



מרצה : פרופסור אבישי מנדלבאום
מתרגל : איתמר זייד

תאריך הבחינה : 26.09.2011

מס' סטודנט _____

שם _____

הנדסת מערכות שירות – 096324

מועד ב' – סמסטר אביב תשע"א 2011

- המבחן נמשך **שלוש** שעות. באחריותך להחזירו במועד.
- **המבחן בחומר סגור.**
- **מותר להשתמש במחשבון.**
- המבחן כולל **27** 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.

ניקוד: שאלה 1	_____ מתוך 10 נקודות	תרגילי בית
שאלה 2	_____ מתוך 10 נקודות	הרצאות\תרגולים\בחינות קודמות
שאלה 3	_____ מתוך 20 נקודות	יישום
שאלה 4	_____ מתוך 10 נקודות	תיאוריה

סה"כ _____ מתוך 50 נקודות

בהצלחה !

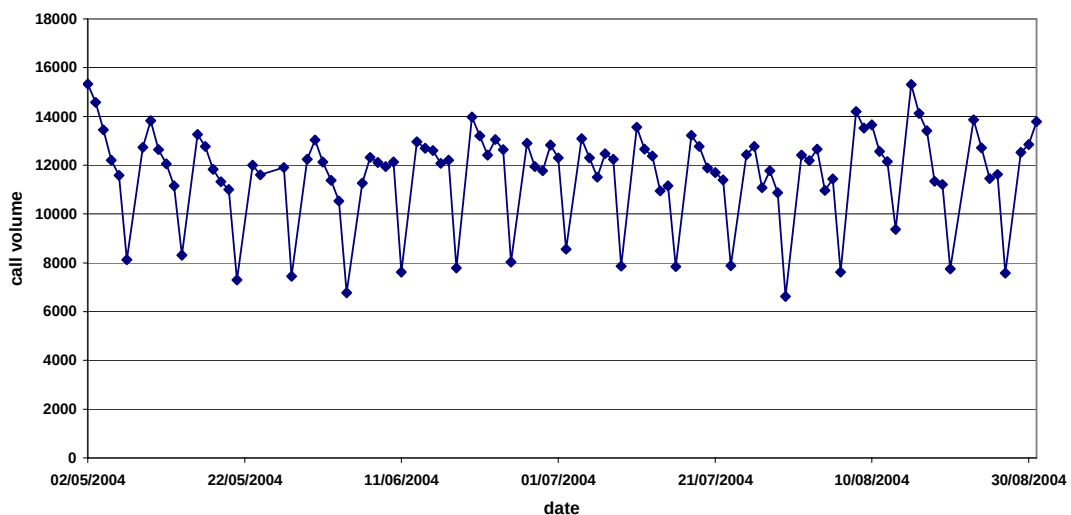
שאלה 1 בחינות קודמות (10 נקודות)

שאלה 1 - חלק 1. מופעים ותחזיות.

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.

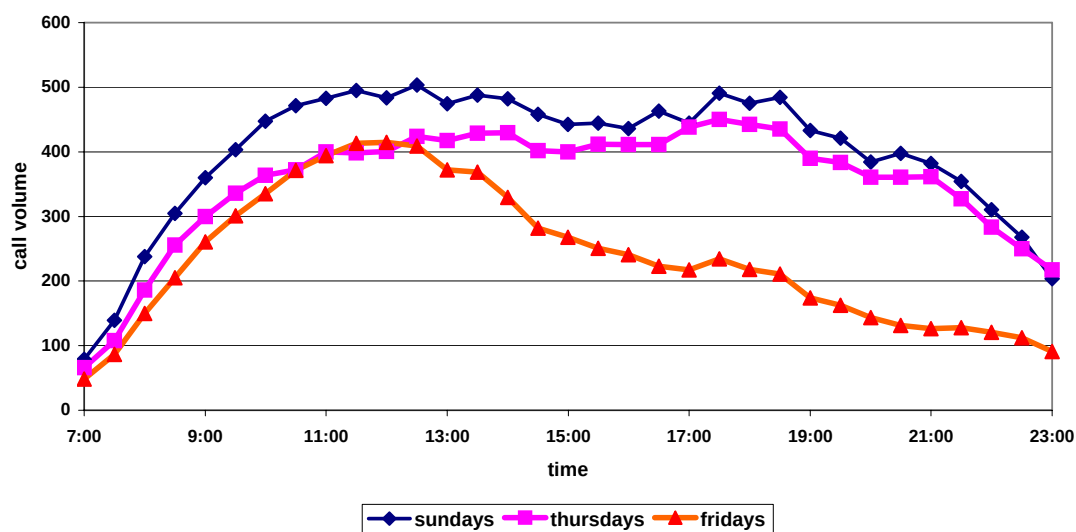
נסמן, כמו בתרגיל הבית, N_{jk} להיות מספר השיחות הנכנסות באינטרוול זמן k ביום j ($j=1, \dots, 6$), ו- F_{jk} להיות תחזית מספר השיחות לאינטרוול זה.

1.1.1 הגרף הבא מציג את סך השיחות הנכנסות ליום בתקופה 02/05/04 - 31/08/04 (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.

1.1.2 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) בצהרים בסביבות 12. יום ראשון הוא העמוס ביותר בשעות אלו (7 עד 12) מהצהרים, ימי חמישי וראשון די דומים בצורתם (עדיין ראשון עמוס יותר), אך ליום שישי צורה שונה: החל מהצהרים 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.

1.1.3 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):

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

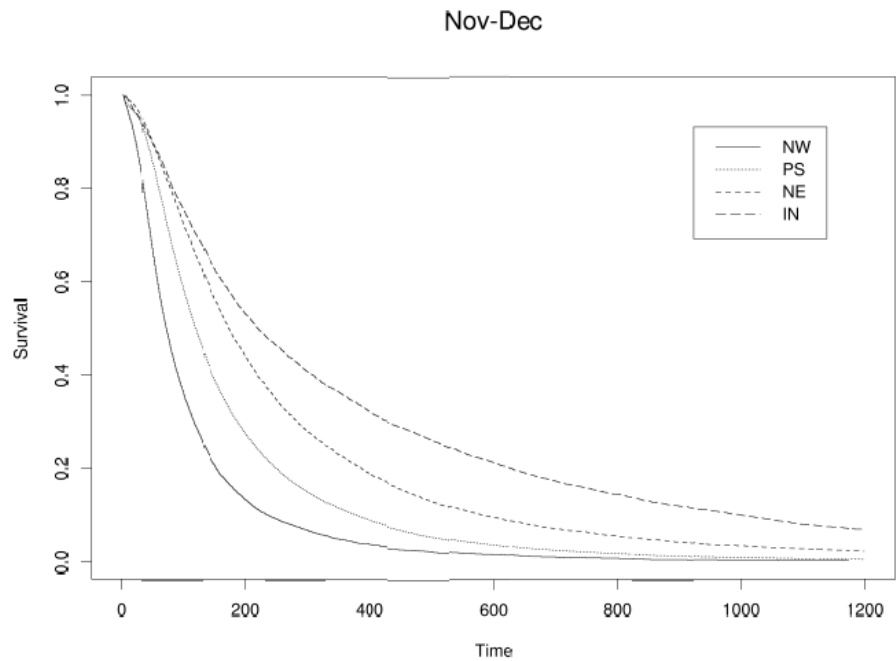
1.1.4 נניח כי המופעים למוקד השרות הם לפי תהליך פואסון עם קצב מופע λ_{jk} . נניח שהצלחנו לחזות את ערכי הפרמטר λ_{jk} במדויק ($F_{jk} = \lambda_{jk}$). האם אתם מצפים שמדדי הטיב RMSE ו- APE יהיו שווים ל- 0? הסבירו תשובתכם.

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

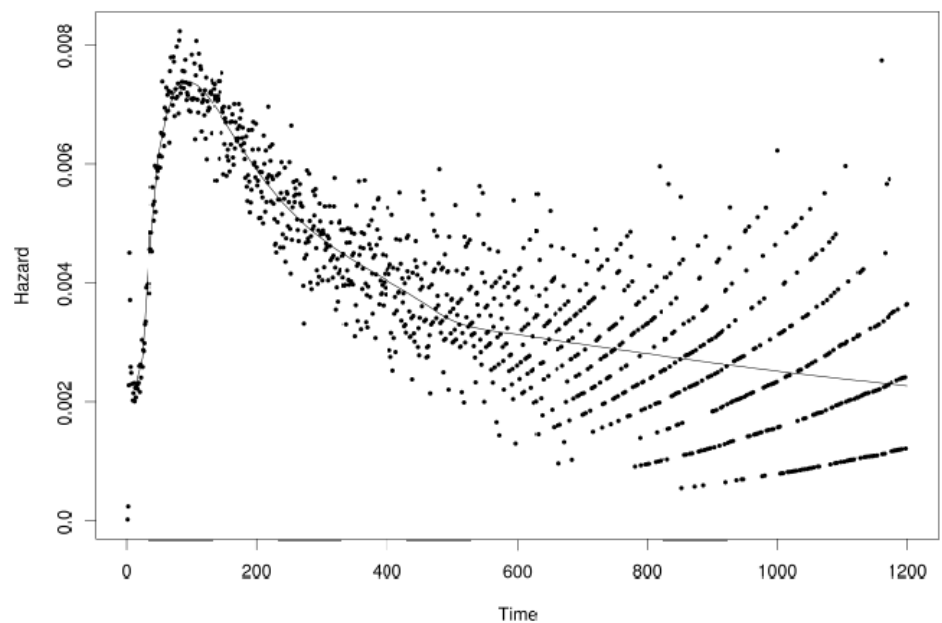
שאלה 1 - חלק 2. פונקציות שרידות ופונקציות סיכון של זמני שירות

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), ברזולוציה של שנייה אחת. הקו בגרף 2 הוא אמד לפונקצית הסיכון אחרי החלקה.

גרף 1



גרף 2



1.2.1 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 $S(x)$ של $0 \leq x$.

1.2.2 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.

1.2.3 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.

1.2.4 הסבירו את הצורה של "נקודות המסודרות על קו עולה", הבולטת בחלק הימני של גרף 2.

הערה:

האמידה של פונקציית 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.
- האמידים לפונקציית הסיכון בשניה k (נקודות בגרף 2) חושבו ע"י $\hat{h}(k) = A_k / \eta_k$.

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- $A_k = 1$).

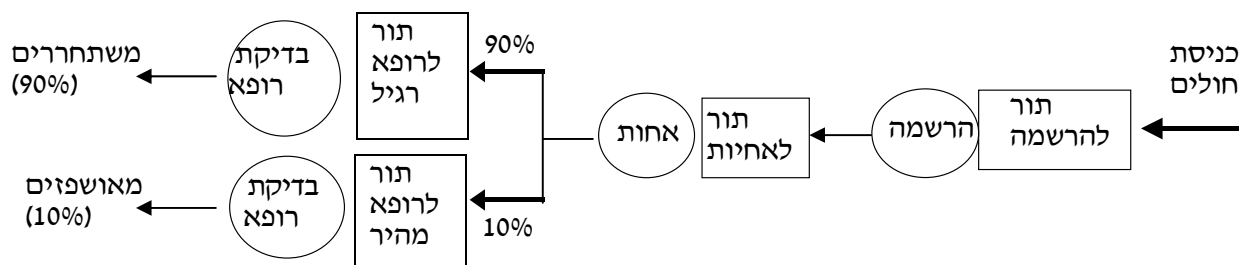
בגרפים אחרים מאותו סוג יתכן שתופענה נקודות על ציר ה- x , המתאימות ל $A_k = 0$ (שפירושו, אין משכי שירות באורך k).

שאלה 2 תרגילי בית (10 נקודות)

חוק Little

חולים הפונים לחדר מיון בבית חולים מסוים עוברים את תהליך הטיפול לפי התרשים המצורף. התהליך מתחיל בהרשמה. אחרי ההרשמה, החולים עוברים בדיקה אצל **אחיות** שמסווגות אותם לחולים **משתחררים פוטנציאליים** (הנשלחים לרופא "רגיל", מקבלים הוראות וטיפול ולא מאושפזים) או **מאושפזים פוטנציאליים** (הנשלחים לטיפול דחוף ובסוף הטיפול יאושפזו). שני סוגי החולים מחכים בתורים נפרדים לקראת בדיקת רופא. הרופא בודק את החולים. לאחר בדיקת הרופא, החולים המאושפזים נשלחים למחלקות בבית-החולים; החולים המשתחררים מקבלים טיפול והוראות להמשכו (למשל מרשמים) ואז משתחררים לביתם. נניח, בשלב זה, שאחיות המיון לא עושות טעויות וממינות 100% של החולים בצורה נכונה.

מדידות מראות שממוצע קצב הגעת חולים למיון הוא 50 חולים בשעה, מתוכם בסופו של דבר מאושפזים 10% מהפונים. תהליך הרשמה של חולה נמשך 2 דקות בממוצע. בדיקת אחות אורכת 3 דקות בממוצע. משך זמן בדיקת רופא משתנה בין שני סוגי החולים: 5 דקות למשתחררים פוטנציאליים ו- 30 דקות למאושפזים פוטנציאליים, בממוצע. נמצא גם כי בממוצע 20 חולים ממתינים בתור להרשמה ו- 5 חולים ממתינים בתור לאחות. בנוסף, תור החולים המשתחררים (הרגילים) לקראת בדיקת רופא הוא 15 חולים בממוצע ותור החולים המאושפזים (הדחופים) הוא חולה אחד בממוצע.



2.1 כמה זמן בממוצע ישהיה חולה בחדר המיון?

תשובה: נחלק את חדר המיון לשש מערכות: 1 – תור להרשמה, 2 – הרשמה, 3 – תור לאחיות, 4 – בדיקת אחיות, 5 –

תור לרופאים, 6 – בדיקת רופא.

נחשב את זמני השהיה בכל שלב ונסכם אותם:

$$\lambda = \frac{5}{6} \text{ customers per min}$$

$$L_1 = 20, W_1 = \frac{L_1}{\lambda} = 24 \text{ min}$$

$$W_2 = 2 \text{ min}$$

$$L_3 = 5, W_3 = \frac{L_3}{\lambda} = 6 \text{ min}$$

$$W_4 = 3 \text{ min}$$

$$L_5 = 15 + 1 = 16, W_5 = \frac{L_5}{\lambda} = 19.2 \text{ min}$$

$$W_6 = 5 \cdot 90\% + 30 \cdot 10\% = 7.5 \text{ min}$$

$$W = W_1 + W_2 + W_3 + W_4 + W_5 + W_6 = 61.7 \text{ min}$$

2.2 כמה חולים בממוצע יהיו סה"כ בחדר המיון?

תשובה: על סמך חוק Little:

$$L = \lambda W = \frac{5}{6} \cdot 61.7 = 51.4$$

2.3 כמה זמן בממוצע ישהו מאושפזים פוטנציאליים בחדר המיון?

תשובה: ההבדל בזמני השהיה של חולים מאושפזים למשתחררים נובע רק משלבים 5 ו-6 ולכן:

$$W_5^{hosp} = \frac{L_5^{hosp}}{\lambda^{hosp}} = \frac{1}{\frac{5}{6} \cdot 10\%} = 12 \text{ min}$$

$$W_6^{hosp} = 30 \text{ min}$$

$$W = W_1 + W_2 + W_3 + W_4 + W_5^{hosp} + W_6^{hosp} = 77 \text{ min}$$

2.4 מהו מספר הפקידים המינימלי בעמדת ההרשמה שיאפשר מצב שבו נצילותם לא תעלה על 80%.

תשובה:

$$R = \lambda E[S] = \frac{5}{6} \cdot 2 = \frac{10}{6} \text{ min offered load}$$

$$\frac{R}{n} = \frac{5}{9} = 55.5\% \text{ לכן, מספר הפקידים המינימלי הוא 3. הנצילות שלהם תהיה}$$

2.5 מהו מספר הרופאים המינימלי שנדרש כדי להבטיח את יציבות המערכת?

תשובה:

נחשב את ה-offered load בנפרד למשתחררים ולמתאשפזים:

$$R_{hosp} = \lambda \cdot 10\% \cdot E[S] = \frac{5}{6} \cdot 0.1 \cdot 30 = 2.5$$

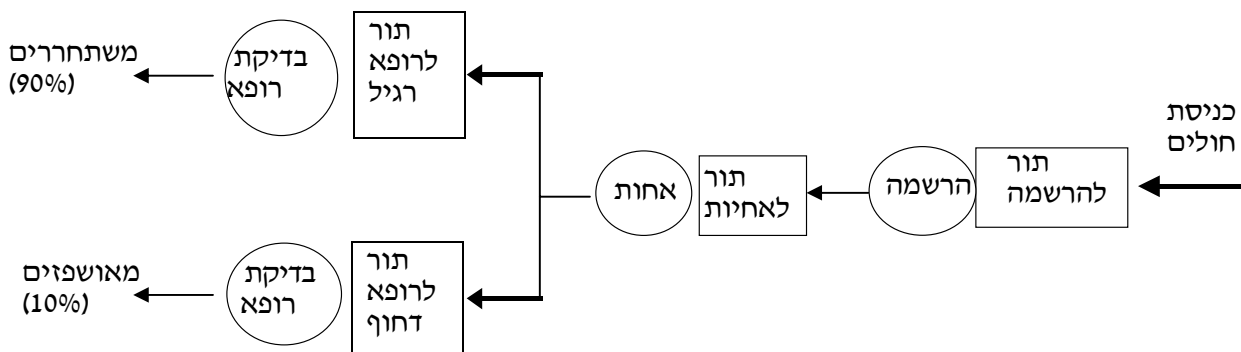
$$R_{dis} = \lambda \cdot 90\% \cdot E[S] = \frac{5}{6} \cdot 0.9 \cdot 5 = 3.75$$

לכן, מספר הרופאים המינימלי הוא 7 רופאים.

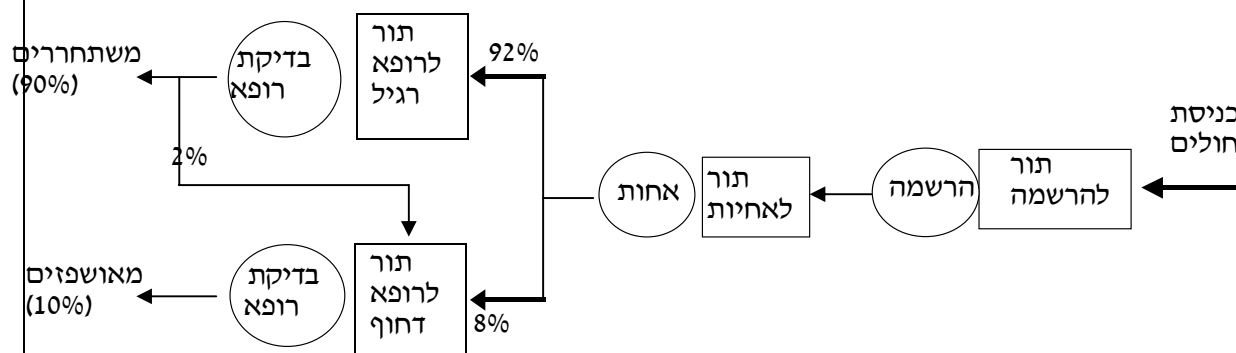
בבדיקה נוספת התגלה כי למרות שרק 90% מהחולים משתחררים (ולכן אינם דחופים), האחות מסווגת 92% מחולים כרגילים. הרופא שמבצע את הבדיקה הרגילה הוא שמגלה את הטעות בסיווג של 2% הנותרים. כתוצאה מכך, תור החולים הרגילים לקראת בדיקת רופא הוא ארוך יותר ומונה 21 חולים בממוצע. (תור החולים הדחופים נשאר חולה אחד בממוצע).

החולים שמסווגים בטעות כרגילים שוהים אצל הרופא הרגיל 10 דקות בממוצע, לאחר מכן הם נשלחים להמשך טיפול אצל הרופא הדחוף. משך זמן בדיקת רופא דחוף נשאר 30 דקות. שימו לב כי החולים שמסווגים באופן שגוי מחכים בפועל בשני תורים (לרופאים). הניחו כי כל שאר הנתונים נשארו זהים למה שנמצא קודם לכן.

2.6 השלם את האיור כך שיתאר את המצב החדש.



תשובה:



2.7 כמה זמן במוצע ישה חולה בחדר המיון?

תשובה:

הזמנים שהשתנו הם זמן ההמתנה לרופא והזמן הממוצע שלוקח ביקור רופא. נחשב אותם ונסכם את הזמנים :

$$W_5^{hosp} = \frac{L_5^{hosp}}{\lambda^{hosp}} = \frac{1}{\frac{5}{6} \cdot 10\%} = 12 \text{ min}$$

$$W_5^{rel} = \frac{L_5^{rel}}{\lambda^{rel}} = \frac{21}{\frac{5}{6} \cdot 92\%} = 27.39 \text{ min}$$

$$W_5 = 27.39 \cdot 90\% + 12 \cdot 8\% + 39.39 \cdot 2\% = 26.4 \text{ min}$$

$$W_6 = 5 \cdot 90\% + 30 \cdot 8\% + 40 \cdot 2\% = 7.7 \text{ min}$$

$$W = W_1 + W_2 + W_3 + W_4 + W_5 + W_6 = 24 + 2 + 6 + 3 + 26.4 + 7.7 = 69.1 \text{ min}$$

להלן כתבה שהתפרסמה באתר האינטרנט The Marker (אוגוסט 2010). שימו לב: עבור פתרון הבחינה אין צורך לקרוא את הכתבה, נתוני הכתבה הרלוונטיים לשאלה יתוארו במפורש בהמשך.

משרדי הממשלה יחויבו להשיב לפונים בטלפון בתוך 20 שניות

בין היעדים: 75% מהפונים ייענו בפנייתם הראשונה, ול-90% מהציבור יינתנו 90% מהשירותים בתוך חצי שעה

08:29 | 08.08.2010 | מאת: צבי זרחיה | הוסף תגובה

שלח לחבר | הדפס | פייסבוק | שתף | הוסף לרשימת קריאה

 [TheMarker Android](#)  [TheMarker iPhone](#) [הורד עכשיו ללא תשלום](#)

<< מהפכה בשיחת הממשלתי לציבור: משרדי הממשלה יחויבו להשיב לפונים בטלפון בתוך 20 שניות ו-75% מהפונים יקבלו פתרון בפנייתם הראשונה - כך עולה מתוכנית לשיפור השיחת הממשלתי שיגיש היום לממשלה השר מיכאל איתן.

סעיפים נוספים בתוכנית: ל-90% מהציבור יינתנו 90% מהשיחות בהמתנה של חצי שעה בלבד, 80% משיחות הממשלה יהיו מקוונים ויהיה אפשר לקבוע תור באינטרנט לפגישה אצל פקיד. בנוסף, יופעל מדד לשיחות הממשלתי לציבור, שיפורסם אחת לשנה או שנתיים, תוקם יחידה לשיחות לקוחות ויופעל פילוט למרכז שיחות ממשלתי (נמ"ל) אשר יכלול מרכז שיחות טלפוני ופורטל שיחות ראשי, במסגרת שער הממשלה (gov.il). לפי ההצעה להחלטת הממשלה תוקם ועדה

השירות במשרדי הממשלה צועד קדימה

עיקרי הרפורמה מבחינת האזרח

משרדי הממשלה יקבלו ציונים - יונהג מדד לשירות הממשלתי שימדוד את שירותי הממשלה המרכזיים וישווה ביניהם על פי קריטריונים כמותיים ואיכותיים

יוחל שער כניסה אחיד לשירותי הממשלה - הפרויקט ייקרא נקודת מפגש לאזרח (נמ"ל). ציבור הפונים יוכל לקבל בטלפון או באתר אינטרנט מידע בנושאים כמו הגורם המטפל בשירות המבוקש, דרכי ההתקשרות עמו והתהליך הנדרש לקבלת השירות

מדיניות השירות במשרדי הממשלה תשובר - תוקם יחידת מטה מרכזית לשירות לקוחות שתגדיר סטנדרטים שירותיים עבור יחידות הממשלה ותפעל לפישוט ולשיפור תהליכי שירות קיימים

הרפורמה אשר תגיש את מסקנותיה עד 1 באוקטובר 2010. התכנית תתקצב ב-16 מיליון שקל ב-2011 וב-34 מיליון שקל ב-2012.

"שיפור השיחת הממשלתי לציבור הוא חלק

עפ"י הכתבה הנ"ל, משרדי הממשלה יחויבו לעמוד בתנאים הבאים :

מוקד שירות טלפוני:

1. מענה תוך 20 שניות (100% מהשיחות).
2. 75% מהשיחות יטופלו לשביעות רצונם של הלקוחות כבר בפנייתם הראשונה (כלומר לא יצטרכו להתקשר שוב).

3.1.1 מדוע האילוץ "מענה תוך 20 שניות" (100% מהלקוחות) לא סביר? הציעו חלופה סבירה.

תשובה

לא ניתן להבטיח ש-100% מהפניות יענו תוך 20 שניות, ללא מספר מוקדנים לא סביר (תיאורטית אין-סוף). חלופה סבירה תהיה "לא יותר מ X% ימתינו יותר מ 20 שניות", כאשר X יקבע ע"י ההנהלה בהתאם לאילוצי רמת-שירות ועלויות.

3.1.2 מהי המשמעות והחשיבות של "יטופלו לשביעות רצונם בפנייה הראשונה"? הציעו שלוש סיבות היכולות לגרום לכך שלקוח לא יטופל לשביעות רצונו בפנייתו הראשונה?

תשובה

מספר השיחות אשר חוזרות מאחר ולא קיבלו מענה בפנייתם הראשונה הינו מדד טבעי לטיב השירות אשר ניתן ע"י המערכת (First time resolution). בכתה נלמד שמדד זה, ביחד עם אחוז הנוטשים, הם מדדים תפעוליים סובייקטיביים, דרכם מביע הלקוח את דעתו על השירות אותו הוא מקבל (להבדיל, למשל, מאחוז הנענים תוך 20 שניות). סיבה לשיחה חוזרת יכולה להיות: נטישה עקב סבלנות שפגה, מתן שירות לא טוב המצריך חזרה של הלקוח, כשל טכני במערכת או אפילו אי מתן שירות ("נטישת" מוקדן).

מענה נתמקד רק בתקן הראשון (מענה תוך 20 שניות ל 100% מהשיחות).

מוקד של משרד ממשלתי מסוים עובד בשעות הבוקר בלבד. אנו נתמקד בשעה ספציפית בה נתוני המוקד הם כדלהלן:

מוקד שירות טלפוני:

- קצב מופע שיחות הוא 100 לקוחות לשעה.
- ממוצע משך השיחות הוא 5 דקות.
- מספר נותני השירות הוא 4.
- סבלנותם הממוצעת של הלקוחות היא 10 דקות.

להלן פלט של תוכנת 4CC, המתאים למתרחש כרגע במוקד השירות הטלפוני. השתמשו בפלט זה כדי לענות על שאלה 3.1.3.

Your Call Center's Parameters

- Number of Agents Answering Calls:
- Average Time to Handle One Call (mm:ss):
- Calls per 60 minute Interval:
- Average Callers' Patience (mm:ss):

Settings

- Features: Abandons
- Basic Interval: 60 minutes
- Target Time: 00:20 (mm:ss)

Change Settings

Compute

Add to Table

Delete Rows

Clear All

Export

Import

Graph

	Basic Interval (minutes)	Target Time to Answer	Number of Agents	Average Handling Time	Calls per Interval	Average Patience	Agent's Occupancy	%Answer	%Abandon	Average Speed of Answer	%Answer within Target	Average Queue Length
Results	60.0	00:20.0	4.0	05:00.0	100.0	10:00.0	99.8%			06:45.6	.9%	8.7

3.1.3 באיזה תחום תפעולי נמצא המוקד? מהו אחוז הנוטשים? נמקו!

תשובה:

נמצא את ה Offered-Load : $R = \frac{\lambda}{\mu} = \frac{100}{12} = 8.33$.

ולכן $\beta = \frac{n-R}{\sqrt{R}} = -1.5$, כלומר המוקד נמצא תחת המשטר ED (יכולנו להסיק כי התחום

התפעולי הוא ED גם מכך שהנצילות קרובה מאוד ל100%)

את אחוז הנוטשים נחשב ע"י נוסחת האיוש בתחום ה ED : $n \approx R(1-\gamma)$ כאשר $\gamma = \%$ הנוטשים.

מנוסחה זו נקבל : $P(Ab) \approx 1 - \frac{n}{R}$.

ולכן $P(Ab) \approx 1 - \frac{n}{R} = 1 - \frac{4}{8.33} = 1 - 0.48 = 0.52$.

במטרה לשפר את רמת השירות, להלן פלט של תוכנת 4CC, המתאים למתרחש כרגע במוקד השירות הטלפוני.
השתמשו בפלט זה כדי לענות על שאלה 3.1.4.

◆

Number of Agents Answering Calls

23

◆

Average Time to Handle One Call (mm:ss)

05:00

◆

Calls per 60 minute Interval

100

◆

Average Callers' Patience (mm:ss)

10:00

Settings

◆

Features:

Abandons

◆

Basic Interval:

60 minutes

◆

Target Time:

00:20 (mm:ss)

Change Settings

Compute

◆

Add to Table

Delete Rows

Clear All

Export

Import

Graph

	Basic Interval (minutes)	Target Time to Answer	Number of Agents	Average Handling Time	Calls per Interval	Average Patience	Agent's Occupancy	%Answer	%Abandon	Average Speed of Answer	%Answer within Target	Average Queue Length
Results	60.0	00:20.0	23.0	05:00.0	100.0	10:00.0	36.2%	100%	.0%	00:00.0	100%	.0
1	60.0	00:20.0	13.0	05:00.0	100.0	10:00.0	63.7%	99.3%	.7%	00:03.9	93.9%	.1
2	60.0	00:20.0	14.0	05:00.0	100.0	10:00.0	59.3%	99.7%	.3%	00:01.9	96.8%	.1
3	60.0	00:20.0	15.0	05:00.0	100.0	10:00.0	55.5%	99.8%	.2%	00:00.9	98.5%	.0
4	60.0	00:20.0	16.0	05:00.0	100.0	10:00.0	52.0%	99.9%	.1%	00:00.4	99.3%	.0
5	60.0	00:20.0	17.0	05:00.0	100.0	10:00.0	49.0%	100%	.0%	00:00.2	99.7%	.0
6	60.0	00:20.0	18.0	05:00.0	100.0	10:00.0	46.3%	100%	.0%	00:00.1	99.9%	.0
7	60.0	00:20.0	19.0	05:00.0	100.0	10:00.0	43.9%	100%	.0%	00:00.0	99.9%	.0
8	60.0	00:20.0	20.0	05:00.0	100.0	10:00.0	41.7%	100%	.0%	00:00.0	100%	.0
9	60.0	00:20.0	21.0	05:00.0	100.0	10:00.0	39.7%	100%	.0%	00:00.0	100%	.0
10	60.0	00:20.0	22.0	05:00.0	100.0	10:00.0	37.9%	100%	.0%	00:00.0	100%	.0
11	60.0	00:20.0	23.0	05:00.0	100.0	10:00.0	36.2%	100%	.0%	00:00.0	100%	.0
12												

3.1.4 לפי הפלט של תוכנת 4CC, מהו מספר המוקדנים המינימאלי הדרוש כדי לעמוד בתקן של מענה תוך 20 שניות לכל הפניות? באיזה תחום תפעולי נמצא המוקד (עבור מספר המוקדנים עבורו המערכת עומדת בתקן)? נמקו!

תשובה:

אנו מחפשים את מספר המוקדנים המינימאלי כך ש $\%Answer\ Within\ Target=100\%$. מספר זה הוא: 20 מוקדנים.
עבור 20 מוקדנים אין נטישות, לכן המוקד יהיה תחת התחום התפעולי QD.

3.1.5 נניח שהמציאות אכן מתוארת ע"י Erlang A. האם ניתן להתחייב שעבור האיוש שהמלצתם בסעיף 3.1.4 אכן לא יהיה אף לקוח אשר ימתין יותר מ 20 שניות? נמקו במדויק ככל שתוכלו.

תשובה

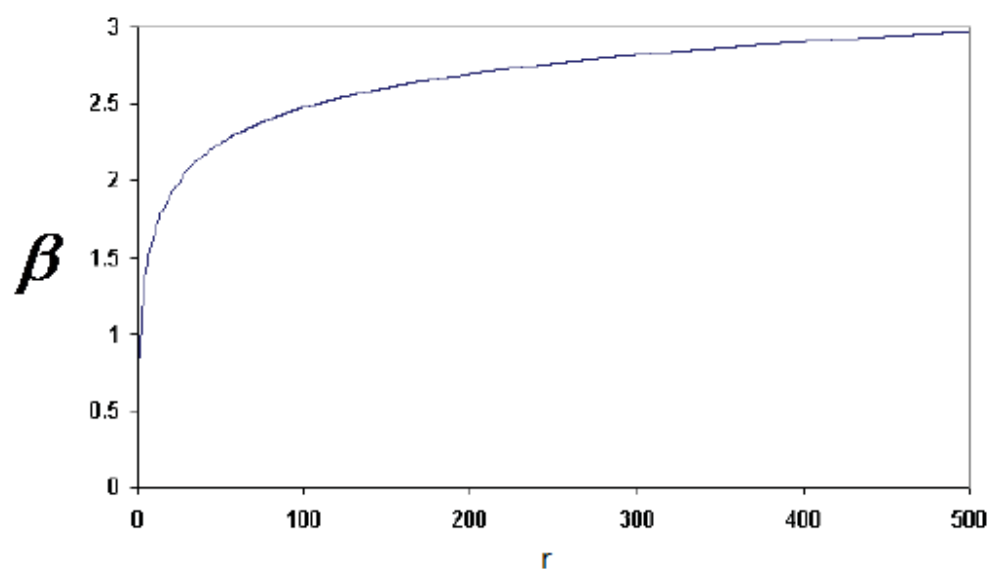
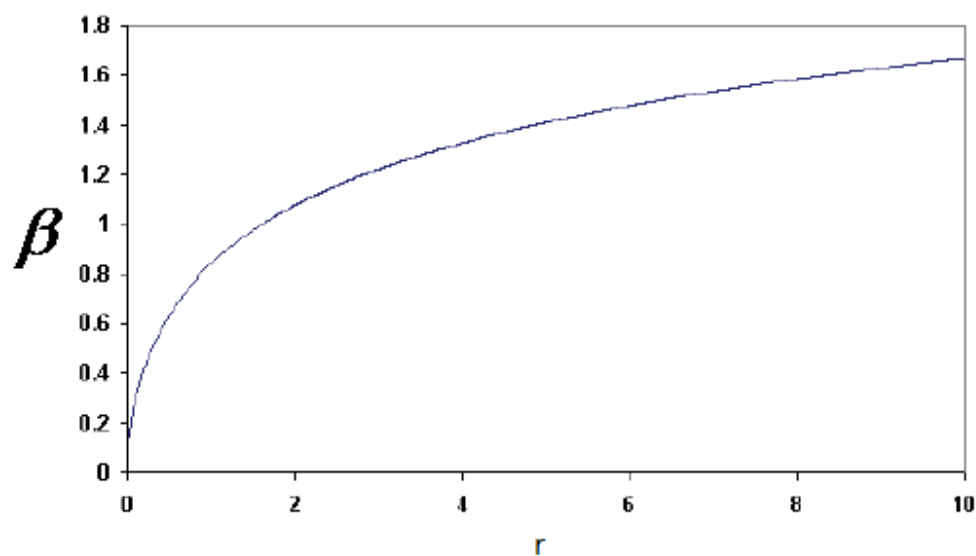
לא ניתן להתחייב. כאשר אין נטישות, Erlang A הוא כמו Erlang C ואז זמן המתנה של לקוח (אשר אכן ממתין בתור) מתפלג לפי התפלגות מעריכית (מ"מ לא חסום). לכן, בהסתברות חיובית ממש, זמן ההמתנה בתור יעלה על 20 שניות.

להלן דפי הרצאה דרכם ניתן לקבוע איוש על פי היחס בין עלות המתנה לבין עלות מוקדן, במודל Erlang C:

Erlang-C: Optimal Square-Root Staffing

$$n = R + \beta^*(r)\sqrt{R}$$

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



3.1.6 כדי לעמוד ביעדים, מנהל שאפתן ממליץ על איוש של 16 מוקדנים. הוא מתייעץ עם בוגרת הקורס "הנדסת מערכות שירות", וזו מסבירה לו את העובדות הבאות:

"מאחר ופרופורציית הנטישה קרובה מאד לאפס, ניתן להניח כי המודל המתאים למערכת עם מספר גדול יחסית של שרתים הוא Erlang C. ואז, שימוש בדף ההרצאה (בעמוד הקודם) מאפשר להמחיש את המשמעות הכלכלית של איוש המוקד ע"י 16 מוקדנים".

השלימו את הסברה של בוגרת הקורס, וזאת ע"י קביעת היחס בין עלות המתנה לבין עלות מוקדן, המתאים לאיוש של 16 מוקדנים. לצורך ההסבר הניחו שעלות מוקדן היא 30 ₪ לשעה.

תשובה:

מספר המוקדנים הוא 16 ולכן $16 = 8.33 + \beta \sqrt{8.33}$ $\beta = 2.65 \Leftarrow$ לכן $r \approx 150$.

כלומר היחס בין עלות ההמתנה לבין עלות המוקדנים הוא 1:150.

אם עלות מוקדן היא 30 ₪ לשעה, עלות שעת המתנה תהיה: $4500 = 30 \cdot 150$ ₪.

זהו כמובן מספר מופרז הממחיש שהחלטה על 16 מוקדנים איננה שקולה.

להלן כתבה שהתפרסמה במגזין "MIT Sloan" (Winter 2011, VOL. 52, NO.2). בכתבה מרואיין מר סמארט, שהוא ראש מחלקת מו"פ בחברה המפתחת אלגוריתמים לשידוך מוקדן-לקוח בזמן אמת.

שימו לב, עבור פתרון הבחינה אין צורך לקרוא את הכתבה. העובדות הרלוונטיות תובאנה בגוף השאלה.

How Analytics Solves the Problem of Conflicting Goals

What do you do when models predicting things such as best CSR match, willingness of a customer to wait and value of a customer to the company all recommend actions that are in conflict?

It sounds like the kind of information you have about customers is not that different from the kind of information you might have had before this whole process began, and that it's really on the CSR side that you have all this new data, plus the data about what happens in each specific interaction between a customer and a CSR. Is that what drives your models?

That's right. There's one other element that goes into the solution that drives revenue: the predicted economic value of a particular customer. Now, there's not a lot of new science in that, and we have models that tell us how to calculate that. But it's important to the solution, because in a call center we sometimes have to decide which customer to focus on. We like the idea that there's a CSR for everyone, but that's not always true because of call volumes and agent availability. So if your goal is long-term revenue, you can use these economic predictors to determine *which* customers we should be focusing on.

There was a problem we didn't quite know how to solve right out of the gate, and that was the fact that the **best matches are almost always not available**. In other words, if we have 50 callers in queue and 1,000 CSRs on the floor, we can create 50,000 different solutions, and we make those calculations 10, 15 times a second. One of the 1,000 CSRs is the

best match, so that's the score to beat — the number that shows how often we make that perfect match.

The vast majority of the time, though, those matches weren't immediately possible because that CSR was on the phone, so we had to factor in another predictive model, and that was **"time to available."** That's not a massively complex model, because the industry has been solving that kind of problem for a long time.

But when you layer "time to available" into the actual scoring engine, you get some interesting results. **If an agent's average handle time is three minutes, 30 seconds, and he or she has been on the phone three minutes, 15 seconds, then we can predict they're about 15 seconds away to available. Then we can weigh in our prediction of customer tolerance or customer survivability — how long they're willing to wait in the queue before just hanging up.**

We know how long we keep customers in queue. We know what the outcomes are when they've been in queue, and we can find out where the curve starts to steepen in terms of abandon rates or bad outcome rates. We connect that information with our CSR's predictive availability curve. If the optimal match is too far away, maybe 45 seconds or three minutes away, then the score for that optimal match becomes dampened and someone else might look more attractive to

כאמור, כתבה זו מדברת על בעיית ניתוב שיחות בשירות טלפוני, בזמן אמת, אתגר קשה הוא העובדה שמשך השיחה תלוי בתהליך השידוך של מוקדן ללקוח המחפש שירות, בתיאור האתגר, מר סמארט מתאר מציאות בה ישנם סוגי לקוחות רבים ומגוונים, וכל סוג עם רוחיות מסוימת עבור המערכת. כל מוקדן יכול לטפל בכל סוג לקוח אבל, בנוסף, ישנם מוקדנים המתמחים בטיפול סוגי לקוחות מסוימים, ולכן בסוגים אלו המוקדנים יטפלו מהר יותר. היעד הוא מקסימום רווח, ולפי זה מחליט אלגוריתם הניתוב איזה מוקדן יטפל באיזה לקוח (כך שהלקוחות אשר מכניסים יותר כסף לחברה יקבלו שירות ממוקדן אשר יכול לטפל בהם בצורה טובה ומהירה, מה שמחייב כמובן שמוקדן זה יתפנה לשירות הנידון במהירות).

3.2.1 מר סמארט מציג, להמחשת סיבוכיות הבעיה, דוגמא בה יש 50 שיחות בתור ו-1000 מוקדנים המשרתים תור זה. כדי לאתר את השיבוץ הטוב ביותר של מוקדנים לשיחות בדוגמא זו, טוען מר סמארט כי מספר כל השיבוצים האפשריים הוא גדול ורב: 50,000 אפשרויות.
עליכם לתקן את הערכתו הלא נכונה. מהו מספר השיבוצים הכולל בדוגמא זו? נמקו!

תשובה:

מאחר וכל שיחה יכולה להיות משובצת לכל אחד מהמוקדנים, לכל שיחה יש 1,000 אפשרויות. (בפרט מוקדן יכול לטפל ביותר משיחה אחת).
ולכן יש $1,000^{50}$ אפשרויות (מספר הגדול בהרבה סדרי גודל מ-50,000).
(אם למשל הייתה מגבלה שמוקדן יכול לקבל לכל היותר שיחה אחת, מס' האפשרויות היה $1,000 \times 999 \times \dots \times 951$ (שהוא קרוב יותר ל-50,000).

אלמנט מרכזי בתהליך השידוך "מוקדן-לקוח" הוא היכולת לאמוד, בכל רגע נתון ועבור כל מוקדן עסוק, את הזמן בו המוקדן יתפנה לשירות חדש. זמן זה נקרא *Time to Available*.
כדי להמחיש את המושג, מר סמארט נותן את הדוגמא הבאה:
"נניח שמשך השירות הממוצע של מוקדן מסוים הוא 3:30 דקות, שיחתו כרגע נמשכת 3:15 דקות והיא עדיין ממשיכה: אזי ניתן לחזות כי השיחה תסתיים בעוד כ-15 שניות, דהיינו *Time to Available* הוא כ-15 שניות".

3.2.2 חוו דעה על הדוגמא הנ"ל של מר סמארט.
רמז: מה אם השיחה בה מדובר נמשכה 3:45 דקות.

תשובה:

עבור שיחה שנמשכה 3:45 דקות, האמד הנ"ל יחזה זמן שלילי עד לסיום השיחה.
ברור כי אמד זה לא סביר. האמד לא מתחשב במידע המדויק על משך השיחה עד כה, אלא רק בממוצע משך כלל השיחות (וידוע ש-"ממוצע לא מספר את כל הסיפור"). אמד ה- *Time to Available* חייב, או ליתר דיוק, רצוי שיחשב את ממוצע (או התפלגות) זמן שארית השירות, בהינתן שהשיחה נמשכה 3:15 דקות: כפי שניראה, אמד זה יהיה גדול מ-3:30 דקות.

3.2.3 תנו דוגמא פשוטה להתפלגות משך שיחה שעבורה נכונה טענתו של מר סמארט, ודוגמא שעבורה היא איננה נכונה.

פתרון

דוגמא נגדית: $T \sim \exp(\lambda)$: מתכונת חוסר הזיכרון $E(T-t) = E(T) - t$ $E(T) = \frac{1}{\lambda}$ $E(\text{Time to Available}) = \frac{1}{\lambda}$.

דוגמא בה הטענה נכונה: T דטרמיניסטי (קבוע).

דוגמא נוספת ומעניינת: $T \sim U(0, D)$: עבור התפלגות זו עבור t הגדול מ $\frac{D}{2}$ דקות נקבל,

$$E(T-t) = E(T) - t = \frac{D}{2} - t < 0$$

הנמצאת בשירות תסתיים בעוד זמן שלילי.

3.2.4 נסמן ב- T מ"מ המתאר משך שיחה. נסחו בעזרת T אמד טבעי ל Time to Available , לאחר שהשיחה נמשכה t יחידות זמן. השוו אמד זה, במדויק ככל שתוכלו, עם האמד אותו מתאר מר סמארט.

רמז: בדקו מה הקשר בין Time to Available לבין $E(T | T > t)$.

פתרון

האמד הטבעי שנציע הוא: $\text{Time to Available} = E(T-t | T > t)$, מאחר ו $T-t$ הוא הזמן שנותר לסיום השיחה. בעוד אנו יודעים כי משך השיחה הוא לפחות t שניות.

האמד בו משתמש ראש מחלקת האנליסטים הוא $E(T-t) = E(T) - t$ $\text{Time to Available}_{\text{Head of Analytics}} = E(T-t)$, שכאמור יכול להיות שלילי.

נשווה בין האמדים:

ברור כי $E(T | T > t) \geq t$ ובנוסף $E(T | T > t) \geq E(T)$. לכן $E(T | T > t) \geq \max\{E(T), t\}$.

$$\text{Time to Available} = E(T | T > t) - t \geq \max\{E(T), t\} - t = [E(T) - t]^+ = [\text{Time to Available}_{\text{Head of Analytics}}]^+$$

כלומר לא רק שהאמד של מר סמארט יכול להיות שלילי, גם התיקון הטבעי שלו מהווה עדיין רק חסם תחתון לאמד המוצע.

עבור הדוגמא בה, $T \sim U(0, D)$ ועבור $\frac{D}{2} < t < D$, מהתיקון הטבעי נקבל כי

$$\text{Time to Available} = [E(T) - t]^+ = 0$$

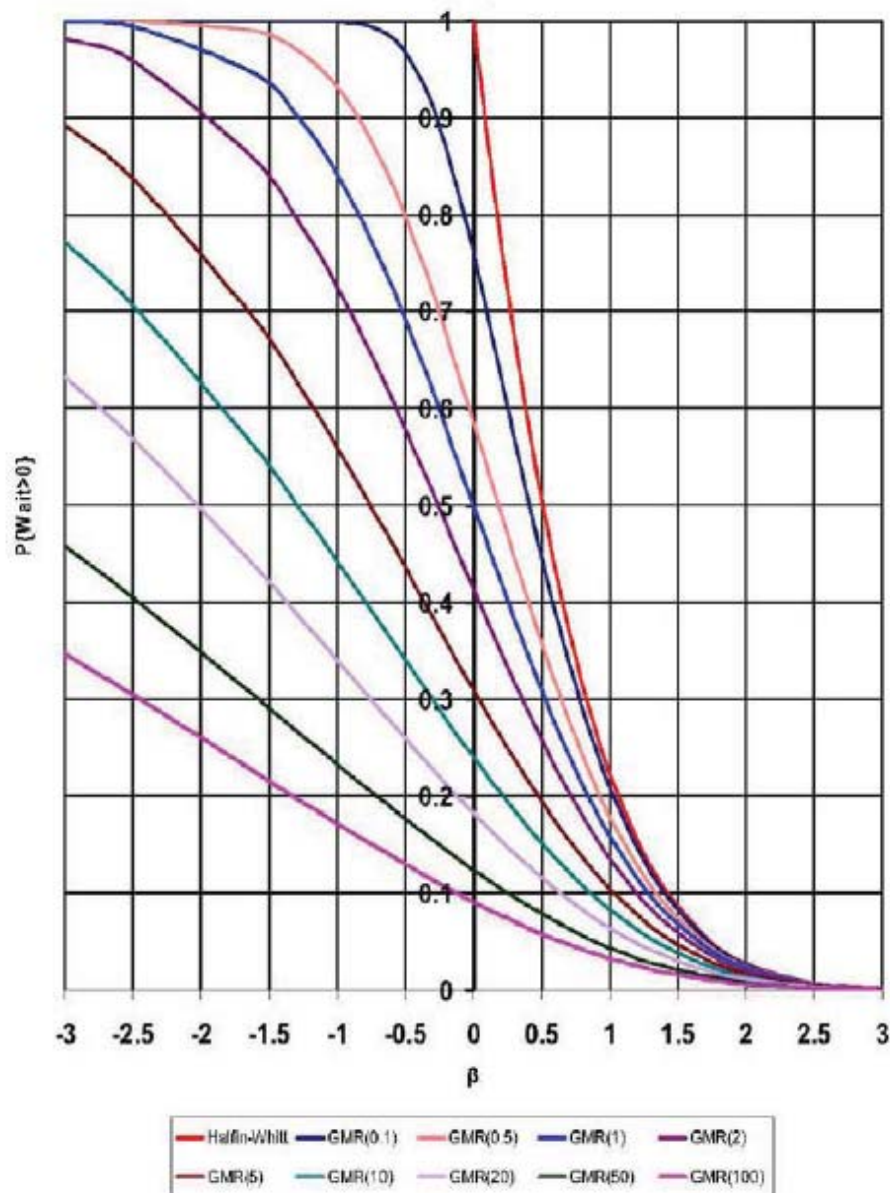
באמידת Time to Available שכן הערך הנאמד הוא לא אפס.

עבור האמד שהצענו מתקבל כי $(T-t | T > t) \sim U(0, D-t)$

ולכן $Time\ to\ Available = E(T - t | T > t) = \frac{D-t}{2} > 0$. כלומר ה $Time\ to\ Available$ הוא אמצע הקטע שבין הזמן שבו השיחה נמצאת כרגע לזמן המקסימלי שהשיחה יכולה להימשך – תוצאה אינטואיטיבית מאד. ואכן דוגמא זו ממחישה היטב את ההבדלים בין האמדים השונים, עד כמה האמד של מר סמארט מטעה וכך שאפילו תיקון האמד אינו מספק במקרים רבים.

Erlang-A: The Garnett Delay-Functions

$P\{W_q > 0\}$ vs. the QOS parameter β , for varying patience θ/μ .



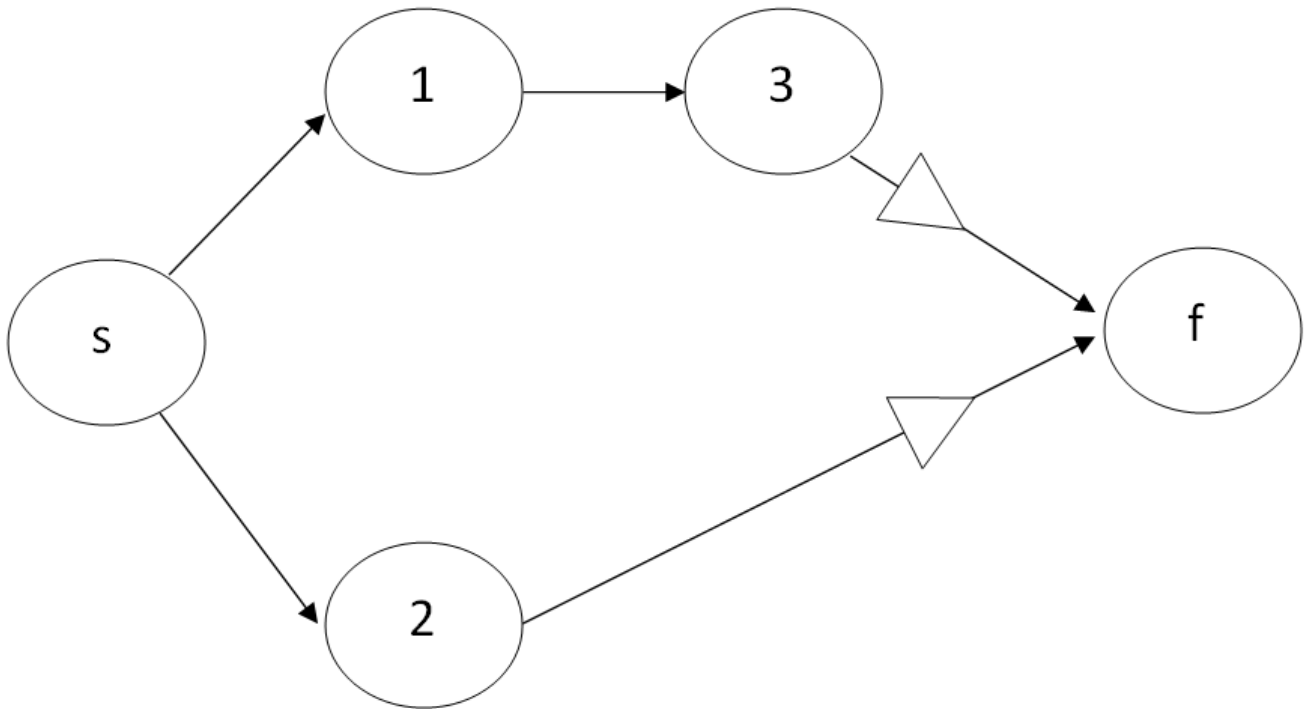
GMR(x) describes the asymptotic probability of delay as a function of β when $\frac{\theta}{\mu} = x$. Here, θ and μ are the abandonment and service rate, respectively.

Note: **Erlang-C** = limit of **Erlang-A**, as patience $\uparrow$ indefinitely.

שאלה 4 תיאוריה (10 נקודות)

חלק 1 – Static Deterministic PERT

להלן גרף PERT המציין את סדר הפעילויות שעל פרויקט לעבור עד לסיומו (3 פעילויות סה"כ):



גרף PERT זה מתאר את הפעולות הנעשות בבית קפה בו לקוח מזמין "קפה ומאפה". ראשית, הלקוח מזמין קפה ומאפה. לאחר ההזמנה מתחילות להתבצע שתי פעולות במקביל: פעולה מספר 2, הכנת קפה ע"י הברמן ופעולה מספר 1 שהיא הכנת המאפה ע"י הטבח. לאחר סיום הכנת המאפה, על הטבח לסדר את הצלחת יפה (פעולה מספר 3). רק לאחר סיום פעולות 1 ו 3 מקבל הלקוח את הקפה והמאפה לשולחנו. בחלק 1 של השאלה הניחו כי משכי הפעילות כולם קבועים. משך פעילות מספר 1 הוא 5 דקות, משך פעילות מספר 2 הוא 7 דקות ומשך פעילות מספר 3 הוא 4 דקות.

4.1.1 כמה זמן ייקח עד שישתיים הפרויקט? מהו המסלול הקריטי? נמקו!

תשובה:

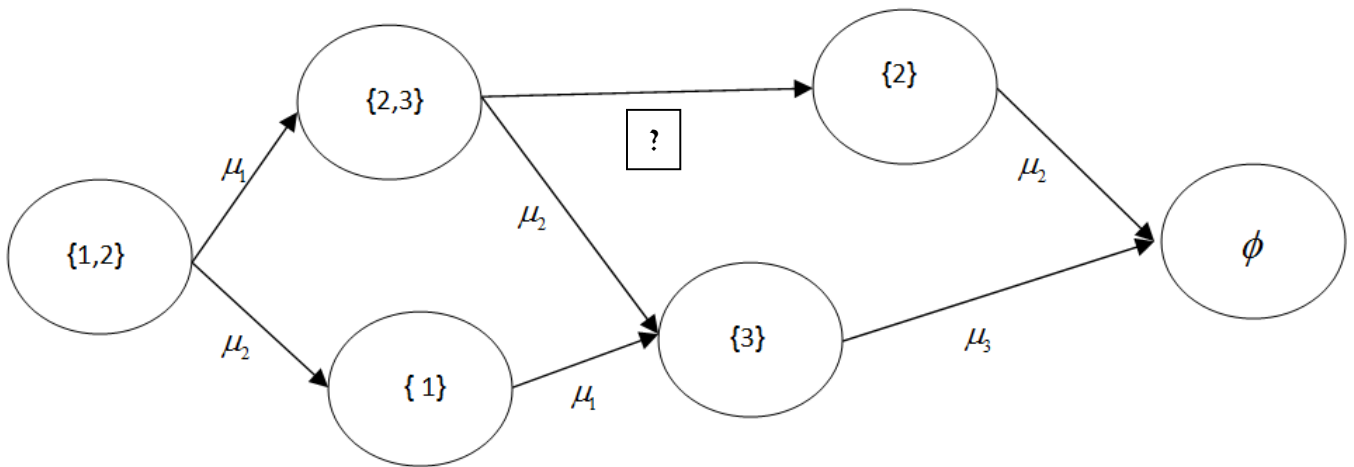
ישנם שני מסלולים:

1. מסלול העובר רק דרך פעילות מספר 2, הזמן שייקח לבצע מסלול זה הוא 7 דקות.
2. מסלול העובר דרך פעילות מספר 1 ואז דרך פעילות מספר 3, הזמן שייקח לבצע מסלול זה הוא 9 דקות.

לכן המסלול הקריטי הוא s-1-3-f והזמן שייקח עד שישתיים הפרויקט הוא $\max \{7, 9\} = 9$ (דקות).

חלק 2 – Static Stochastic PERT

עבור חלק זה הניחו כי משך הזמן שייקח לבצע כל פעולה הוא מערכי, כאשר קצבי הביצוע הם μ_1, μ_2, μ_3 , עבור פעילויות 1,2,3 בהתאמה. כדי לחשב את התפלגות משך הפרויקט, במודל זה שהוא סטטי-סטוכסטי, ניתן להגדיר תק"מ המתאר את תהליך ביצוע הפרויקט. באופן פורמאלי, ניתן לייצג את הפרויקט כתהליך קפיצה מרקובי $\{X_t, t \geq 0\}$, כאשר $X_t =$ קבוצת הפעילויות אותן אנו מבצעים בזמן t . לדוגמא: $X_t = \{1, 2\}$ הוא המצב ההתחלתי בו אנו מבצעים את פעילויות 1 ו-2 במקביל; $X_t = \{2\}$ פירושו שרק פעילות 2 מתבצעת (ובהכרח 1 ו-3 הסתיימו). להלן דיאגרמת קצבי המעבר של $\{X_t, t \geq 0\}$:



4.2.1 הסבירו מה המשמעות של המאורע $\{X_t = \phi\}$.

תשובה

ϕ מייצג מצב שבו אין פעילויות בבצוע, דהיינו הפרויקט הסתיים. במילים אחרות, ϕ הוא מצב סופג, המייצג את העובדה שהפרויקט הסתיים. ולכן המאורע $\{X_t = \phi\}$ מייצג את העובדה שהפרויקט הסתיים בזמן t או בזמן קצר יותר מ- t .

4.2.2 מהו קצב המעבר ממצב $\{2,3\}$ למצב $\{2\}$? הסבירו את תשובתכם.

תשובה:

כדי שנעבור ממצב $\{2,3\}$ למצב $\{2\}$ אנו צריכים שתסתיים פעולה 3, ומאורע זה קורה בקצב μ_3 .

4.2.3 נגדיר $S =$ הזמן עד לסיום הפרויקט (מרגע התחלתו). כיצד מתפלג S ? נמקו!

תשובה

מאחר ו S הוא הזמן עד שנגיע למצב סופג בתק"מ סופג נסיק כי S הוא מהתפלגות Phase-Type.

4.2.4 מהי ההסתברות שפעילות מספר 2 היא האחרונה אותה נסיים לבצע? מהי ההסתברות שהמסלול הקריטי הוא s-1-3-t? בטאו תשובתכם בעזרת הפרמטרים μ_1, μ_2, μ_3 .

תשובה

- אנו צריכים שפעולה מספר 1 תסתיים לפני פעולה מספר 2: עפ"י תחרות בין אקספוננציאלים, זה קורה בהסתברות $\frac{\mu_1}{\mu_1 + \mu_2}$; לאחר מכן, אנו צריכים שפעולה מספר 3 תסתיים לפני פעולה מספר 2. מתכונת חוסר הזיכרון, הזמן שנותר עד לסיום פעולה מספר 2 עדיין מתפלג אקספוננציאלית עם קצב μ_2 , לכן מאורע זה יקרה בהסתברות $\frac{\mu_3}{\mu_3 + \mu_2}$. נסכם שההסתברות שפעילות מספר 2 תהיה אחרונה היא $\frac{\mu_1}{\mu_1 + \mu_2} \cdot \frac{\mu_3}{\mu_3 + \mu_2}$.
- כדי שהמסלול הקריטי יהיה s-1-3-f אנו צריכים שפעילות 2 לא תהיה האחרונה לכן זהו המסלול הקריטי בהסתברות $1 - \frac{\mu_1}{\mu_1 + \mu_2} \cdot \frac{\mu_3}{\mu_3 + \mu_2}$.

4.2.5 שאלת בונוס (חיזרו לשאלה זו רק כאשר סיימתם את הבחינה). מצאו את $E(S)$ (תוחלת הזמן שייקח לסיים את הפרויקט).

תשובה

$$E(S) = \frac{1}{\mu_1 + \mu_2} + \frac{\mu_1}{\mu_1 + \mu_2} \cdot \left[\frac{1}{\mu_2 + \mu_3} + \left(\frac{\mu_2}{\mu_2 + \mu_3} \cdot \frac{1}{\mu_3} + \frac{\mu_3}{\mu_2 + \mu_3} \cdot \frac{1}{\mu_2} \right) \right] + \frac{\mu_2}{\mu_1 + \mu_2} \cdot \left[\left(\frac{1}{\mu_1} + \frac{1}{\mu_3} \right) \right]$$

הזמן שייקח לסיים את הפעולה הראשונה

הזמן שייקח לסיים את הפעולה הראשונה

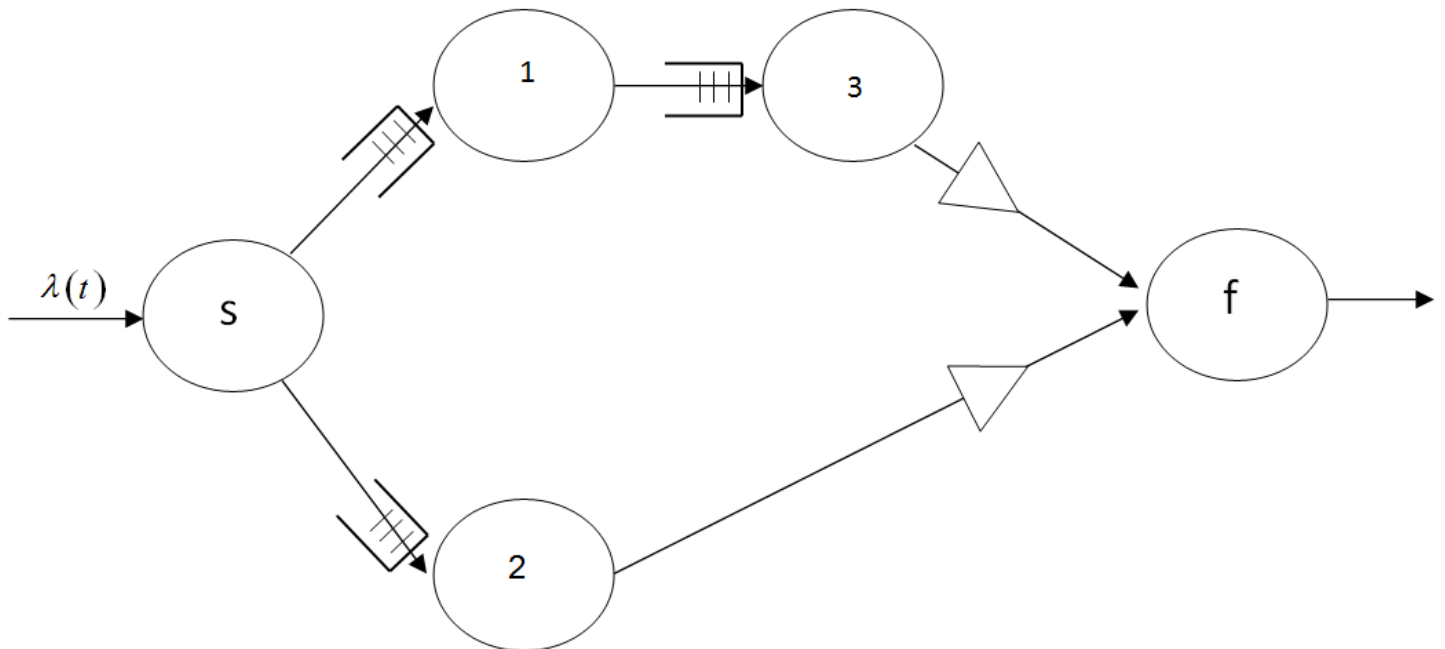
הזמן שייקח לסיים את פעולה 3

ההסתברות ש פעילות מספר 1 היא הראשונה שהסתיימה

ההסתברות ש פעילות מספר 2 היא זו שהסתיימה

חלק 3 – Dynamic Stochastic PERT

לקוחות מגיעים אל בית הקפה עפ"י תהליך פואסון לא הומוגני בזמן עם פונקצית קצב $\lambda(t)$ התלויה בזמן. להלן תרשים הפעילויות של תהליך הכנת ה"קפה ומאפה", במודל סטוכסטי דינאמי זה:



ענו על השאלות בעמוד הבא, המתייחסות למודל הנ"ל.

4.3.1 נתון כי שיא ההגעות למערכת מתרחש בשעה 13:00. נגדיר:

L_1 = שעת שיא העומס בפעילות מספר 1.

L_2 = שעת שיא העומס בפעילות מספר 2.

L_3 = שעת שיא העומס בפעילות מספר 3.

בכל אחד מהסעיפים הבאים עליכם לבחור את התשובה הנכונה (מתוך ארבעת האפשרויות) ולנמק מדוע בחרתם בתשובה זו.

4.3.1.1 א $L_1 < 13:00$, **ב** $L_1 > 13:00$, **ג** $L_1 = 13:00$, **ד** לא ניתן לקבוע

תשובה:
ג

פעילות מספר 1 היא תור $M_t / GI / N_t$ ולכן ה Offered-Load בתחנה זו שווה ל $E(S_1)$ $R_1(t) = E(\lambda(t - S_{1e}))$, לכן שיא העומס יהיה אחרי שיא ההגעות שהוא ב 13:00. ולכן $L_1 > 13:00$

4.3.1.2 א $L_1 < L_3$, **ב** $L_1 > L_3$, **ג** $L_1 = L_3$, **ד** לא ניתן לקבוע

תשובה:
א

קצב הכניסה לפעילות מספר 3 הוא קצב היציאה מפעילות מספר 1. לכן, כמו בסעיף הקודם שיא העומס ידחה ולכן $L_1 < L_3$.

4.3.1.3 א $L_2 < L_3$, **ב** $L_2 > L_3$, **ג** $L_2 = L_3$, **ד** לא ניתן לקבוע

תשובה:
ד

מאחר ויתכן כי זמן השירות בפעילות מספר 2 גבוהה יותר מסכום זמני השירות בפעילויות מספר 1 ו 3. יתכן כי שיא העומס בפעילות מספר 2 יהיה אחרי השיא בפעילות מספר 3.
בנוסף במידה וזמן השירות בפעילות מספר 2 יהיה קצר מזמן השירות בפעילות מספר 1. יתכן כי שיא העומס בפעילות מספר 2 יגיע לפני שיא העומס של פעילות מספר 1 ופרט של פעילות מספר 3.