



מרצה : ד"ר גלית יום-טוב

מתרגל : אריק סנדרוביץ'

תאריך הבחינה : 12.7.2015

מס' סטודנט _____

שם _____

הנדסת מערכות שירות – 096324

מועד א' – סמסטר אביב 2015

פתרון

- המבחן נמשך שלוש שעות. באחריותך להחזירו במועד.
- המבחן בחומר סגור.
- מותר להשתמש במחשבון מדעי רגיל.
- המבחן כולל 29 עמודים, כולל עמוד זה.
- כאשר נדרשים חישובים, ניתן לעגלם באופן שאינו פוגע בהבנת התשובה.
- מלאו את הפרטים הנדרשים בראש העמוד, בכתב ברור בבקשה.
- על התשובות יש לענות במקומות המיועדים, שאמורים להספיק: מלל מיותר יגרע ניקוד.
- יש לתת הסברים או הוכחות רק אם התבקשתם במפורש לעשות זאת.

שאלה 1	_____ מתוך 25 נקודות	בחינות קודמות/הרצאות/תרגולים
שאלה 2	_____ מתוך 20 נקודות	תרגילי בית
שאלה 3	_____ מתוך 35 נקודות	יישום
שאלה 4	_____ מתוך 20 נקודות	תיאוריה ושימושיה

סה"כ _____ מתוך 100 נקודות

בהצלחה !

שאלה 1 - בחינות קודמות/הרצאות/תרגולים (25 נקודות).

חלק 1 – SEESat (15 נקודות)

1.1 (3 נקודות) גרף 1 וגרף 2 (בעמוד הבא), אשר הופקו באמצעות SEESat, מציגים היסטוגרמה של משכי שירות ב – USBank, באפריל 2001. הסבירו מהו ההבדל בין שני הגרפים, מהו היתרון ומהו החיסרון של שימוש בגרף 2 על פני גרף 1.

תשובה:

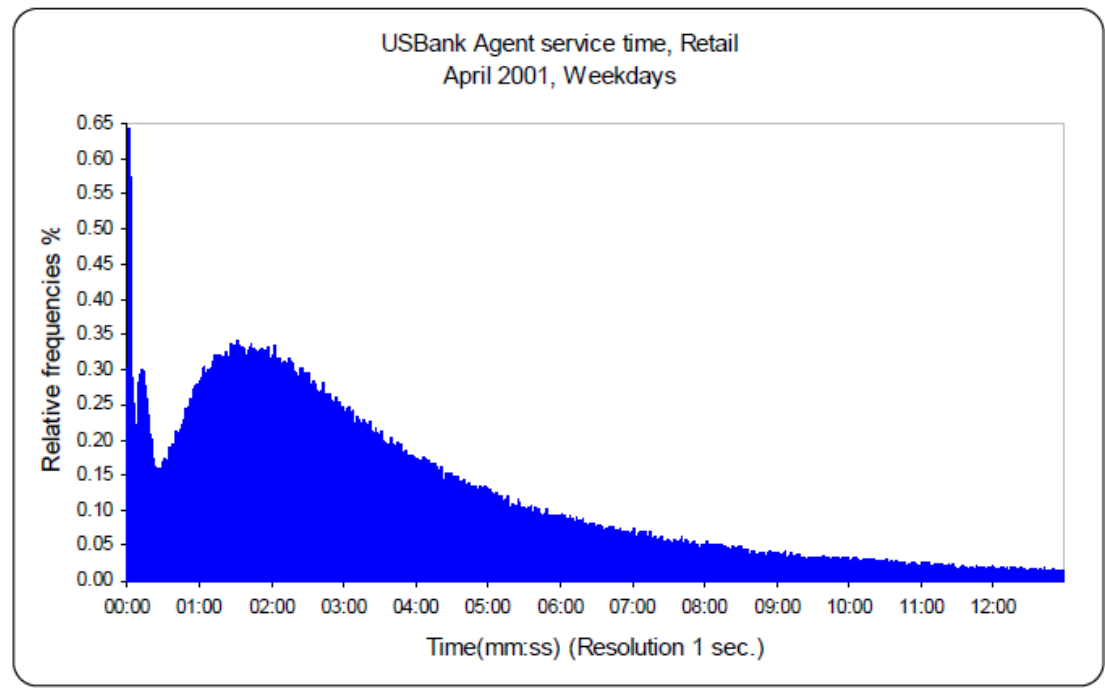
ההבדל בין שני הגרפים הוא ברזולוציה שלהם (אשר מתאימה לרוחב העמודות בהיסטוגרמה). גרף 1 הינו ברזולוציה של שניה אחת ואילו גרף 2 הינו ברזולוציה של 30 שניות. בשל כך, גרף 2 הינו חלק יותר, מה שמאפשר לראות טוב יותר את הצורה הכללית של התפלגות משכי השירות. עם זאת, בגרף 2 אנו מאבדים מידע על אי הסדירות בראשית הצירים.

1.2 (3 נקודות) איזו תופעה מיוחדת באה לידי ביטוי בגרפים? הציעו הסבר לתופעה זו.

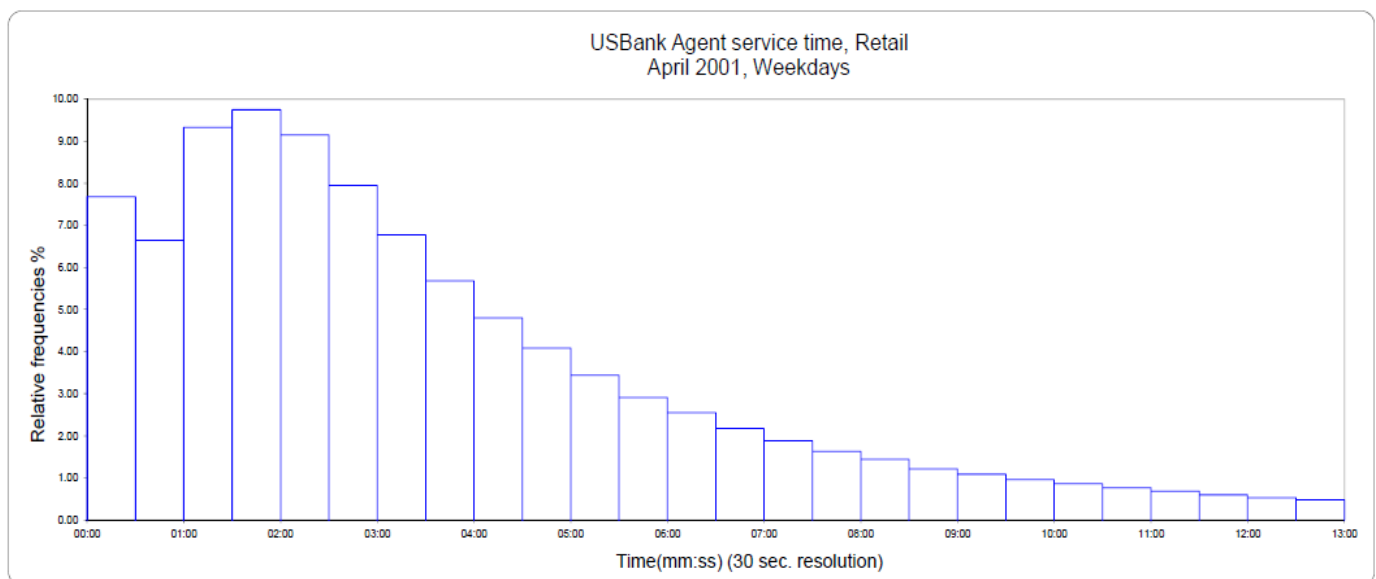
תשובה:

ניתן לראות כי יש אחוז גבוה יחסית של שיחות שנמשכות זמן קצר מאוד, עד כדי כך שברור כי לא הייתה יכולה להתרחש כל אינטראקציה בין המוקדן ללקוח. סיבה אפשרית לכך היא שהמוקדנים עונים לשיחות ומתנקים אותן מיד. התנהגות כזו יכולה לנבוע מכך שמוקדנים מתוגמלים על כמות השיחות להן הם עונים או על קיצור ממוצע משך השיחה. סיבה נוספת יכולה להיות עייפות הנובעת מהעסקת יתר.

גרף 1



גרף 2



1.3 (3 נקודות) הסבירו בקצרה מה מציג גרף 3 (בעמוד הבא).

תשובה:

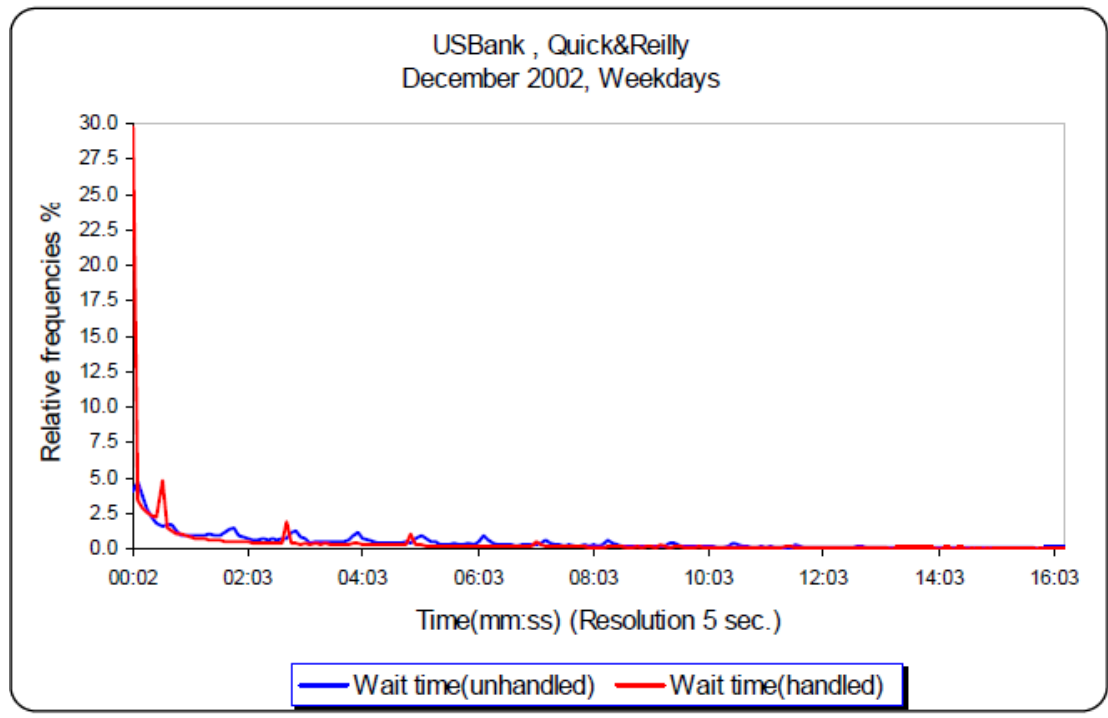
גרף 3 מציג היסטוגרמת משך זמני ההמתנה עבור לקוחות שקיבלו שירות (Handled) ועבור לקוחות שלא קיבלו שירות (Unhandled). מדובר בזמני המתנה לשירות מסוג Quick&Reily ב – USBank, בכל ימי החול של דצמבר 2002.

1.4 (3 נקודות) הסבירו בקצרה מהו ההבדל בין גרף 3 וגרף 4 ומהו היתרון בשימוש בגרף 4 על פני גרף 3.

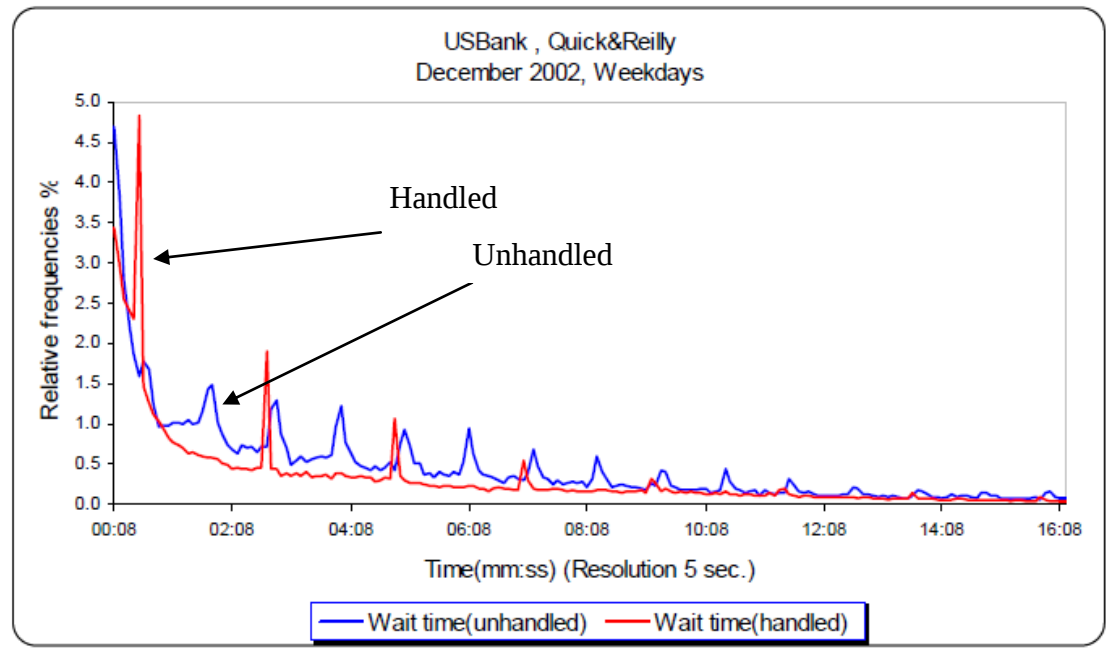
תשובה:

גרף 4 מציג את אותן היסטוגרמות כמו גרף 3, אך מציג אותן החל מ-5 שניות. במקור (גרף 3), כ – 30% מהלקוחות שקיבלו שירות נענו מיד, כך שזמן ההמתנה שלהם היה אפס. צמצום טווח ציר X מקטין את הטווח של ציר ה – Y ומאפשר התמקדות בתופעות המעניינות עבור לקוחות שלא נענו מיד, תופעות שקשה להבחין בהן בגרף 3.

399



499



1.5 (3 נקודות) איזו תופעה מעניינת באה לידי ביטוי בגרף 3 ו-4? הציעו הסבר לתופעה זו.

תשובה:

ניתן לראות בגרפים כי בהיסטוגרמה של זמן ההמתנה של לקוחות שלא קיבלו שירות (Unhandled) יש פיק כל 65 שניות. בהיסטוגרמה של זמן ההמתנה של לקוחות שקיבלו שירות (Handled) יש פיק כל 130 שניות. אנו סבורים כי ניתן להסביר את הפיקים בזמני ההמתנה של לקוחות שלא קיבלו שירות ע"י הסבר "פסיכולוגי" למשל כתגובה של הלקוחות להודעה שנשמעת תוך כדי ההמתנה שלהם כל 65 שניות (כגון הודעה על מיקומם בתור או משך ההמתנה המשוער). את הפיקים בזמני ההמתנה של לקוחות שקיבלו שירות ניתן כנראה להסביר דרך פרוטוקול המערכת, למשל מדיניות ניהול תור שמקדמת בתור לקוחות מסויימים בפרקי זמן קבועים מתחילת המתנתם, כל 130 שניות במקרה שלנו.

חלק 2 - איוש משמרות (10 נקודות).

מוקד שירות הטלפוני של בנק גדול בארץ מופעל 10 שעות ביממה, בין השעות 10:00 בבוקר ל-20:00 בערב. נסמן ב- $N_1, \dots, N_{10}$ את רמות האיוש האופטימליות עבור כל שעה (10:00-20:00). בעיית מנהלת המוקד היא שלא ניתן להזמין מוקדנים רק לשעה מסוימת: כל מוקדן מוזמן למשמרת שאורכה 8 שעות, כולל שעת הפסקה לאחר 4 שעות עבודה.

נסמן עלות מוקדן במשמרת ה- j ב- C_j . על המנהלת לכסות את 10 שעות פעילות המוקד. הנכם מתבקשים לעזור למנהלת לקבוע את רמת האיוש האופטימלית למשמרת תחת האילוצים הנ"ל, וכפי שנלמד בתרגול.

2.1 (3 נקודות) הגדירו מטריצה $A = [a_{ij}]$ המתארת את המשמרות האפשריות (ז"א הגדירו את אברי

המטריצה a_{ij}).

תשובה:

$$a_{ij} = \begin{cases} 1, & \text{if an agent working according to schedule } j \text{ is available during interval } i; \\ 0, & \text{otherwise.} \end{cases}$$

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

2.2 (4 נקודות) נסחו בעיית תכנות בשלמים, שפתרונה נותן איוש אופטימלי למשמרת עבור כל המשמרות האפשריות, וזאת במטרה להביא למינימום את עלויות האיוש של המוקד.

תשובה:

משתנה ההחלטה: X_j - מספר המוקדנים במשמרת ה- j .

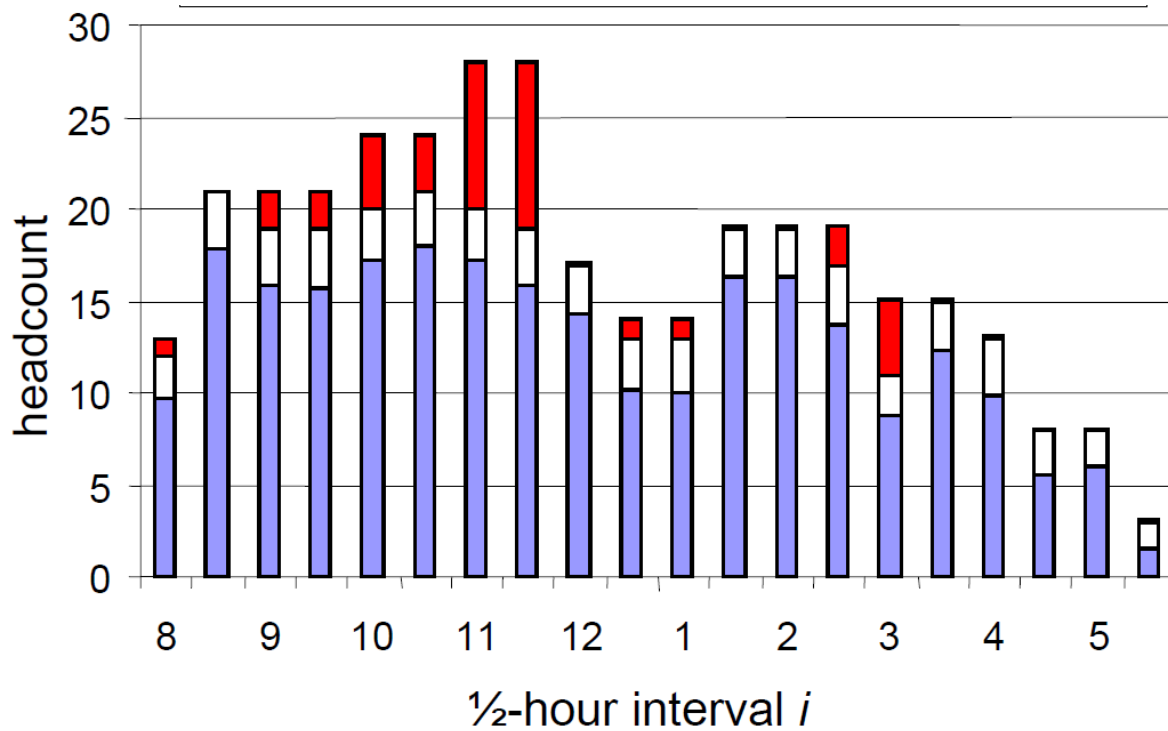
$$\min \sum_{j=1}^3 C_j X_j$$

s.t.

$$\sum_{j=1}^3 a_{ij} X_j \geq N_i \quad i = 1, 2, \dots, 10$$

$$X_j \geq 0 \text{ integer}$$

2.3 (3 נקודות) בתרגול ראינו את הגרף הבא :



הסבר את שלושת המרכיבים של כל עמודה (Bar) בגרף. השתמש בסימונים המתמטיים שלמדנו בקורס על מנת לאפיין כל מרכיב בנפרד.

תשובה:

החלק הראשון הוא העומס המוצע R . זוהי כמות העבודה הממוצעת ליחידת הנכנסת למערכת באינטרוול זמן זה.

החלק השני נובע מאיוש (N) על סמך כלל השורש.

החלק השלישי הוא הפער בין הוקטורים Ax ל- N . זהו האיוש העודף שנובע מאילוצי המשמרות.

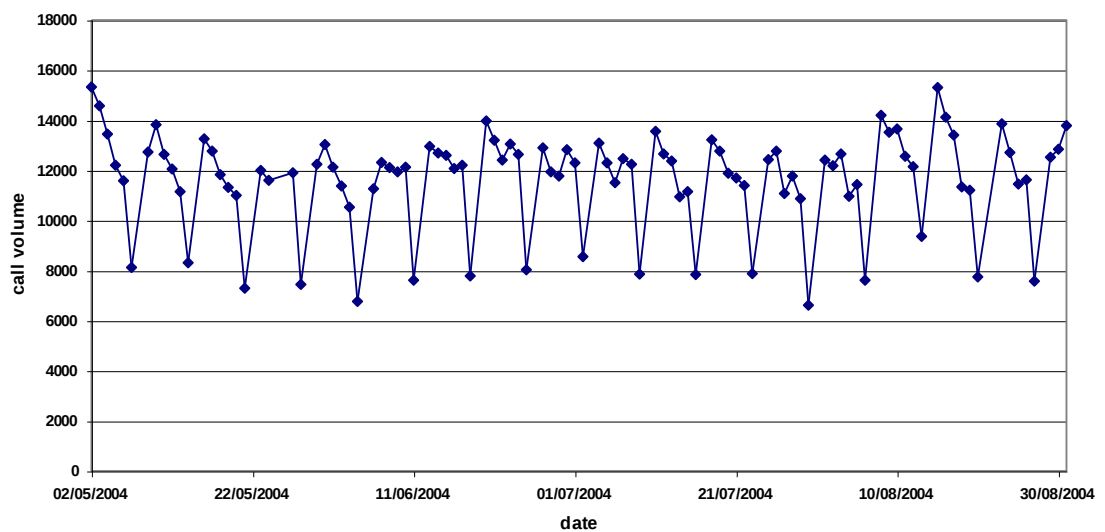
שאלה 2 – תרגילי בית (20 נקודות).

חלק 1. מופעים ותחזיות (12 נקודות).

נתונים סיכומים של נתוני שיחות טלפון, על פני שלושה חודשים, במוקד של חברת תקשורת ישראלית. הנתונים כוללים, לכל אינטרוול זמן של 1/2 שעה, את כמות השיחות הנכנסות למוקד. מוגדרים 33 אינטרוולים במהלך יום עבודה, בין השעות 7:00 ל-23:30.

נסמן, כמו בתרגיל הבית, N_{jk} להיות מספר השיחות הנכנסות באינטרוול זמן k ביום j ($j=1, \dots, 6$), ו- F_{jk} להיות תחזית מספר השיחות לאינטרוול זה.

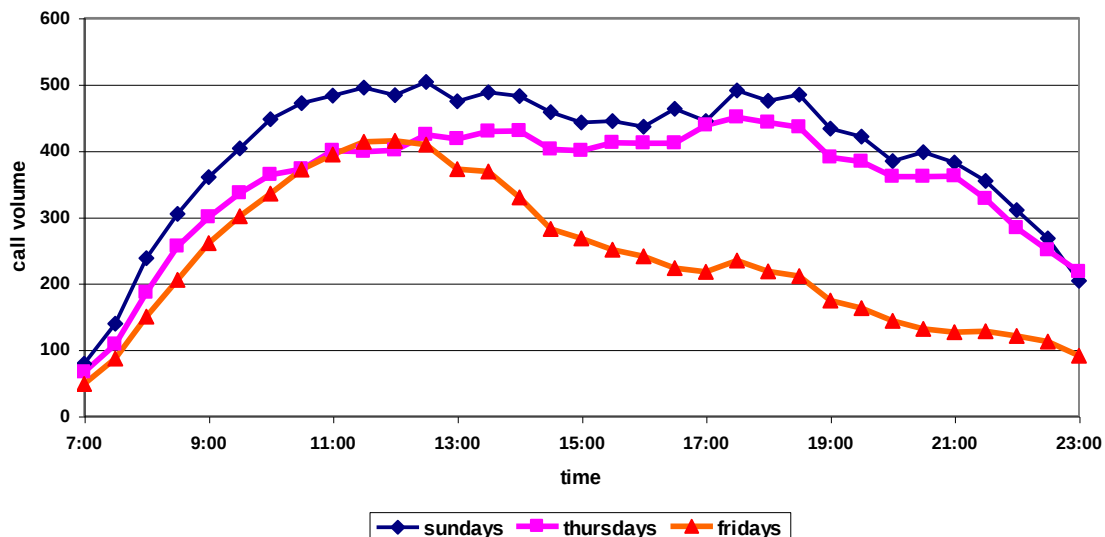
2.1.1 (3 נקודות) הגרף הבא מציג את סך השיחות הנכנסות ליום בתקופה 3108//04 - 02/05/04 (בימים א'-ו', ללא שבת). אילו מסקנות (בעלות משמעות תפעולית) ניתן להסיק מהגרף?



תשובה:

ניתן לראות מבנה שבועי בו יש בימי שישי הרבה פחות שיחות וביום ראשון המספר הגבוה ביותר של שיחות ביחס ליתר ימי השבוע. באופן כללי נראה שמספר השיחות יורד הדרגתית מראשון עד שישי. להבחנות אלו יש כמובן משמעות תפעולית בהקשר של החלטות איוש, או כל היבט אחר בהתמודדות עם עומס.

2.1.2 (3 נקודות) הגרף הבא מציג את ממוצע מספר המופעים בכל אינטרוול זמן (חצי שעה) במהלך היום, בימי ראשון, חמישי ושישי בחודש יוני. השוו בין הגרפים השונים. אילו מסקנות תפעוליות ניתן להסיק מגרף זה?



תשובה:

בכל הימים יש מקסימום קצב מופע (התקופה העמוסה ביותר) בצהריים בסביבות 12. יום ראשון הוא העמוס ביותר בשעות אלו (12 בצהריים עד 7 בערב) מהצהריים, ימי חמישי וראשון די דומים בצורתם (עדיין ראשון עמוס יותר), אך ליום שישי צורה שונה: החל מהצהריים קצב ההגעה הולך וקטן. השלכות תפעוליות: איוש ב"צורה דומה" לימי ראשון וחמישי, עם פחות מוקדנים ביום חמישי; ליום שישי "צורה" שונה. דבר זה יכול להשפיע לדוגמה על מבנה משמרות בימים אלו.

2.1.3 (3 נקודות) תארו במדויק את אחת משיטות התחזית (Forecasting) שהוצגו בכיתה או בתרגילי הבית. ציינו באיזו שיטה בחרתם. (השתמשו בתשובתכם בסימונים N_{jk} ו- F_{jk} מסעיף 1.1.1)

תשובה:

למשל ממוצע נע (4) - נגדיר את התחזית הבאה על סמך ממוצע של ארבעת האינטרוולים האחרונים מאותו סוג (למשל יום ג' בין 10:00 ל- 10:30):

$$F_{j,k} = \frac{N_{j-6,k} + N_{j-12,k} + N_{j-18,k} + N_{j-24,k}}{4}$$

2.1.4 (3 נקודות) נניח כי המופעים למוקד השרות הם לפי תהליך פואסון עם קצב מופע λ_{jk} . נניח שהצלחנו לחזות את ערכי הפרמטר λ_{jk} במדויק ($F_{jk} = \lambda_{jk}$). האם אתם מצפים שמדדי הטיב RMSE ו- APE יהיו שווים ל-0? הסבירו תשובתכם.

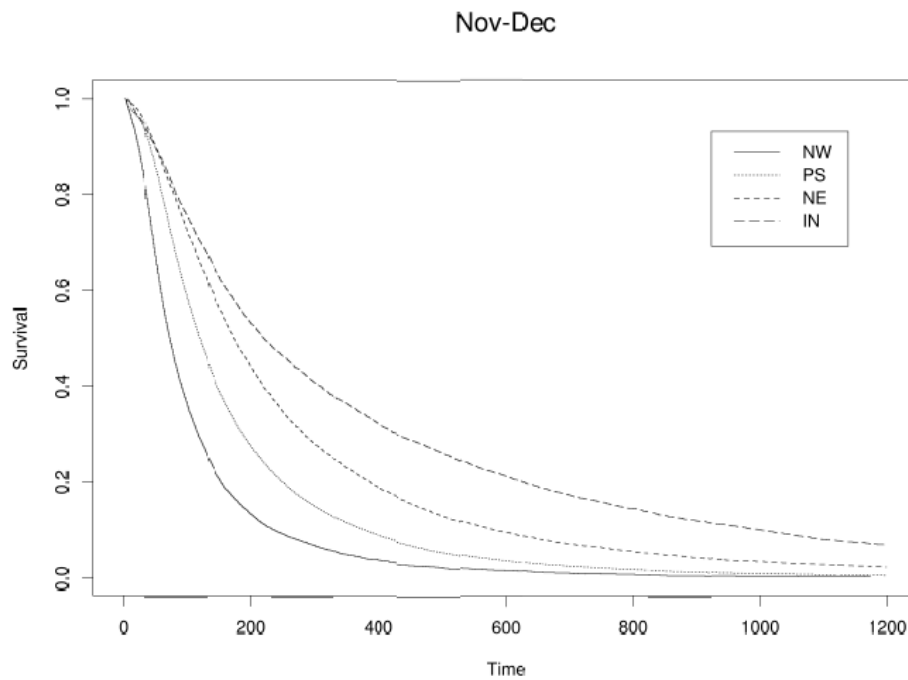
תשובה:

לא. הערכים לא צפויים להיות 0, מכיוון שגם אם הפרמטר ידוע באופן מדויק, לתהליך עצמו יש השתנות סטוכסטית (שונות), והדגימות לא תהיינה זהות לממוצע.

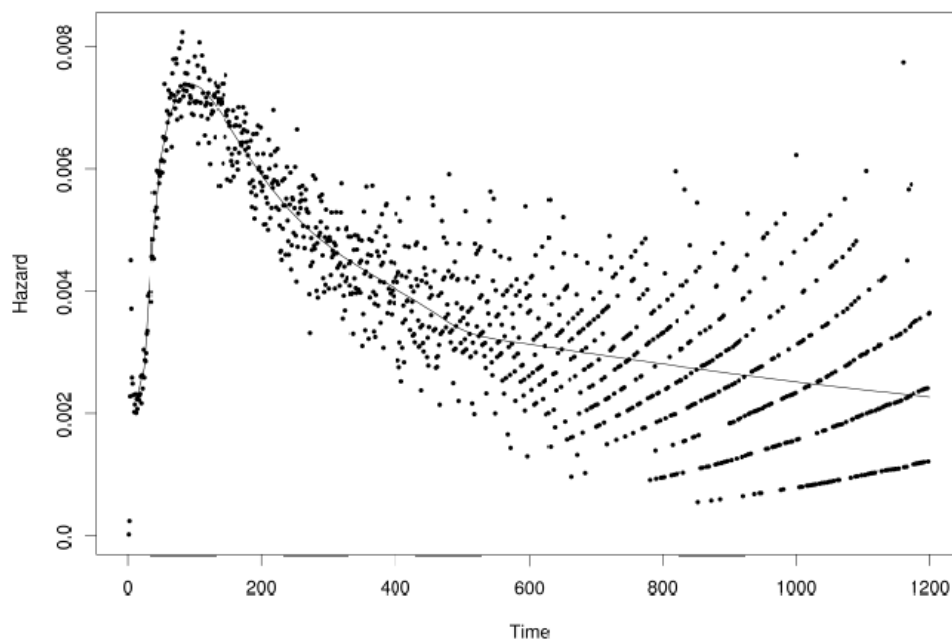
חלק 2. פונקציות שרידות ופונקציות סיכון של זמני שירות (13 נקודות)

שני הגרפים שלהלן מבוססים על נתוני מוקד שירות טלפוני של בנק ישראלי קטן. נותחו ארבעה סוגי שירות עיקריים: NW, NE, IN, PS. גרף 1 מתאר פונקציות שרידות (survival functions) של ארבעת סוגי השירות. נקודות בגרף 2 הן אמדים לפונקצית הסיכון (hazard rate) של זמן השירות הכולל (כל סוגי השירות ביחד), ברזלוציה של שנייה אחת. הקו בגרף 2 הוא אמד לפונקצית הסיכון אחרי החלקה.

גרף 1



גרף 2



2.2.1 (3 נקודות) איך הייתם אומדים, מתמטית או נומרית (במידת הצורך), את תוחלת זמן השירות של סוג שירות מסוים, לו הייתה נתונה לכם רק פונקציית השרידות (survival function) שלו, $S(x), x \geq 0$?

תשובה :

נשתמש בנוסחת הזנב: $E[S] = \int_0^{\infty} S(x) dx$. לכן התשובה תהיה או מתמטית, ע"י חישוב האינטגרל, או נומרית, על ידי חישוב השטח מתחת לגרף של $0 \leq x, S(x)$.

2.2.2 (3 נקודות) איך הייתם אומדים את תוחלת זמן השירות של סוג שירות מסוים, לו הייתה נתונה לכם רק פונקציית הסיכון (hazard rate) שלו $h(x), x \geq 0$?

תשובה :

ניתן לקבל את פונקציית השרידות מפונקציית הסיכון: $S(x) = \exp \left\{ - \int_0^x h(u) du \right\}$ ואחר כך להשתמש בנוסחת הזנב, כמו בסעיף הקודם.

2.2.3 (3 נקודות) סדרו את ארבעת סוגי השירות לפי תוחלתם (מהקצר ביותר לארוך ביותר). האם הנתונים בידיכם מאפשרים לסדר גם את סטיות התקן או השונות של משכי השירות? נמקו בקצרה!

תשובה:

$$E(NW) < E(PS) < E(NE) < E(IN)$$

אי אפשר לסדר את סטיות התקן, ובאותו אופן גם לא את השונות. רואים מהגרף שמשכי השירות מסודרים סטוכסטית, ולכן $E(T)$ ו $E(T^2)$ גם הם מסודרים לפי אותו סדר. אבל השונות היא ההפרש בין $E(T^2)$ לבין $E(T)^2$ ולכן אין דרך לנבא את סדר השונות.

2.2.4 (4 נקודות) הסבירו את הצורה של "נקודות המסודרות על קו עולה", הבולטת בחלק הימני של גרף 2.

הערה:

האמידה של פונקצית הסיכון בגרף 2 נעשתה באופן הבא:

- נסמן ב A_k את מספר התצפיות עם זמן שירות ששווה ל- k שניות.
- נסמן ב η_k את מספר התצפיות עם זמן שירות שגדול או שווה מ- k שניות.
- האמדים לפונקצית הסיכון בשניה k (נקודות בגרף 2) חושבו ע"י $h(k) = A_k / \eta_k$.

הסבר:

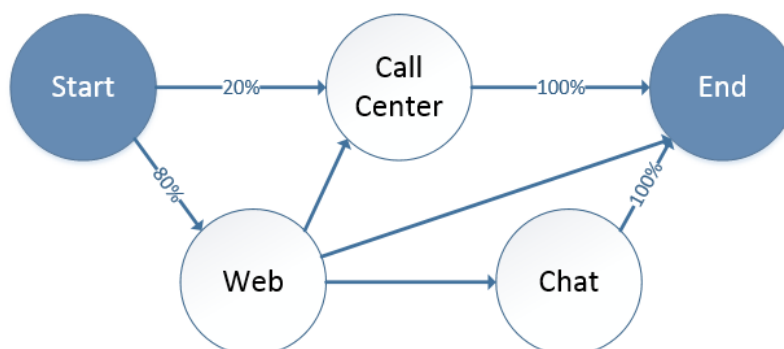
הנקודות המסודרות על קו עולה מתאימות ל- $A_k = 1$, $A_k = 2$, ..., משום ש- η_k יורד כאשר הזמן עולה, קוים אלו עולים. (למשל הקו התחתון מתאים ל- $A_k = 1$). בגרפים אחרים מאותו סוג יתכן שתופענה נקודות על ציר ה- x , המתאימות ל- $A_k = 0$ (שפירושו, אין משכי שירות באורך k).

שאלה 3 – יישום (35 נקודות)

חברת אשראי מעניקה שירות לקוחות בשלושה ערוצים עיקריים: מוקד שירות טלפוני, אתר אינטרנט ומוקד צ'אט (שיחות). לקוח אשר נכנס לאתר האינטרנט, משוטט בו זמן מה עד עזיבת האתר. ללקוחות מסוימים ניתנת האופציה (אחרי זמן מסויים בגלישה) לבחור בצ'אט עם נציג.

חלק 1: ניתוח מצב קיים (13 נקודות)

מנהל התפעול של החברה אפיין את תהליך השירות בעזרת התרשים הבא:



תאור התהליך במילים: כ-20% מהלקוחות של החברה מתקשרים ישירות למוקד השירות הטלפוני. השאר, כ-80%, נכנסים לאתר האינטרנט לקבלת השירות. מתוכם לכ-50% מוצע צ'אט עם נציג. 80% מן הלקוחות להם מוצע צ'אט מסכימים לכך ומסיימים בצ'אט את שירותם, בעוד ש-20% האחרים ממשיכים לגלוש באתר לצורך קבלת השירות. כ-10% מלקוחות האתר שלא קיבלו הצעה לצ'אט או שסירבו להצעה כזו, מתקשרים בתום הגלישה למוקד השירות הטלפוני. זמני השהיה הממוצעים הם: 7 דקות באתר האינטרנט, 10 דקות במוקד השירות ו-5 דקות בצ'אט.

3.1.1 (3 נקודות) אפיינו את התהליך כתהליך קפיצה מרקובי (תק"מ). פרטו את ההנחות המתאימות וכתבו את: מצבי התק"מ, מטריצת המעברים (P), וקטור משכי שירות (m) ואת וקטור המצב ההתחלתי (q).

תשובה:

הנחות – זמן בכל מצב מתפלג מעריכית. המעברים ממצב למצב תלויים רק במצב הנוכחי.

$$S = \{CallCenter, Web, Chat, End\}$$

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0.06 & 0 & 0.4 & 0.54 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$m = (10, 7, 5, 0)$$

$$q = (0.2, 0.8, 0, 0)$$

חישובים:

$$P_{2,3} = 0.5 \cdot 0.8 = 0.4$$

$$P_{2,1} = (0.5 + 0.5 \cdot 0.2) \cdot 0.1 = 0.06$$

3.1.2 (1 נקודה) כיצד מתפלג משך השירות של לקוח? נמק.

תשובה:

משך השירות הוא הזמן עד לספיגת השרשרת במצב End, לכן ההתפלגות היא PH.

3.1.3 (2 נקודות) הסבירו כיצד תחשבו את תוחלת הזמן עד ספיגה (אין צורך להציב את הנתונים באלגוריתם החישוב).

תשובה:

יש לחשב את המטריצה R:

$$R = [I - P]^{-1}$$

משך ממוצע:

$$E[T] = qRm$$

אופציה נוספת – חישוב דרך מ"מ מעריכיים (התניות על מסלולים). כלומר לשקלל את כל המסלולים:

$$0.2 * \text{time in call center} + 0.8 * 0.4 * (\text{time in web} + \text{chat}) + 0.8 * 0.54 * (\text{time in chat})$$

$$+ 0.8 * 0.06 * (\text{time in web} + \text{call center})$$

3.1.4 (2 נקודות) בכל רגע נתון, מספר הלקוחות הממוצע בשירות הציאט הינו 500. מהו קצב מופע הלקוחות הכללי לשעה.

תשובה: נשתמש בחוק ליטל

$$L_{Chat} = 500$$

$$W_{Chat} = 5 \text{ min}$$

$$\lambda_{Chat} = \frac{L_{Chat}}{W_{Chat}} 100 / \text{min} = 6000 / \text{hr}$$

$$\lambda_{Chat} = 0.4 \lambda_{Web}$$

$$\Rightarrow \lambda_{Web} = 15000 / \text{hr} = 0.8 \lambda$$

$$\Rightarrow \lambda = 30000 / \text{hr}$$

מעתה נתמקד בשירות הציאט בלבד. נסמן ב- λ את קצב המופע השעתי הכולל שמצאתם בסעיף 3.3 (במידה ולא חישבתם את λ יש לכתוב אותו כפרמטר).

3.1.5 (1 נקודה) במצב הנוכחי, מהו קצב המופע האפקטיבי (לשעה) לתוך שירות הציאט?

תשובה:

$$\lambda_{Chat} = \frac{L_{Chat}}{W_{Chat}} 100 / \text{min} = 6000 / \text{hr}$$

3.1.6 (4 נקודות) הניחו כי כל נותן שרות בערוץ הציאט יכול לשוחח עם לקוח אחד בלבד בו זמנית בזמן השירות. הציעו מודל תורים מרקובי המתאר את שירות הציאט במצב הנוכחי. הניחו כי לקוח שבחר את האפשרות לציאט מסיים את השירות בציאט (לא נוטש). כתבו את הנחות ופרמטרי המודל. אין צורך לכתוב את ערכי הפרמטרים של המודל.

תשובה:

המודל הינו Erlang-C. משכי שירות אקספ', תהליך מופע פוסאון, זמני שירות ב"ת בזמני המופע, מספר שרתים קבוע, משרתים לקוח אחד בכל פעם.

$$\text{Parameters} = (\lambda_{Chat}, \frac{1}{m_{Chat}}, N)$$

כאשר N הוא מספר השרתים.

חלק 2: איוש (9 נקודות)

3.2.1 (2 נקודות) מנהל התפעול מעוניין בכך שזמן ההמתנה לשרת צ'אט מרגע בחירת הלקוח תשאף ל-0. הצע תחום תפעולי וכלל איוש המתאימים למודל שבחרתם ב-3.1.6. יש לכלול בתשובה טווח עבור פרמטר רמת שירות בו תשתמשו עבור כלל האיוש.

תשובה:

המנהל מעוניין בתחום QD עבור ERLANG-C.

כלל האיוש המתאים הוא:

$$N = R + \delta R = 500 + 500\delta$$

$$0.1 \leq \delta \leq 0.25$$

3.2.2 (2 נקודות) מנהל התפעול למד את הקורס הנדסת מערכות שירות, והוא מעוניין לאייש את מוקד הצ'אט בתחום ה-QED. מה צריך להיות פרמטר האיוש (β), כך שמשך ההמתנה הממוצע בהינתן המתנה לצ'אט יהיה בקירוב 0.1 דקות? ניתן לכתוב משוואה עבור הערך של β , בלי לפתור אותה, כך ש- β היא הנעלם היחיד במשוואה. העזרו בדף העזר בעמוד הבא.

תשובה:

נשתמש בביטוי:

$$E[W_q | W_q > 0] \approx \frac{1}{\sqrt{n}} \cdot \frac{1}{\mu\beta};$$

$$0.1 \approx \frac{1}{\sqrt{500 + \beta\sqrt{500}}} \cdot \frac{1}{\beta} m_{Chat}$$

נחלץ את β .

Theorem (QED Erlang-B; Jagerman, 1974)

As $n \rightarrow \infty$, the following 3 statements are equivalent:

1. **Customers:** $E_{1,n} \approx \frac{\gamma}{\sqrt{n}}$, for some $\gamma > 0$;
2. **Servers:** $\rho \approx 1 - \frac{\beta}{\sqrt{n}}$, for some $-\infty < \beta < \infty$;
3. **Manager:** $n \approx R + \beta\sqrt{R}$ (square-root “staffing”);

in which case

$$\gamma = h(-\beta) = \frac{\phi(-\beta)}{\bar{\Phi}(-\beta)} = \frac{\phi(\beta)}{\Phi(\beta)},$$

where $\phi, \Phi, \bar{\Phi}$ and h are the density, cdf, survival function and hazard rate of $N(0, 1)$ (standard-normal), respectively.

Note: **Servers’ Occupancy** $\approx 1 - \frac{\beta + \gamma}{\sqrt{n}}$, accounting for blocking.

Theorem (QED Erlang-C; Halfin & Whitt, 1981)

As $n \rightarrow \infty$, the following 4 statements are equivalent:

0. **QED:** $E_{2,n} \approx \alpha$, for some $0 < \alpha < 1$;
1. **Manager:** $n \approx R + \beta\sqrt{R}$, for some $0 < \beta < \infty$;
2. **Servers:** $\rho \approx 1 - \frac{\beta}{\sqrt{n}}$;
3. **Customers:** $E[W_q | W_q > 0] \approx \frac{1}{\sqrt{n}} \cdot \frac{1}{\mu\beta}$;

in which case

$$\alpha = \alpha(\beta) = \left[1 + \frac{\beta}{h(-\beta)} \right]^{-1},$$

which we call the **Halfin-Whitt Delay-Function**.

The Erlang-A Queue in the QED-Regime

Theorem (with Garnett & Reiman, 2002)

The following **points of view** are equivalent:

0. **QED:** $P\{W_q > 0\} \approx \alpha$, for some $0 < \alpha < 1$;
1. **Manager:** $n \approx R + \beta\sqrt{R}$, for some $-\infty < \beta < \infty$;
2. **Servers:** $\text{Occupancy} \approx 1 - \frac{\beta + \gamma}{\sqrt{n}}$;
3. **Customers:** $P\{\text{Ab}\} \approx \frac{\gamma}{\sqrt{n}}$, for some $0 < \gamma < \infty$;

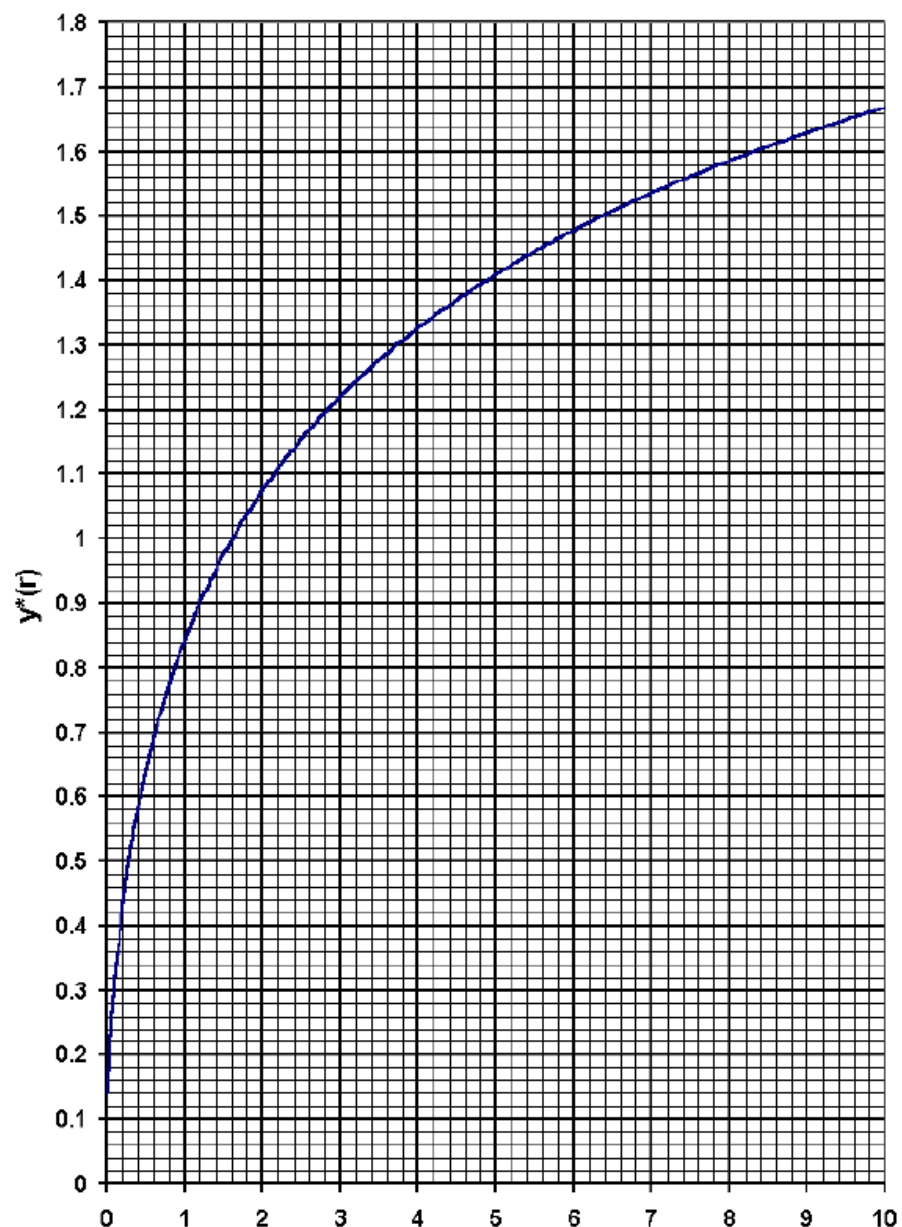
in which case

$$\alpha = \alpha(\beta, \frac{\mu}{\theta}) = \left[1 + \sqrt{\frac{\theta}{\mu}} \cdot \frac{h(\hat{\beta})}{h(-\beta)} \right]^{-1},$$

which we call the **Garnett Delay-Function(s)**;

ועד העובדים של חברת האשראי מוחה נגד האיוש שהוצע על ידי מנהל התפעול בסעיף 3.7. הוא טוען כי אין לפטר/להעסיק (תלוי בתשובה שקיבלתם) עובדים נוספים, וכי מספר נציגי הצ'אט צריך להיות 500. דרך אלטרנטיבית אותה מציע מנהל התפעול לשמר את זמני ההמתנה הנמוכים לנציגי צ'אט הינה דרך "בקרת כניסה". כלומר, הוא מעוניין לקבוע מכסה של מופעים אל תוך מוקד הצ'אט. בקורס הנדסת מערכות שירות, ראה מנהל התפעול את הגרף הבא, הקושר את **יחס עלויות ההמתנה (d) ועלויות האיוש (c)** לפרמטר האיוש β :

$$r = \text{cost-of-delay} / \text{cost-of-staffing}$$



3.2.3 (2 נקודות) כתוב ביטוי לפונקציית העלות המיוצגת על ידי התרשים הנ"ל.

תשובה:

$$\text{Cost} = c \cdot n + d \cdot \lambda E[W_q],$$

c = Staffing cost;

d = Delay cost.

3.2.4 (3 נקודות) המנהל נעזר במחלקת השיווק של החברה וקובע כי עלות המתנה היא 5 ש"ח לדקת המתנה ועלות איוש היא 1 ש"ח לדקת איוש. מה צריך להיות קצב המופע השעתי למוקד הצ'אט, תחת הנחת העלויות של מנהל התפעול.

תשובה:

מהתרשים רואים כי

$$\beta^* = 1.4$$

$$500 = R + 1.4\sqrt{R} = \lambda m_{Chat} + 1.4\sqrt{\lambda m_{Chat}}$$

מחלצים את קצב המופע מן המשוואה לעיל ומקבלים שתי פתרונות:

$$\lambda_1 = 106/\text{min}$$

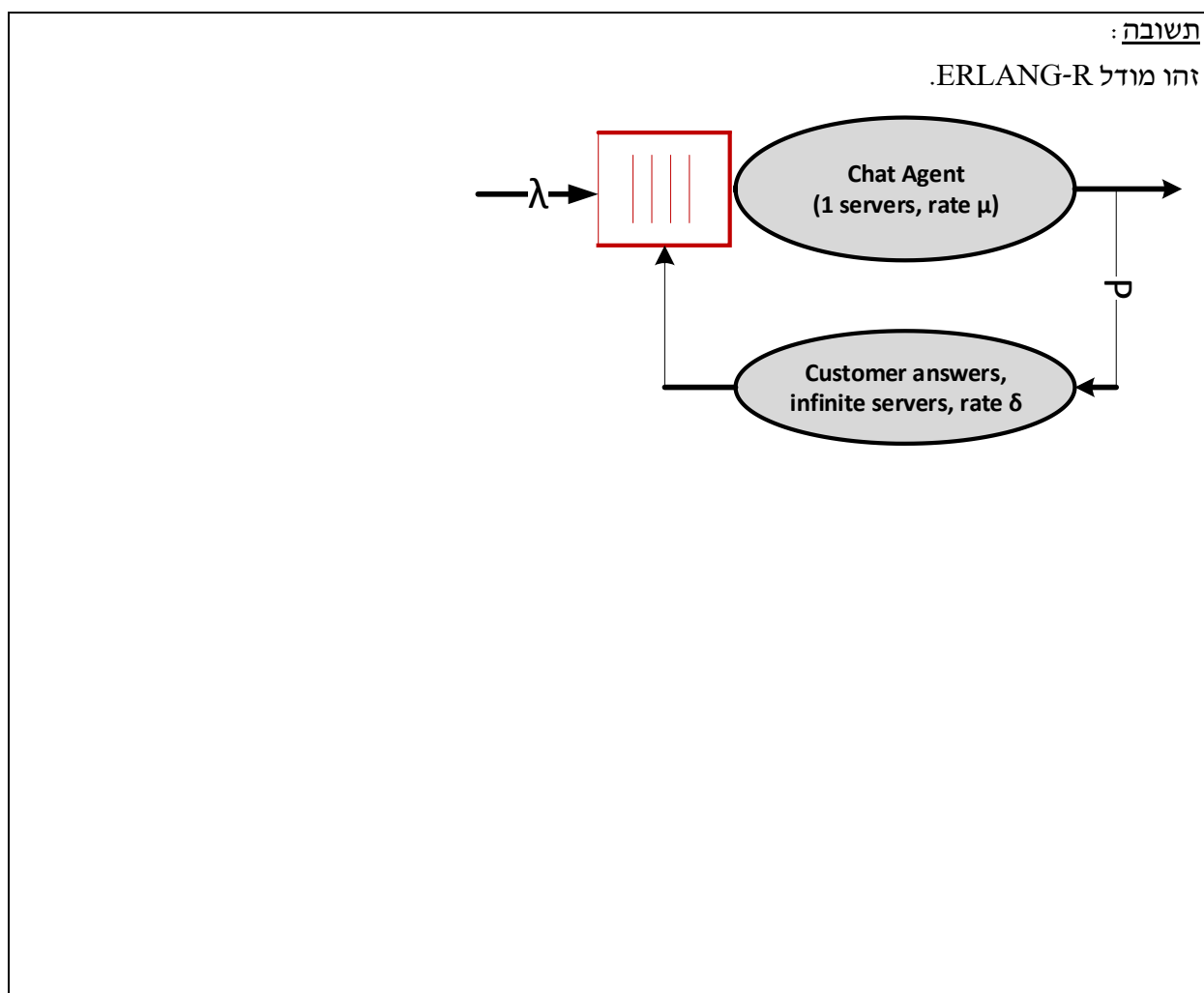
$$\lambda_2 = 94/\text{min}$$

כאשר הפתרון הגדול יותר מפר תנאי סטביליות (נוצר כתוצאה מפתרון המשוואה). לכן הפתרון הוא שקצב המופע החדש הוא 94 לקוחות לדקה (לעומת 100 לקוחות לדקה, בתחילת השאלה).

חלק 3: מידול (13 נקודות)

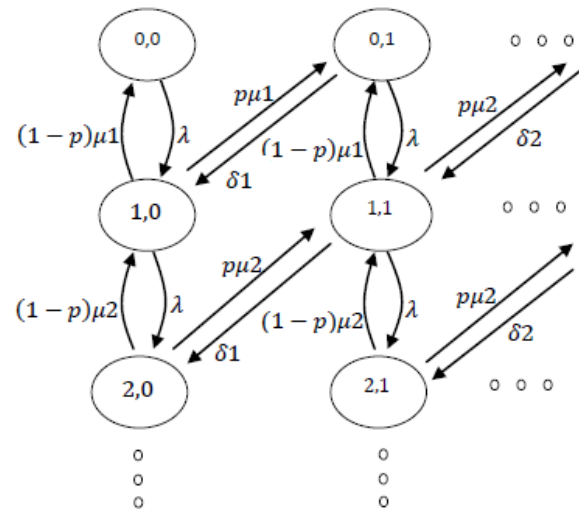
כעת נניח כי נציג צ'אט יכול לטפל בכמה לקוחות בו זמנית. נרצה למדל מערכת הכוללת את השירות שמקבלים מספר לקוחות צ'אט **מנציג יחיד**. שרות בצ'אט מורכב מסדרת שאלות ותשובות (להלן: תגובות) שהלקוח ונציג השירות כותבים זה לזה. ידוע כי בתום תגובה של מוקדן בצ'אט, הסיכוי לסיום הצ'אט הינו $(1-p)$ – כלומר בסיכוי p הלקוח כותב תגובה נוספת ומחכה לתגובה נוספת של נותן השירות. תמיד התגובה הראשונה והאחרונה הן של נותן השירות. תוחלת משך כתיבת תגובה אחת של סוכן צ'אט (זמן השירות) הינו $\frac{1}{\mu}$ ותוחלת משך כתיבת תגובת לקוח חזרה למוקדן הינו $\frac{1}{\delta}$. תהליך המופע הוא פואסון הומוגני בזמן עם קצב λ .

3.3.1 (3 נקודות) הצע את מודל תורים **המתאים ביותר** לייצג את המצב המתואר לעיל? שרטט סכימה של המודל, פרט את ערכי הפרמטרים במודל המוצע וכתוב את הנחות המודל.



3.3.2 (4 נקודות) בהמשך לסעיף הקודם נגדיר את התק"מ הבא: $X(t) = (X_1(t), X_2(t))$ - כאשר $X_1(t)$ הוא מספר האנשים בשירות הציאט עבור מוקדן יחיד (ממתינים ובשירות) ו $X_2(t)$ הוא מספר האנשים שעונים למוקדן (לאחר שקיבלו תשובה). שרטטו סכמה של דיאגרמת קצבי המעבר של התק"מ עבור המצבים בהם: $X_1(t) \leq 2$ AND $X_2(t) \leq 2$.

תשובה:



3.3.3 (3 נקודות) כפי שלמדנו בהרצאות והתרגולים מערכת יציבה היא מערכת שאינה "מתפוצצת" - כלומר מספר הלקוחות נשאר סופי לכל זמן t . תנו תנאי יציבות מערכת מוקדן הציאט, במונחי p .

תשובה:

$$\rho = \frac{\lambda}{(1-p)\mu} < 1$$

$$\Rightarrow p < 1 - \lambda / \mu$$

3.3.4 (3 נקודות) כעת נניח כי המופע הינו תהליך פואסון לא-הומוגני בזמן, עם פונקציית קצב $\lambda(t)$. כתבו את משוואות הנוזלים עבור מספר הלקוחות הכולל במערכת. ניתן להניח כי המערכת ריקה בזמן 0.

תשובה:

$$X(t) = X_1(t) + X_2(t)$$

$$\frac{\partial X_1(t)}{\partial t} = \lambda(t) + \delta X_2(t) - \mu(X_1(t) \wedge N)$$

$$\frac{\partial X_2(t)}{\partial t} = (1-p)\mu(X_1(t) \wedge N) - \delta X_2(t)$$

$$X(0) = 0$$

שאלה 4 – תיאוריה ושימושיה (20 נקודות)

בהרצאה הוגדרו המשתנים הבאים :

τ - הזמן שלקוח מוכן לחכות (Patience time).

v - הזמן שלקוח נדרש לחכות (Offered wait).

4.1. (3 נקודות) בטא בעזרת שני המשתנים לעיל את W_q - הזמן שלקוח יחכה בפועל ואת P_{ab} - ההסתברות שלקוח ינטוש.

תשובה :

$$W_q = \min(\tau, v)$$

$$P_{ab} = P(\tau < v)$$

4.2. אנו מעוניינים בבניית אמד ל- v עבור לקוח שמגיע למערכת. לשם כך נשתמש במודל G/M/N, כלומר מודל עם תהליך מופע כללי, זמן שירות S מתפלג מעריכית (קצב שירות μ) ומספר שרתים N. מדיניות השירות היא FCFS. כמו כן, נניח כי אנו יודעים את מספר הלקוחות במערכת בזמן הגעת לקוח המטרה (נסמן את מספר הלקוחות ב-X). הניחו כי אין נטישות מן התור.

4.2.1. (2 נקודות) מהו ערכו המשתנה $v | X < N$? הסבר.

תשובה :

$$v | X < N = 0 \quad \text{כיוון שיש שרת פנוי ואין צורך להמתין.}$$

4.2.2. (3 נקודות) כיצד מתפלג המשתנה המקרי $v | X = N$?

תשובה :

$$v | X = N \sim \text{Exp}(N\mu)$$

4.2.3. (3 נקודות) כיצד מתפלג המשתנה המקרי $v | X = N + 1$?

תשובה :

$$v | X = N \sim \text{Exp}(N\mu) + \text{Exp}(N\mu) = \text{Erlang}(2, N\mu)$$

4.2.4. (4 נקודות) מציעים כי האמד ל- v יהיה התוחלת של $v | X = N + k$, כלומר עבור הזמן שלקוח נדרש

להמתין, כאשר ישנם k לקוחות בתור (k ידוע). כתוב את הביטוי לתוחלת של $v | X = N + k$.

תשובה :

$$E(v | X = N + k) = E\left(\sum_{i=1}^{k+1} S_i\right) = \sum_{i=1}^{k+1} ES_i = \frac{k+1}{N\mu}$$

4.3 מציעים אמד נוסף ל- v : כאשר לקוח מגיע שואלים את **ראש התור** כמה זמן הוא ממתין. נסמן זמן זה ב- w . נסמן ב- $A(w)$ את מספר הלקוחות שהגיעו במהלך ההמתנה של **ראש התור**.

4.3.1 (3 נקודות) כתבו נוסחה לתוחלת זמן ההמתנה שלקוח נדרש להמתין (v) בהינתן שראש התור ממתין w . רמז: יש להשתמש ב- $A(w)$.

תשובה:

$$E(v | W = w) = E\left(\sum_{i=1}^{A(w)+2} S_i\right)$$

בזמן שראש התור המתין (w) הגיעו $A(w)$ לקוחות. הזמן שלקוח המטרה נדרש להמתין הינו סכום זמני השירות של $A(w)$, זמן השירות של ראש התור, וזמן השירות של לקוח שהיה בשירות כשראש התור נכנס.

בנוסחה נעשה שימוש בנוסחת Wald (גם התשובה $E(W_{HOL} | w) = E\left(\sum_{i=1}^{A(w)+2} S_i\right)$ מתקבלת).

4.3.2 (2 נקודות) הראו תחת אילו תנאים תוחלת ההמתנה של ראש התור (סעיף 4.3.1) שווה לזו שחישבתם בסעיף 4.2.4.

תשובה:

$$A(w) + 2 = k + 1$$

$$\Rightarrow A(w) = k - 1$$

כלומר, אם בזמן שראש התור ממתין (w) יגיעו בדיוק $k-1$ לקוחות, אזי שתי התוחלות שוות.

Table 1: Discrete and Continuous Distributions

Distribution	Notation	Probability Function	Parameters	Support	Moment-Gen. Function	Expectation	Variance	Comments
Binomial	$X \sim \text{Bin}(n, p)$	$p(k) = \binom{n}{k} p^k q^{n-k}$	$0 \leq p \leq 1 - q \leq 1; n \geq 1$	$k = 0, 1, \dots, n$	$[p \cdot e^t + q]^n$	np	npq	
Multinomial	$X \sim \text{Mult}(n, p)$	$p(x) = \frac{n!}{x_1! \cdots x_k!} \prod_{i=1}^k p_i^{x_i}$ $x = (x_1, \dots, x_k)$	$0 \leq p_i; \sum_{i=1}^k p_i = 1; n \geq 1$ $p = (p_1, \dots, p_k)$	$0 \leq x_i; \sum_{i=1}^k x_i = n$		$E(X_i) = np_i$ $i = 1, \dots, k$	$\text{cov}(X_i, X_j) = -np_i p_j$ $(i \neq j)$	$X_i \sim \text{Bin}(n, p_i)$
Negative Binomial	$X \sim NB(n, p)$	$p(k) = \binom{k-1}{n-1} p^n q^{k-n}$	$0 \leq p \leq 1 - q \leq 1; n \geq 1$	$k = n, n+1, \dots$	$\left[\frac{pe^{qt}}{1-qe^{qt}} \right]^n$	$\frac{np}{p}$	$\frac{npq}{p^2}$	Number of trials till n th success
Geometric	$X \sim G(p)$	$p(k) = q^{k-1} \cdot p$	$0 \leq p \leq 1$	$k = 1, 2, \dots$	$\frac{pe^{qt}}{1-qe^{qt}}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$X \sim NB(1, p)$
Poisson	$X \sim P(\lambda)$	$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda \geq 0$	$k = 0, 1, \dots$	$\exp[-\lambda(1 - e^t)]$	λ	λ	
Hypergeometric	$X \sim H(n, M, n)$	$p(k) = \frac{\binom{N}{k} \binom{M-k}{n-k}}{\binom{N+M}{n}}$	$N, M > 0$ $1 \leq n \leq N+M$	$\max(0, n-M) \leq k$ $k \leq \min(n, N)$	$\sum_{k=\max(0, n-M)}^{\min(n, N)} e^{kt} p(k)$	$\frac{n}{N+M}$	$n \cdot \frac{M}{N+M} \cdot \frac{N}{N+M} - \left(\frac{n}{N+M} \right)^2$	N - special element M - ordinary ones n - sample size $N+M$ - population
		Dirichlet Function						
Beta	$X \sim \text{Beta}(\alpha, \beta)$	$f(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1} (1-t)^{\beta-1}$	$\alpha, \beta > 0$	$0 \leq t \leq 1$	(Complicated)	$\frac{\alpha\beta}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\alpha - \beta = 1 \rightarrow \text{Uniform}$
Gamma	$X \sim \text{Gamma}(\alpha, \lambda)$	$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}$	$\alpha, \lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-\alpha}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ $= \Gamma(x-1)\Gamma(x-1)$
Erlang	$X \sim \text{Erlang}(n, \lambda)$	$f(t) = \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$	$n = 1, 2, \dots; \lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-n}$	$\frac{n}{\lambda}$	$\frac{n}{\lambda^2}$	$X \sim \text{Gamma}(n, \lambda)$
Exponential	$X \sim \text{Exp}(\lambda)$	$f(t) = \lambda e^{-\lambda t}$	$\lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-1}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$X \sim \text{Gamma}(1, \lambda)$
Normal	$X \sim N(\mu, \sigma^2)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$	$-\infty < \mu < \infty; \sigma > 0$	$-\infty < t < \infty$	$\exp \left\{ t\mu + \frac{1}{2} t^2 \sigma^2 \right\}$	μ	σ^2	