



מרצה : ד"ר גלית יום-טוב

מתרגלת : נועה זיכלינסקי

תאריך הבחינה : 13.10.17

מס' סטודנט _____

הנדסת מערכות שירות – 096324

מועד ב' – סמסטר אביב תשע"ז 2017

פתרון

- המבחן נמשך שלוש שעות. באחריותך להחזירו במועד.
- המבחן בחומר סגור.
- מותר להשתמש במחשבון.
- המבחן כולל 30 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.

הרצאות\תרגולים\בחינות קודמות	מתוך 25 נקודות	שאלה 1	ניקוד:
תרגילי בית	מתוך 25 נקודות	שאלה 2	
יישום	מתוך 35 נקודות	שאלה 3	
תיאוריה	מתוך 15 נקודות	שאלה 4	

ס ה " כ _____

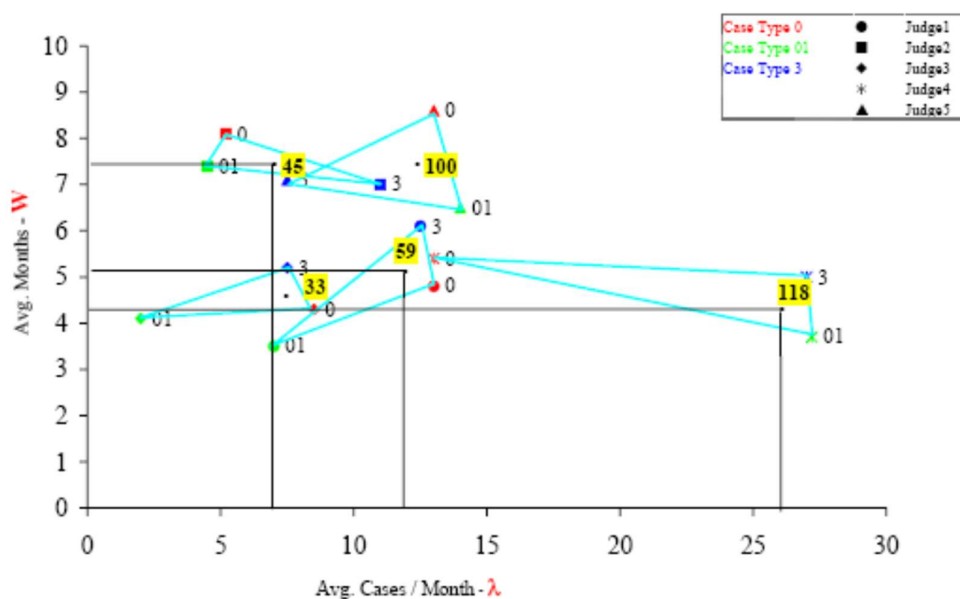
בהצלחה !

שאלה 1 - תרגילי כיתה/הרצאות/מבחנים קודמים (25 נקודות)

חלק 1 (11 נקודות)

גרף 1

Judges: The Best/Worst (Operational) Performer



התרשים הנ"ל שהוצג בהרצאה מתאר את רמת הביצועים התפעוליים של חמישה שופטים בבית הדין לעבודה.

1.1.1. (2 נקודות) מה מתארים שטחי 3 המלבנים שבתרשים? בחרו את התשובה הנכונה (אחת מארבע האפשרויות).

1. מלבן מתאר את סך זמן ההמתנה של לקוחות לכל סוג שופט בתקופות שונות.

2. מלבן מתאר את ממוצע מספר התיקים שנמצאים בטיפול אצל שופט בזמן נתון, וזאת על-פי חוק Little.

3. מלבן מתאר את סך זמן ההמתנה של לקוחות לכל אחד מסוגי התיקים בהם השופט מטפל.

4. אף תשובה לא נכונה.

תשובה: 2

1.1.2. (2 נקודות) ניתן לנסות להשוות בין רמות הביצועים התפעוליים של שלושה שופטים על סמך שטחי שלושת מלבנים שאנו רואים בתרשים (השטחים שווים ל-45, 59 ו-118). מהי מהות הקשר בין שטח המלבן ורמת הביצועים של השופט? בחרו את התשובה הנכונה (אחת מארבע האפשרויות).

1. שופט עם שטח המלבן גדול יותר הוא שופט עם רמת הביצועים הגבוהה יותר.
2. שופט עם שטח המלבן הגדול יותר הוא שופט עם רמת הביצועים נמוכה יותר.
3. לא נכון להשוות בין רמות הביצועים התפעוליים של שופטים על סמך שטחי המלבנים. לשופט עם רמת הביצועים גבוהה צריך להיות λ גבוה ו- W נמוך.
4. אף תשובה לא נכונה.

תשובה: 3

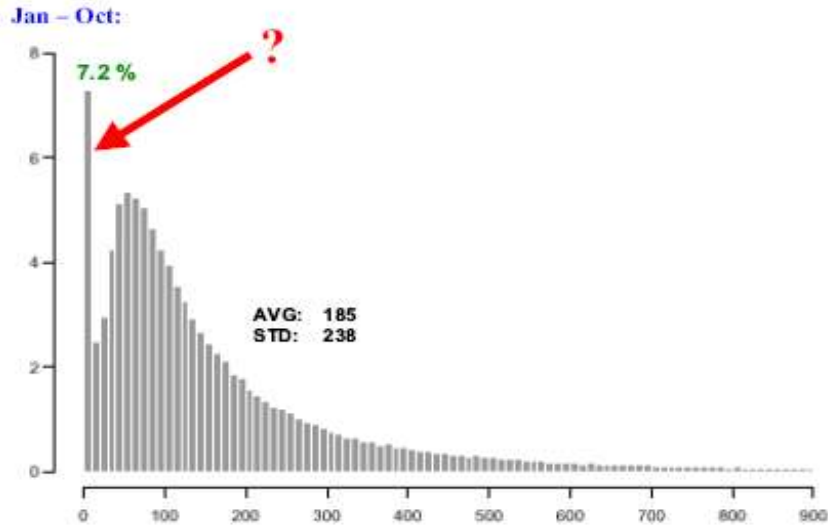
1.1.3. (3 נקודות) מבין שלושת השופטים, מיהו השופט עם רמת הביצועים התפעוליים הגבוהה ביותר? הסבירו בקצרה.

תשובה:

רמת ביצועים גבוהה מאופיינת בקצב מופע גבוה וזמן שהיה נמוך, ולכן רמת הביצועים של שופט 4 עברו $\lambda \approx 26$ ו- $W \approx 4.2$ היא הגבוהה מבין השלושה.

גרף 2

Beyond Data Averages Short Service Times



1.1.4. (2 נקודות) גרף 2 שהוצג בהרצאה מתאר משכי שירות במרכז שירות טלפוני. אנו רואים ש- 7.2% מהשיחות משכן קצר ביותר (עד 10 שניות). בחרו פיסקה שמסכמת בצורה הסבירה ביותר את הסיבות לתופעה ואת השלכותיה הניהוליות.

1. התופעה אינה בעייתית. יש לקוחות רבים שבעצם הגיעו למוקדן הלא נכון, ומשם הם מועברים הלאה, ולכן משך השירות שלהם קצר ביותר.

2. התופעה בעייתית. הגרף מראה כי תמריץ לפי מספר שיחות מטופלות גורם למוקדנים לנתק ללקוח מיד לאחר קבלת השיחה.

3. התופעה אינה בעייתית. השיחות הקצרות הן לא שיחות טלפוניות, כי אם שרות בעזרת צ'אט. בשירות זה השירות ניתן למקוטעין, וכל נותן שירות מטפל בשניים / שלושה לקוחות במקביל. ולכן משך השיחה מחולק למקטעים קצרים בהם השרת עסוק עם הלקוח, וחלק מהם הם אפילו פחות מ- 10 שניות.

4. התופעה נובעת מבעיה טכנית. מדובר כפי הנראה בתקלה במערכת הטלפונים של החברה.

תשובה: 2

1.1.5 (2 נקודות) למדנו בהרצאה שאת חווית ההמתנה של לקוח במרכז טלפוני ניתן לסכם באופן הבא:

- E - הזמן שלקוח מצפה להמתין
- τ - הזמן שלקוח מוכן להמתין
- V - הזמן שלקוח נדרש להמתין
- W - הזמן שלקוח למעשה המתין
- F - הזמן שלקוח הרגיש שהמתין

בחרו שוויון או אי שוויון שתמיד מתקיים לפי הגדרות שהוצגו בהרצאה.

1. $F \geq W$

2. $W = \min(\tau, V)$

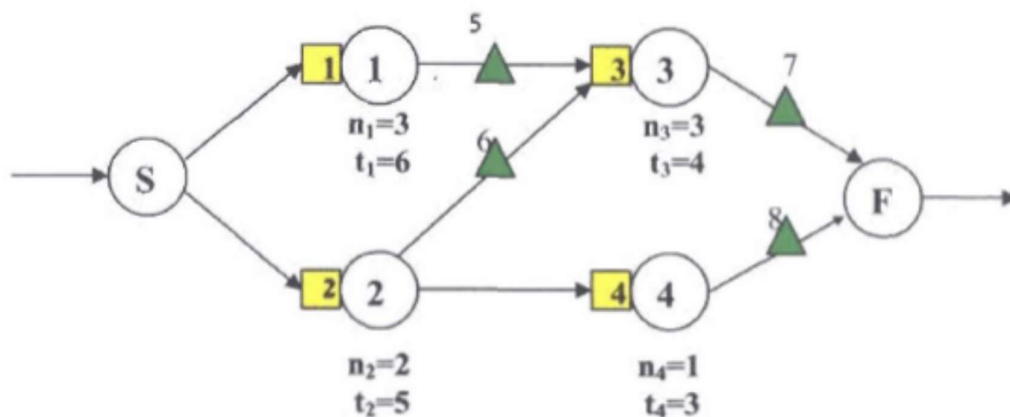
3. $W = \max(\tau, V)$

4. $W = E$

תשובה: 2

חלק 2. רשתות דינמיות סטוכסטיות (14 נקודות)

נתונה רשת PERT/CPM דינמית סטוכסטית, כמתואר בגרף 1 וכפי שנלמד בכיתה: t_i הוא ממוצע זמן השירות בתחנה i , ו- n_i הוא מספר השרתים בתחנה i , $i = 1, 2, 3, 4$. פרויקטים חדשים מופיעים למערכת לפי תהליך פואסון עם ממוצע זמן בין מופעי של 3.5 ימים. נניח שהרשת במצב יציב.



Graph Number 1

1.2.1 (2 נקודות) לרשת 8 תורים הממוספרים מ-1 עד 8. מיהם התורים הנובעים מפערי סינכרון ומי הנובעים ממשאבים מוגבלים?

תשובה:

תורי סינכרון: 5,6,7,8
תורי משאבים: 1,2,3,4

1.2.2 (4 נקודות) מהו המספר המינימלי של שרתים הנדרשים לטפל בפעילות 3 כדי שנצילות כל אחד לא תעלה על 33%?

תשובה:

מספר מינימלי: $\lceil 4 / (3.5 * n_3) \rceil \leq 0.33$. התוצאה היא 4 שרתים.

1.2.3 (4 נקודות) נסמן ב- T_{DS} את משך ביצוע פרויקט שמגיע למערכת במצב יציב. הסבירו במדויק ככל שתוכלו, אבל בשני משפטים לכל היותר, מדוע $ET_{DS} < 10$.

תשובה:

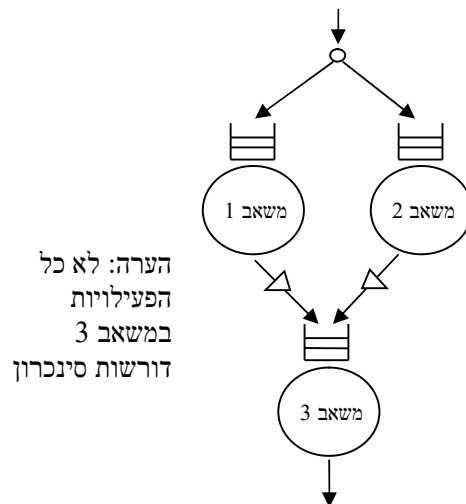
ההסבר כולל שני חלקים. מכיוון שהרשת סטוכסטית ועל סמך מבנה הקדימויות, אנחנו כבר יודעים ש- $10 \leq ET_{DS}$ (ראו הוכחה בהמשך). בנוסף, הדינאמיות של המערכת (ריבוי הפרויקטים) גורמת לתורי משאבים שמאריכים עוד יותר את משך ביצוע הפרויקט.

$$E\left(\max_j \sum_{i \in j} T_i\right) \geq \max_j E\left(\sum_{i \in j} T_i\right) = \max_j \sum_{i \in j} ET_i = 10$$

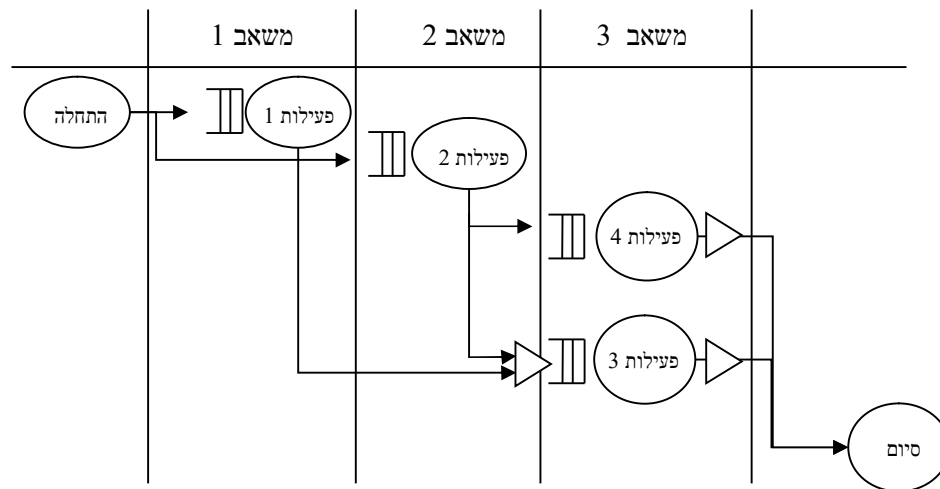
כאשר j מונה את המסלולים האפשריים ברשת.

1.2.4 (4 נקודות) מנהל החברה הכשיר את השרתים שמבצעים את פעילות 3 ו-4 לבצע את שתי הפעילויות. צייר תרשים משאבים ותרשים משולב פעילויות ופעילויות לפרויקט תחת תרחיש חדש זה.

תרשים משאבים:



תרשים משולב פעילויות-משאבים:



שאלה 2 – תרגילי בית (25 נקודות)

איוש בעזרת 4CC

ארבעה עשר מוקדנים עובדים במרכז השירות של GazolCo. רוב השיחות הן של לקוחות המתקשרים לשלם את חשבון הגז שלהם. מנתוני שיחות שניתחנו עולה כי זמן שירות טלפוני ממוצע הוא 4 דקות, קצב השיחות הנכנסות הוא 180 שיחות לשעה בממוצע, והסבלנות הממוצעת היא 3 דקות. אנו רוצים לעזור למנהל השירות לנתח את המערכת ולקבוע איוש מתאים בעזרת תוכנת 4CC.

2.1. (3 נקודות) מנהל המערכת רצה לבדוק מה יקרה אם קצב השיחות הנכנסות יגדל (בעוד מספר השרתים נשאר קבוע). להלן שתי טבלאות המתארות מה יקרה במצבים אלו. הטבלאות מבוססות על פלטים של התוכנה, אחד הופק בעזרת מודל $M/M/N$ והשני בעזרת מודל $M/M/N+M$. איזה מהטבלאות מתאימה לאיזה מודל? הסבירו תשובתכם.

טבלה 1

Calls Per Hour	Average Speed of Answer (secs)	Agent's Occupancy (%)	$P(\text{Wait} > 0)$
180	57.8	85.7%	$1 - 0.518 = 0.482$
185	79.6	88.1%	$1 - 0.447 = 0.553$
190	113.5	90.5%	$1 - 0.37 = 0.63$
195	171.3	92.9%	$1 - 0.286 = 0.714$
200	289.2	95.2%	$1 - 0.197 = 0.803$
205	647.0	97.6%	$1 - 0.101 = 0.899$
210	n/a	n/a	

טבלה 2

Calls Per Hour	Average Speed of Answer (secs)	Agent's Occupancy (%)	$P(\text{Wait} > 0)$
180	9.2	80.7%	$1 - 0.699 = 0.300$
185	10.6	82.2%	$1 - 0.666 = 0.333$
190	12.0	83.7%	$1 - 0.632 = 0.367$
195	13.5	85.1%	$1 - 0.598 = 0.401$
200	15.1	86.3%	$1 - 0.564 = 0.435$
205	16.8	87.5%	$1 - 0.530 = 0.469$
210	18.6	88.7%	$1 - 0.497 = 0.502$

תשובה:

טבלה 1 הופקה בעזרת מודל $M/M/N$ וטבלה 2 בעזרת מודל עם נטישות $M/M/M+M$. אנו מגלים זאת כי המודל עם הנטישות תמיד יציב, בעוד המודל ללא נטישות לא יציב אם קצב הגעת שיחות גדול מ- 210 שיחות לשעה.

2.2. (3 נקודות) הסבירו מדוע בשורה האחרונה של טבלה 1 מופיע הערך μ/a . מהו קצב ההגעה המקסימלי בו תתקבל תשובה מספרית.

תשובה:

מודל Erlang-C יציב כל עוד $\mu > \lambda$, ולכן $60/4 \cdot 14 = 210$ עבור כל קצב הגעה קטן ממש מ-210 לקוחות לשעה נקבל שהמערכת יציבה והתוכנה תחשב מדדי שירות, ועבור כל ערך גדול או שווה ל-210 נקבל שהמערכת לא יציבה.

2.3. (4 נקודות) ציינו מה אתם מצפים שיקרה למדדים הבאים, במידה ופרמטר הסבלנות θ יעלה: אחוז הנוטשים, ממוצע זמן ההמתנה לשירות (ASA), ממוצע אורך התור, נצילות השרתים. סמנו את התשובה הנכונה והסבירו תשובתכם בקצרה.

תשובה: כאשר פרמטר הסבלנות עולה:

אחוז הנוטשים: יעלה / ירד

ממוצע זמן ההמתנה לשירות (ASA): יעלה / ירד

ממוצע אורך התור: יעלה / ירד

נצילות השרתים: תעלה / תרד

הסבר: כאשר פרמטר הסבלנות עולה, הסבלנות הממוצעת של לקוח יורדת ולכן יותר לקוחות נוטשים. כתוצאה מכך פחות לקוחות יישארו לחכות לשירות מה שיגרום לתור לקטון ולעומס לרדת. לכן, גם זמן ההמתנה ונצילות השרתים ירדו.

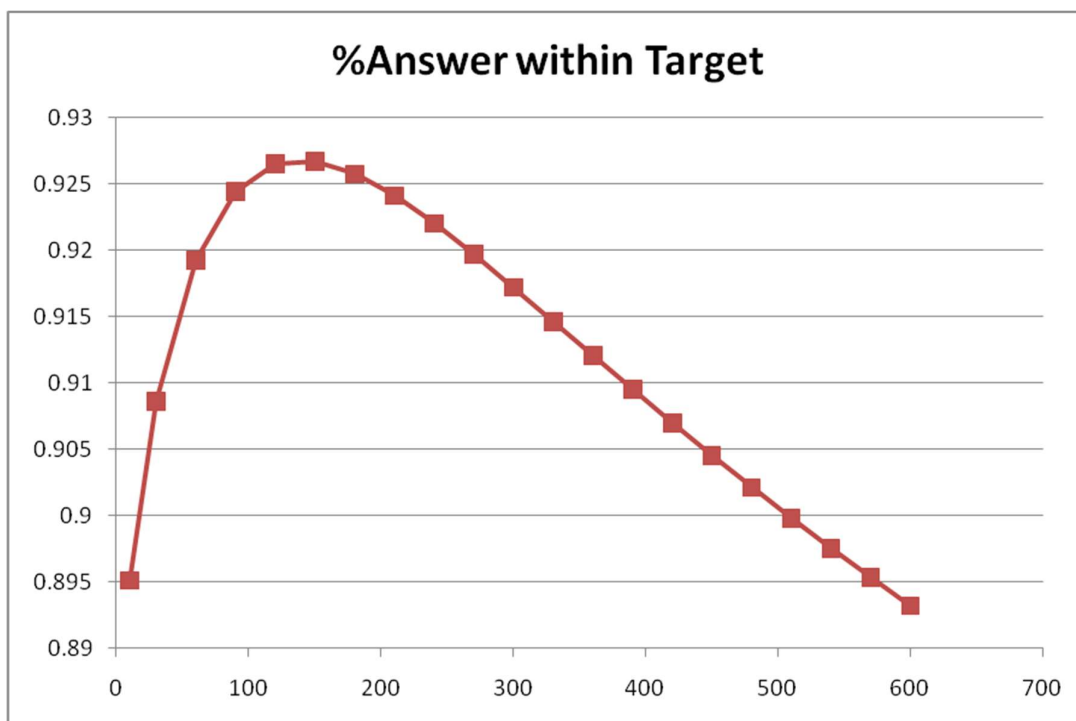
2.4. (4 נקודות) מה יקרה לאחוז הנוטשים אם סבלנות הלקוחות תעלה לאינסוף או תרד ל-0? לאיזה מודל המערכת תשאף במקרים אלו? הסבירו תשובתכם.

תשובה:

$\theta \rightarrow \infty$: כאשר פרמטר הסבלנות עולה לאינסוף, ממוצע הסבלנות שואף ל-0, ולכן כל לקוח שמגיע כאשר אין שרת פנוי עוזב מיד את המערכת. כלומר, % הנוטשים בגבול יהיה שווה לאחוז הלקוחות המגיעים לשרת עסוק - $P(Abandon) \rightarrow P(W > 0)$. במצב זה המערכת שואפת למערכת M/M/N/N (Erlang-B).

$\theta \rightarrow 0$: כאשר פרמטר הסבלנות יורד ל-0, ממוצע הסבלנות שואף לאינסוף, אף לקוח לא נוטש (כלומר, $P(Abandon) \rightarrow 0$), ולכן המערכת שואפת למערכת M/M/N (Erlang-C).

2.5. (7 נקודות) כאשר בודקים את התנהגות של "אחוז הנענים תוך 1.5 דקות" כפונקציה של ממוצע הסבלנות מקבלים את הגרף הבא:

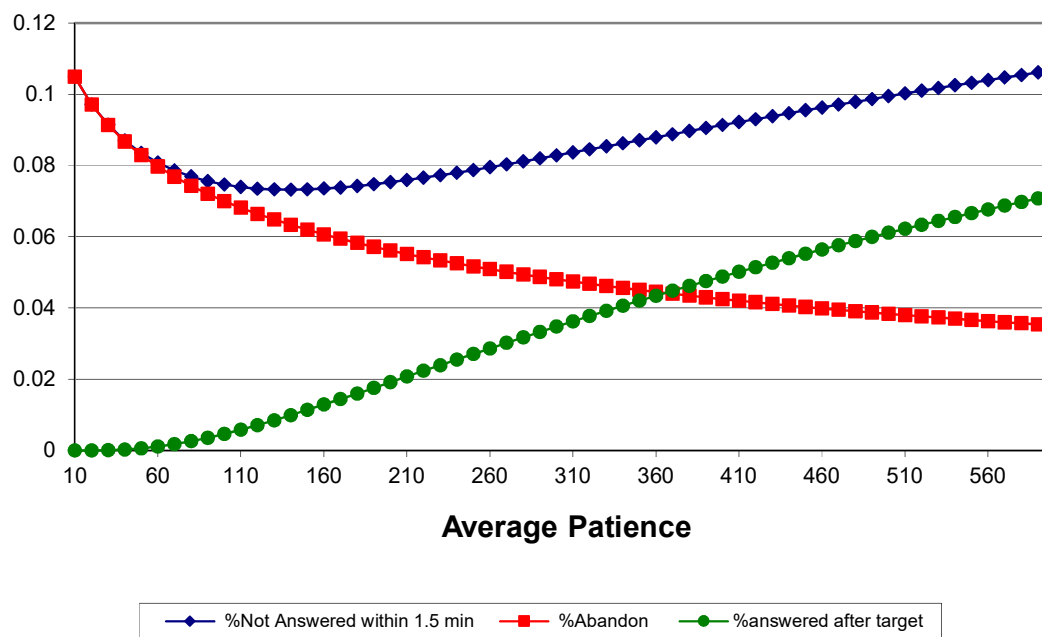


הסבירו מדוע לגרף צורה כזו.

רמז: ניתן לחלק את הלקוחות שלא נענים תוך 1.5 דקות לשתי קבוצות. הגדירו מהן, וכיצד הן עוזרות להסברת התופעה הנ"ל.

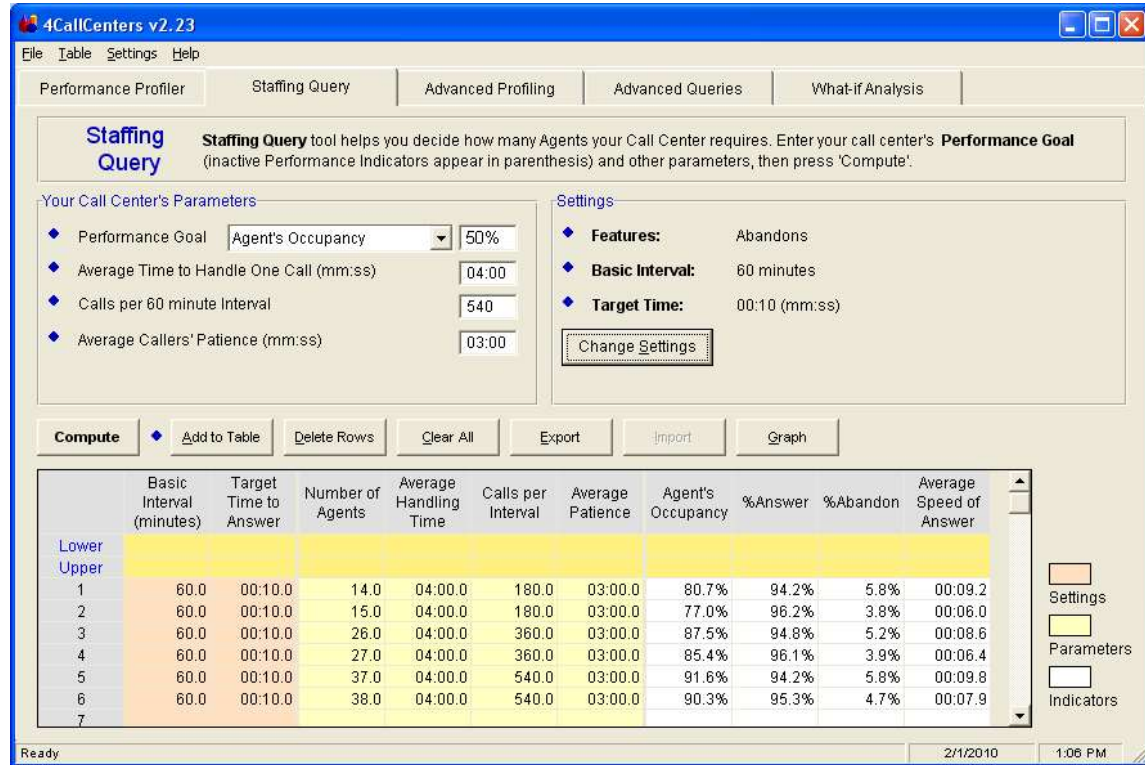
ניתן לחלק את הלקוחות שלא נענים תוך 1.5 דקות, ללקוחות שנטשו (לפני או אחרי 1.5 דקות) וללקוחות שלא נטשו וקיבלו שרות לאחר יותר מ- 1.5 דקות.

אחוז הלקוחות הנוטשים יורד כפונקציה של ממוצע הסבלנות. אחוז הלקוחות שלא נטשו ומקבלים שרות לאחר יותר מ- 1.5 דקות עולה ככל שממוצע הסבלנות עולה (כיוון שאם ללקוחות יש יותר סבלנות יהיו יותר לקוחות בתור $\leq$ זמני ההמתנה יעלו $\leq$ אחוז אלו שלא נוטשים ומקבלים שרות לאחר יותר מ- 1.5 דקות עולה). לכן, חיבור שתי הפונקציות המונוטוניות הנ"ל גורם לכך שאחוז הלקוחות שלא קיבלו שירות תוך 1.5 דקות כפונקציה של ממוצע הסבלנות יראו לפי הגרף למטה. מכאן נובע שאחוז הלקוחות שכן קיבלו שרות תוך 1.5 דקות היא פונקציה כפי שרואים בשאלה.



2.6. (4 נקודות) מנהל המוקד גילה שבימים מסוימים (לאחר שליחת חשבונות ללקוחות), קצב השיחות למוקד גדול פי כמה מימים רגילים. ליתר דיוק, ביום לאחר שליחת החשבונות קצב ההגעה הוא 540 לקוחות לשעה (פי 3 מיום רגיל) וביום למחרת 360 לקוחות לשעה (פי 2 מיום רגיל); לאחר מכן הקצב חוזר לרמתו הרגילה.

מנהל המוקד רוצה לחשב את כמות המוקדנים הדרושה כך שאחוז הנוטשים ישאר ברמתו של יום רגיל (קרי כ- 5%). לשם כך הוא יצר בתוכנת 4CC את הפלט הבא:



הוא טוען שניתן לראות שבמערכת שלנו קיים יתרון לגודל באיזו. האם הוא צודק? הסבירו.

תשובה:

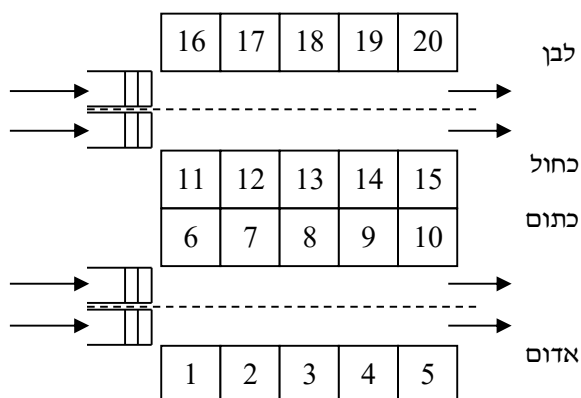
המנהל צודק. כפי שניתן לראות מהנתונים, כאשר קצב ההגעות עולה פי 2 מספר השרתים עולה רק פי 1.8; כאשר קצב ההגעות עולה פי 3, מספר השרתים הדרוש על מנת לשמור על אותה רמת שירות עולה רק פי 2.5. זוהי בדיוק התופעה של "יתרון לגודל".

שאלה 3 – יישום (35 נקודות)

השאלה עוסקת בסופרמרקט בניו-יורק השייך לרשת הסופרמרקטים Whole Foods היוקרתית המוכרת מוצרי בריאות. מצורפות שתי תמונות מהתור לקופות ביציאה.

הקופות ממוקמות במקום אחד בחנות. ישנם ארבעה תורים במתחם הקופות ולכל אחד מהם יש צבע: כתום, כחול, לבן, אדום. כשקופה מתפנה, מופיע מספר הקופה אליה יש לגשת, על גבי מסך מעל התור, וזאת לפי צבע התור בו עומדים הלקוחות (כפי שרואים בתמונה): התור הכתום לקופה 9, התור האדום לקופה 12.

הקופות מסודרות בשתי קבוצות גדולות, כך שבכל קבוצה עשר קופות עם מעבר ביניהן – חמש מכל צד, לפי האירור המצורף.



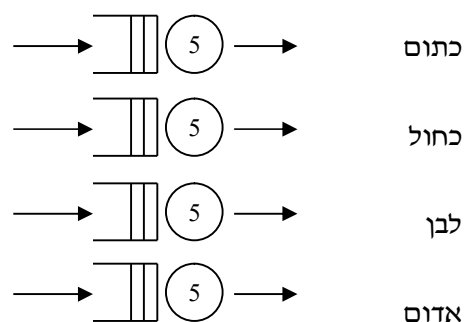
בשאלה זו ננתח תפעולית את תכנון הקופות בסופרמרקט זה.

נניח כי תהליך הגעת לקוחות לתור הוא תהליך פואסון עם קצב λ , ושמשי השירות בקופה מתפלג מעריכית עם קצב μ . כמו כן נניח, בשלב זה, כי תוחלת משך השירות היא 5 דקות. בנוסף נניח כי הזמן שלוקח ללקוח לאתר את הקופה הפנויה ולהגיע אליה הוא זניח.

3.1. (2 נקודות) תארו גרפית את תצורת תורי הקופות בסופרמרקט.

תשובה:

בהנחה שלקוחות לא עוברים בין התורים, התרשים יראה כך:



נתמקד מעתה באחד התורים (לדוגמה התור ה"כחול"). נסמן ב- $Q(t)$ את מספר הלקוחות במערכת השירות בזמן t . $Q(t)$ כולל את הלקוחות הממתינים בתור הכחול, כמו גם את הלקוחות שמקבלים כרגע שירות בעמדות הקשורות לתור זה). אזי, $Q = \{Q(t), t \geq 0\}$ הוא תהליך לידה ומוות תורי, תחת הנחות מתאימות, בפרט זמן שירות מעריכי. נניח מעתה שהנחות אלו אכן מתקיימות.

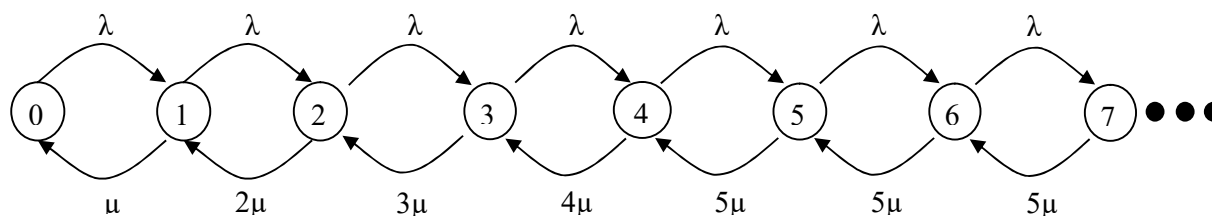
3.2. (2 נקודות) איזה מודל תורי שנלמד בכיתה מתאים לתיאור התור הנ"ל?

תשובה:

בהנחה שלקוחות לא נוטשים את התור (לא כדי לעזוב את הסופר ולא כדי לעבור לתור אחר), כל אחד מהתורים יכול להיות מתואר ע"י מודל Erlang-C, כלומר תור M/M/5.

3.3. (2 נקודות) שרטטו את דיאגרמת קצבי המעבר של Q . היעזרו בסימונים הבאים: λ – קצב מופע לקוחות לתחנה. μ – קצב שירות בתחנה. N – מספר קופאים הנותנים שירות לתור הכחול.

תשובה:



3.4. (2 נקודות) מדוע לדעתכם נוצר הצורך לקבוע לכל תור צבע ולהכריז במסך על מספר העמדה שהתפנתה? כיצד מערכת זו משפיעה על תפעול המערכת?

תשובה:

הגדרת התורים והמסך אמורים לשפר את זרימת הקונים. כך שלקונה יהיה קל יותר לשים לב שהתפנתה עמדה ואיזו עמדה פנויה כרגע. באופן זה זמן הבטלה של הקופאים במעבר בין קונה אחד למשנהו יורד.

3.5. (3 נקודות) הסידור המקובל של קופות בסופרמרקטים בישראל הוא שלכל קופה יש תור משלה. נניח כעת כי לסופרמרקט הישראלי מגיעים אותו מספר אנשים לשעה, משך השירות בקופה דומה ומספר הקופאות זהה לסופרמרקט האמריקאי. על פי עיקרון ה-pooling, האם בתצורה האמריקאית צפויה רמת שרות טובה יותר ללקוחות? הסבירו. בפרט, האם יהיה משך ההמתנה לשירות קצר יותר, ובכמה?

תשובה:

בסופר האמריקאי צפויה רמת שירות טובה יותר. זה נובע מתופעת ה-pooling. ברגע שלכל קבוצת קופות יש תור אחד במקום תורים נפרדים זמני ההמתנה בתור יקטנו, וההסתברות להמתין תקטן, מכיוון שכאשר מאחדים תורים לא יתכן ששרת אחד יהיה בטל בזמן שבתור אחר מישהו מחכה. על פי עקרון זה זמן ההמתנה הממוצע בתור יקטן בערך פי 5. (ליתר דיוק ב- $\frac{5\rho}{E_{2,5}}$, אבל ρ ו- $E_{2,5}$ הם באותו סדר גודל).

3.6. (2 נקודות) האם בתצורה האמריקאית צפוי להיות בממוצע מספר שונה של לקוחות הממתינים בתור (בכל המערכת)? הסבירו.

תשובה:

כן. ע"ס חוק Little, אם תוחלת זמן ההמתנה קטן, גם תוחלת מספר האנשים בתור קטן.

3.7. (3 נקודות) בניסוי שבוצע בסופר שפועל בדומה לסופרמרקטים בישראל, הראו שמשך ההמתנה **התאריך** כשאיחדו תורים של מספר קופות. אחת הסיבות שהועלו להסבר התופעה היתה, שלאחר איחוד הקופות נוסף זמן הליכה בין התור לקופה. בהנחה שמשך ההליכה הממוצע מהתור לקופה ידוע, כיצד ניתן לחשב את משך ההמתנה במערכת החדשה על מנת להשוותו למשך ההמתנה במערכת הקודמת?

תשובה :

במערכת החדשה משך השירות יחושב על ידי סכימת משך השירות בפועל למשך ההליכה לקופה (כיוון שבזמן המעבר, השרת בטל). שאר הפרמטרים ישארו ללא שינוי. חישוב משך ההמתנה יתבצע על פי הנוסחאות המתאימות ל Erlang-C או באמצעות 4CC.

3.8. (2 נקודות) באותו ניסוי נראה כי למרות שמשך ההמתנה התאריך הלקוחות היו יותר מרוצים במערכת עם תור אחד. הסבר על סמך העקרונות במאמר "הפסיכולוגיה של ההמתנה" מאת לרסון, מדוע תופעה זו הגיונית.

תשובה :

במאמר על הפסיכולוגיה של ההמתנה נטען שלקוחות תופסים המתנה צודקת יותר כהמתנה קצרה יותר. ההמתנה בתור אחד היא FCFS והסיכוי שמישהו יעקוף אותך נמוך, ולכן היא יכולה להיתפס כצודקת יותר ולכן קצרה יותר. כתוצאה מכך, הגיוני שלקוחות יהיו יותר מרוצים.

3.9. (3 נקודות) אמדו בעזרת הגרף המצורף מה יהיה אחוז הלקוחות המשורתיים מיידית עבור אינטרוול זמן שבו : קצב הגעה – 60 לקוחות בשעה, זמן שרות ממוצע 10 דקות, 12 קופאיות נותנות שירות.

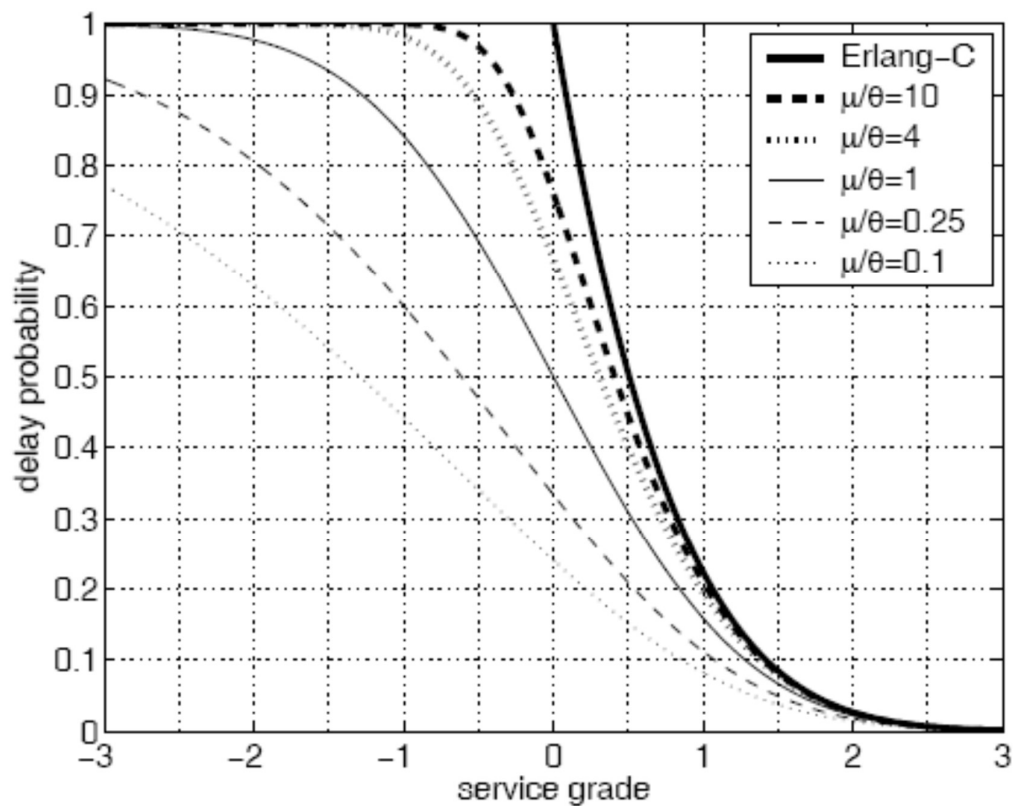
תשובה:

תחילה נחשב את ה- Offered Load : $R=60/6=10$

עתה נחשב את פרמטר השירות :

$$\beta = \frac{N-R}{\sqrt{R}} = \frac{12-1}{\sqrt{10}} = 0.63$$

על פי הגרף ההסתברות להמתנה תהיה כ- 0.4, ולכן אחוז המקבלים שרות מידי הוא 60%.



3.10. (4 נקודות) מנהל הסופר רצה לחשב תחזית למספר הקונים בחנות בכל שעה. מניסיונו הוא יודע שמספר האנשים משתנה בין שעות היום ובין ימות השבוע. המנהל יכול לאסוף בקלות מידע מהקופות על מספר הקונים שרכשו בחנות בכל שעה בשבוע ברבעון האחרון. הוא החליט לחלק את היממה לאינטרוולים של חצי שעה.

נגדיר: $N_{i,j}$ – הגעות בפועל ביום i באינטרוול j , ו- $F_{i,j}$ – תחזית ליום i באינטרוול j . המנהל רוצה להשתמש בשיטת הממוצע הנע, על פני 4 המדידות האחרונות, כאשר ככל שהמדידה מעודכנת יותר (קרובה יותר לזמן התחזית) היא תקבל משקל גבוה יותר בחישוב התחזית. הגדירו והדגימו כיצד יש לבצע את החישוב לחיזוי מספר הקונים הצפוי בכל שעה על פי מודל זה. השתמשו בסימונים: $F_{i,j}$, $N_{i,j}$.

תשובה:

בשיטת הממוצע הנע המשוקלל, נוכל לדוגמה להשתמש בנוסחה הבאה:

$$F_{i,j} = 0.5 * N_{i-7,j} + 0.3 * N_{i-14,j} + 0.15 * N_{i-21,j} + 0.05 * N_{i-28,j}$$

יש לוודא שסכום המשקולים הוא 1 ושהם מסודרים בסדר יורד: המדידה הקרובה ביותר תקבל את המשקל הגבוה ביותר. כדי לחשב את מספר הקונים בשעה יש לסכום שני אינטרוולים.

הידיעה הבאה פורסמה בעיתון "הארץ" לאחרונה (אין צורך לקרוא את כל הכתבה על מנת לענות על השאלות).

חדשות חינוך וחברה

השרים אישרו: קשישים מעל גיל 80 לא יעמדו בתור לקבלת שירות במקומות ציבוריים

ההצעה תחול בין היתר על תורים לקולנוע, לאצטדיונים ולחנויות ממכר מזון. בתורים לבדיקה רפואית לא יחול החוק

שמור 82

21:04 | 18:17 19.06.2016
יהונתן לוי | התראות במייל

ועדת השרים לחקיקה החליטה היום (ראשון) לתמוך בהצעת חוק שלפיה קשישים מעל גיל 80 לא יעמדו בתור לקבלת שירות במקומות ציבוריים. ההצעה לא תחול על תור לבדיקה רפואית, כדי שלא לקבוע הקדמת תור על סמך גילו של אדם במקום שמצבו הבריאותי יהיה השיקול הקובע. קיצור התורים לא יחול גם על המתנה ברכב בכניסה לגן לאומי או לתחנת דלק.

על פי הנוסח שיובא כעת להצבעה במליאת הכנסת, אזרחים ותיקים מעל גיל 80 יוכלו, אם ירצו בכך, לקבל מגוון שירותים ציבוריים בלא צורך להמתין בתור ככלל האזרחים. המקומות בהם יינתן הפטור בין היתר הם רשות הדואר, בנקים, בתי קולנוע, אולמות מופעים, אצטדיונים, גנים לאומיים ושמורות טבע, חנויות גדולות לממכר מזון ועוד. בנוסף, לשרה לשווין חברתי תינתן האפשרות להוסיף בעתיד מקומות נוספים בהם ייקבלו האזרחים והאזרחיות מעל גיל 80 פטור מעמידה בתורים.

בעליו סופרמרקט ישראלי מעוניין להעריך לקראת השינוי המדובר בכתבה. עפ"י ניתוח נתוני הקניות בסופר, הוערך ש- 15% מהקונים בסופר הם קשישים מעל גיל 80 והם יקבלו עדיפות מסוג non-preemptive בתור לקופה. בניתוח הבא נתייחס רק לקופה אחת ותור אחד. לקוחות מגיעים לקופה לפי תהליך פואסון בקצב של 6 לקוחות בשעה. משך השירות בקופה מתפלג אקספוננציאלית עם ממוצע של 9 דקות ללקוח רגיל ו- 12 דקות ללקוח קשיש. העזרו בדפי העזר בסוף השאלה כדי לענות על השאלות הבאות.

3.11 (2 נקודות) חשבו ובידקו האם המערכת תשאר יציבה לאורך זמן.

תשובה:

נסמן את הלקוחות הקשישים כלקוחות מסוג 1 ואת שאר הלקוחות כלקוחות מסוג 2.
ידוע ש -

$$\lambda_1 = 0.015 \text{ per minute}, \quad m_1 = 12 \text{ minutes}, \quad \rho_1 = 0.18$$

$$\lambda_2 = 0.085 \text{ per minute}, \quad m_2 = 9 \text{ minutes}, \quad \rho_2 = 0.765$$

$$\bar{\rho} = \rho_1 + \rho_2 = 0.945 < 1$$

ולכן, המערכת תהיה יציבה.

3.12 (6 נקודות) חשבו מהו ממוצע משך ההמתנה לקופה.

תשובה:

נחשב:

$$\text{Var}(S_1) = 144, \quad E(S_1^2) = 288, \quad \text{Var}(S_2) = 81, \quad E(S_2^2) = 162.$$

תוחלת שארית משך השירות היא

$$E(R) = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2} = 9.045$$

ועל כן, ממוצע משך ההמתנה עבור לקוח מסוג 1:

$$E(W_q^1) = \frac{E(R)}{1 - \rho_1} = 11.03 \text{ minutes}$$

ועבור סוג לקוח 2 :

$$E(W_q^2) = \frac{E(R)}{(1 - \rho_1)(1 - \rho_1 - \rho_2)} = 200.55 \text{ minutes}$$

ממוצע משך ההמתנה הכולל מתקבל עפ"י הממוצע המשוקלל הבא :

$$E(W_q) = \frac{\lambda_1 E(W_q^1) + \lambda_2 E(W_q^2)}{\lambda_1 + \lambda_2} = 172.122 \text{ minutes}$$

3.12 (2 נקודות) חשבו מהו אורך התור הממוצע לקופה.

תשובה:

על פי חוק ליטל

$$E(L_q) = \lambda E(W_q) = 17.2$$

M/G/1 without priorities

$$E(W_q) = \frac{E(R)}{1-\rho}.$$

R - Expected residual service time

ρ - Prob. of arriving to a busy server

$$E(R) = (1-\rho) \cdot 0 + \rho \cdot E(S) \cdot \frac{1+C^2(S)}{2}$$

Khinchine-Pollaczek Formula:

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

M/G/1 with priorities

Steady state $\Leftrightarrow \rho \triangleq \rho_1 + \dots + \rho_K < 1$, where $\rho_k = \lambda_k m_k$.

Convenient notation: $\bar{\rho}_k = \rho_1 + \dots + \rho_k$, $1 \leq k \leq K$.

$$\bar{\rho}_0 \equiv 0$$

Note: ρ_k = fraction of time allocated by server to class k .

$1 - \rho$ = idleness/availability.

- $E(W_q^k)$ - expected waiting time of class k customer.
- $E(L_q^k)$ - expected number of waiting class k customers.
- $E(R)$ - expected residual service time.

Non- Preemptive Regime

$$1. \quad E(W_q^1) =$$

$$E(R) + m_1 E(L_q^1) = E(R) + m_1 \lambda_1 E(W_q^1) = E(R) + \rho_1 E(W_q^1)$$

$$\Rightarrow E(W_q^1) = \frac{E(R)}{(1 - \rho_1)}, \text{ as before } (K = 1).$$

$$2. \quad E(W_q^2) = E(R) + \underbrace{m_1 E(L_q^1) + m_2 E(L_q^2)}_{\text{wait due to class 1 \& 2 in queue}} + \underbrace{m_1 \lambda_1 E(W_q^2)}_{\text{wait due to class 1, arriving during wait of 2.}}$$

$$\Rightarrow E(W_q^2) = E(R) + \rho_1 E(W_q^1) + \rho_2 E(W_q^2) + \rho_1 E(W_q^2)$$

$$\Rightarrow E(W_q^2) = [E(R) + \rho_1 E(W_q^1)] / (1 - \rho_1 - \rho_2) =$$

$$\begin{aligned} \Rightarrow E(W_q^2) &= \frac{E(R)}{[(1 - \rho_1)(1 - \rho_1 - \rho_2)]} \\ \uparrow \\ &\text{substitute } E(W_q^1) \end{aligned}$$

$$k. \quad EW_q^k =$$

$$ER + m_1 \cdot EL_q^1 + \dots + m_k \cdot EL_q^k + \lambda_1 m_1 EW_q^k + \dots + \lambda_{k-1} m_{k-1} EW_q^k$$

$$\Rightarrow = ER + \rho_1 EW_q^1 + \dots + \rho_{k-1} EW_q^{k-1} + (\rho_1 + \dots + \rho_k) EW_q^k$$

$$E(W_q^k) = \frac{E(R) + \rho_1 E(W_q^1) + \dots + \rho_{k-1} E(W_q^{k-1})}{(1 - \rho_1 - \rho_2 - \dots - \rho_k)} = \frac{E(R)}{(1 - \bar{\rho}_{k-1})(1 - \bar{\rho}_k)}$$

Where

$$E(R) = (1 - \rho) \cdot 0 + \sum_{j=1}^K \rho_j \cdot m_j \cdot \frac{1 + C_j^2(S)}{2} =$$

$$= \sum_{j=1}^K \lambda_j m_j^2 \cdot \frac{E(S_j^2)}{2m_j^2} = \frac{1}{2} \sum_{j=1}^K \lambda_j E(S_j^2)$$

שאלה 4 – תיאוריה ושימושיה (15 נקודות)

4.1. (5 נקודות) בכיתה הוסבר המשפט הבא המכסה את מודל Erlang-B בתחום התפעול QED.
בפרט הוכחנו שסעיף 3 במשפט (איוש לפי כלל השורש), גורר את 1. דהיינו

$$\lim_{n \rightarrow \infty} \sqrt{n} E_{1,n} = h(-\beta)$$

כאשר h היא פונקצית קצב הסיכון של ההתפלגות הנורמאלית הסטנדרטית.
הוכיחו טענה זו תוך השלמת החלקים החסרים בהוכחה הנ"ל. (ההוכחה ניתנת בעמוד הבא).

The Erlang-B Queue in the QED-Regime

Recall the **Erlang-B Formula**:

$$E_{1,n} \triangleq \text{P}\{\text{Blocked}\} = \frac{R^n}{n!} \bigg/ \sum_{j=0}^n \frac{R^j}{j!}$$

Consider a sequence of **M/M/ n/n** queues,
indexed by the number of servers **$n = 1, 2, \dots$** .

- λ_n = arrival-rate, varies with n ;
- μ = service-rate, fixed (independent of n).
- $R_n = \lambda_n/\mu$ (Offered Load) ; $\rho_n = R_n/n$ (Load per Server);
We shall use R and ρ , without the subscript n , for simplicity.

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.

QED Erlang-B: Proof

Proof:

2 $\iff$ **3** is straightforward algebra from the definitions.

3 $\Rightarrow$ **1**. Assume $n = R + \beta\sqrt{R}$. The key observation is a **Poisson-Representation** of the Erlang-B Formula:

$$E_{1,n} = \frac{\mathbb{P}\{X_R = n\}}{\mathbb{P}\{X_R \leq n\}},$$

where $X_R \stackrel{d}{=} \text{Poisson}(R)$.

$$\begin{aligned} \mathbb{P}\{X_R \leq n\} &= \mathbb{P}\left\{\frac{X_R - R}{\sqrt{R}} \leq \frac{n - R}{\sqrt{R}}\right\} \\ &\stackrel{CLT,3}{\approx} \mathbb{P}\{N(0, 1) \leq \beta\} = \Phi(\beta). \end{aligned}$$

$$\begin{aligned} \mathbb{P}\{X_R = n\} &= \mathbb{P}\{n - 1 < X_R \leq n\} \\ &= \mathbb{P}\left\{\frac{n - R - 1}{\sqrt{R}} < \frac{X_R - R}{\sqrt{R}} \leq \frac{n - R}{\sqrt{R}}\right\} \\ &\approx \mathbb{P}\left\{\beta - \frac{1}{\sqrt{R}} \leq N(0, 1) \leq \beta\right\} \\ &\approx \frac{1}{\sqrt{R}} \cdot \phi(\beta) \approx \frac{1}{\sqrt{n}} \cdot \phi(\beta). \end{aligned}$$

Finally, $\frac{\phi(\beta)}{\Phi(\beta)} = \frac{\phi(-\beta)}{1 - \Phi(-\beta)} = h(-\beta)$, by the symmetry of $N(0, 1)$.

4.2. נניח קצב מופע λ , וקצב שירות μ .
 בכיתה נאמר שחישוב נאיבי של איוש n , ממליץ:

$$n = \left\lceil \frac{\lambda}{\mu} \right\rceil = \lceil R \rceil$$

(עיגול של R כלפי מעלה לשלם)

4.2.1. (3 נקודות) למה תוביל המלצה זו במודל Erlang-C? בפרט מה תהיה (בערך) ההסתברות להמתין?

תשובה:

המלצה זו תוביל לאיוש בתחום ה-ED. בפרט, ההסתברות להמתין תשאף ל-1.
 אם R שלם בעצמו המערכת לא תהיה ארגודית, ולכן תתפוצץ. אם R אינו שלם, אזי
 $n = R + \gamma$, $\gamma = \lceil R \rceil - R$, ולכן ממוצע זמן ההמתנה יהיה $\gamma \cdot E[S]$.

4.2.2. (4 נקודות) למה תוביל המלצה זו במודל Erlang-A? בפרט, נניח $\mu = \theta$, מה תהיה (בערך) ההסתברות להמתין במקרה זה? הסבר ללא שימוש בגרף מהשאלה הקודמת.

תשובה:

המלצה זו תוביל לאיוש בתחום ה-QED. בפרט, ההסתברות להמתין תשאף לערך בין 0 ל-1, הערך המדויק תלוי ביחס בין μ ל- θ . במקרה ש- $\mu = \theta$, הערך ישאף ל-0.5.
 הסבר ל-0.5: במקרה ש- $\mu = \theta$, התפלגות סך מספר הלקוחות במערכת במודל Erlang-A זהה להתפלגות סך מספר הלקוחות במערכת במודל $M|M|\infty$ עם קצב מופע λ וקצב שירות μ , כאשר התפלגות סך מספר הלקוחות במערכת במודל $M|M|\infty$ היא $Poisson(R = \lambda / \mu)$.
 מהחוק החזק של המספרים הגדולים, עבור R מספיק גדול, ניתן לקרב את ההתפלגות $Poisson(R)$ על-ידי התפלגות $Normal(R, R)$. ומכאן,

$$\begin{aligned} P(W_q = 0) &\stackrel{PASTA}{=} P(L \leq R) = P\left(\frac{L-R}{\sqrt{R}} \leq \frac{R-R}{\sqrt{R}}\right) \approx \\ &= P(N(0,1) \leq 0) = 0.5 \end{aligned}$$

4.2.3. (3 נקודות) בכיתה הוסבר למה שאלה 4.2.2 מהווה דוגמה למקרה שבו "מנהלים מקבלים תשובה נכונה מהסיבות הלא נכונות" – הסבירו במדויק ככל שתוכלו.

תשובה:

המנהל פועל בצורה נאיבית הוא מכפיל את מספר האנשים שמגיעים בשעה בממוצע זמן השירות וכך מחליט כמה מוקדנים לאיש. הסיבה שזה עובד טוב למרות השגיאה היא שיש נטישות, שמייצבות את המערכת ושומרות על רמת שרות טובה.

Table 1: Discrete and Continuous Distributions

Distribution	Notation	Probability Function	Parameters	Support	Moment Gen. Function	Expectation	Variance	Comments
Binomial	$N \sim \text{Bin}(n, p)$	$p(k) = \binom{n}{k} p^k q^{n-k}$	$0 \leq p = 1 - q \leq 1; n \geq 1$	$k = 0, 1, \dots, n$	$[p \cdot e^t + q]^n$	np	npq	
Multi-nomial	$X \sim \text{Mult}(n, p)$	$p(x) = \frac{n!}{x_1! \dots x_k!} \prod_{i=1}^k p_i^{x_i}$ $x = (x_1, \dots, x_k)$	$0 \leq p_i; \sum_{i=1}^k p_i = 1; n \geq 1$ $p = (p_1, \dots, p_k)$	$0 \leq x_i; \sum_{i=1}^k x_i = n$		$E(X_i) = np_i$ $i = 1, \dots, k$	$\text{cov}(X_i, X_j) = -np_i p_j$ $(i \neq j)$	$X_i \sim \text{Bin}(n, p_i)$
Negative Binomial	$X \sim NB(m, p)$	$p(k) = \binom{k-1}{m-1} p^m q^{k-m}$	$0 \leq p = 1 - q \leq 1; m \geq 1$	$k = m, m+1, \dots$	$\left[\frac{p \cdot e^t}{1 - q \cdot e^t} \right]^m$	$\frac{mp}{p}$	$\frac{mq}{p^2}$	Number of trials till m -th success
Geometric	$X \sim G(p)$	$p(k) = q^{k-1} \cdot p$	$0 \leq p \leq 1$	$k = 1, 2, \dots$	$\frac{p \cdot e^t}{1 - q \cdot e^t}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$X \sim NB(1, p)$
Poisson	$X \sim \mathcal{P}(\lambda)$	$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda \geq 0$	$k = 0, 1, \dots$	$\exp\{-\lambda(1 - e^t)\}$	λ	λ	
Hyper-geometric	$X \sim HG(N, M, n)$	$p(k) = \frac{\binom{n}{k} \binom{M-k}{n-k}}{\binom{N+M}{n}}$	$N, M > 0;$ $1 \leq n \leq N + M$	$\max(0, n - M) \leq k$ $k \leq \min(n, N)$	$\sum_k e^{tk} p(k)$	$\frac{p \cdot N}{N+M}$	$n \cdot \frac{N}{N+M} \cdot \frac{M}{N+M} \cdot \frac{N}{N+M-1}$	N – special elements M – ordinary ones n – sample size k – special in sample
		Density 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	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\alpha = \beta = 1 \rightarrow \text{Uniform}$
Gamma	$X \sim \text{Gamma}(\alpha, \lambda)$	$f(t) = \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)}$	$\alpha, \lambda > 0$	$t \geq 0$	$(1 - \frac{t}{\lambda})^{-\alpha}$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$	$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ $= (x-1)\Gamma(x-1)$
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	$X \sim N(\mu, \sigma^2)$	$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\}$	$-\infty < \mu < \infty; \sigma > 0$	$-\infty < t < \infty$	$\exp\left\{t\mu + \frac{1}{2}t^2\sigma^2\right\}$	μ	σ^2	