είναι η συνέλιξη μιας εκθετικής κατανομής με μέση τιμή 1/(cμ-λ) και μιας εκθετικής κατανομής με μέση τιμή 1/μ.

$$P(T \le t) = \frac{c(1-\rho)}{c(1-\rho)-1}(1 - e^{-\mu t}) - \frac{1}{c(1-\rho)-1}(1 - e^{-(c\mu-\lambda)t})$$

Επομένως η ολική συνάρτηση κατανομής του συστήματος M/M/c ουράς μπορεί να γραφεί ως εξής:

$$\begin{aligned} W(t) = {} & W_q(0)(1 - e^{-\mu t}) \\ & + \left(1 - W_q(0)\right)\left[\frac{c(1-\rho)}{c(1-\rho)-1}(1 - e^{-\mu t}) - \frac{1}{c(1-\rho)-1}\left(1 - e^{-(c\mu-\lambda)t}\right)\right] \\ = {} & \frac{c(1-\rho) - W_q(0)}{c(1-\rho)-1}(1 - e^{-\mu t}) - \frac{1 - W_q(0)}{c(1-\rho)-1}(1 - e^{-(c\mu-\lambda)t}) \end{aligned}$$

2.3.2 Επιλογή του Αριθμού των Μονάδων Εξυπηρέτησης

Κατά τον σχεδιασμό ενός συστήματος ουράς είναι συχνά επιθυμητός ο καθορισμός ενός ικανοποιητικού αριθμού εξυπηρετητών c για το σύστημα. Ένας μεγάλος αριθμός εξυπηρετητών βελτιώνει την ποιότητα της εξυπηρέτησης των πελατών αλλά επιβαρύνει με υψηλότερο κόστος των ιδιοκτήτη του συστήματος. Το πρόβλημα μας είναι να βρούμε των αριθμό των εξυπηρετητών που εξισορροπούν ικανοποιητικά την ποιότητα και το

κόστος της υπηρεσίας. Σ' αυτή την ενότητα θα δώσουμε μια απλή προσέγγιση η οποία είναι ιδιαίτερα χρήσιμη στην επιλογή του αριθμού των εξυπηρετητών μιας M/M/c ουράς.

Παρατηρούμε ότι σε κατάσταση ισορροπίας ο αριθμός των εξυπηρετητών πρέπει να είναι μεγαλύτερος από το προσφερόμενο φορτίο r. Αυτό γράφεται:

c=r+Δ

όπου Δ>0 είναι ο αριθμός των επιπλέον εξυπηρετητών που χρησιμοποιήθηκαν πέραν του προσφερόμενου φορτίου (το Δ μπορεί να χρειαστεί να είναι ένα κλάσμα έτσι ώστε ο αριθμός c να προκύπτει ακέραιος).

Η παρακάτω είναι μια απλή προσεγγιστική σχέση για την επιλογή του c:

$$c \approx r + \beta\sqrt{r} \text{ ή } \Delta \approx \beta\sqrt{r}$$

Όπου β είναι μια σταθερά.

Η παραπάνω σχέση βασίζεται στο παρακάτω θεώρημα:

Θεώρημα Halfin και Whitt 1981

Θεωρούμε μια ακολουθία M/M/c ουρών με δείκτη την παράμετρο n=1,2,…. Υποθέτουμε ότι η ουρά n έχει c_n=n εξυπηρετητές και προσφερόμενο φορτίο r_n. Τότε:

$$\lim_{n\to\infty} C\left(c_{n=n,} r_n\right) = a, \quad 0 < a < 1$$

Αν και μόνο αν:

$$\lim_{n\to\infty} \frac{n - r_n}{\sqrt{n}} = \beta, \quad \beta > 0$$

Όπου C(c,r) είναι η φόρμουλα-C του Erlang και α, β συσχετιζόμενες σταθερές από την:

$$\alpha = \frac{\varphi(\beta)}{\varphi(\beta) + \beta\Phi(\beta)}$$

Διαισθητικά: C(c,r)=1-W_q(0) είναι κατά προσέγγιση σταθερά

$n - r_n \approx \beta\sqrt{n}$ ή $n \approx r_n + \beta\sqrt{n}$

Οι παράμετροι α και β μπορούν να ερμηνευτούν ως σταθερές που αντιπροσωπεύουν την ποιότητα της εξυπηρέτησης: α είναι η πιθανότητα μη μηδενικής καθυστέρησης στην ουρά α=1-W_q(o) και β η συσχετιζόμενη σταθερά που υπολογίζεται από τον τύπο του

θεωρήματος:

$$\alpha = \frac{\varphi(\beta)}{\varphi(\beta) + \beta\Phi(\beta)}$$

2.4 Η ουρά M/M/c/K

Η ουρά M/M/c/K είναι ουρά οι οποία έχει c εξυπηρετητές, αλλά το σύστημα έχει έναν περιορισμό και μπορεί να δεχθεί μέχρι K πελάτες. Η προσέγγιση είναι ίδια με αυτήν της ουράς M/M/c με την διαφορά ότι ο ρυθμός αφίξεων λ_n πρέπει να είναι μηδέν όταν $n \geq K$.

Οι πιθανότητες p_n υπολογίζονται ως εξής:

$$p_n = \begin{cases} \frac{\lambda^n}{n!\,\mu^n} p_0 & 0 \leq n < c \\ \frac{\lambda^n}{c^{n-c} c!\,\mu^n} p_0 & c \leq n \leq K \end{cases}$$

όπως και στην M/M/c ουρά, το p_n, έχει μορφή Poisson για $0 \leq n < c$ και μορφή γεωμετρικής για $c \leq n \leq K$. Ο υπολογισμός του p_0 είναι ίδιος με τη M/M/c ουρά με τη διαφορά ότι επειδή οι σειρές είναι πεπερασμένες δεν υπάρχει απαίτηση το ρ<1. Έτσι:

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{\lambda^n}{n!\,\mu^n} + \sum_{n=c}^{K} \frac{\lambda^n}{c^{n-c} c!\,\mu^n} \right)^{-1}$$

Για r=λ/μ και ρ=r/c έχουμε:

$$\sum_{n=c}^{K} \frac{r^n}{c^{n-c} c!} = \frac{r^c}{c!} \sum_{n=c}^{K} \rho^{n-c} = \begin{cases} \frac{r^c}{c!} \left(\frac{1 - \rho^{K-c+1}}{1-\rho} \right), & (\rho \neq 1) \\ \frac{r^c}{c!} (K - c + 1), & (\rho = 1) \end{cases}$$

Έτσι:

$$p_0 = \begin{cases} [\frac{r^c}{c!} \left(\frac{1 - \rho^{K-c+1}}{1-\rho} \right) + \sum_{n=0}^{c-1} \frac{r^n}{n!}]^{-1} & (\rho \neq 1) \\ [\frac{r^c}{c!} (K - c + 1) + \sum_{n=0}^{c-1} \frac{r^n}{n!}]^{-1}, & (\rho = 1) \end{cases}$$

2.4.1 Μήκος της Ουράς:

Για ρ≠1:

$$L_q = \sum_{n=c+1}^{K}(n-c)p_n = \sum_{n=c+1}^{K}(n-c)\frac{\lambda^n}{c^{n-c}c!\,\mu^n}p_0 = \frac{p_0 r^c \rho}{c!}\sum_{n=c+1}^{K}(n-c)\frac{r^{n-c}}{c^{n-c}}$$

$$= \frac{p_0 r^c \rho}{c!}\sum_{n=c+1}^{K}(n-c)\rho^{n-c-1} = \frac{p_0 r^c \rho}{c!}\sum_{i=1}^{K-c} i\rho^{i-1} = \frac{p_0 r^c \rho}{c!}\frac{d}{d\rho}\left(\sum_{i=0}^{K-c}\rho^i\right)$$

$$= \frac{p_0 r^c \rho}{c!}\frac{d}{d\rho}\left(\frac{1-\rho^{K-c+1}}{1-\rho}\right)$$

$$= \frac{p_0 r^c \rho}{c!\,(1-\rho)^2}[1-\rho^{K-c+1}-(1-\rho)(K-c+1)\rho^{K-c}]$$

2.4.2 Μέσος Αριθμός Πελατών στο σύστημα:

Από την ουρά M/M/c έχουμε ότι:

$$L = L_q + r$$

Στην ουρά M/M/c/K θα χρειαστεί να προσαρμόσουμε τον παραπάνω τύπο, μιας και ένα μέρος p_K των αφίξεων δεν μπαίνουν στο σύστημα, επειδή ο πελάτης έφτασε όταν δεν είχε μείνει χώρος αναμονής. Έστω λ_{eff} ο αποτελεσματικός ρυθμός άφιξης, τότε αυτός είναι $\lambda(1- p_K)$. Έτσι η παραπάνω σχέση γίνεται:

$$L = L_q + \frac{\lambda_{eff}}{\mu} = L_q + \frac{\lambda(1-p_K)}{\mu} = L_q + r(1-p_K)$$

Γνωρίζουμε ότι η ποσότητα $r(1-p_K)$ πρέπει να είναι μικρότερη από c, μιας και ο μέσος αριθμός πελατών που εξυπηρετούνται πρέπει να είναι μικρότερος από τον συνολικό αριθμό των εξυπηρετητών. Γι αυτό τον λόγο θα πρέπει:

$$\rho_{eff} = \frac{\lambda_{eff}}{c\mu} < 1$$

2.4.3 Μέσος Χρόνος Αναμονής Ενός Πελάτη Στο Σύστημα:

Από το θεώρημα του Little: $W = \frac{L}{\lambda_{eff}} = \frac{L}{\lambda(1-p_K)}$

2.4.4 Μέσος Χρόνος Αναμονής Ενός Πελάτη Στην Ουρά:

$$W_q = W - \frac{1}{\mu} = \frac{L_q}{\lambda_{eff}}$$

Τα μέτρα αποδοτικότητας για την πιο απλή περίπτωση μιας M/M/1/K ουράς είναι:

$$p_0 = \begin{cases} \dfrac{1-\rho}{1-\rho^{K+1}}, (\rho \neq 1) \\ \dfrac{1}{K+1}, \qquad (\rho = 1) \end{cases}$$

$$p_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{K+1}}, \quad \rho \neq 1 \\ \dfrac{1}{K+1}, \qquad \rho = 1 \end{cases}$$

$$L_q = \begin{cases} \dfrac{\rho}{1-\rho} - \dfrac{\rho(K\rho^K+1)}{1-\rho^{K+1}}, \qquad \rho \neq 1 \\ \dfrac{K(K-1)}{2(K+1)}, \qquad \rho = 1 \end{cases}$$

$$L = L_q + (1 - p_0)$$

Από την τελευταία σχέση προκύπτει ότι:

$$1 - p_0 = \frac{\lambda(1-p_K)}{\mu} \Rightarrow \mu(1-p_0) = \lambda(1-p_K)$$

Που σημαίνει ότι ο ρυθμός των αποδοτικών εξόδων από το σύστημα είναι ίσος με τον ρυθμό των αποδοτικών εισόδων.

Είναι πλέον απαραίτητο να καθοριστούν οι πιθανότητες $\{q_n\}$ του σημείου άφιξης, από τη στιγμή που οι είσοδοι δεν είναι πλέον Poisson, λόγω της μείωσης του μεγέθους σε K και $q_n \neq p_n$. Για να καθορίσουμε το q_n χρησιμοποιούμε το Θεώρημα του Bayes:

$$q_n = P(n\ στο\ σύστημα|αφιξη\ είναι\ να\ συμβει)$$
$$= \frac{P(αφιξη\ είναι\ να\ συμβεί|n\ στο\ σύστημα)p_n}{\sum_{n=0}^{K} P(αφιξη\ είναι\ να\ συμβεί|n\ στο\ σύστημα)p_n}$$
$$= \lim_{\Delta t\to 0}\left\{\frac{[\lambda\Delta t + o(\Delta t)]p_n}{\sum_{n=0}^{K-1}[\lambda\Delta t + o(\Delta t)]p_n}\right\} = \lim_{\Delta t\to 0}\left\{\frac{\left[\lambda + \frac{o(\Delta t)}{\Delta t}\right]p_n}{\sum_{n=0}^{K-1}\left[\lambda + \frac{o(\Delta t)}{\Delta t}\right]p_n}\right\} = \frac{\lambda p_n}{\lambda\sum_{n=0}^{K-1} p_n}$$
$$= \frac{p_n}{1-p_K}, \quad n \le K-1$$

Από την παραπάνω παρατηρούμε ότι όσο το K τείνει στο άπειρο, το p_Kτείνει στο μηδέν και οδηγούμαστε στην ουρά M/M/c/∞ και στο ότι $q_n = p_n$.

Για να βρούμε τη συνάρτηση κατανομής $W_q(t)$ έχουμε:

$$W_q(t) = P\big(T_q \le t\big)$$
$$= W_q(0)$$
$$+ \sum_{n=c}^{K-1} P(n-c+1\ ολοκληρώσεις\ σε\ χρόνο \le t|η\ άφιξη\ βρίσκει\ n\ στο\ σύστημα)q_n$$

Και από τη στιγμή που δεν μπορούν να γίνουν αφίξεις όταν εξυπηρετούν K πελάτες, έχουμε ότι:

$$W_q(t) = W_q(0) + \sum_{n=c}^{K-1} q_n \int_0^t \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x}dx = W_q(0)$$
$$+ \sum_{n=c}^{K-1} q_n\left(1 - \int_t^{\infty} \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x}dx\right.$$

Και επειδή:

$$\int_t^{\infty} \frac{\lambda(\lambda\chi)^m}{m!} e^{-\lambda\chi}dx = \sum_{i=0}^{m} \frac{(\lambda t)^i e^{-\lambda t}}{i!}$$

Για m=n-c και λ=cμ έχουμε:

$$\int_t^{\infty} \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x}dx = \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!}$$

Και έτσι:

$$W_q(t) = W_q(0) + \sum_{n=c}^{K-1} q_n - \sum_{n=c}^{K-1} q_n \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!} = 1 - \sum_{n=c}^{K-1} q_n \sum_{i=0}^{n-c} \frac{(c\mu t)^i e^{-c\mu t}}{i!}$$

2.5 Η φόρμουλα του Erlang M/M/c/c

Είναι μια ειδική περίπτωση της M/M/c/K ουράς όπου K=c. Αντικαθιστώ στον τύπο της M/M/c/K και έχω:

$$p_n = \frac{\frac{(\frac{\lambda}{\mu})^n}{n!}}{\sum_{i=0}^{c} \frac{(\frac{\lambda}{\mu})^i}{i!}}, \quad 0 \le n \le c$$

Γνωστή ως πρώτη φόρμουλα του Erlang.

Για n=c, η φόρμουλα p_c που προκύπτει καλείται φόρμουλα απώλειας του Erlang ή αλλιώς B-φόρμουλα του Erlang. Αυτή είναι η πιθανότητα ενός γεμάτου στιστήματος κάθε στιγμή σε κατάσταση ισορροπίας.

Από τη στιγμή που οι είσοδοι στο σύστημα της M/M/c/c γίνονται με Poisson, p_c είναι επίσης το μέρος των πελατών που φτάνουν και βρίσκουν γεμάτο το σύστημα και χάνονται από το σύστημα.

$$B(c,r) = p_c = \frac{\frac{r^c}{c!}}{\sum_{i=0}^{c} \frac{r^i}{i!}}, \quad r = \frac{\lambda}{\mu}$$

Όπου το c είναι ανεξάρτητο από το r.

Σχόλιο:

Όταν ο αριθμός c των εξυπηρετητών είναι μεγάλος, ο όρος c! είναι αρκετά μεγάλος και έτσι η B-φόρμουλα του Erlang προκαλεί αριθμητικά προβλήματα σ' έναν υπολογιστή.

Η B-φόρμουλα ικανοποιεί την εξής σχέση:

$$B(c,r) = \frac{rB(c-1,r)}{c + rB(c-1,r)}, \quad c \ge 1$$

Με αρχική κατάσταση B(c=0,r)=1.

Έτσι για τον υπολογισμό του B(c,r) για δεδομένη τιμή του c, ξεκινάει από την B(0,r)=1 και αντικαθιστά στην παραπάνω σχέση μέχρι να φτάσει το c. Αυτή η μέθοδος αποφεύγει τα αριθμητικά προβλήματα.

Η Β-φόρμουλα του Erlang εφαρμόζεται στην M/M/c/c και επίσης χρησιμοποιείται για την εύρεση μέτρων της M/M/c ουράς.

Αν $C(c,r) = 1 - W_q(0)$ η πιθανότητα Erlang-C της αργοπορίας σε μια M/M/c ουρά, τότε το C(c,r) μπορεί να γραφεί:

$$C(c,r) = \frac{cB(c,r)}{c - r + rB(c,r)}$$

Έτσι για παράδειγμα το L_q μπορεί να γραφεί:

$$L_q = C(c,r)\frac{\rho}{1-\rho} = C(c,r)\frac{r}{c-r}$$

2.6 Ουρές Με Απεριόριστη Εξυπηρέτηση M/M/∞

Όταν αναφερόμαστε σε ουρές με απεριόριστη εξυπηρέτηση εννοούμε ουρές με άπειρο αριθμό εξυπηρετητών.

Χρησιμοποιούμε τους τύπους της ανέλιξης γέννησης-θανάτου με $\lambda_n=\lambda$ και $\mu_n=n\mu$ για όλα τα n, τότε:

$$p_n = \frac{r^n}{n!}p_0$$

$$p_0 = (\sum_{n=0}^{\infty}\frac{r^n}{n!})^{-1}$$

Και με αντικατάσταση έχουμε:

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

Που είναι κατανομή Poisson με μέση τιμή r=λ/μ.

Ο αναμενόμενος αριθμός πελατών στο σύστημα είναι η μέση τιμή της Poisson κατανομής και είναι L=r=λ/μ. Από τη στιγμή που έχουμε τόσους περισσότερους εξυπηρετητές σε σχέση με τους πελάτες στο σύστημα $L_q=0=W_q$. Ο μέσος χρόνος αναμονής στο σύστημα, είναι μόνο ο μέσος χρόνος εξυπηρέτησης δηλαδή: W=1/μ και η συνάρτηση κατανομής του χρόνου αναμονής W(t) είναι ίδια με την συνάρτηση κατανομής του χρόνου εξυπηρέτησης, δηλαδή ακολουθούν την εκθετική με μέση τιμή 1/μ.

2.7 Ουρές με Πεπερασμένη Πηγή

Στα προηγούμενα μοντέλα είχαμε θεωρήσει ότι ο πληθυσμός από τον οποίο έρχονται οι αφίξεις είναι άπειρος, από τη στιγμή που ο αριθμός των αφίξεων σε κάθε χρονικό διάστημα είναι μια τυχαία μεταβλητή Poisson. Έστω ότι ο πληθυσμός από τον οποίο προέρχονται οι αφίξεις (καλούμενος πληθυσμός) είναι πεπερασμένος και ίσος με M. Μια τυπική εφαρμογή αυτού του μοντέλου είναι η επιδιόρθωση μηχανών, όπου ο καλούμενος πληθυσμός είναι οι μηχανές, μια άφιξη αντιστοιχεί σε βλάβη της μηχανής, και οι τεχνικοί που τις επιδιορθώνουν είναι οι εξυπηρετητές. Υποθέτουμε ότι c εξυπηρετητές είναι διαθέσιμοι, ότι οι χρόνοι εξυπηρέτησης είναι τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με μέση τιμή 1/μ, και ότι η διαδικασία άφιξης περιγράφεται ως εξής: Αν μια καλούμενη μονάδα δεν είναι στο σύστημα την χρονική στιγμή t, η πιθανότητα να ενταχθεί σε χρόνο t+Δt είναι λΔt+o(Δt), δηλαδή ο χρόνος που περνά μια καλούμενη μονάδα εκτός συστήματος είναι εκθετική κατανομή με μέση τιμή 1/λ.

Με βάση τις παραπάνω υποθέσεις χρησιμοποιούμε την θεωρία της γέννησης-θανάτου: Αν n ο αριθμός των πελατών στο σύστημα, τότε:

$$\lambda_n = \begin{cases} (M-n)\lambda, & 0 \le n < M \\ 0, & n \ge M \end{cases}$$

$$\mu_n = \begin{cases} n\mu, & 0 \le n < c \\ c\mu, & n \ge c \end{cases}$$

Και για r=λ/μ:

$$p_n = \begin{cases} \dfrac{M!/(M-n)!}{n!} r^n p_0, & 1 \le n < c \\ \dfrac{M!/(M-n)!}{c^{n-c}c!} r^n p_0, & c \le n \le M \end{cases}$$

Ή αλλιώς:

$$p_n = \begin{cases} \binom{M}{n} r^n p_0 \,, & 1 \le n < c \\ \binom{M}{n} \frac{n!}{c^{n-c} c!} r^n p_{0,} & c \le n \le M \end{cases}$$

Η αλγεβρική μορφή του p_n δεν επιτρέπει τον εύκολο υπολογισμό του p_0, έτσι αν α_n συντελεστής τέτοιος ώστε: $p_n = a_n p_0$, τότε:

$$p_0 = \frac{1}{1 + a_1 + a_2 + \cdots + a_M}$$

Για να βρούμε τον μέσο αριθμό πελατών στο σύστημα (στην περίπτωση των μηχανών, ενδιαφερόμαστε για τις μηχανές που χαλάνε και πανε για επιδιόρθωση) έχουμε:

$$L = \sum_{n=1}^{M} n p_n = p_0 \sum_{n=1}^{M} n a_n$$

Για να βρούμε τα $L_{q,}$ W, και W_q πρέπει πρώτα να βρούμε τον αποτελεσματικό μέσο ρυθμό αφίξεων στο σύστημα. Ο μέσος ρυθμός αφίξεων όταν το σύστημα είναι σε κατάσταση n είναι (M-n)λ. Ο συνολικός ρυθμός αφίξεων είναι:

$$\lambda_{eff} = \sum_{n=0}^{M} (M - n) \lambda p_n = \lambda (M - L)$$

Η παραπάνω προκύπτει διαισθητικά από τη στιγμή που κατά μέσο όρο L είναι στο σύστημα, συνεπώς M-L είναι έξω από αυτό, και καθένα έχει μέσο ρυθμό άφιξης λ. Από τη φόρμουλα του Little έχουμε:

$$L_q = L - \frac{\lambda_{eff}}{\mu} = L - r(M - l)$$

$$W = \frac{L}{\lambda (M - L)}$$

$$W_q = \frac{L_q}{\lambda(M - L)}$$

Σχόλιο:

Στην περίπτωση που έχουμε έναν εξυπηρετητή, τότε:

$$p_n = \binom{M}{n} n!\, r^n p_0, \qquad 0 \le n \le M$$

2.8 Μοντέλα με Ανταλλακτικά

Το μοντέλο που περιγράψαμε με πεπερασμένη πηγή μπορεί να γενικευτεί με τη χρήση ανταλλακτικών. Υποθέτουμε ότι υπάρχουν Μ μηχανές σε λειτουργία και επιπλέον Υ ανταλλακτικές. Όταν μια μηχανή σε λειτουργία χαλάει, μια ανταλλακτική αμέσως την αντικαθιστά (αν είναι διαθέσιμη). Αν δεν υπάρχει διαθέσιμη ανταλλακτική όταν χαλάσει η μηχανή τότε το σύστημα γίνεται λειψό (short). Όταν η μηχανή διορθωθεί, γίνεται ανταλλακτική εκτός αν το σύστημα είναι λειψό, οπότε σ' αυτή την περίπτωση η επιδιορθωμένη μηχανή μπαίνει αμέσως στην εξυπηρέτηση. Σε κάθε χρονική στιγμή υπάρχουν το πολύ Μ μηχανές σε λειτουργία, έτσι ο ρυθμός των αποτυχιών(δηλαδή να χαλάσουν) είναι το πολύ Μλ (οι ανταλλακτικές που δεν είναι σε λειτουργία δεν συμβάλουν στον ρυθμό αποτυχίας). Γι' αυτό το μοντέλο το λ_n είναι:

$$\lambda_n = \begin{cases} M\lambda, & 0 \le n < Y \\ (M - n + Y)\lambda, & Y \le n < Y + M \\ 0, & n \ge Y + M \end{cases}$$

Όπου n ο αριθμός των μηχανών που αποτυγχάνουν. Έτσι για c τεχνικούς έχουμε:

$$\mu_n = \begin{cases} n\mu, & 0 \le n < c \\ c\mu & n \ge c \end{cases}$$

Αν $c \leq Y$ και $r = \frac{\lambda}{\mu}$, τότε:

$$p_n = \begin{cases} \frac{M^n}{n!} r^n p_0, & 0 \leq n < c \\ \frac{M^n}{c^{n-c} c!} r^n p_0, & c \leq n < Y \\ \frac{M^Y M!}{(M-n+Y) c^{n-c} c!} r^n p_0, & Y \leq n \leq Y + M \end{cases}$$

Αν το Υ είναι πάρα πολύ μεγάλο αντιλαμβανόμαστε ότι έχουμε έναν άπειρο καλούμενο πληθυσμό με ρυθμό Μλ. Έχουμε δηλαδή τη σχέση της M/M/c/∞ ουράς με Μλ αντί για λ.

Αν c>Y έχουμε:

$$p_n = \begin{cases} \frac{M^n}{n!} r^n p_0, & 0 \leq n \leq Y \\ \frac{M^Y M!}{(M-n+Y)!\, n!} r^n p_0, & Y+1 \leq n < c \\ \frac{M^Y M!}{(M-n+Y)!\, c^{n-c} c!} r^n p_0, & c \leq n \leq Y + M \end{cases}$$

Αν Υ=0 τότε προφανώς οδηγούμαστε στην απλή περίπτωση πεπερασμένης πηγής χωρίς ανταλλακτικά.

Για Μ,Υ πολύ μεγάλα η παραπάνω σχέση είναι σχετικά μπερδεμένη αν θέλουμε να βρούμε τους συντελεστές $\{\alpha_n\}$. Ευτυχώς αυτό μπορούμε να το αποφύγουμε κάνοντας χρήση του αναδρομικού τύπου που συνδέει το p_n με το p_{n+1}:

$$p_{n+1} = (\frac{\lambda_n}{\mu_{n+1}}) p_n$$

Έτσι για την περίπτωση του προβλήματος χωρίς ανταλλακτικά έχουμε:

$$p_{n+1} = \begin{cases} \frac{M-n}{n+1} r p_n, & 0 \leq n \leq c-1 \\ \frac{M-n}{c} r p_n, & c \leq n \leq M-1 \end{cases}$$

Ο υπολογισμός του λ_{eff} γίνεται ως εξής:

$$\lambda_{eff} = \sum_{n=0}^{Y-1} M\lambda p_n + \sum_{n=Y}^{Y+M} (M-n+Y)\lambda p_n = \lambda(M - \sum_{n=Y}^{Y+M} (n-Y)p_n)$$

Στη συνέχεια θα βρούμε την κατανομή του χρόνου αναμονής. Η συνήθης διαδικασία μέχρι στιγμής ήταν να βρούμε το χρόνο αναμονής ενός πελάτη που φτάνει δεδομένου ότι υπάρχουν n στο σύστημα τη στιγμή της άφιξης και στη συνέχεια λαμβάναμε υπόψη την κατανομή των $\{q_n\}$ όπου $\{q_n\}$ είναι οι πιθανότητες όταν συμβαίνει μια άφιξη. Έχουμε ότι $q_n \neq p_n$ και για την περίπτωση του προβλήματος χωρίς ανταλλακτικά έχουμε:

$$q_n = \frac{(M-n)p_n}{k}$$

Όπου κ μια σταθερά, η οποία καθορίζεται αν πάρουμε το άθροισμα των $\{q_n\}$ να είναι ίσο με 1. Χρησιμοποιούμε το θεώρημα του Bayes και έχουμε:

$$\begin{aligned} q_n &= P(n \text{ στο σύστημα} | \text{μια άφιξη πρόκειται να συμβεί}) \\ &= \frac{P(n \text{ στο σύστημα}) P(\text{μια άφιξη είναι να συμβεί} | n \text{ στο σύστημα})}{P(\text{μια άφιξη είναι να συμβεί})} \\ &= \frac{P(n \text{ στο σύστημα}) P(\text{μια άφιξη είναι να συμβεί} | n \text{ στο σύστημα})}{\sum_n (P(n \text{ στο σύστημα}) P(\text{μια άφιξη είναι να συμβεί} | n \text{ στο σύστημα}))} \\ &= \lim_{\Delta t \to 0} \frac{p_n[(M-n)\lambda\Delta t + o(\Delta t)]}{\sum_n p_n[(M-n)\lambda\Delta t + o(\Delta t)} = \frac{(M-n)p_n}{M-L} \end{aligned}$$

Στην περίπτωση που έχουμε Μ μηχανές και δεν έχουμε ανταλλακτικά, η πιθανότητα του σημείου άφιξης $q_n(M)$ είναι ίση με $p_n(M-1)$ την πιθανότητα γενικού χρόνου με Μ-1 μηχανές.

Στην περίπτωση που υπάρχουν ανταλλακτικά το $q_n(M)$ είναι:

$$q_n = \begin{cases} \dfrac{Mp_n}{M - \sum_{i=Y}^{Y+M} (i-Y)p_i}, & 0 \leq n \leq Y-1 \\ \dfrac{(M-n+Y)p_n}{M - \sum_{i=Y}^{Y+M} (i-Y)p_i}, & Y \leq n \leq Y+M-1 \end{cases}$$

Η κατανομή των χρόνων αναμονής είναι όπως και στη περίπτωση της M/M/c/K ουράς είναι Poisson:

$$W_q(t) = P(T_q \le t)$$
$$= W_q(0)$$
$$+ \sum_{n=c}^{Y+M-1} [P(n-c+1\ ολοκληρώσεις\ σε\ χρόνο \le t | η\ άφιξη\ βρίσκει\ n\ στο\ σύστημα) q_n]$$
$$= W_q(0) + \sum_{n=c}^{Y+M-1} q_n \int_0^t \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx$$
$$= W_q(0) + \sum_{n=c}^{Y+M-1} q_n [1 - \int_t^{\infty} \frac{c\mu(c\mu x)^{n-c}}{(n-c)!} e^{-c\mu x} dx] = 1 - \sum_{n=c}^{Y+M-1} q_n \sum_{i=0}^{n-c} \frac{(c\mu t)^i}{i!} e^{-c\mu t}$$

2.9 Εξυπηρέτηση Εξαρτώμενη από τον Αριθμό των Πελατών

Σ' αυτή την ενότητα θα εξετάσουμε ουρές στις οποίες η εξυπηρέτηση εξαρτάται από την κατάσταση του συστήματος, δηλαδή τον αριθμό των πελατών, αυτό συμβαίνει όταν ο μέσος ρυθμός εξυπηρέτησης εξαρτάται από τον αριθμό των πελατών στο σύστημα. Σε πραγματικές καταστάσεις ένας ή περισσότεροι εξυπηρετητές μπορούν να αυξήσουν τον ρυθμό εξυπηρέτησης και να εξυπηρετούν γρηγορότερα όταν δουν να σχηματίζεται μεγάλη ουρά. Αντίθετα μπορεί ένας εξυπηρετητής να μην έχει μεγάλη εμπειρία και να αγχωθεί με αποτέλεσμα ο μέσος ρυθμός εξυπηρέτησης να μειώνεται όσο προκαλείται συμφόρηση στο σύστημα.

Το πρώτο μοντέλο με το οποίο θα ασχοληθούμε είναι όταν ένας εξυπηρετητής έχει δυο μέσους ρυθμούς , γρήγορο και αργό. Η εργασία εκτελείται με αργό ρυθμό μέχρι να υπάρχουν k πελάτες στο σύστημα, όπου σ' αυτό το σημείο γίνεται η αλλαγή σε γρήγορο ρυθμό. Υποθέτουμε ότι οι χρόνοι εξυπηρέτησης είναι Μαρκοβιανοί, αλλά ο μέσος ρυθμός μ_n σαφώς πλέον εξαρτάται από τον αριθμό των πελατών n. Επιπλέον δεν υπάρχει περιορισμός στον αριθμό πελατών στο σύστημα. Τότε:

$$\mu_n = \begin{cases} \mu_1, & 1 \le n < k \\ \mu, & n \ge k \end{cases}$$

Υποθέτουμε ότι η διαδικασία άφιξης είναι Poisson παραμέτρου λ και έχουμε:

$$p_n = \begin{cases} {\rho_1}^n p_0, & 0 \le n < k \\ {\rho_1}^{k-1} \rho^{n-k+1} p_0, & n \ge k \end{cases}$$

Όπου: $\rho_1=\lambda/\mu_1$ και $\rho=\lambda/\mu<1$. Και επειδή το άθροισμα των πιθανοτήτων πρέπει να είναι 1 έχουμε:

$$p_0 = (\sum_{n=0}^{k-1} {\rho_1}^n + \sum_{n=k}^{\infty} {\rho_1}^{k-1}\rho^{n-k+1})^{-1}$$

Και έτσι:

$$p_0 = \begin{cases} (\frac{1-{\rho_1}^k}{1-\rho_1} + \frac{\rho{\rho_1}^{k-1}}{1-\rho})^{-1}, & \rho_1 \neq 1, \rho < 1 \\ (k + \frac{\rho}{1-\rho})^{-1}, & \rho_1 = 1, \rho < 1 \end{cases}$$

Αν $\mu_1=\mu$ τότε οδηγούμαστε στις εξισώσεις της M/M/1 ουράς.

Για να βρούμε τον αναμενόμενο αριθμό πελατών στο σύστημα, υποθέτουμε ότι $\rho_1\neq1$ και έχουμε:

$$\begin{aligned} L = \sum_{n=0}^{\infty} np_n &= p_0\left(\sum_{n=0}^{k-1} n{\rho_1}^n + \sum_{n=k}^{\infty} n{\rho_1}^{k-1}\rho^{n-k+1}\right) \\ &= p_0\left[\rho_1\sum_{n=0}^{k-1} n{\rho_1}^{n-1} + \rho_1\left(\frac{\rho_1}{\rho}\right)^{k-2}\sum_{n=k}^{\infty} n\rho^{n-1}\right] \\ &= p_0\left[\rho_1\frac{d}{d\rho_1}\sum_{n=0}^{k-1}{\rho_1}^n + \rho_1\left(\frac{\rho_1}{\rho}\right)^{k-2}\frac{d}{d\rho}\sum_{n=k}^{\infty}\rho^n\right] \\ &= p_0[\rho_1\frac{d}{d\rho_1}\left(\frac{1-{\rho_1}^k}{1-\rho_1}\right) + \rho_1\left(\frac{\rho_1}{\rho}\right)^{k-2}\frac{d}{d\rho}\left(\frac{1}{1-\rho} - \frac{1-\rho^k}{1-\rho}\right)] \end{aligned}$$

Έτσι τελικώς έχουμε:

$$L = p_0(\frac{\rho_1[1+(k-1){\rho_1}^k - k{\rho_1}^{k-1}]}{(1-\rho_1)^2} + \frac{\rho{\rho_1}^{k-1}[k-(k-1)\rho]}{(1-\rho)^2})$$

Από αυτήν μπορούμε να βρούμε το L_q από την:

$$L_q = L - (1-p_0)$$

Και από το Θεώρημα του Little βρίσκουμε τα W, W_q:

$$W = \frac{L}{\lambda} \qquad \kappa\alpha\iota \qquad W_q = \frac{L_q}{\lambda}$$

Παρατηρούμε ότι η σχέση :$W = W_q + \frac{1}{\mu}$ δεν μπορεί να χρησιμοποιηθεί από τη στιγμή που το μ δεν είναι σταθερά και εξαρτάται από το σημείο αλλαγής της κατάστασης του συστήματος k.

Παρόλα αυτά έχουμε ότι:

$$W = W_q + \frac{1 - p_0}{\lambda}$$

Που σημαίνει ότι ο αναμενόμενος χρόνος εξυπηρέτησης είναι: $\frac{1-p_0}{\lambda}$

2.10 Ουρές Με Ανυπόμονους Πελάτες

Ο σκοπός αυτής της ενότητας είναι να αναλύσουμε τα φαινόμενα που δημιουργεί η ανυπομονησία των πελατών σε μια ουρά τύπου M/M/c. Ανυπόμονους λέμε τους πελάτες οι οποίοι έχουν την τάση να μπαίνουν στην ουρά όταν αναμένεται ελάχιστη αναμονή και έχουν την τάση να μένουν στην ουρά όταν η αναμονή έχει γίνει επαρκώς μικρή. Η ανυπομονησία ως αποτέλεσμα της υπερβολικής αναμονής είναι τόσο σημαντική στην συνολική διαδικασία της ουράς όσο οι αφίξεις και οι αναχωρήσεις. Όταν η ανυπομονησία γίνει επαρκώς έντονη και οι πελάτες αποχωρούν πριν να εξυπηρετηθούν ο διευθυντής της επιχείρησης πρέπει να λάβει μέτρα ώστε να μειώσει το πρόβλημα της συμφόρησης σε επίπεδα που οι πελάτες να μπορούν να ανεχθούν. Τα μοντέλα που θα αναπτυχθούν στην συνέχεια βρίσκουν πρακτική εφαρμογή σ' αυτήν την απόπειρα του διευθυντή να παρέχει επαρκή εξυπηρέτηση για τους ανυπόμονους πελάτες.

Η ανυπομονησία έχει τρεις μορφές, η πρώτη μορφή είναι το balking δηλαδή η απροθυμία ενός πελάτη να μπει στην ουρά κατά την άφιξη, η δεύτερη μορφή είναι η υπαναχώρηση δηλαδή η απροθυμία του πελάτη να παραμείνει στην ουρά αφού έφτασε και περίμενε και η τρίτη μορφή είναι να κάνει ελιγμούς μεταξύ ουρών, όταν κάθε παράλληλος εξυπηρετητής έχει τη δική του ουρά.

2.10.1 M/M/1 Balking

Σε πρακτική εφαρμογή συχνά συμβαίνει οι πελάτες που φτάνουν να αποκαρδιώνονται όταν η ουρά είναι μεγάλη και δεν επιθυμούν να αναμένουν. Ένα τέτοιο μοντέλο είναι το M/M/c/K στο οποίο αν οι πελάτες βλέπουν K μπροστά από αυτούς στο σύστημα δεν εισέρχονται σε αυτό. Γενικά εκτός αν το K είναι το αποτέλεσμα φυσικού περιορισμού όπως ο χώρος αναμονής όπου οι άνθρωποι δεν αντιδρούν έτσι οικιοθελώς. Σπάνια πάντως έχουν όλοι οι πελάτες το ίδιο όριο δυσανασχέτησης.

Μια άλλη προσέγγιση για το balking είναι η χρήση μονοτονικά φθινουσών συναρτήσεων του αριθμού των πελατών πολλαπλασιασμένων με τον μέσο ρυθμό λ.

Έστω b_n μια τέτοια συνάρτηση όπου: $\lambda_n=b_n\lambda$ και

$0 \leq b_{n+1} \leq b_n \leq 1, \quad n > 0\,, b_0 \equiv 1$

Αν πάρουμε τον τύπο μιας ανέλιξης γέννησης-θανάτου για το p_n με c=1, έχουμε:

$$p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} = p_0 (\frac{\lambda}{\mu})^n \prod_{i=1}^{n} b_{i-1}$$

Οι πελάτες δεν κουράζονται μόνο από το μέγεθος της ουράς, αλλά μπορεί να προσπαθήσουν να εκτιμήσουν πόση ώρα θα χρειαστεί να περιμένουν.

Αν η ουρά κινείται γρήγορα, τότε ένας πελάτης μπορεί να μπει σε μια μεγάλη ουρά. Αντίθετα αν η ουρά κινείται αργά ένας πελάτης μπορεί να χάσει την υπομονή του ακόμα και αν η ουρά είναι μικρή.

Αν n πελάτες είναι στο σύστημα μια εκτίμηση για τον μέσο χρόνο αναμονής είναι n/μ.

Μια εύλογη συνάρτηση είναι η $b_n = e^{-\frac{an}{\mu}}$.

Η ουρά M/M/1/K είναι μια ειδική περίπτωση όπου $b_i=1$ για $0 \leq i \leq K-1$ αλλιώς είναι μηδέν.

2.10.2 M/M/1 Reneging

Οι πελάτες οι οποίοι έχουν την τάση να είναι ανυπόμονοι δεν χάνουν πάντα την υπομονή τους από το υπερβολικά μεγάλο μέγεθος της ουράς, αντίθετα μπορεί να μπουν στην ουρά για να δουν πόσο μπορεί να διαρκέσει η αναμονή, διατηρώντας το προνόμιο να μπορούν να υπαναχωρήσουν αν η εκτίμησή τους είναι ότι ο συνολικός χρόνος αναμονή είναι ανυπόφορος.

Θεωρούμε ένα μοντέλο γέννησης-θανάτου με έναν εξυπηρετητή και μια συνάρτηση r(n) που ορίζεται ως εξής:

$$r_n = \lim_{\Delta t \to \infty} \frac{P(\mu\iota\alpha\ \mu o\nu\acute{\alpha}\delta\alpha\ \upsilon\pi\alpha\nu\alpha\chi\omega\rho\epsilon\acute{\iota}\ \kappa\alpha\tau\acute{\alpha}\ \tau\eta\ \delta\iota\acute{\alpha}\rho\kappa\epsilon\iota\alpha\ \tau o\upsilon\ \Delta t\ \acute{o}\tau\alpha\nu\ \upsilon\pi\acute{\alpha}\rho\chi o\upsilon\nu\ n\ \pi\epsilon\lambda\acute{\alpha}\tau\epsilon\varsigma\ )}{\Delta t}$$

$$r(0) = r(1) \equiv 0$$

Αυτή εξακολουθεί να είναι διαδικασία γέννησης-θανάτου με $\mu_n = \mu + r(n)$. Έτσι έχουμε:

$$p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i} = p_0 \lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{\mu + r(i)}, \quad n \geq 1$$

Όπου:

$$p_0 = (1 + \sum_{n=1}^{\infty} \lambda^n \prod_{i=1}^{n} \frac{b_{i-1}}{\mu + r(i)})^{-1}$$

Μια καλή περίπτωση για την συνάρτηση υπαναχώρησης r(n) είναι η $e^{an/\mu},\ n \geq 2$.

Ένας πελάτης που περιμένει θα μπορούσε να εκτιμήσει τον μέσο όρο αναμονής στο σύστημα ως n/μ αν n-1 πελάτες ήταν μπροστά από αυτόν, και η πιθανότητα υπαναχώρησης θα μπορούσε να εκτιμηθεί από συνάρτηση της μορφής $e^{an/\mu}$.

ΚΕΦΑΛΑΙΟ 3: ΟΥΡΕΣ ΜΕ ΓΕΝΙΚΗ ΚΑΤΑΝΟΜΗ

3.1 Η Ουρά M/G/1

Σ' αυτή την ενότητα θα ασχοληθούμε με μια ουρά ενός εξυπηρετητή με αφίξεις σύμφωνα με μια διαδικασία Poisson και οι χρόνοι εξυπηρέτησης ακολουθούν μια γενική κατανομή. Ως συνήθως υποθέτουμε ότι η πειθαρχεία της ουράς είναι η FCFS και όλοι οι χρόνοι εξυπηρέτησης και οι ενδιάμεσοι χρόνοι είναι ανεξάρτητοι. Έστω λ ο ρυθμός άφιξης και S μια τυχαία μεταβλητή με γενική κατανομή. Έστω μ=1/E(S) ο ρυθμός εξυπηρέτησης και θεωρούμε ότι ρ=λ/μ<1 (κατάσταση ισορροπίας).

Pollaczek-Khintchine Formula

Θα εξετάσουμε δύο τρόπους από τους οποίους προκύπτουν οι αναμενόμενες τιμές των μέτρων αποδοτικότητας για την M/G/1. Ο πρώτος τρόπος λαμβάνει αποτελέσματα εξετάζοντας το σύστημα με βάση χρόνους άφιξης στο σύστημα, ενώ ο δεύτερος τρόπος λαμβάνει αποτελέσματα εξετάζοντας το σύστημα με βάση τους χρόνους αποχώρησης από αυτό.

3.1.1 Με βάση τους Χρόνους Άφιξης

Έστω ένας πελάτης ο οποίος φτάνει στο σύστημα της ουράς, η καθυστέρησή του καθορίζεται από τους πελάτες που είναι ήδη στο σύστημα όταν αυτός φτάνει. Ειδικότερα μπορεί να υπάρχουν πελάτες στην ουρά και μπορεί να υπάρχει ήδη ένας πελάτης στην εξυπηρέτηση.

Αρχικά εξετάζουμε τους πελάτες που είναι στην ουρά τη στιγμή της άφιξης. Κάθε πελάτης που είναι μπροστά από τον πελάτη μας, συμβάλει κατά μέσο όρο E(S) στην καθυστέρησή του. Υπάρχουν κατά μέσο όρο L_q πελάτες στην ουρά όταν αυτός φτάνει, έτσι η μέση αργοπορία του πελάτη μας με βάση αυτούς τους πελάτες είναι $L_qE(S)$.

Τώρα ο πελάτης που είναι στην εξυπηρέτηση (αν υπάρχει τέτοιος πελάτης), όταν ο πελάτης μας φτάνει, συμβάλει ένα άλλο ποσό στην αργοπορία του πελάτη μας. Αυτός ο πελάτης έχει ολοκληρώσει ένα μέρος της εξυπηρέτησής του έτσι η συμβολή του στην αργοπορία είναι ο υπολειπόμενος χρόνος εξυπηρέτησής του όχι ο συνολικός χρόνος εξυπηρέτησής του.

Επομένως ο μέσος χρόνος αναμονής στην ουρά του πελάτη μας είναι:

$$W_q = L_q E(S) + P(απασχολημ. εξυπηρετητής) E(υπολειπόμ. χρόνος\ εξυπηρέτησης | εξυπηρετητής\ απασχολημενος)$$

Και επειδή: $L_q = \lambda W_q$

Έχουμε:

$$W_q = \frac{P(απασχολημ. εξυπηρετητής) E(υπολειπόμ. χρόνος\ εξυπηρέτησης | εξυπηρετητής\ απασχολημένος)}{1-\rho}$$

Το P(απασχολημένος εξυπηρετητής) είναι η πιθανότητα ένας πελάτης που φτάνει να βρίσκει τον εξυπηρετητή απασχολημένο, που είναι το ίδιο με το μέρος του χρόνου που ο εξυπηρετητής είναι απασχολημένος, έτσι P(απασχολημένος εξυπηρετητής)=ρ.

Η αναμενόμενη τιμή του υπολειπόμενου χρόνου εξυπηρέτησης, δεδομένου ότι κατά την άφιξη ο εξυπηρετητής είναι απασχολημένος είναι:

$$E(υπολειπόμενος\ χρόνος\ εξυπηρέτησης | εξυπηρετητής\ απασχολημένος) = \frac{E(S^2)}{2E(S)} = \frac{1+{C^2}_B}{2} E(S)$$

Όπου:

$${C^2}_B = \frac{Var(S)}{E^2(S)}$$

Παρατηρούμε ότι: $\frac{1+{C^2}_B}{2} E(S) > \frac{E(S)}{2}$, αυτό σημαίνει ότι ο αναμενόμενος χρόνος εξυπηρέτησης όπως τον αντιλαμβάνεται ένας πελάτης που φτάνει σ' έναν απασχολημένο εξυπηρετητή είναι περισσότερος από τον μισό αναμενόμενο χρόνο εξυπηρέτησης ενός πελάτη. Η ισότητα στην προηγούμενη σχέση επιτυγχάνεται μόνο όταν ${C^2}_B = 0$. Ο λόγος που ο αναμενόμενος υπολειπόμενος χρόνος είναι περισσότερο από αυτό θα έπρεπε να είναι διαισθητικά αναμενόμενο είναι ότι οι πελάτες είναι

περισσότερο σύνηθες να φτάνουν κατά τη διάρκεια μεγάλων διαστημάτων εξυπηρέτησης και αυτό αυξάνει το αναμενόμενο κατά E(S)/2.

Αντικαθιστούμε στο τύπο του W_q και έχουμε:

$$W_q = \frac{1 + {C^2}_B}{2} \frac{\rho}{1-\rho} E(S)$$

Ο παραπάνω τύπος έχει 3 όρους:

Ο πρώτος όρος $\frac{1+{C^2}_B}{2}$ περιλαμβάνει τον συντελεστή της διασποράς υψωμένο στο τετράγωνο ${C^2}_B$ της κατανομής της εξυπηρέτησης. Όταν η κατανομή εξυπηρέτησης είναι η εκθετική τότε ${C^2}_B = 1$ και έτσι $\frac{1+{C^2}_B}{2} = 1$, σ' αυτή την περίπτωση οδηγούμαστε στον ανάλογο τύπο της M/M/1 ουράς. Έτσι:

$$W_q = \frac{1 + {C^2}_B}{2} \{W_q \ για \ ανάλογη \ M/M/1 \ \ ουρά\}$$

Όσο η μεταβλητότητα της κατανομής εξυπηρέτησης αυξάνεται, αυξάνεται και η αναμενόμενη αναμονή στην ουρά, έτσι για μεγάλο ${C^2}_B$, το W_q είναι κατά προσέγγιση γραμμικό στο ${C^2}_B$.

Ο δεύτερος όρος $\frac{\rho}{1-\rho}$ τείνει στο άπειρο όσο το ρ τείνει στο 1.

Ο τρίτος όρος E(S) περιλαμβάνει μονάδες χρόνου και μπορεί να θεωρηθεί ως παράγοντας χρονικής κλίμακας.

Έτσι το W_q είναι προϊόν δύο χρονικών ποσοτήτων που είναι ανεξάρτητες από την επιλεγμένη χρονική κλίμακα και την χρόνο-εξαρτώμενη ποσότητα E(S).

Για τον υπολογισμό του W_q χρειαζόμαστε μόνο 3 παραμέτρους: τον ρυθμό άφιξης λ, τη μέση τιμή E(S)=1/μ της κατανομής της εξυπηρέτησης, και το τετράγωνο του συντελεστή διακύμανσης ${C^2}_B$ της κατανομής εξυπηρέτησης. Χρήσιμοι είναι οι εξής τύποι:

$$ {C^2}_B = \frac{Var(S)}{E^2(S)} \qquad και \qquad Var(S) = E(S^2) - E^2(S)$$

Σ’ ένα πραγματικό σύστημα η πληροφορία που αφορά το μηχανισμό εξυπηρέτησης είναι εύκολα διαθέσιμη, έτσι αυτές οι παράμετροι μπορούν εύκολα να εκτιμηθούν.

Τα υπόλοιπα μέτρα αποδοτικότητας εκφράζονται με τρεις διαφορετικές μορφές, στην πρώτη στήλη εκφράζονται με τη χρήση του $C^2{}_B$, στη δεύτερη στήλη με τη χρήση του $E(S^2)$ και στη Τρίτη στήλη με τη χρήση της διασποράς της κατανομής εξυπηρέτησης $\sigma^2{}_B$.

Μέτρα Αποδοτικότητας της M/G/1 Ουράς

$$L_q = \frac{1 + C^2{}_B.}{2}\frac{\rho^2}{1-\rho} \qquad = \frac{\lambda^2 E(S^2)}{2(1-\rho)} \qquad = \frac{\rho^2 + \lambda^2\sigma^2{}_B}{2(1-\rho)}$$

$$W_q = \frac{1 + C^2{}_B.}{2}\frac{\rho}{\mu-\lambda} \qquad = \frac{\lambda E(S^2)}{2(1-\rho)} \qquad = \frac{\frac{\rho^2}{\lambda} + \lambda\sigma^2{}_B}{2(1-\rho)}$$

$$W = \frac{1 + C^2{}_B.}{2}\frac{\rho}{\mu-\lambda} + \frac{1}{\mu} \qquad = \frac{\lambda E(S^2)}{2(1-\rho)} + \frac{1}{\mu} \qquad = \frac{\frac{\rho^2}{\lambda} + \lambda\sigma^2{}_B}{2(1-\rho)} + \frac{1}{\mu}$$

$$L = \frac{1 + C^2{}_B.}{2}\frac{\rho^2}{1-\rho} + \rho \qquad = \frac{\lambda^2 E(S^2)}{2(1-\rho)} + \rho \qquad = \frac{\rho^2 + \lambda^2\sigma^2{}_B}{2(1-\rho)} + \rho$$

3.1.2 Με βάση τους Χρόνους Αναχώρησης

Σ’ αυτή την ενότητα θα μελετήσουμε την ουρά, αμέσως μετά την αποχώρηση του πελάτη από το σύστημα. Έστω X_n ο αριθμός των πελατών που μένουν στο σύστημα αμέσως μόλις ο n-οστός πελάτης αποχωρήσει (έτσι ο πελάτης που αποχωρεί δεν υπολογίζεται στο μέτρημα) και έστω A_n ο αριθμός των πελατών που φτάνουν κατά τη διάρκεια της εξυπηρέτησης του n-οστού πελάτη, τότε για όλα τα $n \geq 1$ έχουμε:

$$X_{n+1} = \begin{cases} X_{n-1} + A_{n+1}, & X_n \geq 1 \\ A_{n+1}, & X_n = 0 \end{cases}$$

Ή αλλιώς:

$$X_{n+1} = X_n - U(X_n) + A_{n+1}$$

Όπου U συνάρτηση:

$$U(X_n) = \begin{cases} 1, & X_n > 0 \\ 0, & X_n = 0 \end{cases}$$

Υποθέτουμε ότι το n είναι αρκετά μεγάλο έτσι ώστε το σύστημα είναι σε κατάσταση ισορροπίας. Έχουμε:

$$E(X_{n+1}) = E(X_n) = L^D$$

Το L^D δηλώνει την αναμενόμενη κατάσταση του συστήματος ισορροπίας στα σημεία αναχώρησης (αντίθετα το L δηλώνει την αναμενόμενη κατάσταση του συστήματος σ' αυθαίρετα χρονικά σημεία).

Έτσι έχουμε:

$$L^{(D)} = L^{(D)} - E[U(X_n)] + E[A_{n+1}]$$

και αυτό συνεπάγεται:

$$E[U(X_n)] = E[A_{n+1}]$$

Αν S η τυχαία μεταβλητή του χρόνου εξυπηρέτησης του (n+1) πελάτη, τότε:

$$E[A_{n+1}] = \int_0^\infty E[A_{n+1}|S = t]dB(t) = \int_0^\infty \lambda t dB(t) = \lambda E[S] = \frac{\lambda}{\mu} = \rho$$

Η δεύτερη ισότητα προκύπτει από το γεγονός ότι $\{A_{n+1}|S = t\}$ είναι τυχαία μεταβλητή που ακολουθεί την Poisson με μέση τιμή λt. Έτσι $E[U(X_n)] = E[A_{n+1}] = \rho$

Στη συνέχεια υψώνουμε στο τετράγωνο την:

$$X_{n+1} = X_n - U(X_n) + A_{n+1}$$

Και έχουμε:

$$X^2{}_{n+1} = X^2{}_n + U^2(X_n) + A^2{}_{n+1} - 2X_n U(X_n) - 2A_{n+1}U(X_n) - 2A_{n+1}X_n$$

Παίρνοντας μέσες τιμές και λαμβάνοντας υπόψη ότι: $E[X^2{}_{n+1}] = E[X^2{}_n]$ έχουμε:

$$0 = E[U^2(X_n)] + E[A^2{}_{n+1}] - 2E[X_n U(X_n)] - 2E[A_{n+1}U(X_n)] + 2E[A_{n+1}X_n]$$

Έχουμε ότι $U^2(X_n) = U(X_n)$ και $X_n U(X_n) = X_n$

Επίσης A_{n+1} είναι ανεξάρτητο από τα X_n και $U(X_n)$, και αυτό διότι ο αριθμός των πελατών που φτάνουν κατά τη διάρκεια της εξυπηρέτησης του (n+1) πελάτη δηλαδή το A_{n+1}, είναι ανεξάρτητος ενός νωρίτερου συμβάντος, δηλαδή ο αριθμός των πελατών παραμένει X_n αφού φύγει και ο n-οστός πελάτης. Από αυτό:

$$0 = E[U(X_n)] + E[A^2{}_{n+1}] - 2E[X_n] - 2E[A_{n+1}]E[U(X_n)] + 2E[A_{n+1}]E[X_n]$$

Και έχουμε δείξει ότι: $E[U(X_n)] = E[A_{n+1}] = \rho$, αντικαθιστούμε και έχουμε:

$$0 = \rho + E[A^2{}_{n+1}] - 2L^{(D)} - 2\rho^2 + 2\rho L^{(D)}$$

Ή αλλιώς:

$$L^{(D)} = \frac{\rho - 2\rho^2 + E[A^2{}_{n+1}]}{2(1-\rho)}$$

Έχουμε επίσης:

$$E[A^2{}_{n+1}] = Var[A_{n+1}] + E^2[A_{n+1}]$$

Όπου το $Var[A_{n+1}]$ υπολογίζεται ως εξής:

$$Var[A_{n+1}] = E[Var[A_{n+1}|S]] + Var[E[A_{n+1}|S]]$$

Τώρα η $\{A_{n+1}|S\}$ είναι τυχαία μεταβλητή που ακολουθεί Poisson με μέση τιμή λS, η διασπορά της είναι επίσης λS, έτσι η παραπάνω σχέση γίνεται:

$$Var[A_{n+1}] = E[\lambda S] + Var[\lambda S] = r + \lambda {\sigma^2}_B$$

Όπου ${\sigma^2}_B$ η διασπορά της κατανομής του χρόνου εξυπηρέτησης.

Αντικαθιστούμε τα παραπάνω και έχουμε:

$$L^{(D)} = \rho + \frac{\rho^2 + \lambda^2 {\sigma^2}_B}{2(1-\rho)}$$

Το $L^{(D)}$ εκτός από αναμενόμενο μέγεθος του συστήματος στα σημεία αναχώρησης είναι και αναμενόμενο μέγεθος συστήματος σ' οποιοδήποτε αυθαίρετο χρονικό σημείο δηλαδή L, έτσι:

$$L = \rho + \frac{\rho^2 + \lambda^2 {\sigma^2}_B}{2(1-\rho)}$$

3.1.2 Departure-Point System-Size Probabilities

Έστω π_n η πιθανότητα να έχουμε n στο σύστημα σ' ένα σημείο αναχώρησης (ένα χρονικό σημείο λίγο μετά αφού ένας πελάτης έχει ολοκληρώσει την εξυπηρέτησή του), μετά την επίτευξη της κατάστασης ισορροπίας. Οι πιθανότητες $\{\pi_n\}$ γενικά δεν είναι ίδιες με τις πιθανότητες σε κατάσταση ισορροπίας $\{p_n\}$, για το μοντέλο ουράς M/G/1 όμως είναι.

Θα δείξουμε ότι αν η ουρά M/G/1 εξεταστεί μόνο ως προς τους χρόνους αναχώρησης οδηγεί σε μια διακριτή Μαρκοβιανή αλυσίδα. Έστω $t_1,t_2,t_3,\ldots$ οι χρόνοι εξυπηρέτησης από την ουρά, και έστω $X_n=X(t_n)$ ο αριθμός των πελατών που μένουν στο σύστημα μετά την αναχώρηση του πελάτη τη χρονική στιγμή t_n, έχουμε:

$$X_{n+1} = \begin{cases} X_n - 1 + A_{n+1}, & X_n \geq 1 \\ A_{n+1,} & X_n = 0 \end{cases}$$

Όπου A_{n+1} είναι ο αριθμός των πελατών που φτάνουν μεταξύ του χρόνου εξυπηρέτησης του n+1 πελάτη.

Για να δείξουμε ότι $X_1, X_2, X_3, \ldots$ είναι Μαρκοβιανή αλυσίδα πρέπει να ισχυριστούμε ότι οι μελλοντικές καταστάσεις της ουράς εξαρτώνται μόνο από την παρούσα κατάσταση και ειδικότερα πρέπει να δείξουμε ότι η δεδομένη παρούσα κατάσταση X_n, η μέλλουσα κατάσταση X_{n+1}, είναι ανεξάρτητες από τις προηγούμενες καταστάσεις $X_{n-1}, X_{n-2,\ldots}$

Για να το δείξουμε αυτό παρατηρούμε ότι από την παραπάνω σχέση το X_{n+1} εξαρτάται από το X_n και το A_{n+1}. Όσο το A_{n+1} είναι ανεξάρτητο των $X_{n-1}, X_{n-2,\ldots}$ τότε $\{X_n\}$ είναι Μαρκοβιανή αλυσίδα, αυτό αληθεύει διότι A_{n+1} είναι ο αριθμός των πελατών που φτάνουν κατά τη διάρκεια της εξυπηρέτησης του (n+1) πελάτη που εξαρτάται από το μήκος του χρόνου εξυπηρέτησής του, αλλά δεν εξαρτάται από γεγονότα που έγιναν νωρίτερα όπως το μήκος της ουράς στο προηγούμενα σημεία αναχώρησης $X_{n-1}, X_{n-2,\ldots}$

Έτσι η διακριτού χρόνου διαδικασία $X_1, X_2, X_3, \ldots$ είναι διακριτού χρόνου Μαρκοβιανή αλυσίδα:

$$p_{ij} = P(X_{n+1} = j | X_n = i)$$

Οι πιθανότητες μετάβασης εξαρτώνται από την κατανομή του αριθμού των πελατών που φτάνουν κατά τη διάρκεια της εξυπηρέτησης. Από τη στιγμή που αυτή η κατανομή δεν εξαρτάται από δείκτη της εξυπηρέτησης ενός πελάτη, τον αφήνουμε αυτόν τον δείκτη. Ειδικότερα, αν S ένας τυχαίος χρόνος εξυπηρέτησης και A ο τυχαίος αριθμός πελατών που φτάνουν κατά τη διάρκεια της εξυπηρέτησης, ορίζουμε:

$$k_i \equiv P(i \text{ αφίξεις κατά τη διάρκεια μιας εξυπηρέτησης}) = P(A = i)$$
$$= \int_0^\infty P(A = i | S = t) dB(t)$$

Όπου (A|S=t) είναι μία τυχαία μεταβλητή που ακολουθεί την Poisson με μέση τιμή λt, έτσι:

$$P(A = i | S = t) = \frac{e^{-\lambda t} (\lambda t)^i}{i!}$$

Έτσι:

$$k_{i=}\int_0^{\infty}\frac{e^{-\lambda t}(\lambda t)^i}{i!}dB(t)$$

Έτσι από τη σχέση:

$$X_{n+1}=\begin{cases}X_n-1+A_{n+1}, & X_n\geq 1\\ A_{n+1,} & X_n=0\end{cases}$$

Έχουμε:

$$P(X_{n+1}=j|X_n=i)=\begin{cases}P(A=j-i+1), & i\geq 1\\ P(A=j), & i=0\end{cases}$$

Συνοπτικά έχουμε τον παρακάτω πίνακα μετάβασης:

$$P=\{p_{ij}\}=\begin{pmatrix}k_0 & k_1 & k_2\ldots\\ k_0 & k_1 & k_2\ldots\\ 0 & k_0 & k_1\ldots\\ 0 & 0 & k_0\ldots\\ 0 & 0 & 0\ldots\\ \vdots & \vdots & \vdots\end{pmatrix}$$

Υποθέτοντας ότι έχουμε κατάσταση στατιστικής ισορροπίας έχουμε:

Το διάνυσμα των πιθανοτήτων σε κατάσταση ισορροπίας: π={π_n}

Και πP=π, έτσι έχουμε:

$$\pi_i=\pi_0k_i+\sum_{j=1}^{i+1}\pi_j k_{i-j+1},\qquad i=1,2,\ldots$$

Ορίζουμε τις γεννήτριες συναρτήσεις:

$$\Pi(z)=\sum_{i=0}^{\infty}\pi_i z^i\qquad \kappa\alpha\iota\qquad K(z)=\sum_{\iota=0}^{\infty}k_i z^i,\qquad |z|\leq 1$$

Από όπου προκύπτει:

$$\Pi(z) = \frac{\pi_0(1-z)K(z)}{K(z)-z}$$

Χρησιμοποιώντας το Π(1)=1 με κανόνα L'Hopital και επειδή K(1)=1 και K΄(1)=λ(1/μ), βρίσκουμε ότι π_0=1-ρ, όπου ρ=λΕ(χρόνου εξυπηρέτησης), έτσι:

$$\Pi(z) = \frac{(1-\rho)(1-z)K(z)}{K(z)-z}$$

3.1.3 Απόδειξη $\pi_n=p_n$

Θα δείξουμε ότι το π_n η πιθανότητα να έχουμε n πελάτες στο σύστημα σε κατάσταση ισορροπίας σ' ένα σημείο αναχώρησης, είναι ίσο με το p_n την πιθανότητα να έχουμε n πελάτες στο σύστημα σε μία αυθαίρετο χρονικό σημείο. Ξεκινάμε λαμβάνοντας υπόψη μια συγκεκριμένη υλοποίηση μιας πραγματικής διαδικασίας για μεγάλο χρονικό διάστημα (0,T). Έστω X(t) το μέγεθος του συστήματος τη χρονική στιγμή t, και έστω $A_n(t)$ ο αριθμός των μονάδων που κάνουν άλμα προς τα πάνω (αφίξεις) από την κατάσταση n που συμβαίνουν στο χρονικό διάστημα (0,t) και $D_n(t)$ ο αριθμός τον μονάδων που κάνουν άλμα προς τα κάτω (αναχωρήσεις) στην κατάσταση n στο χρονικό διάστημα. Έτσι αφού οι αφίξεις γίνονται κατ' άτομο και το ίδιο και οι εξυπηρετήσεις τότε έχουμε:

$$|A_n(T) - D_n(T)| \leq 1$$

Επιπλέον ο συνολικός αριθμός των αναχωρήσεων D(T) συνδέεται με τον συνολικό αριθμό των αφίξεων A(T) με τη σχέση:

$$D(T) = A(T) + X(0) - X(T)$$

Και οι πιθανότητες του σημείου αναχώρησης είναι:

$$\pi_n = \lim_{T\to\infty} \frac{D_n(T)}{D(T)}$$

Από τις δύο παραπάνω σχέσεις και προσθαφαιρώντας το $A_n(T)$ έχουμε:

$$\frac{D_n(T)}{D(T)} = \frac{A_n(T) + D_n(T) - A_n(T)}{A(T) + X(0) - X(T)}$$

Έτσι έχουμε:

$$\lim_{T\to\infty} \frac{D_n(T)}{D(T)} = \lim_{T\to\infty} \frac{A_n(T)}{A(T)}$$

Με πιθανότητα 1.

Από τη στιγμή που οι αφίξεις συμβαίνουν στα σημεία μιας διαδικασίας Poisson που λειτουργεί ανεξάρτητα από την κατάσταση της διαδικασίας, έχουμε ότι οι Poisson αφίξεις βρίσκουν μέσους χρόνους. Έτσι η γενικού χρόνου πιθανότητα p_n είναι ίδια με την πιθανότητα στα σημεία άφιξης q_n όπου:

$$q_n = \lim_{T\to\infty} \frac{A_n(T)}{A(T)}$$

Που είναι ίδια με την π_n.

3.1.4 Χρόνοι Αναμονής

Σ' αυτή την ενότητα θα παρουσιάσουμε διάφορα σημαντικά αποτελέσματα σχετικά με τους χρόνους καθυστέρησης. Έχουμε ήδη δείξει ότι η μέση αναμονή του συστήματος σχετίζεται με τον μέσο αριθμό πελατών από τον τύπο του Little W=L/λ.

Έχουμε ότι η στάσιμη πιθανότητα για την M/G/1 μπορεί να γραφεί ως:

$$p_n = \pi_n = \frac{1}{n!}\int_0^\infty (\lambda t)^n e^{-\lambda t} dW(t), \quad n \geq 0$$

Από τη στιγμή που το μέγεθος το συστήματος με πειθαρχία FCFS θα είναι ίσο με n σ' ένα αυθαίρετο σημείο αναχώρησης αν έχουν γίνει n (Poisson) αφίξεις κατά τη διάρκεια της αναμονής στο σύστημα από την αναχώρηση.

Ορίζουμε την παρακάτω γεννήτρια συνάρτηση:

$$P(z) = \sum_{n=0}^{\infty} p_n z^n = \int_0^{\infty} e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t z)^n}{n!} dW(t) = \int_0^{\infty} e^{-\lambda t(1-z)} dW(t) = W^*[\lambda(1-z)]$$

Από τον κανόνα της αλυσίδας έχουμε:

$$\frac{d^k P(z)}{dz^k} = (-1)^k \lambda^k \frac{d^k W^*(u)}{du^k}\Big|_{u=\lambda(1-z)} = (-1)^k \lambda^k (-1)^k E[T^k e^{-Tu}]\Big|_{u=\lambda(1-z)}$$

Έτσι αν $L_{(k)}$ ονομάσουμε την κ-οστή παραγοντική στιγμή του μεγέθους του συστήματος και W_k την κανονική κ-οστή στιγμή του χρόνου αναμονής του συστήματος τότε:

$$L_{(k)} = \frac{d^k P(z)}{dz^k}\Big|_{z=1} = \lambda^k W_k$$

Το αποτέλεσμα μας παρέχει μια καλή γενίκευση της φόρμουλας του Little, από τη στιγμή που οι πιο συνηθισμένες στιγμές μπορούν να ληφθούν από τις παραγοντικές στιγμές.

Στην ουρά M/M/1 είμαστε σε θέση εύκολα να λάβουμε έναν απλό τύπο για την κατανομή του χρόνου αναμονής από την άποψη της κατανομής εξυπηρέτησης, έτσι:

$$W(t) = (1-\rho) \sum_{n=0}^{\infty} \rho^n B^{(n+1)}(t)$$

Όπου B(t) είναι η αθροιστική συνάρτηση εκθετικής κατανομής και η $B^{(n+1)}(t)$ είναι η (n+1) συνέλιξη. Η παραγωγή αυτών των αποτελεσμάτων απαιτεί την ιδιότητα μη-μνήμης της εκθετικής εξυπηρέτησης, από τη στιγμή που οι αφίξεις βρίσκουν τον εξυπηρετητή στη μέση της περιόδου εξυπηρέτησης με πιθανότητα ίση με ρ. Ωστόσο πρέπει να χάσουμε την ιδιότητα της μη-μνήμης και επιπλέον να αναζητήσουμε μια εναλλακτική προσέγγιση ώστε να βρούμε ένα συγκρίσιμο αποτέλεσμα για την M/G/1 ουρά.

Έχουμε ότι:

$$P(z) = W^*[\lambda(1-z)]$$

Έτσι έχουμε:

$$P(z) = \Pi(z) = \frac{(1-\rho)(1-z)K(z)}{K(z)-z}$$

Επιπλέον:

$$K(z) = \int_0^\infty e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t z)^n}{n!} dB(t) = \int_0^\infty e^{-\lambda t(1-z)} dB(t) = B^*[\lambda(1-z)]$$

Από τις τρείς παραπάνω έχουμε:

$$W^*[\lambda(1-z)] = \frac{(1-\rho)(1-z)B^*[\lambda(1-z)]}{B^*[\lambda(1-z)]-z}$$

ή αλλιώς:

$$W^*(s) = \frac{(1-\rho)sB^*(s)}{s-\lambda[1-B^*(s)]}$$

Και επειδή:

$$W^*(s) = W_q^*(s)B^*(s)$$

Από τη στιγμή που:

$$T = T_q + S$$

Έτσι:

$$W_q^*(s) = \frac{(1-\rho)s}{s-\lambda[1-B^*(s)]}$$

Αναλύουμε το δεύτερο μέλος σαν γεωμετρική σειρά, από τη στιγμή που:

$$\frac{\lambda}{s}[1 - B^*(s)] < 1$$

Έτσι:

$$W_q^*(s) = (1-\rho)\sum_{n=0}^{\infty}\left(\frac{\lambda}{s}[1 - B^*(s)]\right)^n = (1-\rho)\sum_{n=0}^{\infty}(\rho\frac{\mu}{s}[1 - B^*(s)])^n$$

Έχουμε ότι:

$$R(t) \equiv \mu\int_0^t [1 - B(x)]dx$$

Διαισθητικά R(t) είναι η αθροιστική συνάρτηση κατανομής του υπολειπόμενου χρόνου εξυπηρέτησης ενός πελάτη που εξυπηρετείται τη στιγμή μιας αυθαίρετης άφιξης, δεδομένου ότι η άφιξη εμφανίζεται όταν ο εξυπηρετητής είναι απασχολημένος.

Ο τύπος για το R(t) μπορεί να παραχθεί από την θεωρία ανανέωσης (renewal theory, Ross 2007). Έτσι έχουμε:

$$W_q^*(s) = (1-\rho)\sum_{n=0}^{\infty}[\rho R^*(s)]^n$$

Η οποία δίνει μετά από αναστροφή για κάθε όρο χρησιμοποιώντας την ιδιότητα της συνέλιξης:

$$W_q(t) = (1-\rho)\sum_{n=0}^{\infty}\rho^n R^{(n)}(t)$$

Αυτό σημαίνει ότι αν ο χρόνος αναδιαταχθεί με τον υπόλοιπο χρόνο εξυπηρέτησης σαν βασική μονάδα, κάθε άφιξη σε κατάσταση στατιστικής ισορροπίας βρίσκει n τέτοιου χρόνου μονάδες ενδεχόμενης εξυπηρέτησης μπροστά της με πιθανότητα $(1-\rho)\rho^n$, δίνοντας ένα αποτέλεσμα σημαντικά ίδιο με αυτό της M/M1 ουράς.

Χρησιμοποιούμε επίσης αυτά τα αποτελέσματα προκειμένου να βρούμε μια σχέση που συνδέει επαναληπτικά τις υψηλότερες στιγμές της αναμονής , έστω $W_{q,k}$.
Ξαναγράφουμε την βασική εξίσωση ως εξής:

$$W_q^*(s)\{s-\lambda[1-B^*(s)]\}=(1-\rho)s$$

Και στη συνέχεια παίρνουμε τη κ-οστή παράγωγο (k>1) της προηγούμενης εξίσωσης εφαρμόζοντας τον κανόνα του Leibniz :

$$\sum_{i=0}^{k}\binom{k}{i}\left(\frac{d^{k-i}W_q^*(s)}{ds^{k-i}}\right)\left(\frac{d^i\{s-\lambda[1-B^*(s)]\}}{ds^i}\right)=\frac{d^k[(1-\rho)s]}{ds^k}$$

Μπορούμε να θεωρήσουμε ότι k>1 από τη στιγμή που επιδιώκουμε τις υψηλότερες στιγμές. Έτσι το δεξί μέλος της παραπάνω εξίσωσης είναι μηδέν, έτσι:

$$0=\frac{d^kW_q^*(s)}{ds^k}\{s-\lambda[1-B^*(s)]\}+k\frac{d^{k-1}W_q^*(s)}{ds^{k-1}}\left[1+\lambda B^{*'}(s)\right]+\sum_{i=2}^{k}\binom{k}{i}\left(\frac{d^{k-i}W_q^*(s)}{ds^{k-i}}\right)\lambda\frac{d^iB^*(s)}{ds^i}$$

Θέτουμε s=0 και έχουμε:

$$0=k(-1)^{k-1}W_{q,k-1}(1-\rho)+\sum_{i=2}^{k}\binom{k}{i}(-1)^{k-i}W_{q,k-i}E[S^i](-1)^i$$

Ή αλλιώς:

$$W_{q,k-1}=\frac{\lambda}{k(1-\rho)}\sum_{i=2}^{k}\binom{k}{i}W_{q,k-i}E[S^i]$$

Μπορούμε να ξαναγράψουμε θέτοντας K=k-1 και j=i-1 από την οποία παίρνουμε:

$$W_{q,K}=\frac{\lambda}{1-\rho}\sum_{j=1}^{K}\binom{K}{j}W_{q,K-j}\frac{E[S^{j+1}]}{j+1}$$

3.1.5 Πεπερασμένη M/G/1 Ουρά

Η ανάλυση μιας ουράς M/G/1/K περιορισμένης χωρητικότητας, είναι κατά ένα τρόπο ίδια με περίπτωση που έχουμε απεριόριστο χώρο αναμονής.

Σ' αυτή την περίπτωση δεν κρατάμε τον τύπο PK, από τη στιγμή που ο αναμενόμενος αριθμός των αφίξεων κατά τη διάρκεια της περιόδου εξυπηρέτησης εξαρτάται από το μέγεθος του συστήματος. Ο καλύτερος τρόπος για να πάρουμε νέο αποτέλεσμα είναι άμεσα από τις πιθανότητες στατιστικής ισορροπίας από τη στιγμή που είναι πεπερασμένος ο αριθμός τους.

Ο πίνακας μετάβασης ενός βήματος πρέπει να περικοπεί στο K-1, έτσι:

$$P = \begin{pmatrix} k_0 & k_1 & k_2 & \dots & 1-\sum_{n=0}^{K-2} k_n \\ k_0 & k_1 & k_2 & \dots & 1-\sum_{n=0}^{K-2} k_n \\ 0 & k_0 & k_1 & \dots & 1-\sum_{n=0}^{K-3} k_n \\ 0 & 0 & k_0 & \dots & 1-\sum_{n=0}^{K-4} k_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1-k_0 \end{pmatrix}$$

Συνεπάγεται ότι η stationary εξίσωση είναι:

$$\pi_i = \begin{cases} \pi_0 k_i + \sum_{j=1}^{i+1} \pi_j \, k_{i-j+1}\,, & i = 0,1,2,\dots,K-2 \\ 1-\sum_{j=o}^{K-2} \pi_j\,, & i = K-1 \end{cases}$$

Αυτές οι K (συνεπής) εξισώσεις με K αγνώστους μπορούν να λυθούν για όλες τις πιθανότητες, και το μέσο μέγεθος του συστήματος στα σημεία αναχώρησης δίνεται τότε από την $L = \sum_{i=0}^{K-1} i\pi_i$. (Σημειώνουμε ότι η μέγιστη κατάσταση της Μαρκοβιανής αλυσίδας δεν είναι το K, από τη στιγμή που παρατηρούμε μετά από μια αναχώρηση, και υποθέτουμε ότι K>1)

Το πρώτο μέλος της stationary εξίσωσης είναι παρόμοιο με αυτό της μη πεπερασμένης M/G/1. Επομένως οι αντίστοιχες stationary πιθανότητες {πᵢ} για την M/G/1/K και {π*ᵢ} για την M/G/1/∞ πρέπει να είναι στην χειρότερη ανάλογες για $i \leq K-1$, έτσι:$\pi_i = C\pi_i^*$, i=0,1,…K-1. Η συνήθης συνθήκη που οι πιθανότητες αθροίζουν στο ένα συνεπάγεται:

$$C = \frac{1}{\sum_{i=0}^{K-1} \pi_i^*}$$

Επιπλέον παρατηρούμε ότι η κατανομή πιθανότητας του συστήματος που αντιμετωπίζει μια άφιξη είναι διαφορετική από $\{\pi_i\}$, από τη στιγμή που τώρα ο χώρος πρέπει να διευρυνθεί ώστε να χωράει K.

Έστω $q_n^{'}$ η πιθανότητα ένας πελάτης που φτάνει να βρίσκει ένα σύστημα με n πελάτες. (εδώ μιλάμε για την κατανομή των πελατών που φτάνουν ανεξάρτητα αν μπαίνουν ή όχι στην ουρά σε αντίθεση με αυτούς που φτάνουν και μπαίνουν στην ουρά , έστω q_n. Έχουμε ότι:

$$\pi_n = P(\eta\ \acute{\alpha}\varphi\iota\xi\eta\ \beta\rho\acute{\iota}\sigma\kappa\epsilon\iota\ n | o\ \pi\epsilon\lambda\acute{\alpha}\tau\eta\varsigma\ \acute{\epsilon}\chi\epsilon\iota\ \acute{o}\nu\tau\omega\varsigma\ \epsilon\nu\tau\alpha\chi\theta\epsilon\acute{\iota}) = q_n = \frac{q_n^{'}}{1-q_K^{'}}, 0 \leq n \leq K-1$$

Άρα: $$q_n^{'} = (1-q_K^{'})\pi_n, \quad 0 \leq n \leq K-1$$

Για να βρούμε το $q_K^{'}$ παίρνουμε ότι ο ρυθμός αποτελεσματικών αφίξεων είναι ίσος με τον ρυθμό των αποτελεσματικών αναχωρήσεων, έτσι έχουμε: $\lambda(1-q_K^{'}) = \mu(1-p_0)$

Άρα:

$$q_n^{'} = \frac{(1-p_0)\pi_n}{\rho}, \qquad 0 \leq n \leq K-1$$

$$q_K^{'} = \frac{\rho - 1 + p_0}{\rho}$$

Αλλά από τη στιγμή που οι αφίξεις γίνονται σύμφωνα με μια ανέλιξη Poisson $q_n^{'} = p_n$για όλα τα n, έτσι:

$$q_0^{'} = p_0 = \frac{(1-p_0)\pi_0}{\rho} => \ p_0 = \frac{\pi_0}{\pi_0 + \rho}$$

Και τελικά:

$$q_n^{'} = \frac{\pi_n}{\pi_0 + \rho}$$

3.2 Γενική Εξυπηρέτηση, Πολλοί εξυπηρετητές (M/G/c/., M/G/∞)

3.2.1 Κάποια Αποτελέσματα για την M/G/c/∞

Για την M/G/c/∞ το βασικό γενικό αποτέλεσμα που μπορεί να βρεθεί είναι:

$$L_{(k)} = \lambda^k W_k$$

Το οποίο αποδεικνύεται παρακάτω.

Γνωρίζουμε ότι για την M/G/1 :

$$\pi_n = P(n\ στο\ σύστημα\ μετά\ απο\ μια\ αναχώρηση) = \frac{1}{n!}\int_0^\infty (\lambda t)^n e^{-\lambda t} dW(t)$$

Η παραπάνω ισχύει για την M/G/c αν θεωρήσουμε τα πάντα σε όρους για την ουρά και όχι για το σύστημα:

$$\pi_n^q \equiv P(nστο\ σύστημα\ μετά\ απο\ μια\ αναχώρηση) = \frac{1}{n!}\int_0^\infty (\lambda t)^n e^{-\lambda t} dW_q(t)$$

Με μέσο μήκος ουράς στα σημεία αναχώρησης:

$$L_q^{(D)} = \sum_{n=1}^{\infty} n\pi_n^q = \int_0^\infty \lambda t dW_q(t) = \lambda W_q$$

Που είναι ο τύπος του θεωρήματος του Little.

Έστω $L_{q(k)}^{(D)}$, η κ-παραγόντων στιγμή του μεγέθους ουράς στο σημείο αναχώρησης, τότε:

$$L_{q(k)}^{(D)} = \sum_{n=1}^{\infty} n(n-1)\dots(n-k+1)\pi_n^q = \int_0^\infty dW_q(t) \sum_{n=1}^{\infty} \frac{n(n-1)\dots(n-k+1)(\lambda t)^n e^{-\lambda t}}{n!}$$

Όπου το άθροισμα είναι κ-παραγόντων στιγμή Poisson και είναι $(\lambda t)^k$.

Έτσι:

$$L_{q(k)}^{(D)} = \lambda^k W_{q,k}$$

Όπου $W_{q,k}$ είναι η κ-οστή στιγμή του χρόνου αναμονής στην ουρά.

3.2.2 Η M/G/∞ Ουρά και M/G/c/c Loss System

Σ' αυτή την ενότητα θα ασχοληθούμε με την παραγωγή δύο σημαντικών αποτελεσμάτων για την M/G/∞ ουρά, την transient κατανομή του αριθμού των πελατών κατά τη στιγμή t και την transient κατανομή για τον αριθμό των πελατών που έχουν ολοκληρώσει την εξυπηρέτησή τους μέχρι την χρονική στιγμή t, αυτή είναι μια διαδικασία καταμέτρησης αναχωρήσεων.

Έστω η N(t) η συνολική διαδικασία του μεγέθους του συστήματος και Y(t) η διαδικασία αναχωρήσεων και η διαδικασία των εισόδων X(t)=Y(t)+N(t).

Έχουμε ότι:

$$P(N(t)=n)=\sum_{i=n}^{\infty}P(N(t)=n|X(t)=i)\frac{e^{-\lambda t}(\lambda t)^{i}}{i!}$$

Από τη στιγμή που οι είσοδοι είναι Poisson. Η πιθανότητα ένας πελάτης που φτάνει τη χρονική στιγμή χ να είναι παρών την χρονική στιγμή t δίνεται από την 1-B(t-x), όπου B(u) είναι η αθροιστική συνάρτηση κατανομής του χρόνου εξυπηρέτησης. Έτσι έχουμε ότι η πιθανότητα ένας αυθαίρετος πελάτης από αυτούς να είναι ακόμα στην εξυπηρέτηση δίνεται από την:

$$q_t=\int_0^t P(\chi\rho\acute{o}\nu os\ \epsilon\xi\upsilon\pi\eta\rho\acute{\epsilon}\tau\eta\sigma\eta s>t-x|\acute{\alpha}\varphi\iota\xi\eta\ \tau\eta\ \chi\rho.\sigma\tau\iota\gamma\mu\acute{\eta}\ x)P(\acute{\alpha}\varphi\iota\xi\eta\ \ \chi\rho.\sigma\tau\iota\gamma\mu\acute{\eta}\ \chi)dx$$

Από τη στιγμή που οι είσοδοι είναι Poisson, P(άφιξή τη χρονική στιγμή χ) είναι ομοιόμορφη στο (0,t) και είναι 1/t, έτσι:

$$q_t=\frac{1}{t}\int_0^t[1-B(t-x)]dx=\frac{1}{t}\int_0^t[1-B(x)]dx$$

Και είναι ανεξάρτητη από κάθε άλλη άφιξη. Επομένως από τον διωνυμικό νόμο:

$$P\{N(t)=n|X(t)=i\}=\binom{i}{n}q_t^n(1-q_t)^{i-n},\quad n\geq 0$$

Και η transient κατανομή είναι:

$$P\{N(t)=n\}=\sum_{i=n}^{\infty}\binom{i}{n}\frac{q_t^n(1-q_t)^{i-n}e^{-\lambda t}(\lambda t)^i}{i!}=\frac{(\lambda q_t t)^n e^{-\lambda t}}{n!}\sum_{i=n}^{\infty}\frac{[\lambda t(1-q_t)]^{i-n}}{(i-n)!}$$
$$=\frac{(\lambda q_t t)^n e^{-\lambda t}e^{-\lambda t-\lambda q_t t}}{n!}=\frac{(\lambda q_t t)^n e^{-\lambda q_t t}}{n!}$$

Που λέγεται μη-ομογενής Poisson με μέση τιμή $\lambda q_t t$.

Για να βρούμε τη λύση ισορροπίας παίρνουμε το όριο του t→∞ και προκύπτει:

$$\lim_{t\to\infty}(\lambda q_t t)=\lambda\int_0^{\infty}[1-B(\chi)]dx=\frac{\lambda}{\mu}$$

Και έτσι η λύση ισορροπίας είναι Poisson με μέση τιμή λE(S)=λ/μ.

Η κατανομή της διαδικασίας καταμέτρησης των αναχωρήσεων Y(t) βρίσκεται με την ίδια διαδικασία χρησιμοποιώντας: $1-q_t=\int_0^t B(x)dx/t$ αντί για q_t. Το αποτέλεσμα όπως αναμένεται είναι:

$$P\{Y(t)=n\}=\frac{[\lambda(1-q_t)t]^n e^{-\lambda(1-q_t)t}}{n!}$$

Αν πάρουμε το όριο του t→∞ βλέπουμε ότι το q_t τείνει στο μηδέν, και έτσι η interdeparture διαδικασία είναι Poisson σε κατάσταση ισορροπίας που είναι ακριβώς ίδια με την διαδικασία αφίξεων.

Σχετικά με την M/G/c/c έχουμε ότι η κατανομή του συστήματος σε κατάσταση ισορροπίας:

$$p_n=\frac{\left(\frac{\lambda}{\mu}\right)^n/n!}{\sum_{i=0}^{c}\left(\frac{\lambda}{\mu}\right)^i/i!},\qquad 0\le n\le c$$

Η συγκεκριμένη τιμή από αυτή για την p_c όπως έχουμε δει καλείται Erlang's Loss Formula και το αποτέλεσμα επεκτείνεται για την M/G/∞, όπου: $p_n=\frac{e^{-\frac{\lambda}{\mu}}(\frac{\lambda}{\mu})^n}{n!}$.

Για την M/G/c/c:

Όταν c=1 δηλαδή την M/G/1/1 είναι απλή περίπτωση διότι έχουμε:

$p_0 = 1 - \rho_{eff} = 1 - p_1$ για κάθε G/G/1/1 ουρά, από τη στιγμή που $\rho_{eff} = \lambda(1 - p_1)/\mu$ έχουμε:

$$p_1 = \frac{\lambda/\mu}{1 + \lambda/\mu}, \qquad p_0 = \frac{1}{1 + \lambda/\mu}$$

Όταν c>1:

$$p_n = p_0 \frac{(\frac{\lambda}{\mu})^n}{n!}$$

$$p_0 = (\sum_{n=0}^{c} \frac{(\frac{\lambda}{\mu})^n}{n!})^{-1}$$

Και έτσι:

$$p_n = \frac{\left(\frac{\lambda}{\mu}\right)^n / n!}{\sum_{i=0}^{c} \left(\frac{\lambda}{\mu}\right)^i / i!}, \quad 0 \le n \le c$$

Που είναι ακριβώς ίδιο με τα αποτελέσματα της M/M/c/c ουράς.

Παρατηρούμε ότι το γεγονός ότι οι πιθανότητες σε κατάσταση ισορροπίας για την M/G/c/c είναι ανεπηρέαστες από την επιλογή της G σημαίνει ότι αυτές οι πιθανότητες πάντα θα ικανοποιούν την σχέση γέννησης-θανάτου της M/M/c/c:

$$\lambda p_n = (n+1)\mu p_{n+1}, \quad n = 0, \dots, c-1$$

<u>Αποτελέσματα για μοντέλα M/G/c/K</u>:

- *M/G/c/c vs. M/M/c/c* : πιθανότητες σε κατάσταση ισορροπίας και διαδικασία εξόδων ανεξάρτητες από τη μορφή του G.
- *M/G/∞ vs. M/M∞* : πιθανότητες σε κατάσταση ισορροπίας και διαδικασία εξόδων ανεξάρτητες από τη μορφή του G.
- *M/G/c/∞ vs. M/M/c/∞* : διαδικασίες εξόδων ίδιες αν και μόνο αν G≡M.

3.3 Η Ουρά G/M/1

Θα εξετάσουμε την ουρά που έχει έναν εξυπηρετητή , οι χρόνοι εξυπηρέτησης ακολουθούν την εκθετική με μέση τιμή 1/μ, οι ενδιάμεσοι χρόνοι άφιξης ακολουθούν μια γενική κατανομή και υποθέτουμε ότι είναι ανεξάρτητες και ισόνομες τ.μ.. Υποθέτουμε ότι η ουρά λειτουργεί σε κατάσταση στατιστικής ισορροπίας. Σ' αυτή την περίπτωση εξετάζουμε το σύστημα στους χρόνους άφιξης. Έστω X_n ο αριθμός των πελατών στο σύστημα ακριβώς πριν την άφιξη του n-οστού πελάτη, τότε:

$$X_{n+1} = X_n + 1 - B_n, \quad B_n \leq X_n + 1, \quad X_n \geq 0$$

Όπου B_n ο αριθμός των πελατών που εξυπηρετήθηκαν κατά τη διάρκεια του ενδιάμεσου χρόνου $T^{(n)}$ ανάμεσα στις n-στή και (n+1)-στη αφίξεις. Από τη στιγμή που έχουμε υποθέσει ότι οι ενδιάμεσοι χρόνοι είναι ανεξάρτητες και ισόνομες τ.μ. η τ.μ. $T^{(n)}$ μπορεί να γραφεί Τα και έστω A(t) η αθροιστική συνάρτηση πιθανότητας αυτής.

Επιπλέον η τ.μ. B_n δεν εξαρτάται από το ιστορικό της ουράς δεδομένου του τρέχοντα αριθμού X_n στο σύστημα κατά τη χρονική στιγμή της n-οστής άφιξης. Έτσι $\{X_0, X_1, \dots\}$ είναι Μαρκοβιανή αλυσίδα.

Ορίζουμε τις παρακάτω πιθανότητες:

$$b_k \equiv \int_0^\infty \frac{e^{-\mu t}(\mu t)^k}{k!}\, dA(t)$$

Η ερμηνεία είναι ότι η b_k είναι η πιθανότητα να γίνουν ακριβώς k ολοκληρώσεις εξυπηρετήσεων μεταξύ δυο διαδοχικών αφίξεων (δεδομένου ότι υπάρχουν τουλάχιστον k στο σύστημα ακριβώς πριν την πρώτη άφιξη), έτσι έχουμε:

$$b_k = P(B_n = k | X_n \geq k)$$

Έτσι ο ενός βήματος πίνακας μετάβασης των πιθανοτήτων για την Μαρκοβιανή αλυσίδα $\{X_0, X_1, \dots\}$ είναι:

$$P = \{p_{ij}\} = \begin{pmatrix} 1-b_0 & b_0 & 0 & 0 & 0 & \dots \\ 1-\sum_{k=0}^{1} b_k & b_1 & b_0 & 0 & 0 & \dots \\ 1-\sum_{k=0}^{2} b_k & b_2 & b_1 & b_0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Υποθέτοντας ότι υπάρχει λύση σε κατάσταση ισορροπίας και έστω το διάνυσμα πιθανοτήτων μία άφιξη να βρίσκει n στο σύστημα, q={q_n}, n=0,1,2,… έχουμε τις εξής ισότητες:

$$qP = q \; \kappa\alpha\iota \; qe = 1$$

Από τις οποίες:

$$q_i = \sum_{k=0}^{\infty} q_{i+k-1} b_k \,, \;\; i \geq 1$$

$$q_0 = \sum_{j=0}^{\infty} q_j (1 - \sum_{k=0}^{j} b_k)$$

Έστω: Dq_i=q_{i+1} για $i \geq 1$ η παραπάνω γράφεται:

$$q_i - (q_{i-1} b_0 + q_i b_1 + q_{i+1} b_2 + \cdots) = 0$$

Ώστε:

$$q_{i-1}(D - b_0 - Db_1 - D^2 b_2 - D^3 b_3 - \cdots) = 0$$

Η χαρακτηριστική εξίσωση για αυτή τη διαφορική εξίσωση είναι:

$$z - b_0 - zb_1 - z^2 b_2 - z^3 b_3 - \cdots = 0$$

Η οποία γράφεται ως εξής:

$$\sum_{n=0}^{\infty} b_n z^n = z$$

Από τη στιγμή που b_n είναι πιθανότητα ο όρος στα αριστερά είναι απλώς η probability generating συνάρτηση των $\{b_n\}$, την οποία καλούμε β(z). Τότε έχουμε:

$$\beta(z) \equiv \sum_{n=0}^{\infty} b_n z^n = z$$

Έχουμε ότι:

$$\beta(z) = A^*[\mu(1-z)]$$

Έτσι έχουμε:

$$z = A^*[\mu(1-z)]$$

Σκοπός μας είναι να βρούμε λύσεις της χαρακτηριστικής εξίσωσης που μπορούν να χρησιμοποιηθούν για να καθορίσουμε τα $\{q_i\}$. Έχουμε ότι η :

$$\beta(z) \equiv \sum_{n=0}^{\infty} b_n z^n = z$$

Έχει μία ρίζα στο (0,1) (υποθέτοντας ότι λ/μ<1) και δεν υπάρχουν άλλες ρίζες με απόλυτη τιμή μικρότερη του 1, έστω r_0 αυτή η ρίζα, έτσι έχουμε:

$$q_i = Cr_0^i, \quad i \geq 0$$

Η σταθερά C καθορίζεται από τη συνθήκη που αθροίζονται οι πιθανότητες στο ένα, έτσι: $C=1-r_0$.

Για να δείξουμε ότι υπάρχει μία και μόνο θετική ρίζα στο (0,1) παίρνουμε τα δυο μέλη της ισότητας ξεχωριστά ώστε:

$$y = \beta(z) \quad \kappa\alpha\iota \quad y = z$$

πρώτα παρατηρούμε ότι:

$$0 < \beta(0) = b_0 < 1 \ \ \kappa\alpha\iota \quad \beta(1) = \sum_{n=0}^{\infty} b_n = 1$$

Μπορούμε εύκολα να δείξουμε ότι η β(z) είναι γνησίως αύξουσα και κυρτή διότι:

$$\beta^{'}(z) = \sum_{n=1}^{\infty} n b_n z^{n-1} \geq 0$$

$$\beta^{''}(z) = \sum_{n=2}^{\infty} n(n-1) b_n z^{n-2} \geq 0$$

Επιπλέον από τη στιγμή που οι χρόνοι εξυπηρέτησης ακολουθούν την εκθετική κάθε b_n είναι αυστηρά θετικό. Έτσι $b_n>0$ για $n\geq0$ έτσι η β(z) είναι αυστηρά κυρτή.

Υπάρχουν δύο πιθανές περιπτώσεις για τα γραφήματα των:

$$y = \beta(z) \quad \kappa\alpha\iota \quad y = z$$

Είτε δεν υπάρχουν τομές στο διάστημα (0,1) είτε υπάρχει ακριβώς μία τομή στο (0,1). Η δεύτερη περίπτωση συμβαίνει όταν:

$$\beta^{'}(1) = E[\alpha\rho\iota\theta\mu\acute{o}\varsigma \ \pi o\upsilon \ \varepsilon\xi\upsilon\pi\eta\rho\varepsilon\tau o\acute{\upsilon}\nu\tau\alpha\iota \ \sigma\tau o\nu \ \varepsilon\nu\delta\iota\acute{\alpha}\mu\varepsilon\sigma o \ \chi\rho\acute{o}\nu o \ \alpha\varphi\acute{\iota}\xi\varepsilon\omega\nu] = \frac{\mu}{\lambda} > 1$$

Έτσι όταν λ/μ<1 υπάρχει ακριβώς μια ρίζα στο (0,1).

Ακόμα δεν έχουμε δείξει ότι είναι η μόνη κυρτή ρίζα με απόλυτη τιμή μικρότερη του 1, για να το δείξουμε χρησιμοποιούμε το θεώρημα του Rouche.

Θεωρούμε $\beta^{'}(1) = \frac{1}{\rho} > 1$ και έστω $f(z) \equiv -z \ \ \kappa\alpha\iota \ \ g(z) \equiv b(z)$, επειδή

$g(1) = 1 \ \kappa\alpha\iota \ \ g^{'}(1) > 1 \ \tau\acute{o}\tau\varepsilon \ g(1-\varepsilon) < 1 - \varepsilon \ \gamma\iota\alpha \ \alpha\rho\kappa\varepsilon\tau\acute{\alpha} \ \mu\iota\kappa\rho\acute{o} \ \varepsilon$

Έστω z τέτοιο ώστε $|z| = 1 - \varepsilon$ από την τριγωνική ανισότητα έχουμε:

$$|g(z)| \leq \sum_{n=0}^{\infty} b_n |z|^n = g(1-\varepsilon) < 1-\varepsilon = |f(z)|$$

Από το θεώρημα του Rouche $f(z) = -z \; \kappa\alpha\iota \; f(z) + g(z) = -z + b(z)$ έχουν τον ίδιο αριθμό ριζών μέσα στο όρια του $|z| = 1-\varepsilon$. Από τη στιγμή που το ε μπορεί να φτιαχτεί αυθαίρετα μικρό υπάρχει ακριβώς μία κυρτή ρίζα της z=b(z) της οποίας η απόλυτη τιμή είναι μικρότερη του 1, έτσι πρέπει να είναι η πραγματική ρίζα r_0 που βρήκαμε νωρίτερα.

Η εύρεση της ρίζας r_0 γενικά περιλαμβάνει αριθμητικές διαδικασίες που είναι σχετικά εύκολες. Για παράδειγμα η μέθοδος διαδοχικών αντικαταστάσεων (successive substitution):

$$z^{(k+1)} = \beta\left(z^{(k)}\right), \quad k = 0,1,2,\ldots. \quad , \quad 0 < z^{(0)} < 1$$

Είναι σίγουρο ότι συγκλίνει λόγω της μορφής της β(z).

Απ' όλα τα παραπάνω προκύπτει ότι:

$$q_n = (1-r_0)r_0^n, \quad n \geq 0, \quad \rho < 1$$

Παρατηρούμε ότι την αναλογία που υπάρχει ανάμεσο στον παραπάνω τύπο και την πιθανότητα στην M/M/1 ουρά που δίνεται από τον τύπο:$p_n = (1-\rho)\rho^n$. Μπορούμε επομένως να χρησιμοποιήσουμε όλα τα αναμενόμενα μέτρα αποδοτικότητας που έχουμε βρει για την M/M/1 απλώς αντικαθιστώντας το ρ με r_0. Ωστόσο, σημειώνουμε ότι το q_n είναι η πιθανότητα σε κατάσταση ισορροπίας να έχουμε n στο σύστημα λίγο πριν την άφιξη και όχι η γενικού χρόνου πιθανότητα σε κατάσταση ισορροπίας δηλαδή η p_n, έτσι τα αναμενόμενα μέτρα ισχύουν μόνο για τα arrival points. Σε αντίθεση με το M/G/1 μοντέλο, δεν ισχύει εδώ ότι $q_n = p_n$. Η ισότητα ισχύει γι αυτό το μοντέλο αν και μόνο αν οι αφίξεις είναι Poisson, δηλαδή $q_n = p_n$ για το G/M/1 αν και μόνο αν G=M.

Υπό αυτές τις συνθήκες χρησιμοποιούμε έναν εκθέτη (A) που υποδηλώνει το γεγονός ότι ένα συγκεκριμένο μέτρο αποδοτικότητας έχει ληφθεί σχετικά με τα arrival points μόνο και έτσι:

$$L^{(A)} = \frac{r_0}{1-r_0} \quad \kappa\alpha\iota \quad L_q^{(A)} = \frac{r_0^2}{1-r_0}$$

Η συναρτήσεις κατανομής του χρόνου στην ουρά και του χρόνου αναμονής στο σύστημα $W_q(t)$ και W(t) μπορούν επίσης να ληφθούν από την M/M/1 αντικαθιστώντας το ρ με r_0 έτσι έχουμε:

$$W_q(t) = 1 - r_0 e^{-\mu(1-r_0)t}, \quad t \geq 0$$

$$W(t) = 1 - e^{-\mu(1-r_0)t}, \quad t \geq 0$$

Και έτσι έχουμε:

$$W_q = \frac{r_0}{\mu(1-r_0)} \quad \kappa\alpha\iota \quad W = \frac{1}{\mu(1-r_0)}$$

Αυτά τα αποτελέσματα αναφέρονται στην κατανομή των χρόνων αναμονής όπως παρατηρούνται από πελάτες που φτάνουν στο σύστημα. Αυτή, σε αντίθεση με την κατανομή των εικονικών (virtual) χρόνων αναμονής που αντιστοιχούν στους χρόνους αναμονής που θα μπορούσαν να παρατηρηθούν για εικονικούς πελάτες που φτάνουν σε τυχαίες χρονικές στιγμές μάλλον σαν τις χρονικές στιγμές των πραγματικών πελατών. Αν κάποιο παραπέμψει στις σχέσεις του $W_q(t)$ και του $W(t)$ για την M/M/1 η σχέση του $W_q(t)$ εξαρτάται από το γεγονός ότι $q_n = p_n$ το οποίο δικαιολογείται από τις Poisson αφίξεις. Εδώ $q_n \neq p_n$ οπότε η εικονική και η πραγματική συναρτήσεις κατανομής των χρόνων αναμονής είναι διαφορετικές.

ΚΕΦΑΛΑΙΟ 4: ΠΡΟΣΟΜΟΙΩΣΗ

4.1 Discrete-Event Stochastic Simulation

Όταν μελετάται ένα σύστημα με τη χρήση της προσομοίωσης, τότε πρέπει κανείς να βασιστεί σε μεθόδους για την πειραματική αναζήτηση αποτελεσμάτων. Αυτές οι μέθοδοι αναζήτησης συχνά δεν είναι τόσο επιτυχημένες όσο άλλες μέθοδοι βελτιστοποίησης. Συχνά ο πειραματιστής απλά θα προσπαθήσει κάποιες εναλλακτικές και απλά θα επιλέξει την καλύτερη από αυτές, μπορεί όμως καμία από την εναλλακτικές να μην είναι βέλτιστη ούτε καν κοντά στην βέλτιστη. Το πόσο κοντά φτάνουμε στη βέλτιστη σε μια μελέτη προσομοίωσης εξαρτάται συχνά από το κατά πόσο έξυπνα ο αναλυτής θα λάβει υπόψη τις εναλλακτικές που θα ερευνηθούν. Λόγω αυτών των πιθανών μειονεκτημάτων, η ανάλυση προσομοίωσης συχνά έχει αναφερθεί ως «τέχνη». Η προσομοίωση έχει βρει κύρια εφαρμογή σε μοντέλα μεταφορών, κατασκευών και συστημάτων επικοινωνιών, τέτοια συστήματα είναι συνήθως στοχαστικά, με διάφορες τυχαίες διαδικασίες που αλληλεπιδρούν με σύνθετους τρόπους.

4.1.1 Στοιχεία Ενός Μοντέλου Προσομοίωσης

Ένα μοντέλο προσομοίωσης αποτελείται από τρία βασικά στοιχεία:

- Επιλογής της κατανομής εισόδου (input modeling) και παραγωγή της
- Bookkeeping
- Ανάλυση εξόδου.

Από τη στιγμή που ενδιαφερόμαστε για μοντελοποίηση στοχαστικών συστημάτων είναι απαραίτητη η επιλογή και μετά η παραγωγή των στοχαστικών φαινομένων στο computer. Για παράδειγμα ένα τηλεπικοινωνιακό κέντρο μπορεί να αποτελείται από ένα δίκτυο ουρών με διάφορες διαφορετικές κατανομές των ενδιάμεσων χρόνων και των χρόνων εξυπηρέτησης. Εμείς πρέπει να αποφασίσουμε με ποια συνάρτηση πιθανότητας επιθυμούμε να αναπαραστήσουμε αυτούς τους μηχανισμούς άφιξης και εξυπηρέτησης (μερικές φορές μπορεί να χρησιμοποιήσουμε μια εμπειρική κατανομή η οποία έχει κατασκευαστεί από πραγματικά δεδομένα που έχουμε συλλέξει). Έτσι τυχαίες μεταβλητές από αυτές τις διαφορετικές κατανομές πρέπει να παραχθούν έτσι ώστε το σύστημα να μπορεί να παρατηρηθεί σε δράση. Όταν γίνει η επιλογή κατανομών και παραχθούν και οι τυχαίες μεταβλητές η φάση bookkeeping παρακολουθεί τις συναλλαγές που κινούνται γύρω από το σύστημα και διατηρεί μετρητές για τις τρέχουσες διαδικασίες προκειμένου να υπολογίσει τα κατάλληλα μέτρα απόδοσης. Η ανάλυση εξόδου σχετίζεται με τον υπολογισμό των μέτρων της αποτελεσματικότητας και

χρησιμοποιεί τις κατάλληλες στατιστικές τεχνικές που απαιτούνται για να κάνει έγκυρες δηλώσεις σχετικά με την απόδοση του συστήματος.

4.1.2 Input Modeling and Random-Variate Generation

Σ' αυτή την ενότητα θα αντιμετωπίσουμε ζητήματα σχετικά με την επιλογή των κατάλληλων κατανομών που εκπροσωπούν τους στοχαστικούς μηχανισμούς του συστήματος (input modeling) και την παραγωγή τυχαίων μεταβλητών των επιλεγμένων κατανομών, και επίσης περιλαμβάνεται μια σύντομη επεξεργασία γεννητριών ψευδο-τυχαίων αριθμών.

- **4.1.2.1 Input Modeling***:* το input modeling δεν είναι κατάλληλο μόνο για την προσομοίωση αλλά είναι απαραίτητο επίσης για κάθε πιθανολογική μοντελοποίηση, συμπεριλαμβανομένων των αναλυτικών και αριθμητικών επεξεργασιών. Είναι πολύ σημαντικό, από τη στιγμή που το αποτέλεσμα (έξοδος) κάθε μοντέλου μπορεί να είναι τόσο καλό όσο η είσοδος (input). Τα δύο βασικά προβλήματα στο input modeling είναι η επιλογή μιας οικογένειας κατανομών (πχ εκθετική, Erlang, κανονική) και από τη στιγμή που η οικογένεια επιλεγεί , η εκτίμηση των παραμέτρων της.
- **4.1.2.2 Εκτίμηση Παραμέτρων***:* Υποθέτουμε ότι έχουμε επιλέξει την οικογένεια κατανομών και έχουμε διαθέσιμα δεδομένα για τις πραγματικές συναλλαγές (ενδιάμεσοι χρόνοι άφιξης ή χρόνοι εξυπηρέτησης). Θεωρούμε ότι έχουμε τυχαίο δείγμα μεγέθους n, έστω $t_1, t_2, \ldots, t_n$. Δύο κλασσικές μέθοδοι είναι η μέθοδος μεγίστης πιθανοφάνειας από την οποία προκύπτουν οι MLEs εκτιμητές και η μέθοδος των moment εκτιμητών MOM.
  Θεωρούμε την περίπτωση όπου πιστεύουμε ότι η υποκείμενη κατανομή είναι εκθετική παραμέτρου θ, και θέλουμε να εκτιμήσουμε το θ από τα δεδομένα του δείγματος. Από τον τύπο της συνάρτησης μεγίστης πιθανοφάνειας και θεωρώντας ότι οι παρατηρήσεις είναι ανεξάρτητες έχουμε:

$$L(\theta) = \prod_{i=1}^{n} \theta e^{-\theta t_i} = \theta^n e^{-\theta \sum_{i=1}^{n} t_i}$$

  Όπου t_i είναι η i-οστη παρατήρηση του δείγματος (ενδιάμεσου χρόνου άφιξης ή χρόνου εξυπηρέτησης). Ο MLE του θ είναι αυτή η τιμή που μεγιστοποιεί την L. Είναι συχνά βολικό να βρίσκουμε το μέγιστο της lnL=Λ. Έτσι $\hat{\theta}$ είναι MLE του θ είναι η τιμή για την οποία επιτυγχάνεται η:

$$\max \Lambda(\theta) = \max(n - ln\theta - \theta \sum_{i=1}^{n} t_i)$$

Παίρνοντας dΛ/dθ=0 έχουμε:

$$\frac{n}{\hat{\theta}} - \sum_{i=1}^{n} t_i = 0 \quad => \quad \hat{\theta} = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{1}{\bar{t}}$$

Όπου: $\bar{t} = \frac{\sum_{i=1}^{n} t_i}{n}$ είναι ο δειγματικός μέσος.

Στη συνέχεια θα βρούμε τους MLE και MOM εκτιμητές για δύο παραμέτρους που ακολουθούν την Erlang.

Έχουμε:

$$f(t) = \frac{\varphi(\varphi t)^{k-1} e^{-\varphi t}}{(k-1)!}$$

η συνάρτηση πιθανοφάνειας είναι:

$$L(\varphi, k) = \prod_{i=1}^{n} \frac{\varphi(\varphi t_i)^{k-1} e^{-\varphi t_i}}{(k-1)!}$$

Και:

$$\Lambda(\varphi, k) = nk ln\varphi - \varphi \sum_{i=1}^{n} t_i + (k-1) \sum_{i=1}^{n} lnt_i - nln(k-1)!$$

$$\frac{\partial \Lambda}{\partial \varphi} = \frac{nk}{\varphi} - \sum_{i=1}^{n} t_i \quad => \quad \hat{\varphi} = \frac{\hat{k}}{\bar{t}}$$

Όπου $\bar{t}$ ο δειγματικός μέσος.

Τώρα για να πάρουμε το ζευγάρι $(\hat{\varphi}, \hat{k})$, θεωρούμε ότι το k είναι μια συνεχής μεταβλητή (έστω χ) , θα βρούμε τον MLE $\hat{\chi}$ του χ ως εξής:

$$\frac{\partial \Lambda}{\partial x} = 0 = nln\hat{\varphi} + \sum_{i=1}^{n} lnt_i - n\psi(\hat{x}) = n(ln\hat{x} - lnt) + \sum_{i=1}^{n} lnt_i - n\psi(\hat{x})$$

$$\text{όπου } \psi(\chi) \equiv dln\Gamma(x)/dx$$

Μια καλή προσέγγιση της ψ(χ) όταν το χ είναι αρκετά μικρό είναι:

$$\psi(x) \approx \ln\left(x - \frac{1}{2}\right) + \frac{1}{24(x - \frac{1}{2})^2}$$

Επομένως ο MLE του κ είναι είτε $[\hat{x}]$ ή $[\hat{x}] + 1$(όπου [χ] ο μεγαλύτερος ακέραιος του χ) εξαρτάται από ποιο ζεύγος $(\frac{[\hat{x}]}{\bar{t}}, [\hat{x}])$ ή $(\frac{[\hat{x}]+1}{\bar{t}}, [\hat{x}] + 1)$ δίνει τη μεγαλύτερη τιμή στον λογάριθμο της πιθανοφάνειας.

Στην Erlang η περίπτωση MOM είναι αρκετά πιο απλή:

Από τη στιγμή που έχουμε δυο παραμέτρους, δυο εξισώσεις χρειάζονται:

$$\bar{t} = \frac{\tilde{k}}{\tilde{\varphi}} \quad \kappa\alpha\iota \quad s^2 = \frac{\tilde{k}}{\tilde{\varphi}^2}$$

Όπου s^2 η διασπορά του δείγματος. Έτσι έχουμε τους MOM εκτιμητές:

$$\tilde{\varphi} = \frac{\bar{t}}{s^2} \quad \kappa\alpha\iota \quad \tilde{k} = \left[\frac{\bar{t}^2}{s^2}\right] \quad \text{ή} \quad \left[\frac{\bar{t}^2}{s^2}\right] + 1$$

Για την περίπτωση της Erlang (και για τις περισσότερες κατανομές), οι MOM και ML δίνουν διαφορετικούς εκτιμητές για τις παραμέτρους. Αν η μέση τιμή και η διασπορά με τη χρήση του MLE δίνει πολύ διαφορετική μέση τιμή και διασπορά από αυτές των δεδομένων του δείγματος, θα πρέπει να είμαστε διστακτικοί με τη χρήση του MLE και να προχωρήσουμε με τους MOM.

- **4.1.2.3 Επιλογή της Συνάρτησης Κατανομής**: Το πιο δύσκολο αλλά και πιο σημαντικό θέμα είναι η επιλογή της οικογένειας κατανομών που αντιπροσωπεύει τις κατανομές εισόδων (ενδιάμεσοι χρόνοι και χρόνοι εξυπηρέτησης). Η επιλογή των κατάλληλων υποψήφιων κατανομών πιθανότητας εξαρτάται από τη γνώση όσων περισσότερων γίνεται για τα χαρακτηριστικά των ενδεχόμενων κατανομών και για τη φυσική της κατάστασης που θα μοντελοποιηθεί. Γενικά έχουμε πρώτα να αποφασίσουμε ποιες συναρτήσεις πιθανότητας είναι κατάλληλες για χρήση για τις διαδικασίες άφιξης και εξυπηρέτησης. Υποθέτουμε ότι επιθυμούμε να επιλέξουμε συνάρτηση κατανομής μιας συνεχούς τυχαίας μεταβλητής T (ενδιάμεσου χρόνου ή χρόνου εξυπηρέτησης) με CDF F(t). Η συνάρτηση πυκνότητας $f(t) = dF(t)/dt$ μπορεί να ερμηνευθεί ως εξής:

$$f(t)dt \approx \Pr\{t \le T \le t + dt\}$$

αυτή είναι η προσεγγιστική πιθανότητα ότι ο τυχαίος χρόνος θα είναι σε μια γειτονιά του t. Η CDF F(t) είναι η πιθανότητα ο χρόνος να είναι μικρότερος ή ίσος του t. Έτσι έχουμε:

$$h(t)dt \approx \Pr\{t \le T \le t + dt | T \ge t\}$$

αυτή είναι η προσεγγιστική πιθανότητα ότι ο χρόνος θα είναι σε μια γειτονιά του t, δεδομένου ότι ο χρόνος είναι ήδη t. Για παράδειγμα στους ενδιάμεσους χρόνους αφίξεων είναι η προσεγγιστική πιθανότητα μια άφιξη να συμβεί σ' ένα ενδιάμεσο dt δεδομένου ότι ήταν t από τη στιγμή της τελευταίας άφιξης. Στους χρόνους εξυπηρέτησης h(t) είναι η προσεγγιστική πιθανότητα ένας πελάτης να εξυπηρετηθεί σε dt δεδομένου ότι ο πελάτης έχει ήση μπει για εξυπηρέτησης για ένα χρονικό διάστημα t. Έχουμε:

$$h(t) \approx \Pr\{t \le T \le t + dt | T \ge t\} = \frac{\Pr\{t \le T \le t + dt \text{ και } T \ge t\}}{\Pr\{T > t\}} = \frac{f(t)dt}{1 - F(t)}$$

Επομένως:

$$h(t) = \frac{f(t)}{1 - F(t)}$$

Αυτός ο κίνδυνος ή αλλιώς συνάρτηση ρυθμού αποτυχίας μπορεί να αυξάνεται στο t (IFR), να μειώνεται στο t(DFR) να είναι σταθερός (θεωρώντας να είναι και τα δύο IFR και DFR).
Αν πιστεύουμε ότι η εξυπηρέτηση είναι αρκετά σταθερή ώστε όσο περισσότερο είναι ο πελάτης στην εξυπηρέτηση το πιο πιθανό είναι η εξυπηρέτηση να ολοκληρωθεί στο επόμενο dt, έτσι επιθυμούμε μια f(t) τέτοια ώστε h(t) αυξάνεται στο t, που είναι μια IFR κατανομή. Από την παραπάνω σχέση μπορούμε να βρούμε την h(t) από την f(t). Έτσι ο ρυθμός κινδύνου είναι ακόμα μια σημαντική πηγή (όπως και το σχήμα της f(t)) για να λάβουμε γνώση θεωρώντας υποψήφιες f(t) που μπορούν να μελετηθούν στην μοντελοποίηση προτύπων αφίξεων και εξυπηρέτησης.
Θεωρούμε την εκθετική κατανομή και επιθυμούμε να βρούμε το h(t), έχουμε:

$$f(t) = \theta e^{-\theta t}$$

Αντικαθιστώ και έχω:

$$h(t) = \frac{\theta e^{-\theta t}}{e^{-\theta t}} = \theta$$

έτσι η εκθετική κατανομή έχει σταθερό ρυθμό αποτυχίας (κίνδυνο).

Υποθέτουμε ότι διαισθανόμαστε ότι για μια συγκεκριμένη κατάσταση ουράς χρειαζόμαστε μια IFR κατανομή για να περιγράψει τους χρόνους εξυπηρέτησης. Προκύπτει ότι η Erlang (k>1) έχει αυτή την ιδιότητα. Η συνάρτηση πυκνότητας της με θ=kμ είναι:

$$f(t) = \frac{\theta^k t^{k-1} e^{-\theta t}}{(k-1)!}$$

Και:

$$F(t) = \frac{\theta^k}{(k-1)!} \int_0^t e^{-\theta t} x^{k-1} dx = 1 - \sum_{i=0}^{k-1} \frac{(\theta t)^i e^{-\theta t}}{i!}$$

Έτσι:

$$h(t) = \frac{\theta^k t^{k-1} e^{-\theta t}}{(k-1)! \sum_{i=0}^{k-1} \frac{(\theta t)^i e^{-\theta t}}{i!}} = \frac{\theta (\theta t)^{k-1}}{(k-1)! \sum_{i=0}^{k-1} \frac{(\theta t)^i}{i!}}$$

Η h(t) μπορεί να γραφεί:

$$h(t) = \frac{1}{\int_0^{\infty} (1 - \frac{u}{t})^{k-1} e^{-\theta u} du}$$

έτσι είναι πιο εύκολο να δούμε ότι για k>1 όσο το t αυξάνεται το ολοκλήρωμα στον παρανομαστή μειώνεται , και έτσι μειώνεται ο παρανομαστής και επομένως αυξάνεται το h(t) όσο αυξάνεται το t (IFR).

Επιπλέον έχει μια ασύμπτωτη στο θ όσο το t τείνει στο άπειρο και h(0)=0. Από τη στιγμή που η h(t) έχει ασύμπτωτη ακόμα και αν η h(t) αυξάνεται με το t, το κάνει συνέχεια με έναν πιο αργό ρυθμό και σταδιακά προσεγγίζει την σταθερά θ.

Αν υποθέσουμε ότι επιθυμούμε μια αντίθετη IFR συνθήκη, αυτή είναι να επιταχύνουμε τον ρυθμό αύξησης σε σχέση με το t. Αυτή η κατανομή καλείται Weibull (με $1 - F(t) = e^{-\theta t^{\alpha}}$) για την οποία μπορούμε να πάρουμε αυτές τις συνθήκες.

Επομένως φτάνουμε πάλι σ' αυτό που είχαμε ήδη αναφέρει, ότι η επιλογή του καλύτερου μοντέλου είναι ένας συνδυασμός γνώσης όσο περισσοτέρων γίνεται από τα χαρακτηριστικά της κατανομής πιθανότητας της οικογένειας και της γνώσης όσο περισσοτέρων για την πραγματική κατάσταση που μοντελοποιείται. Στις περισσότερες περιπτώσεις είναι διαθέσιμα ή μπορούν να συλλεχθούν δεδομένα της διαδικασίας που θέλουμε να μοντελοποιήσουμε, αυτά τα δεδομένα θα μας βοηθήσουν να βρούμε την κατάλληλη οικογένεια κατανομών.

Έστω ότι έχουμε παρατηρήσεις διαθέσιμες για τους ενδιάμεσους χρόνους άφιξης και για τους χρόνους εξυπηρέτησης. Επίσης υποθέτουμε ότι τα δεδομένα προέρχονται από μια ομοιογενή χρονική περίοδο(πχ η διαδικασία από την οποία λήφθηκαν τα δεδομένα δεν αλλάζει με την αλλαγή του χρόνου-έστω ένα σύστημα κατά τη διάρκεια ωρών αιχμής της κυκλοφορίας). Υποθέτουμε ότι όλες οι παρατηρήσεις είναι ανεξάρτητες . Είναι πολύ σημαντικό να ελέγξουμε τις παρατηρήσεις αν ικανοποιούν αυτές τις συνθήκες (IID).

Υποθέτουμε έχουμε δεδομένα από ένα IID δείγμα. Μερικά από τα πρώτα πράγματα που μπορούμε να κάνουμε στην προσπάθεια μας να αποφασίσουμε

ποια οικογένεια κατανομών είναι κατάλληλη είναι να υπολογίσουμε των δειγματικό μέσο, την δειγματική τυπική απόκλιση και το δειγματικό C. Ένα ιστόγραμμα παρατηρήσεων είναι πολύ χρήσιμο παρότι το σχήμα του ιστογράμματος που προκύπτει εξαρτάται από το τη μήκος του διαστήματος που χρησιμοποιούνται για την συσσώρευση των συχνοτήτων(ο αριθμός των παρατηρήσεων που πέφτει σε κάθε διάστημα).

Υποθέτουμε ότι αφού κοιτάξουμε τα ιστογράμματα και λάβουμε υπόψη τα φυσικά χαρακτηριστικά του συστήματος που μοντελοποιούμε, όπως και τα χαρακτηριστικά της οικογένειας κατανομών επιλέγουμε μια ενδεχόμενη υποψήφια συνάρτηση. Υπάρχει μια ποικιλία στατιστικών tests που μπορούμε να κάνουμε για να δούμε αν η υποψήφια κατανομή είναι μια λογική επιλογή.

Το πιο συνηθισμένο στη χρήση (αλλά όχι πάντα το καλύτερο) test είναι το x^2 goodness-of-fit tests. Η στατιστική του x^2 είναι:

$$x_k^2 = \sum_{i=1}^{n} \frac{(o_i - c_i)^2}{e_i}$$

Όπου o_i η παρατηρηθείσα τιμή της συχνότητας της i, e_i η αναμενόμενη τιμή της συχνότητας της i, αν η κατανομή που έχουμε υποθέσει είναι σωστή, και k οι βαθμοί ελευθερίας, πάντα ίσοι με των συνολικό αριθμό κατηγοριών μείων ένα και μείων κάθε παράμετρο που εκτιμάται. Για το e_i ολοκληρώνουμε την θεωρητική κατανομή στο διάστημα, πχ αν i^- το κατώτερο όριο για το i διάστημα και i^+ το ανώτερο, τότε αν για την εκθετική κατανομή έχουμε:

$$e_i = n \int_{i^-}^{i^+} \theta e^{-\theta t} dt = n(e^{-\theta i^-} - e^{-\theta i^+})$$

Όπου το θ αντικαθίσταται από το MLE $\hat{\theta}$ και n είναι το μέγεθος του δείγματος. Κάνουμε το test και αναλόγως αποδεχόμαστε ή απορρίπτουμε την υπόθεση που έχουμε κάνει για την κατανομή. Η βασική αδυναμία του x^2 test είναι απαίτησή του για μεγάλα δείγματα.

Ένα άλλο δημοφιλές goodness-of-fit test είναι το Kolmogorov-Smirnov (KS) test, το οποίο συγκρίνει τις αποκλίσεις της εμπειρικής συνάρτησης κατανομής από την θεωρητική συνάρτηση κατανομής και χρησιμοποιεί την εξής στατιστική:

$$K = \max_{\mathrm{j}} \max \left\{ \left| \frac{j}{n} - F(t_j) \right|, \left| \frac{j-1}{n} - F(t_j) \right| \right\}$$

Όπου t_j είναι η j-κοστή διατεταγμένη (αύξουσα) παρατήρηση και $F(t_j)$ είναι η τιμή της συνάρτησης κατανομής που έχουμε υποθέσει της j παρατήρησης. Δυστυχώς όπως βλέπουμε το συγκεκριμένο τεστ προϋποθέτει ότι η αθροιστική συνάρτηση κατανομής F είναι εντελώς γνωστή.

Παρόλο που τα goodness-of –fit tests απαιτούν μεγάλα μεγέθη δείγματος για να κάνουν διάκριση και πολλά είναι περιορισμένα αν έχουμε να εκτιμήσουμε παραμέτρους από το δείγμα, υπάρχουν ειδικά tests για την εκθετική κατανομή. Ένα πολύ ισχυρό τεστ (ισχυρό σε σχέση με την ικανότητα να διακρίνει την λάθος υπόθεση) για την εκθετική κατανομή είναι το F test. Για να κάνουμε ένα F test r(≈n/2) και n-r ενός σετ n ενδιάμεσων χρόνων t_i τυχαία διατεταγμένων. Έτσι η ποσότητα:

$$F = \frac{\sum_{i=1}^{r} \frac{t_i}{r}}{\sum_{i=r+1}^{n} \frac{t_i}{(n-r)}}$$

Είναι ο λόγος δυο Erlangs κατανομών και κατανέμεται σαν μια κατανομή F με 2r και 2(n-r) βαθμούς ελευθερίας όταν η υπόθεση της εκθετικής κατανομής είναι αληθής.

- **4.1.2.4 Genaration of Random Variates**: όταν επιλεχθούν οι κατάλληλες κατανομές πιθανότητας που αντιστοιχούν στις διαδικασίες εισόδων (ενδιάμεσοι χρόνοι αφίξεων και χρόνοι εξυπηρέτησης) είναι απαραίτητο να παράγουμε τυπικές παρατηρήσεις από αυτές για να τρέξουμε το σύστημα προσομοίωσης. Η διαδικασία παραγωγής IID τυχαίων μεταβλητών από μία δεσομένη συνάρτηση κατανομής έστω f(x) γενικά συνίσταται από δύο φράσεις:
(1) παραγωγή ψευδοτυχαίων αριθμών που κατανέμονται ομοιόμορφα στο (0,1) και
(2) χρησιμοποιώντας τους ψευδοτυχάιους αριθμούς να αποκτήσουμε μεταβλητές (παρατηρήσεις) από την f(x).
Οι περισσότερες γλώσσες προγραμματισμού περιλαμβάνουν μια γεννήτρια ψευδοτυχαίων αριθμών. Είναι <<ψευδο>> επειδή είναι αναπαραγόμενοι από έναν μαθηματικό αλγόριθμο, αλλά τυχαίοι με την έννοια ότι έχουν περάσει στατιστικά tests με ίδια πιθανότητα όλων των τιμών και στατιστικής ανεξαρτησίας. Οι περισσότερες computer routines βασίζονται σ' ένα αναδρομικό αλγόριθμο της μορφής:

$$r_{n+1} = (kr_n + a) mod m$$

Όπου k, α και m είναι θετικοί ακέραιοι (k<m, α<m). Έτσι r_{n+1} το υπόλοιπο της διαίρεσης του $kr_n + a$ με το m. Πρέπει να επιλέξουμε μια αρχική τιμή r_0 η οποία καλείται σπόρος και πρέπει να είναι μικρότερος του m.

Στη συνέχεια επιθυμούμε να παράγουμε αντιπροσωπευτικές παρατηρήσεις από μια συγκεκριμένη συνάρτηση κατανομής με αθροιστική συνάρτηση κατανομής έστω F(x). Η βασική μέθοδος που θα ασχοληθούμε συχνά αναφέρεται σαν αντίστροφη μέθοδος ή μέθοδος μετασχηματισμού πιθανότητας ή παραγωγή με αντιστροφή. Μπορεί να περιγραφεί γραφικά θεωρώντας τη γραφική παράσταση μιας αθροιστικής συνάρτησης κατανομής από την οποία θέλουμε να παράγουμε τυχαίες μεταβλητές. Η διαδικασία είναι πρώτα να παράγουμε ομοιόμορφα στο (0,1) τυχαίες μεταβλητές, έστω $r_1, r_2, \ldots$ Για να αποκτήσουμε την χ_1 την πρώτη τυχαία μεταβλητή που αντιστοιχεί στην F(x) γραφικά από το r_1 φέρνουμε παράλληλη στον άξονα των χ και παίρνουμε την τετμημένη του σημείου τομής της παράλληλης με την F(x). Κάνουμε την ίδια διαδικασία με τα $r_2, r_3, \ldots$ και παράγουμε τα $x_2, x_3, \ldots$. Για να αποδείξουμε ότι αυτή η διαδικασία δουλεύει θέλουμε να δείξουμε ότι η τυχαία μεταβλητή έστω X_i που έχει παραχθεί από αυτή τη διαδικασία υπακούει στην συνθήκη: $P(X_i \leq x) = F(x)$.
Έχουμε θεωρήσει ότι:

$$P(X_i \leq x) = P\big(R_i \leq F(x)\big)$$

Από τη στιγμή που R_i ομοιόμορφη (0,1), μπορούμε να γράψουμε:

$$P\big(R_i \leq F(x)\big) = F(x)$$

έτσι:

$$P(X_i \leq x) = F(x)$$

Για κάποιες θεωρητικές κατανομές, η αντιστροφή μπορεί να αποκτηθεί αναλυτικά σε κλειστή μορφή. Για παράδειγμα ας θεωρήσουμε την εκθετική κατανομή παραμέτρου θ. Η αθροιστική συνάρτηση κατανομής της είναι η:

$$F(x) = 1 - e^{-\theta x}, x \geq 0$$

Έχουμε:

$$r = 1 - e^{-\theta x} => \ e^{-\theta x} = 1 - r$$

Από τη στιγμή που η r είναι ομοιόμορφη στο (0,1) είναι ασήμαντο αν χρησιμοποιήσουμε r ή r-1 σαν τον τυχαίο μας αριθμό, έτσι μπορούμε να γράψουμε: $e^{-\theta x} = r$. Και παίρνοντας λογαρίθμους έχουμε:

$$x = \frac{-lnr}{\theta}$$

Δυστυχώς η αναλυτική αντιστροφή δεν είναι δυνατή για όλες τις συναρτήσεις

κατανομής. Στην περίπτωση κάποιων συνεχών κατανομών, μπορούμε να βρούμε εναλλακτικούς τρόπους για να παράγουμε μεταβλητές. Μπορούμε πχ να εξετάσουμε την κατανομή Erlang και αντί να επιχειρήσουμε αντιστροφή της αθροιστικής συνάρτησης κατανομής της Erlang μπορούμε απλώς να πάρουμε άθροισμα των k εκθετικών τυχαίων μεταβλητών. Έτσι αν επιθυμούμε να παράγουμε τυχαίες μεταβλητές από μια Erlang τύπου k, με μέση τιμή 1/μ μπορούμε να πάρουμε αυτόν τον τύπο τυχαίας μεταβλητής έστω χ από τις ομοιόμορφες στο (0,1) τυχαίες μεταβλητές $r_1,r_2,\ldots,r_k$ από την:

$$x = \sum_{i=1}^{k}(-\frac{lnr_i}{k\mu}) = -\frac{ln\prod_{i=1}^{k} r_i}{k\mu}$$

4.1.3 Bookkeeping Aspects of Simulation Analysis

Όπως αναφέραμε νωρίτερα η φάση bookkeeping ενός μοντέλου προσομοίωσης πρέπει να παρακολουθεί τις συναλλαγές που γίνονται στο σύστημα και να στήσει μετρητές στις εξελισσόμενες διαδικασίες προκειμένου να υπολογίσει διάφορα μεγέθη επίδοσης του συστήματος. Ο σχεδιαστής του μοντέλου προσομοίωσης έχει μια μεγάλη ποικιλία γλωσσών και πακέτων προγραμματισμού από τα οποία μπορεί να διαλέξει. Αυτά μπορεί να είναι γενικής χρήσης γλώσσες προγραμματισμού όπως οι C++, Java και Visual Basic που επιτρέπουν μεγαλύτερη ευελιξία στη μοντελοποίηση, αλλά απαιτούν περισσότερη προσπάθεια ή διάφορα άλλα πακέτα γλωσσών προσομοίωσης.

4.1.4 Output Analysis

Η επίτευξη αξιόπιστων συμπερασμάτων από τα αποτελέσματα της προσομοίωσης απαιτεί καλό συνδυασμό σκέψης και προσοχής. Όταν προσομοιάζουμε στοχαστικά συστήματα ένα και μόνο τρέξιμο αποδίδει τιμές εξόδου που είναι στατιστικές στη φύση έτσι πολύ καλός πειραματικός σχεδιασμός και στατιστική ανάλυση είναι απαραίτητα για έγκυρα συμπεράσματα. Σε αντίθεση με τη δειγματοληψία από ένα πληθυσμό με την κλασσική έννοια όπου μεγάλη προσπάθεια γίνεται ώστε να έχουμε τυχαία δείγματα με ανεξάρτητες παρατηρήσεις, συχνά σκοπίμως προκαλούμε συσχέτιση στην μοντελοποίηση προσομοίωσης σαν μέθοδο μείωσης της διασποράς, έτσι κλασσικές έτοιμες στατιστικές τεχνικές για ανάλυση του δείγματος συχνά δεν είναι κατάλληλες. Στη συνέχεια θα παρουσιάσουμε κάποιες βασικές διαδικασίες για την ανάλυση των τιμών εξόδου της προσομοίωσης.

Υπάρχουν δύο ειδών μοντέλα προσομοίωσης: αυτά που τερματίζουν (terminating) και τα συνεχή (continuing). Ένα terminating μοντέλο έχει μια φυσική ώρα έναρξης και

τερματισμού, για παράδειγμα μια τράπεζα ανοίγει στις 08:00 και κλείνει στις 14:30. Ένα μοντέλο continuing δεν έχει ώρα έναρξης και λήξης, για παράδειγμα μια διαδικασία κατασκευής που στην αρχή της περιόδου εργασίας τα πράγματα παίρνονται ακριβώς όπως είχαν αφεθεί στην προηγούμενη περίοδο εργασίας έτσι κατά μια έννοια η διαδικασία τρέχει συνεχώς.

Θεωρούμε αρχικά μια terminating προσομοίωση, από ένα τρέξιμο δεν μπορούμε να κάνουμε στατιστικές δηλώσεις. Για παράδειγμα ο μέγιστος χρόνος αναμονής είναι μία μόνο παρατήρηση. Αυτό που μπορούμε να κάνουμε είναι να επαναλάβουμε το πείραμα (επαναλαμβανόμενα τρεξίματα) χρησιμοποιώντας διαφορετικό τυχαίο αριθμό streams για την εντολή άφιξης και τους χρόνους επεξεργασίας για κάθε τρέξιμο και έτσι παράγουμε ένα δείγμα ανεξάρτητων παρατηρήσεων στον οποίο η κλασσική στατιστική μπορεί να εφαρμοστεί. Υποθέτουμε ότι έχουμε επαναλάβει n φορές, τότε έχουμε n τιμές για τον μέγιστο χρόνο αναμονής έστω $w_1,w_2,\ldots,w_n$. Υποθέτουμε ότι το n είναι αρκετά μεγάλο ώστε να χρησιμοποιηθεί το κεντρικό οριακό θεώρημα, μπορούμε να πάρουμε ένα 100(1-α)% διάστημα εμπιστοσύνης υπολογίζοντας τον δειγματικό μέσο και την δειγματική τυπική απόκλιση ως εξής:

$$\overline{w} = \frac{\sum_{i=1}^{n} w_i}{n}$$

$$s_w = \sqrt{\frac{\sum_{i=1}^{n}(w_i - \overline{w})^2}{n-1}}$$

Και έτσι παίρνουμε το διάστημα εμπιστοσύνης:

$$[\overline{w} - \frac{t\left(n-1,1-\frac{\alpha}{2}\right)s_w}{\sqrt{n}}, \overline{w} + \frac{t\left(n-1,1-\frac{\alpha}{2}\right)s_w}{\sqrt{n}}]$$

Όπου t(n-1,1-α/2) είναι η μεγαλύτερη 1-α/2 κρίσιμη τιμή της t κατανομής με n-1 βαθμούς ελευθερίας.

Για την περίπτωση της continuing προσομοίωσης στην οποία μας ενδιαφέρουν αποτελέσματα σε κατάσταση ισορροπίας γνωρίζουμε από το εργοδικό θεώρημα μιας στοχαστικής διαδικασίας ότι:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T X^n(t)dt = E[X^n]$$

Έτσι αν το τρέξουμε μεγάλο χρονικό διάστημα θα πρέπει να φτάσουμε κοντά στην οριακή μέση τιμή. Αλλά δεν είναι ξεκάθαρο πόσο χρονικό διάστημα είναι μεγάλο χρονικό διάστημα και συχνά θα θέλαμε να είμαστε σε θέση να βρούμε ένα διάστημα εμπιστοσύνης. Έτσι έχουμε δύο παραπάνω προβλήματα: να καθορίσουμε πότε φτάνουμε σε κατάσταση στατιστικής ισορροπίας και να αποφασίσουμε πότε να τερματίσουμε το τρέξιμο της προσομοίωσης. Υποθέτουμε για την ώρα ότι αυτά τα προβλήματα έχουν λυθεί και αποφασίζουμε να τρέξουμε την προσομοίωση για n συναλλαγές αφού φτάσει στην κατάσταση στατιστικής ισορροπίας και μετράμε το χρόνο που ένας πελάτης ξοδεύει περιμένοντας σε μια συγκεκριμένη ουρά για εξυπηρέτηση, μπορούμε να πάρουμε n τιμές αναμονής στην ουρά που τις συμβολίζουμε πάλι με w_i (τώρα αυτές είναι πραγματικές αναμονές και όχι μέγιστες αναμονές). Θα ήταν δελεαστικό να υπολογίζαμε μέση τιμή και τυπική απόκλιση και να παίρναμε διάστημα εμπιστοσύνης, όμως αυτά τα w_i είναι συσχετισμένα και αν χρησιμοποιήσουμε τον τύπο του s_w σε μεγάλο βαθμό υποεκτιμάται η πραγματική διακύμανση.

Για να ξεφύγουμε από το πρόβλημα της συσχέτισης μπορούμε να επαναλάβουμε το τρέξιμο m φορές χρησιμοποιώντας διαφορετικό τυχαίο αριθμό σπόρου κάθε φορά όπως κάναμε στην περίπτωση terminating. Για κάθε τρέξιμο υπολογίζουμε τον μέσο των w_i ονομάζοντας τον μέσο για την j επανάληψη $\overline{w_j}$ έτσι:

$$\overline{w_j} = \frac{\sum_{i=1}^{n} w_{ij}}{n}$$

Όπου w_{ij} είναι ο χρόνος αναμονής της συναλλαγής i στην επανάληψη j, i=1,2,…,n και j=1,2,…,m. Οι $\overline{w_j}$ είναι ανεξάρτητες και μπορούμε να βρούμε ένα 100(1-α)% διάστημα εμπιστοσύνης υπολογίζοντας τα:

$$\overline{w} = \frac{\sum_{j=1}^{m} \overline{w_j}}{m}$$

$$s_{\overline{w_j}} = \sqrt{\frac{\sum_{j=1}^{m} (\overline{w_j} - \overline{w})^2}{m-1}}$$

Και έτσι το διάστημα εμπιστοσύνης:

$$[\overline{w} - \frac{t\left(m-1,1-\frac{\alpha}{2}\right) s_{\overline{w_j}}}{\sqrt{m}}, \overline{w} + \frac{t\left(m-1,1-\frac{\alpha}{2}\right) s_{\overline{w_j}}}{\sqrt{m}}$$

Θα επανέλθουμε στα δύο προβλήματα που αναφέραμε νωρίτερα για continuing προσομοίωση. Το μήκος τρεξίματος n και ο αριθμός των επαναλήψεων m και τα δυο επηρεάζουν το τυπικό σφάλμα $s_{\overline{w_j}}/\sqrt{m}$ που απαιτείται για να φτιάξουμε το παραπάνω διάστημα εμπιστοσύνης. Όσο πιο μικρό είναι το τυπικό σφάλμα τόσο πιο ακριβές είναι το διάστημα εμπιστοσύνης για δεδομένη εμπιστοσύνη (1-α). Επίσης από τον τύπο του τυπικού σφάλματος έχουμε ότι όσες περισσότερες επαναλήψεις κάνουμε το τυπικό σφάλμα μικραίνει επομένως το διάστημα εμπιστοσύνης είναι πιο ακριβές. Επίσης είναι αναμενόμενο όσο μεγαλώνει το μήκος n τρεξίματος η υπολογιζόμενη τιμή $s_{\overline{w_j}}$ θα γίνεται μικρότερη για δεδομένο αριθμό επαναλήψεων, έτσι μακρυά μήκη τρεξίματος επίσης αυξάνουν την ακρίβεια του διαστήματος εμπιστοσύνης.

Η περίοδος προθέρμανσης (το διάστημα του χρόνου που απαιτείται να φτάσει η διαδικασία κοντά σε κατάσταση στατιστικής ισορροπίας) δεν είναι εύκολο θέμα να καθοριστεί. Συχνά στην εφαρμογή διαισθητικά αγνοείται με την ελπίδα ότι υπάρχουν αρκετά σημεία δεδομένων στο τρέξιμο έτσι ώστε οι παροδικές επιδράσεις δεν υπάρχουν από το τμήμα των δεδομένων που παίρνεται αφού έχουν επιτευχθεί συνθήκες κοντά στην κατάσταση στατιστικής ισορροπίας. Ο πιο συνηθισμένη και πιο επιστημονική προσέγγιση είναι να χωρίσουμε το τρέξιμο σε δυο τμήματα: μια παροδική περίοδο και μια περίοδο στατιστικής ισορροπίας. Η βασική ιδέα είναι αν μπορούσαμε να υπολογίσουμε το χρόνο ή τον αριθμό των συναλλαγών που χρειάζονται μέχρι να φτάσουμε κοντά στην στατιστική ισορροπία απλά δεν θα αρχίζαμε να καταγράφουμε δεδομένα για τον υπολογισμό μέτρων εξόδου μέχρι το κεντρικό ρολόι να περάσει αυτό το σημείο. Το να βρούμε που βρίσκεται αυτό το σημείο είναι μια από τις πιο δύσκολες εφαρμογές στην ανάλυση εξόδου και έχει αναπτυχθεί μια ποικιλία μεθόδων.

Συγκρίνοντας δυο εναλλακτικά σχέδια συστήματος η τεχνική που πιο συχνά χρησιμοποιείται είναι paired t διάστημα εμπιστοσύνης ενός δεδομένου μέτρου επίδοσης για κάθε σχέδιο. Για παράδειγμα αν προσομοιώνουμε ένα σύστημα ουράς και το ένα σχέδιο έχει δυο εξυπηρετητές που εξυπηρετούν με συγκεκριμένο ρυθμό σ' ένα σταθμό εξυπηρέτησης και το ανταγωνιστικό σχέδιο αντικαθιστά τους δυο εξυπηρετητές με αυτόματες μηχανές, τότε ίσως μας ενδιαφέρει ο μέσος χρόνος συναλλαγής. Τρέχουμε το σχέδιο 1 και υπολογίζουμε το μέσο χρόνο, τρέχουμε και το 2 και υπολογίζουμε και τον δικό του χρόνο συναλλαγής, και υπολογίζουμε τη διαφορά τους. Τότε επαναλαμβάνουμε το ζεύγος τρεξίματος m φορές βρίσκοντας m διαφορές d_i, i=1,…,m. Υπολογίσουμε τον μέσο και την τυπική απόκλιση και έτσι έχουμε ένα 100(1-α)% διάστημα εμπιστοσύνης:

$$[\bar{d} - \frac{t\left(m-1{,}1-\frac{a}{2}\right)s_{d_i}}{\sqrt{m}}, \bar{d} + \frac{t\left(m-1{,}1-\frac{a}{2}\right)s_{d_i}}{\sqrt{m}}]$$

Οποτεδήποτε είναι δυνατό το ίδιο stream τυχαίων αριθμών θα πρέπει να χρησιμοποιηθεί για κάθε σχέδιο εντός μιας αντιγραφής έτσι ώστε η διαφορά που παρατηρείται εξαρτάται μόνο από την αλλαγή του σχεδιασμού παραμέτρων και όχι της διακύμανσης λόγω της τυχαιότητας των τυχαίων μεταβλητών που έχουν παραχθεί. Φυσικά διαφορετικά streams τυχαίων αριθμών χρησιμοποιούνται μεταξύ των αντιγραφών .

4.1.5 Model Validation

Η επικύρωση του μοντέλου είναι πολύ σημαντικό βήμα σε μια μελέτη προσομοίωσης που συχνά παραλείπεται από τους σχεδιαστές. Πριν ξεκινήσει την ανάπτυξη ενός μοντέλου προσομοίωσης επιβάλλεται στον αναλυτή να γίνει πολύ οικείος με το σύστημα που μελετάται, για τη συμμετοχή των στελεχών και του προσωπικού του λειτουργικού συστήματος και, συνεπώς, να συμφωνήσουν σχετικά με το επίπεδο λεπτομερειών που απαιτούνται για την επίτευξη του στόχου της μελέτης. Ένα πρόβλημα με την μοντελοποίηση προσομοίωσης είναι ότι από τη στιγμή που κάθε επίπεδο λεπτομέρειας μπορεί να μοντελοποιηθεί, τα μοντέλα που αναπτύσσονται έχουν περισσότερες λεπτομέρειες απ’ όσες χρειάζονται και αυτό μπορεί να είναι μη αποδοτικό και αντιπαραγωγικό.

Η εγκυρότητα συνδέεται στενά με τον έλεγχο και την αξιοπιστία. Ο έλεγχος έχει να κάνει με το program debugging να διασφαλίσει ότι το πρόγραμμα κάνει αυτό για το οποίο προορίζεται. Αυτό είναι γενικά ο πιο απλός από τους τρεις στόχους να επιτευχθεί, καθώς υπάρχουν γνωστές και καθιερωμένες μέθοδοι για το debugging computer programs. Η εγκυρότητα ασχολείται με το πόσο ακριβής είναι η παρουσίαση της πραγματικότητας που παρέχει το μοντέλο και η αξιοπιστία ασχολείται με το πόσο πιστευτό είναι το μοντέλο στους χρήστες. Για να επιτύχουν εγκυρότητα και αξιοπιστία οι χρήστες πρέπει να συμμετέχουν στην μελέτη νωρίς και συχνά. Οι στόχοι της μελέτης, τα κατάλληλα μέτρα επίδοσης και το επίπεδο των πληροφοριών πρέπει να συμφωνηθούν και να διατηρηθούν όσο πιο απλά γίνεται.

Όταν είναι εφικτό η έξοδος ενός μοντέλου προσομοίωσης πρέπει να ελέγχεται έναντι της πραγματικής απόδοσης του συστήματος αν το σύστημα που μοντελοποιείται είναι σε λειτουργία. Αν το μοντέλο μπορεί να αντιγράφει πραγματικά δεδομένα και η εγκυρότητα και η αξιοπιστία είναι προηγμένες. Αν δεν υπάρχει σύστημα αυτή τη στιγμή, τότε αν το σύστημα μπορεί να τρέξει κάτω από συνθήκες που θεωρητικά αποτελέσματα είναι γνωστά και αν η προσομοίωση αντιγράφει θεωρητικά αποτελέσματα τότε η αξιοπιστία επιβεβαιώνεται. Το μοντέλο μπορεί να τρέξει κάτω από μια ποικιλία συνθηκών και τα αποτελέσματα εξετάζονται από τους χρήστες για την αληθοφάνειά τους.

Βιβλιογραφία

1. Τρύφων Ι. Δάρας – Παναγιώτης Θ. Σύψας, Στοχαστικές Ανελίξεις, Θεωρία και Εφαρμογές, Εκδόσεις Ζήτη, 2003
2. Cinlar E., Introduction to Stochastic Processes, Pentice-Hall, Engelwood Cliffs, NJ 1975
3. De Smit, J.H.A, Some general results for many server queues. Advances in Applied Probability 5, 153-169, 1973
4. Donald Gross – John F. Shortle – James M. Thomson – Carl M. Harris, Fundamentals of Queueing Theory, Wiley series 2008
5. Harris C.M., Some new Results in the statistical analysis of queues. In Mathematical Methods in Queuing Theory, A.B. Clarke, Ed. Springer-Verlag, Berlin, 1974
6. Little, J.D.C., A Proof for the queuing formula L=λW. Operations Research 9, 383-387, 1961
7. Papoulis A., Probability, Random Variables and Stochastic Processes, 2nd ed. McGraw-Hill, New York, 1991
8. Ross S.M. Stochastic Processes, 2nd ed. Wiley, New York, 1996
9. Saaty T. L., Elements of Queuing Theory with Applications, McGraw-Hill, New York, 1961

Πηγές από το Διαδίκτυο

1. http://en.wikipedia.org/wiki/Queueing_theory
2. http://en.wikipedia.org/wiki/Queueing_model