



מרצה : ד"ר גלית יום-טוב
מתרגלת : נועה זיכלינסקי

תאריך הבחינה : 30.7.17

מס' סטודנט _____

הנדסת מערכות שירות – 096324

מועד א' – סמסטר אביב 2017

פתרון

- המבחן נמשך שלוש שעות. באחריותך להחזירו במועד.
- המבחן בחומר סגור.
- מותר להשתמש במחשבון מדעי רגיל.
- המבחן כולל 26 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 25 נקודות	תרגילי בית
שאלה 3	_____ מתוך 35 נקודות	יישום
שאלה 4	_____ מתוך 15 נקודות	תיאוריה
סה"כ	_____ מתוך 100 נקודות	

בהצלחה !

שאלה 1 – הרצאות\תרגולים\בחינות קודמות (25 נקודות).

חלק 1 – SEESat (15 נקודות)

1.1.1 (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:

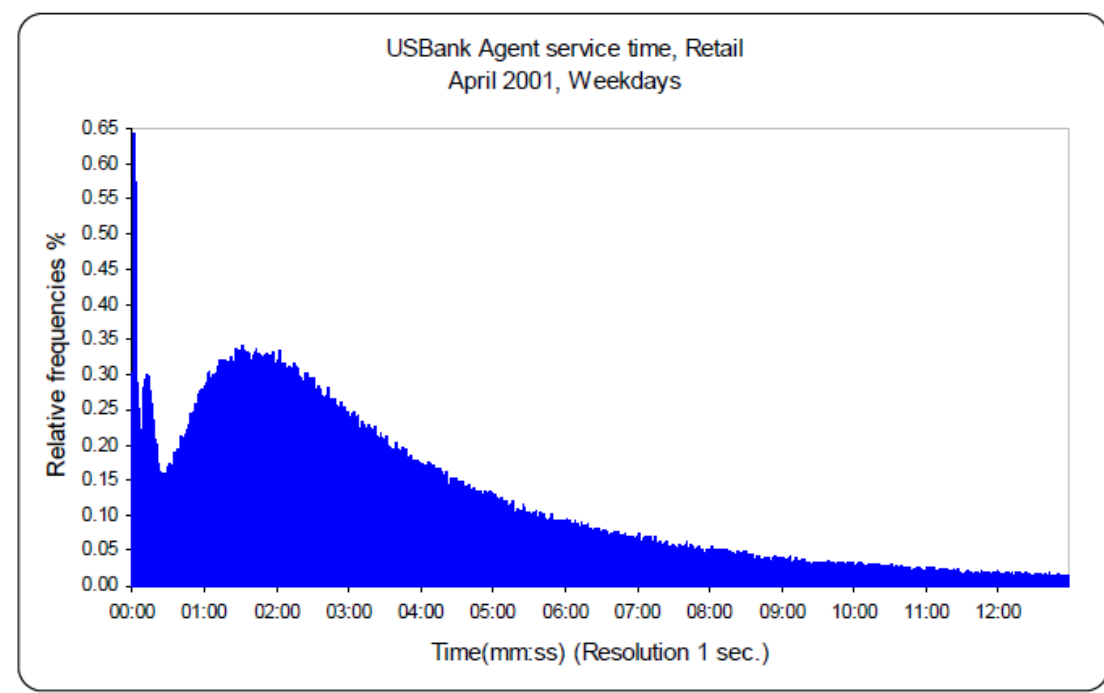
ההבדל בין שני הגרפים הוא ברזולוציה שלהם (אשר מתאימה לרוחב העמודות בהיסטוגרמה). גרף 1 הינו ברזולוציה של שניה אחת ואילו גרף 2 הינו ברזולוציה של 30 שניות. בשל כך, גרף 2 הינו חלק יותר, מה שמאפשר לראות טוב יותר את הצורה הכללית של התפלגות משכי השירות. עם זאת, בגרף 2 אנו מאבדים מידע על אי הסדירות בראשית הצירים.

1.1.2 (3 נקודות) איזו תופעה מיוחדת באה לידי ביטוי בגרפים? הציעו הסבר לתופעה זו.

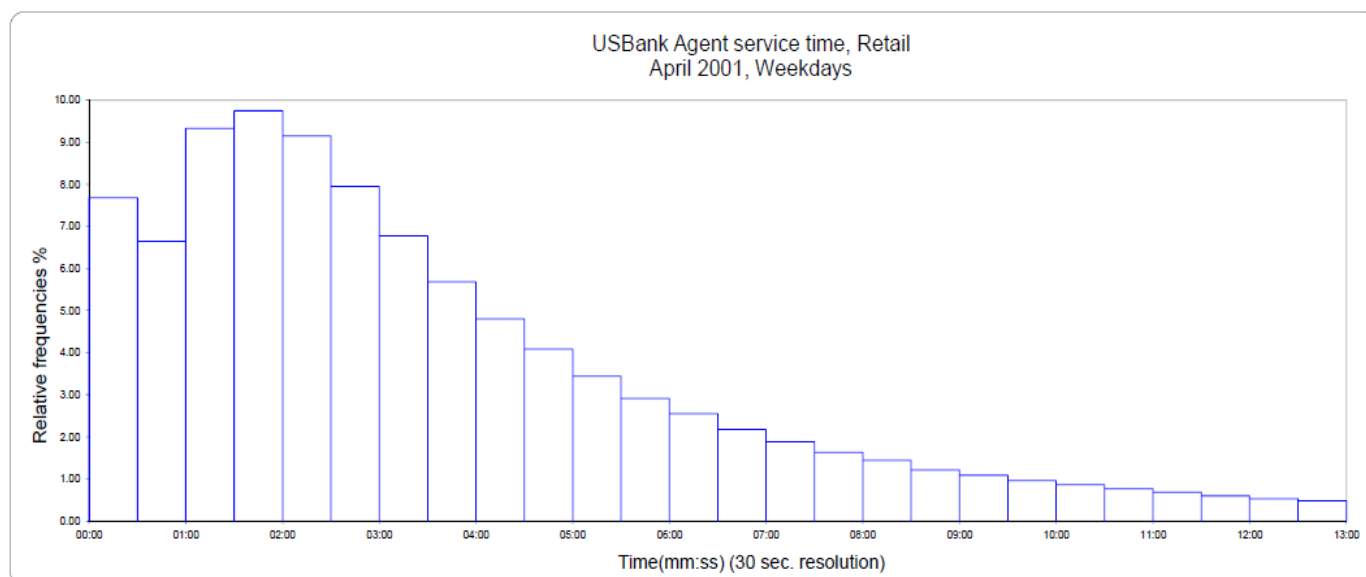
תשובה:

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

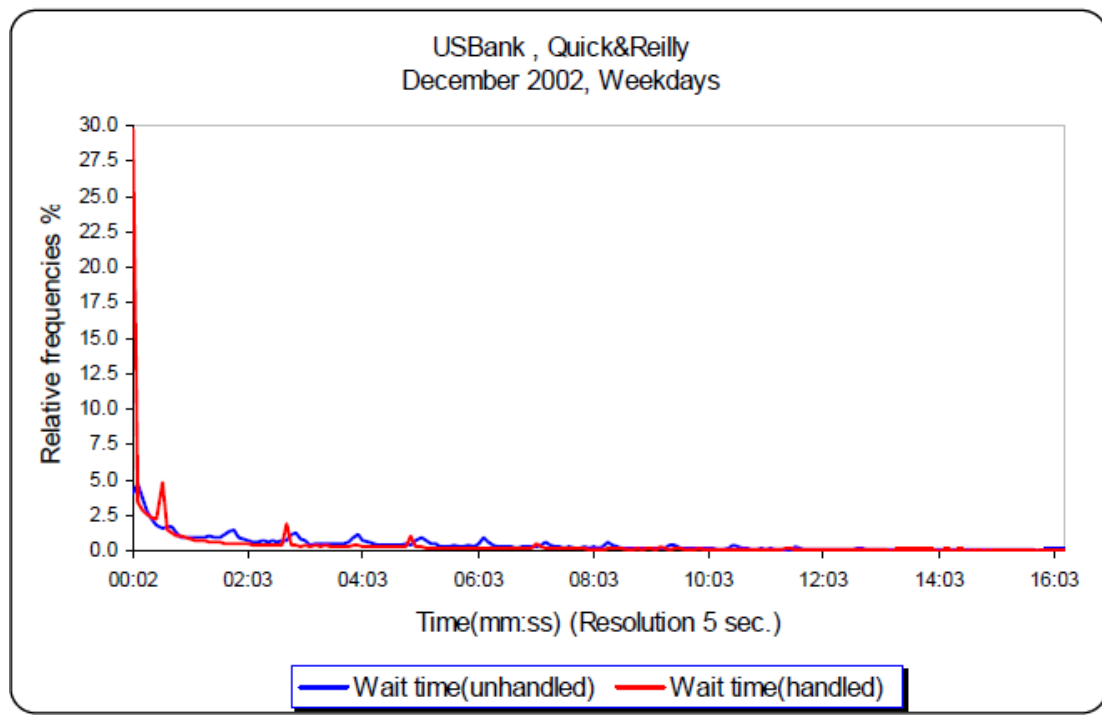


1.1.3 (5 נקודות) הסבירו בקצרה מה מתארים גרף 3 וגרף 4. הסבירו מה ההבדל בין שני הגרפים ומה היתרון בשימוש בגרף 4 לעומת גרף 3

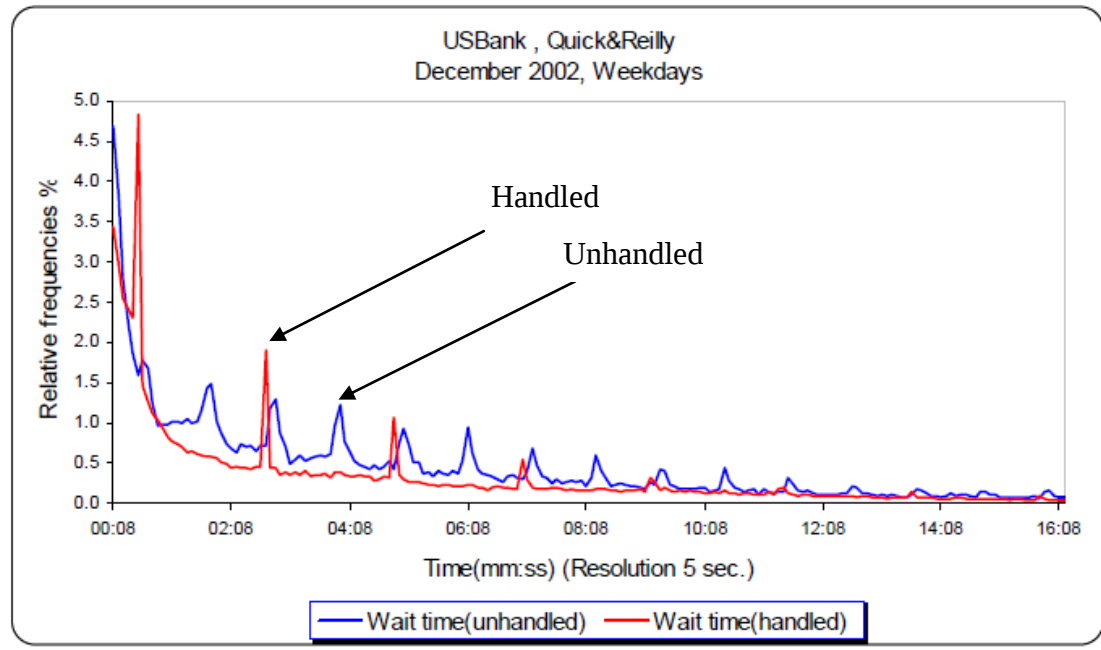
תשובה:

גרפים 3 ו-4 מציגים היסטוגרמת משך זמני ההמתנה עבור לקוחות שקיבלו שירות (Handled) ועבור לקוחות שלא קיבלו שירות (Unhandled). מדובר בזמני המתנה לשירות מסוג Quick&Reilly ב – USBank, בכל ימי החול של דצמבר 2002. גרף 4 מציג את אותה היסטוגרמה החל מ-5 שניות. 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.

גרף 3



גרף 4



1.1.4 (4 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 מסוימים 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 או 3 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.

1.2.1 (3 נקודות) הגדירו מטריצה

$$A = [a_{ij}]$$

המתארת את המשמרות האפשריות (ז"א הגדירו את אברי המטריצה).

תשובה:

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

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

1.2.2 (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:

משתני ההחלטה:

X_j - מספר המוקדנים במשמרת ה- j .

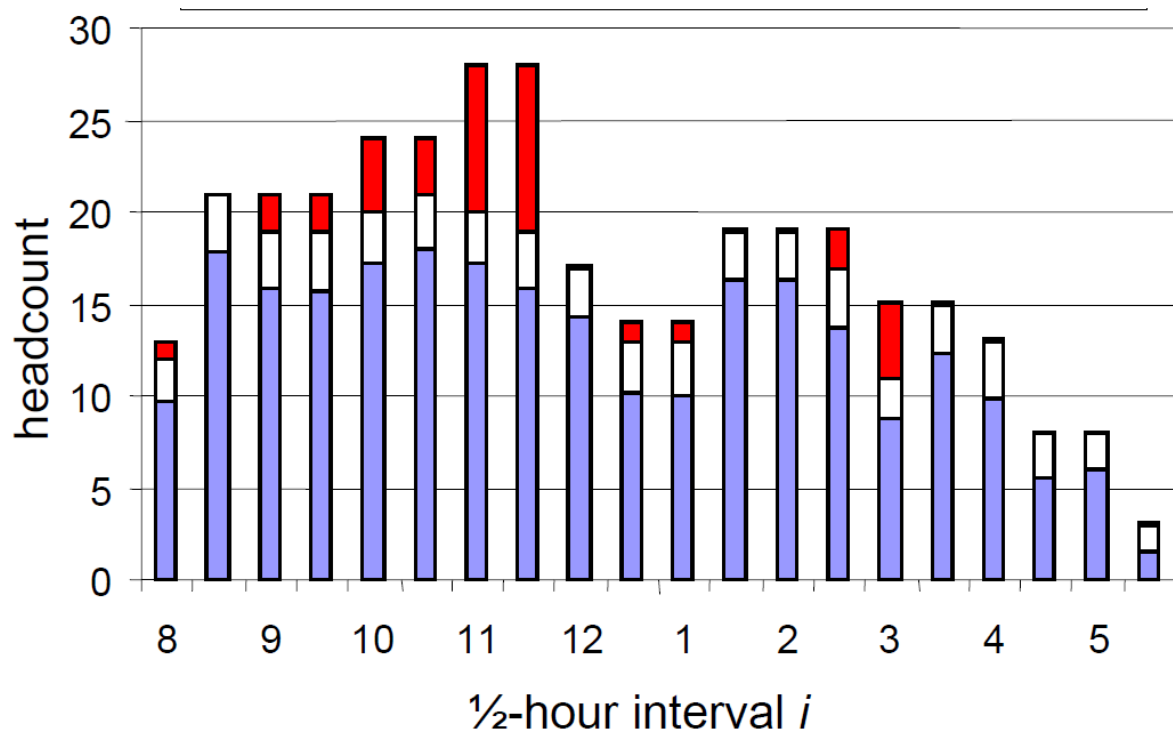
$$\min \sum_{j=1}^6 C_j X_j$$

s.t.

$$\sum_{j=1}^6 a_{ij} X_j \geq N_i \quad i=1,2,\dots,10$$

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

1.2.3 (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.

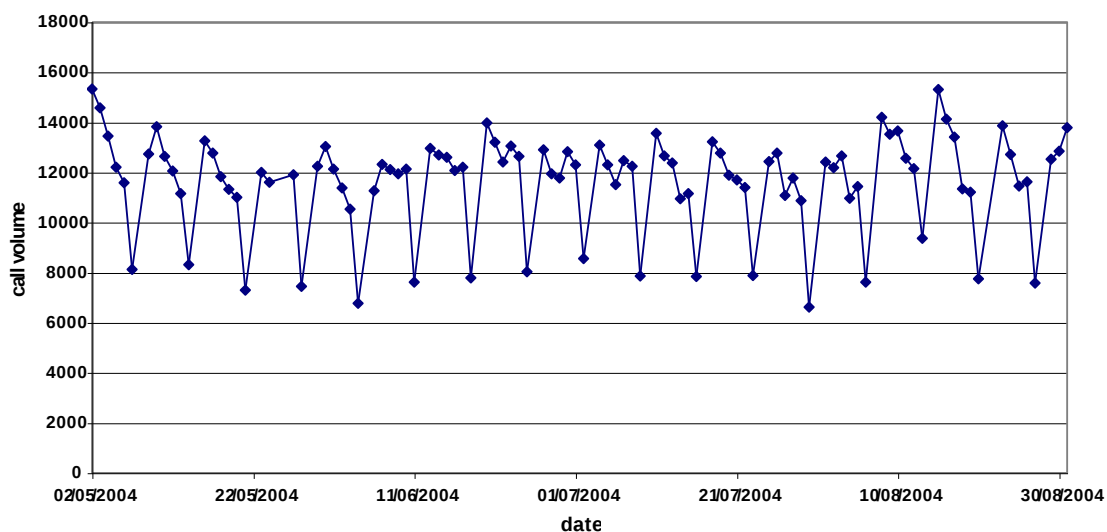
תשובה:

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 (25 נקודות).

חלק 1 – תחזיות

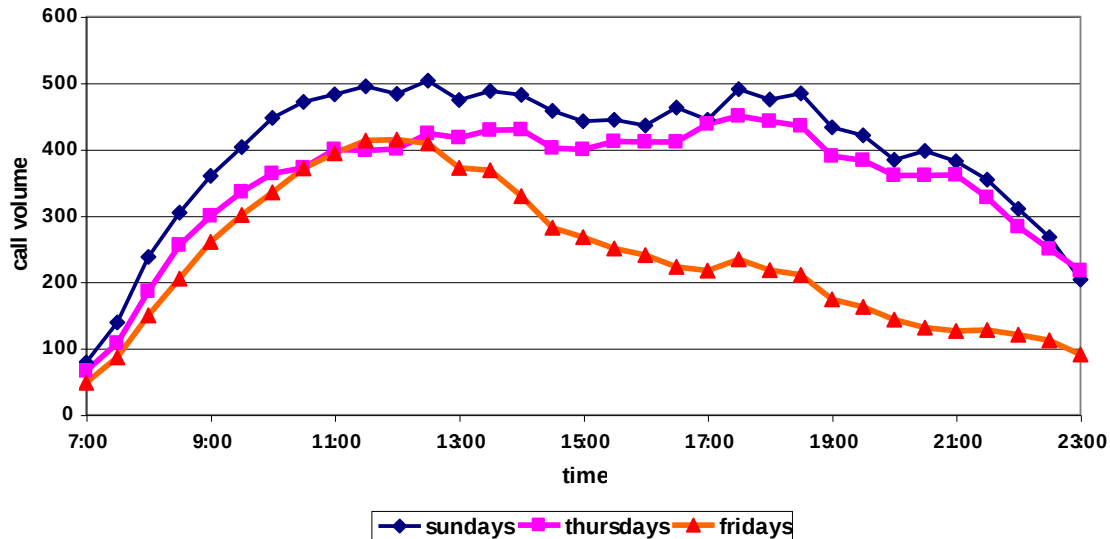
2.1.1 (2 נקודות) הגרף הבא מציג את סך השיחות הנכנסות ליום בתקופה 02/05/04 - 31/08/04 (בימים א'-ו', ללא שבת). אילו מסקנות (בעלות משמעות תפעולית) ניתן להסיק מהגרף?



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.

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 העקומות השונות. 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. נסמן, כמו בתרגיל הבית, את N_{jk} להיות מספר השיחות הנכנסות באינטרוול זמן k ביום j ו- F_{jk} להיות תחזית מספר השיחות לאינטרוול זה. נניח שבוע עבודה של 6 ימים.

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) בקורס הוגדרו 2 מדדים לקביעת טיב התחזית F_{jk} :

RMSE (Root Mean-Square Error) ו-APE (Average Percent Error). תארו (במשפט אחד) כיצד מסיקים באיזו שיטת תחזית לבחור, בהתבסס על מדדים אלו. איזה מדד מהשניים יותר אינטואיטיבי לשימוש? הסבירו מדוע.

תשובה :

ככל שהמדדים קטנים יותר כך התחזית טובה יותר. ולכן נבחר במודל עם הערכים המינימליים. מדד APE יותר אינטואיטיבי כיוון שקל יחסית להגדיר מהו אחוז הטעות שאנו מוכנים לסבול במערכת כלשהי.

2.1.5 (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 :

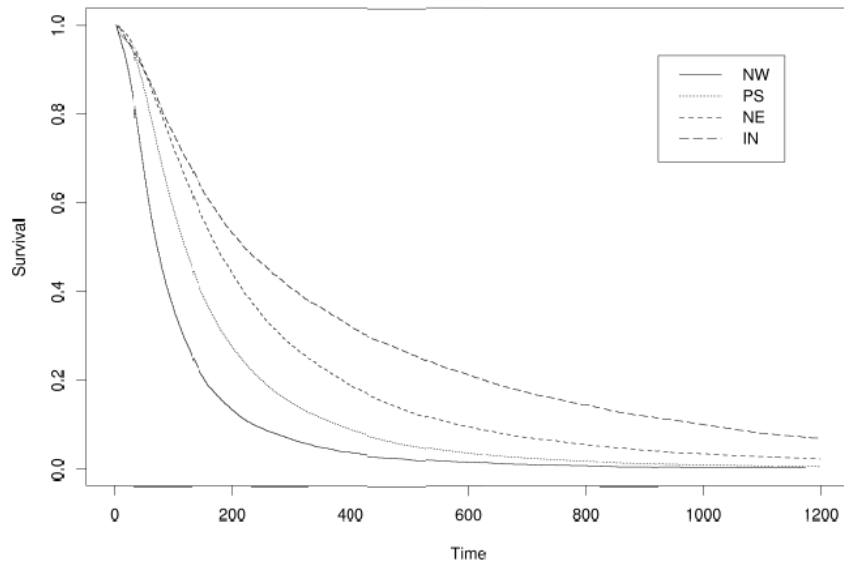
לא. הערכים לא צפויים להיות 0, מכיוון שגם אם הפרמטר ידוע באופן מדויק, לתהליך עצמו יש השתנות סטוכסטית (שונות), והדגימות לא תהיינה זהות לממוצע.

חלק 2 – פונקציות שרידות ופונקציות סיכון של זמני שירות

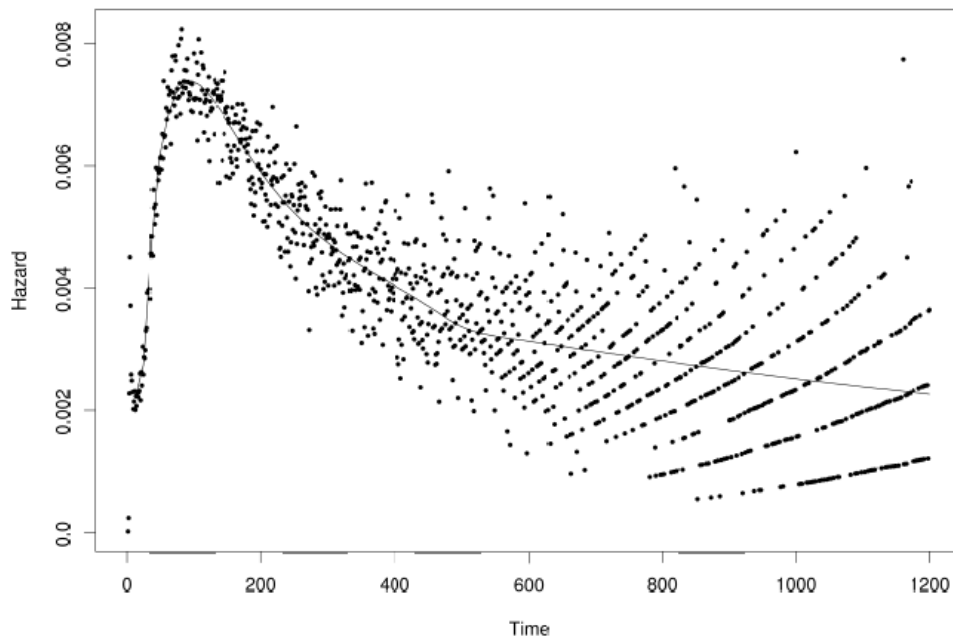
שני הגרפים שלהלן מבוססים על נתוני מוקד שירות טלפוני של בנק ישראלי קטן. נותחו ארבעה סוגי שירות עיקריים: PS, IN, NE, NW. גרף 1 מתאר פונקציות שרידות (survival functions) של ארבעת סוגי השירות. נקודות בגרף 2 הן אמדים לפונקצית הסיכון (hazard rate) של זמן השירות הכולל (כל סוגי השירות ביחד), ברזולוציה של שנייה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

גרף 1

Nov-Dec



גרף 2



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

תשובה:

נשתמש בנוסחת הזנב:

$$E[S] = \int_0^{\infty} S(x) dx$$

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.

2.2.3 (2 נקודות) סדרו את ארבעת סוגי השירות לפי תוחלתם (מהקצרה ביותר לארוכה ביותר)

תשובה:

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

2.2.4 (3 נקודות) 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 שווה 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 $h(k) = A_k / \eta_k$.

הסבר:

בחלק הימני של הגרף מוצג זנב ההתפלגות ובו מעט דגימות. 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- $A_k = 1$)

שאלה 3 – יישום (35 נקודות)

גבי, בוגר הפקולטה להנדסת תעשייה וניהול ובוגר הקורס "הנדסת מערכות שירות" בטכניון, התקבל לעבודה בבית חולים תל השומר. המשימה הראשונה שהוטלה עליו על ידי מנהל בית החולים היא לבחון ולהציע דרכים לשיפור זמני ההמתנה הארוכים בחדר המיון. גבי היה נחוש בדעתו להוכיח את עצמו וליישם את הכלים שלמד במהלך התואר. אחרי שהתייעץ עם רופאים ואחיות במיון ואף ביצע ניתוח נתונים של זרימת חולים במיון, גיבש גבי את הצעתו. ההצעה כוללת הקמת חדר מיון מהיר לטיפול במקרים קלים, שיעבוד לצד חדר המיון הכללי של חדר המיון. במיון החדש יטופלו פונים מתחת לגיל 60, המסוגלים ללכת וסובלים מבעיות פשוטות.

3.1 (1 נקודה) ציינו סוג הפרדה נוסף הקיים בחדרי מיון אחרים.

תשובה:

הפרדה לפי סוג הטיפול הנדרש לדוגמה: פנימי, אורתופדי, ילדים ויולדות או הפרדה לפי המשאבים הנדרשים: מיון מהלכים (חולים עצמאיים – דרוש כיסא לכל חולה) ומיון שוכבים (דרושה מיטה לכל חולה).

גבי מודע לנושא היתרון לגודל. לכן, לפני יישום הרעיון הכרוך בהשקעה כספית גדולה, הוא מציע ליישם מדיניות של עדיפויות בחדר המיון הכללי. מההיבט התפעולי, הפרדת החולים לשתי הקטגוריות מתבססת על משך השרות הנדרש לחולה. חולים "קלים" זקוקים לטיפול קצר יחסית לחולים "רגילים". גבי טוען שכדי לקצר את זמן השהייה הממוצע במיון יש לתת עדיפות לחולים הקלים.

3.2 (2 נקודות) הסבירו את טענתו של גבי. האם היא נכונה?

תשובה:

כאשר המטרה היא קיצור משך השהיה הממוצע, אזי ההערכה היא שעלות ההמתנה של כל החולים זהה. אזי לפי כלל ה- C_μ , כדאי לתת עדיפות לחולים עם קצב שרות גבוה יותר, שהם החולים הקלים, ואז תוחלת זמן ההמתנה של כלל החולים תקטן לעומת כל מדיניות שרות אחרת. לכן, טענתו של גבי נכונה.

3.3 (2 נקודות) כיצד תשפיע מדיניות התעדוף החדשה על משך ההמתנה של החולים בשתי הקבוצות?

תשובה :

משך ההמתנה של החולים בעדיפות גבוהה ("קלים") יתקצר ומשך ההמתנה של חולים בעדיפות נמוכה ("רגילים") יתארך.

3.4 (2 נקודות) מנהל המיון הנוכחי סירב לתת עדיפות דווקא לחולים קלים. טענתו היא שהנזק הנגרם כתוצאה מהמתנה ארוכה לחולים רגילים גבוה יותר מחולים קלים, ולכן אם רוצים להפעיל מדיניות עדיפויות, צריך לתעדף חולים רגילים. כיצד ניתן לקחת בחשבון את נזקי ההמתנה בקביעת מדיניות התיעדוף בין שתי הקבוצות?

תשובה :

ניתן להגדיר "עלות" המתנה לכל סוג חולה, כאשר עלות ההמתנה של חולה רגיל גבוהה מעלות ההמתנה של חולה קל. באופן זה כאשר מפעילים את כלל ההחלטה C_μ , יתכן שדווקא החולים הרגילים יקבלו עדיפות. (קיימת הכללה לכלל ה- C_μ לעלויות לא ליניאריות, שהיא נראית רלוונטית יותר במקרה זה)

גבי הצליח לשכנע את הנהלת בית החולים בצורך לקיצור משכי שהיה. כדי שחדר המיון לא יפעל בניגוד לדחיפות הרפואית, הוחלט להפריד את שתי הקבוצות לשני חדרי מיון נפרדים. ננסה עתה לבחון, **תפעולית**, מה יקרה בחדר המיון החדש. לשם כך דרושים לנו נתונים נוספים. בדיקה ראשונה שביצע גבי מעלה שאחוז החולים המתאימים להפניה למיון המהיר הוא 10% מכלל החולים המגיעים. בנוסף, מתקבלים למיון כ- 180,000 חולים בשנה ומשך השהייה הממוצע במיון כיום הוא 4 שעות (לכל החולים).

3.5 גבי מעוניין להעריך את גודל השטח הדרוש למיון החדש, ולהשוותו למיון הנוכחי. הוא מכיר את חוק Little, והחליט בעזרתו לחשב כמה חולים קלים היו בממוצע במיון הרגיל, בכל שעה, לפני השינוי, וכמה יהיו במיון החדש. בשני הסעיפים הבאים, שחזרו את החישובים שביצע גבי.

3.5.1 (2 נקודות) מהו מספר החולים הקלים ששוהים במיון הישן, בממוצע שנתי?

תשובה :

$$E[L] = \lambda * E[W] = \left(\frac{180,000}{365 * 24} * 0.1 \right) * 4 = 2.05 * 4 = 8.2$$

על סמך חוק ליטל: 8.2

3.5.2 (2 נקודות) כמה חולים בממוצע מצפה מנהל בית החולים שיהיו במיון החדש, בכל רגע נתון, בהנחה שהיעד התפעולי שהוצב הוא שהייה של שתיים בממוצע?

תשובה :

$$E[L] = \lambda * E[W] = \left(\frac{180,000}{365 * 24} * 0.1 \right) * 2 = 2.05 * 2 = 4.1$$

3.5.3 (3 נקודות) כאמור, גבי מבקש להשתמש בחישוביו, בשני הסעיפים הקודמים, כדי לתכנן את גודל המיון החדש. האם דרכו מוצדקת? הסבירו תשובתכם. אם לא, הציעו עקרונות (ללא פירוט) דרך חלופית.

תשובה :

השימוש בחוק Little הוא בעייתי מכמה סיבות :

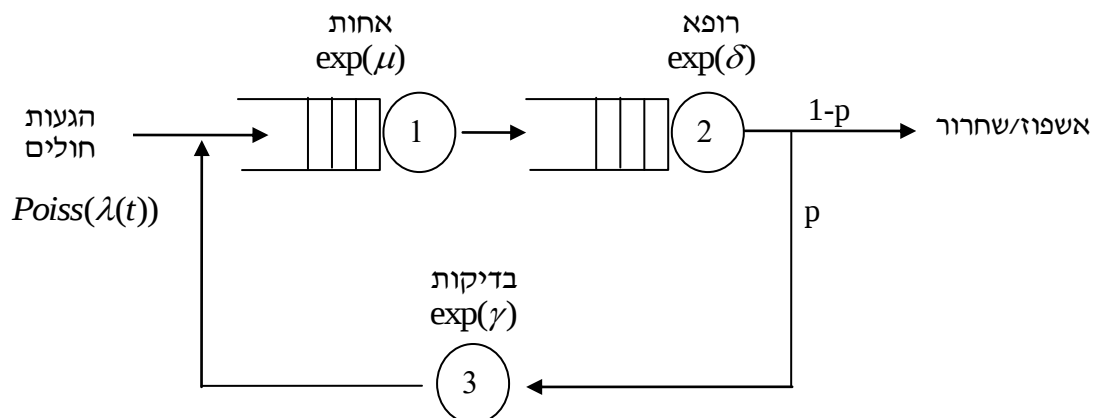
1. מספר החולים בחדר המיון משתנה, בין שעות היממה (הרבה ביום ומעט בלילה), בין ימי השבוע (הרבה בימים רגילים ומעט בסופי שבוע) ולפי עונות השנה.
2. חדרי מיון צריכים לתת מענה גם למקרים של אירוע רב נפגעים, שקורה לעיתים רחוקות, ושטח חדר המיון צריך לאפשר התמודדות עם מקרים אלו.
3. חדר מיון נבנה לתקופה ארוכה, ולכן יש להתחשב גם בפרמטרים נוספים כמו גידול צפוי באוכלוסיה.
4. לעיתים מפרידים את המיון לסוגי חולים שונים הזקוקים לשטח שונה לפי סוגם (לדוגמה : טראומה / מתהלכים / שוכבים).

מסיבות אלו, הסתכלות על ממוצע שנתי כולל תיתן ערך שסביר שיהיה נמוך משמעותית ממה שצריך בפועל.

שיטה נכונה יותר תהיה להסתכל על התפלגות מספר החולים במיון, ולתת מענה לאחוז גבוה מהמקרים, כך שההסתברות שלחולה לא יהיה מקום במיון תהיה קטנה עד זניחה. בנוסף, יש לבצע ניתוחי רגישות, לגידול בכמות החולים הנכנסים, ולבדוק את השפעת הנהלים להפניית חולים במקרים של אירועים חריגים על הנעשה במיון.

הטענה שלפיה זמן השהיה יתקצר מארבע שעות לשעתיים מבוססת על העובדה שזמן הטיפול הכולל "נטו" בחולה קל קטן כמובן משעתיים, וכל יתר זמן שהייתו בחדר המיון נובע מהמתנות לרופא, לאחות ו/או לבדיקות.

נתאר בצורה מפורטת יותר את תהליך הטיפול בחולה בחדר המיון :
 כאשר חולה מגיע למיון, הוא צריך להיבדק ע"י אחות ורופא. לעיתים הרופא שולח את החולה לבדיקות וכאשר מגיעות תוצאותיהן הוא חוזר לרופא ; לעיתים עדיין אין אבחנה והחולה נשלח לסדרת בדיקות נוספות, שלאחריהן הוא שוב צריך להיבדק ע"י רופא. כאשר הבדיקות מושלמות ויש אבחנה, הרופא מחליט על צורת הטיפול ומחליט באם לאשפז את החולה או לשלחו לביתו. ניתן לתאר (בקירוב שימושי) את התהליך המתואר לעיל בעזרת רשת התורים הבאה :



תחנה 1 מתארת את הביקור אצל האחות. בתחנה זו מספר ה"שרתים" N_1 הוא מספר האחיות במיון, והתור לפני התחנה הוא התור לביקור אצל האחות. נניח כי זמן הביקור אצל אחות מתפלג מעריכית עם פרמטר μ .

תחנה 2 מתארת את הביקור אצל הרופא. בתחנה זו מספר ה"שרתים" N_2 הוא מספר הרופאים במיון, והתור לפני התחנה הוא התור לביקור אצל הרופא. נניח כי זמן הביקור אצל רופא מתפלג מעריכית עם פרמטר δ .

תחנה 3 מתארת את תהליך הבדיקות שהחולה צריך לעבור. אנו לא נמדל במפורט תהליך זה, אלא רק את העיכוב הכללי שנוצר כתוצאה מבדיקות אלו. לכן נניח כי בתחנה 3 אינסוף שרתים, וזמן הבדיקות מתפלג מעריכית עם פרמטר γ .

בזמן כל ביקור אצל הרופא, מחליט הרופא אם יש צורך בבדיקות נוספות או לא. אחוז החולים שעבורם נדרשות בדיקות נוספות הוא $100 \cdot p$. (לכן, מספר הביקורים הכולל אצל הרופא/האחות מתפלג גיאומטרית עם פרמטר "הצלחה" $(1-p)$).

מדידות בחדר המיון מראות שערכי הפרמטרים הם כדלקמן : נמצא כי בממוצע חולה קל עובר 3 ביקורים אצל הרופא לפני החלטה על אשפוז/שחרור. עוד נמצא כי ביקור בודד של חולה אצל רופא מתפלג מעריכית עם ממוצע של 15 דקות, זמן ביקור אצל אחות מתפלג מעריכית עם ממוצע של 10 דקות, וזמן הבדיקות מתפלג מעריכית עם ממוצע של 15 דקות.

3.6 (2 נקודות) חשבו את הסתברות המעבר בין תחנה 2 ל-3 (ערכו של P)

תשובה:

נגדיר X – מספר הטיפולים שעובר חולה קל אצל הרופא לפני החלטה על אשפוז/שחרור.

$$X \sim \text{Geo}(1-P)$$

$$E(X) = \frac{1}{1-P} = 3 \rightarrow P = \frac{2}{3}$$

3.7 (3 נקודות) כיצד מתפלג זמן הטיפול הכולל (טיפול אחות, רופא, ובדיקות ללא המתנות) בחולה קל בחדר המיון? הגדירו במדויק ככל שתוכלו את הפרמטרים של ההתפלגות.

תשובה:

נגדיר T – משך שירות כולל המתפלג Phase-type, עם פאזות מעריכיות במבנה התואם את הרשת הנ"ל.

$$q = (1, 0, 0), \quad m = \left(\frac{1}{\mu}, \frac{1}{\delta}, \frac{1}{\gamma} \right) = (10, 15, 15)$$

$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & p \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2/3 \\ 1 & 0 & 0 \end{bmatrix}$$

3.8 (2 נקודות) מהו ממוצע משך הטיפול הכולל בחולה קל?

תשובה:

ממוצע זמן הטיפול הוא :

$$E(T) = (15+10)*3 + 15*2 = 25*3 + 30 = 105 \text{ min} = 1.75 \text{ hour}$$

3.9 (5 נקודות) מהי שונות משך הטיפול הכולל בחולה קל? העזר בנוסחאות בסוף השאלה.

תשובה:

נגדיר T – משך שירות כולל,

באמצעות שימוש בנוסחאות עבור התפלגות Phase-Type נקבל:

$$R = [1-p]^{-1} = \begin{bmatrix} 3 & 3 & 2 \\ 2 & 3 & 2 \\ 3 & 3 & 3 \end{bmatrix}, \quad M = \begin{bmatrix} 10 & 0 & 0 \\ 0 & 15 & 0 \\ 0 & 0 & 15 \end{bmatrix}$$

$$E(T^2) = 2 * q * (R * M)^2 * 1 = 22,050$$

$$\text{Var}(t) = E(T^2) - E^2(T) = 11,025$$

הערה – בגלל האופן בו מוגדר q , אפשר לפשט את ההכפלה של המטריצות.

3.10 (4 נקודות) מהו מספר הרופאים והאחיות המינימלי הדרוש לטיפול כך שמספר החולים במיון לא "יתפוצץ"? בססו את תשובתכם על שיקולי מצב יציב (כלומר, מופע בקצב קבוע).

תשובה:

זמן הטיפול בחולה ע"י צוות רופא או אחות הוא:

$$\text{רופא: } 15 * 3 = 45 \text{ min} = 0.75 \text{ hour}$$

$$\text{אחות: } 10 * 3 = 30 \text{ min} = 0.5 \text{ hour}$$

$$\text{מספר החולים הנכנסים למיון הוא בממוצע } \bar{\lambda} = \left(\frac{180,000}{365 * 24} * 0.1 \right) = 2.05 \text{ חולים בשעה.}$$

$$\text{ולכן תנאי היציבות לאחיות הוא: } \frac{\lambda * E[S_1]}{N_1} < 1 \Leftrightarrow 2.05 * 0.5 < N_1 \text{ כלומר, } N_1 \geq 1.03.$$

$$\text{ולכן תנאי היציבות לרופאים הוא: } \frac{\lambda * E[S_2]}{N_2} < 1 \Leftrightarrow 2.05 * 0.75 < N_2 \text{ כלומר, } N_2 \geq 1.54.$$

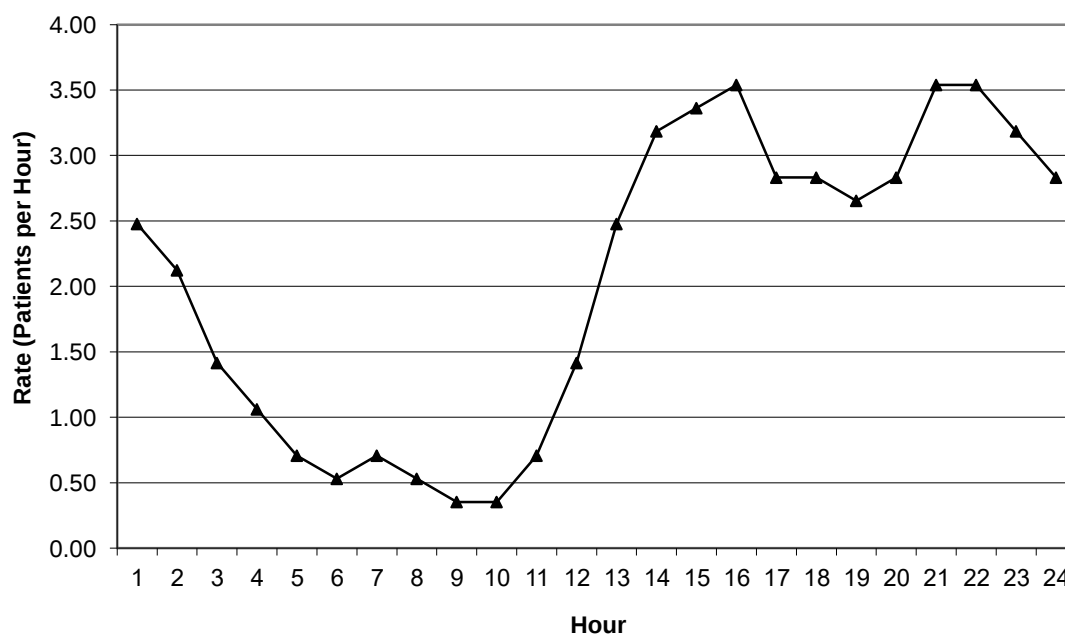
3.11 (2 נקודות) הסבירו מדוע תשובתכם לשאלה 3.10 רלוונטית גם כאשר המופע הוא משתנה בזמן אך מחזורי (לדוגמה עם מחזור של יממה).

תשובה:

התשובה נכונה גם במצב מחזורי כיוון שבמקרה זה, על פני יממה, יש מספיק קיבולת, ולכן חולים ימתינו בשעות עמוסות (בהן קצב הכניסה גדול מהממוצע היומי) ויקבלו שירות בשעות לא עמוסות (בהן קצב הכניסה קטן מהממוצע היומי). כלומר, כל החולים שנכנסו גם יצאו, כלומר המצב יהיה יציב, אם כי זמני ההמתנה עלולים להיות מאוד ארוכים בשעות מסוימות (זה יהיה תלוי באמפליטודה של קצב המופע על פני המחזור).

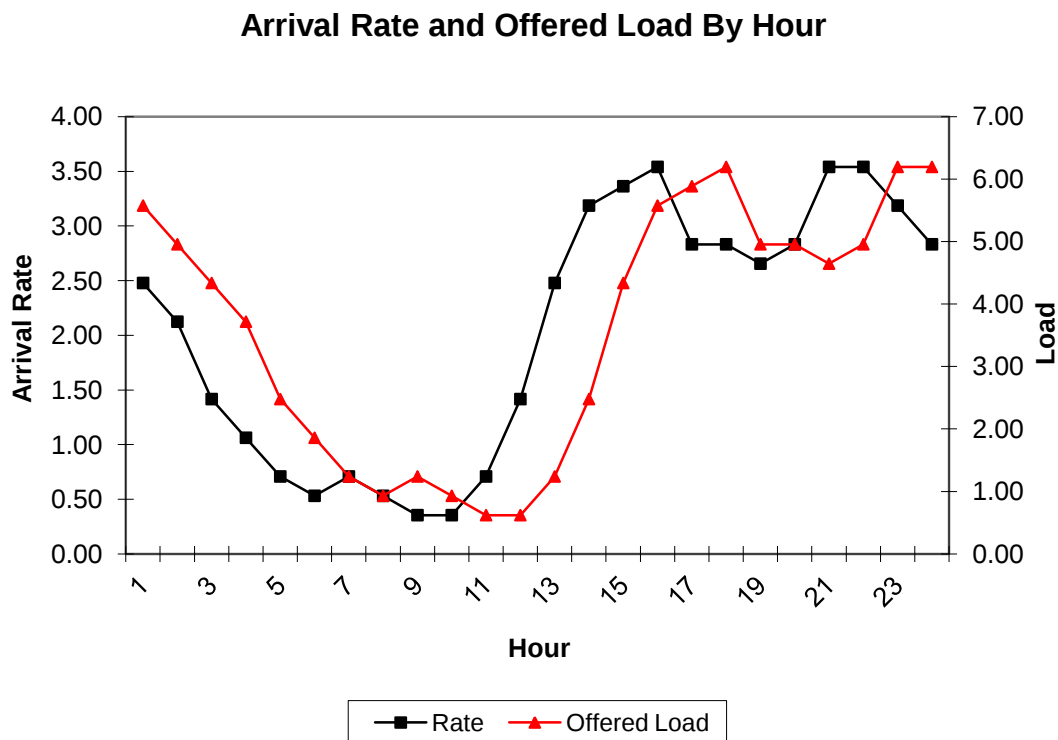
הגרף הבא מתאר את קצב הגעות חולים קלים למיון, לפי שעה במהלך יום חול ממוצע:

Arrival Rate per Hour



3.12 (3 נקודות) ציירו על הגרף הבא כיצד אתם מצפים שיראה הגרף של "העומס המוצע" (Offered load) במיון. הסבירו (תוך שימוש בסקלה בצד ימין) במידת הצורך.

תשובה :



הסבר : בתור יחיד העומס המוצע נתון לפי הנוסחה : $R(t) = E[\lambda(t - S_e)]E[S]$, ובקירוב

$R(t) = \lambda(t - E[S_e])E[S]$. לכן, גרף העומס המוצע אמור לעקוב אחרי גרף קצב ההגעות עם הפרש אופקי שהוא פחות או יותר תוחלת זמן שארית שרות (כולל) ולכן במקרה שלנו פחות או יותר עיכוב של 1.75 שעה (מסעיף 3.6). ומבחינת הערכים (ציר ה-Y) הוא צריך להיות בסדר גודל של קצב ההגעה * 1.75.

Phase-Type Service Times (Durations).

Service-Time = a sequence/collection of tasks, of an *exponential* duration.
There are K types of tasks, indexed by $k = 1, \dots, K$.

m_k = expected duration of task k ; $m = (m_k)$
 q_k = % of services in which k is first; $q = (q_k)$
 P_{jk} = % of incidences in which task j is immediately followed by k . $P = [P_{jk}]$

$1 - \sum_{\ell=1}^K P_{k\ell}$ = probability to end service at k .

Representation directly in terms of (q, P, m) .

Denote here $R = [I - P]^{-1}$ (as in Mandelbaum & Reiman).
Average work content $E(T) = qRm$ ($= \sum_j q_j R_{jk} m_k$).

$$\text{Moments: } E(T^n) = n! q(RM)^{n-1}, \quad \text{where } M = \begin{bmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_K \end{bmatrix}$$

$$\frac{E(T^2)}{2(E(T))^2} = \frac{1 + C^2(T)}{2} = \frac{q(RM)^2 1}{(qRM1)^2}$$

לנוחיותכם :

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1/3 \\ -1 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1.5 & 1.5 & 0.5 \\ 0.5 & 1.5 & 0.5 \\ 1.5 & 1.5 & 1.5 \end{bmatrix},$$

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -2/3 \\ -1 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 3 & 2 \\ 2 & 3 & 2 \\ 3 & 3 & 3 \end{bmatrix},$$

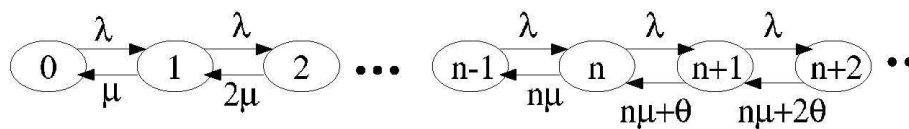
$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1/2 \\ -1 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 2 \end{bmatrix}$$

שאלה 4 – תיאוריה ושימושיה (15 נקודות)

בשאלה זו נדון במודל Erlang-A עם פרמטרים λ, μ, θ, N המתאר מערכת שירות עם לקוחות קצרי רוח. נסמן ב- $L(t)$ את מספר האנשים במערכת בזמן t . אזי $L = \{L(t), t \geq 0\}$ הוא תהליך לידה ומוות.

4.1 (3 נקודות) ציירו את דיאגרמת קצבי המעבר של L , וכתבו את משוואות המצב היציב. (נסמן כרגיל את הפתרון למשוואות ב- $\pi_0, \pi_1, \pi_2, \dots$)

תשובה:



משוואות המצב היציב הן:

$$\lambda \pi_i = i \mu \pi_{i+1} \quad i \leq n$$

$$\lambda \pi_i = (n \mu + (i - n) \theta) \pi_{i+1} \quad i > n$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

4.2 (3 נקודות) הוכיחו מתמטית שמודל Erlang-A הוא תמיד ארגודי (יציב). דהיינו, הוכיחו שהטור הרלוונטי ממשוואות המצב היציב אכן מתכנס.

פתרון:

מערכת משוואות המצב היציב מ-4.1, שקולה ל-

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} \pi_0 \quad i \leq n$$

$$\pi_i = \prod_{k=n+1}^i \frac{\lambda}{(n\mu + (i-n)\theta)} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \pi_0 \quad i > n$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

מכיוון שהתהליך אי-פריק, כדי להוכיח ארגודיות, הכרחי ומספיק להוכיח שלמערכת המשוואות הנ"ל יש

פתרון. כלומר, צריך להוכיח שהטור $\pi_0^{-1} = \sum_{i=0}^n \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} + \sum_{i=n+1}^{\infty} \prod_{k=n+1}^i \frac{\lambda}{(n\mu + (i-n)\theta)} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!}$ מתכנס.

מכיוון ש- $n\mu + (i-n)\theta \geq k \min(\mu, \theta)$ לכל $k \geq n$ ו- $k\mu \geq k \min(\mu, \theta)$ לכל $k < n$,

מתקיים:

$$\pi_i = \left(\frac{\lambda}{\mu}\right)^i \frac{1}{i!} \pi_0 \leq \left(\frac{\lambda}{\mu \wedge \theta}\right)^i \frac{1}{i!} \pi_0 \quad i \leq n$$

$$\pi_i = \prod_{k=n+1}^i \frac{\lambda}{(n\mu + (i-n)\theta)} \left(\frac{\lambda}{\mu}\right)^n \frac{1}{n!} \pi_0 \leq \left(\frac{\lambda}{\mu \wedge \theta}\right)^i \frac{1}{i!} \pi_0 \quad i > n$$

מאחר ו- $\sum_{i=0}^{\infty} \pi_i = 1$, נובע כי $\pi_0 \geq \sum_{i=0}^{\infty} \left(\frac{\lambda}{\mu \wedge \theta}\right)^i \frac{1}{i!} = e^{-\frac{\lambda}{\mu \wedge \theta}} < \infty$

שכמובן שקול להתכנסות הטור הרלוונטי.

ולכן, לכל ערכי פרמטרים λ, μ, θ, N מודל Erlang-A ארגודי.

הערה: החסמים הנ"ל "הומצאו" בעזרת האטת קצבי המוות, עובדה שנלמדה בכיתה כדי לחסום מלמעלה

(סטוכסטית) את Erlang-A על ידי מודל $M/M/\infty$ מתאים (שהוא תמיד ארגודי).

4.3 (3 נקודות) נניח בנוסף כי $\mu = \theta$. ניתן להראות שלתהליך התפלגות סטציונארית פואסונית (אין צורך להוכיח). נסמן את הפרמטר ההתפלגות הפואסונית ב- R . מה משמעותו של R כמדד תפעולי של מערכת השירות? מהו ערכו של R (כפונקציה של הפרמטרים בבעיה)? חשבו ונמקו.

תשובה: כאשר $\mu = \theta$, Erlang A מתנהג כמו $M/M/\infty$ ולכן, כאמור בשאלה, ההתפלגות הסטציונארית היא פואסונית. $\text{Offered Load} = R$, העומס המוצע למערכת.

4.4 (3 נקודות) מהו מספר השרתים המינימלי שנדרש כדי להבטיח שלפחות 50% מהלקוחות יקבלו שירות מיידי (ללא המתנה בתור). הוכיחו את תשובתכם ללא שימוש בגרפים.

תשובה:

נסמן $L(\infty) = \text{Poisson}(R)$ ההתפלגות הסטציונארית הנ"ל. רוצים N מינימאלי עבורו $P(L(\infty) \geq N) < 0.5$.

$$P(L(\infty) \geq N) = P\left(\frac{L(\infty) - R}{\sqrt{R}} \geq \frac{N - R}{\sqrt{R}}\right) \approx P\left(N(0,1) \geq \frac{N - R}{\sqrt{R}}\right) \approx 0.5$$
$$\Leftrightarrow \frac{N - R}{\sqrt{R}} \approx 0$$

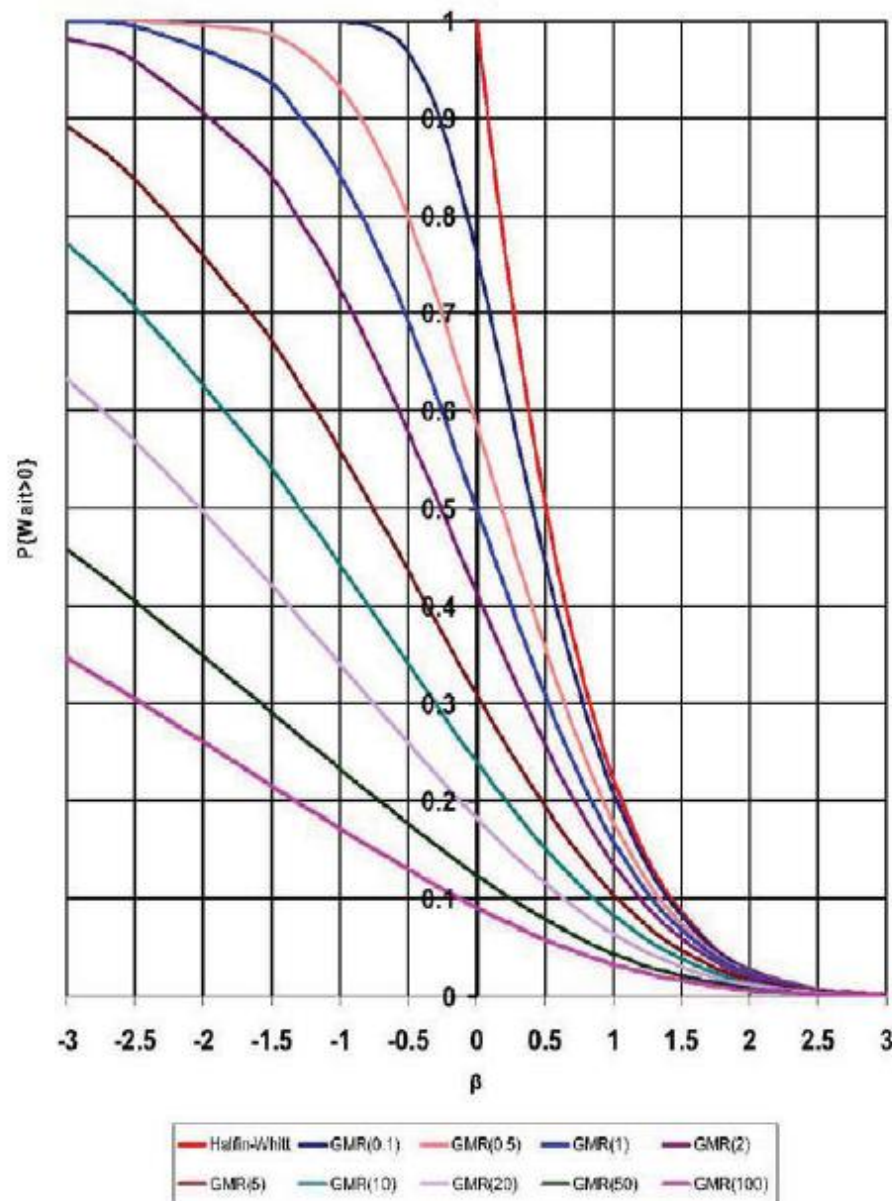
זאת אומרת $N = [R] = \lceil \lambda/\mu \rceil$ הוא המינימום הנדרש (מספר שרתים שווה לעומס המוצע).

4.5 (3 נקודות) כיצד תשתנה תשובתכם לסעיף 4.4 אם ללקוחות המערכת יש סבלנות אינסופית (אין נטישות)? הסבירו תשובתכם תוך שימוש בגרף פונקציות Garnett המופיעות בעמוד הבא.

תשובה: סבלנות אינסופית (שפירושה $\theta = 0$ או $\frac{1}{\theta} = \infty$) מביאה למודל Erlang-C. הפונקציה בגרף של Garnett המתאימה היא הפונקציה של Halfin-Whitt (העליונה בגרף). לשירות מיידי של 50% מהפונים נדרשת רמת איוש של $N = R + 0.5 \sqrt{R}$ ($\beta = \frac{1}{2}$). ז"א נדרשים $0.5 \sqrt{R}$ יותר שרתים כדי להבטיח שירות מיידי בהסתברות 50% (מכיוון שלקוחות אינם נוטשים בעצמם).

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 $\theta/\mu = x$. Here, θ and μ are the abandonment and service rate, respectively.

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

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	$X \sim NB(m, p)$	$p(k) = \binom{k-1}{m-1} p^m q^{k-m}$	$0 \leq p = 1 - q \leq 1; m \geq 1$	$k = m, m+1, \dots$	$\left[\frac{p \cdot e^t}{1 - q \cdot e^t} \right]^m$	$\frac{mp}{p}$	$\frac{mq}{p}$	Number of trials till m -th success
Geometric	$X \sim G(p)$	$p(k) = q^{k-1} \cdot p$	$0 \leq p \leq 1$	$k = 1, 2, \dots$	$\frac{p \cdot e^t}{1 - q \cdot e^t}$	$\frac{1}{p}$	$\frac{q}{p}$	$X \sim NB(1, p)$
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	$X \sim HG(N, M, n)$	$p(k) = \frac{\binom{N}{k} \binom{M-k}{n-k}}{\binom{N+M}{n}}$	$N, M > 0;$ $1 \leq n \leq N + M$	$\max(0, n - M) \leq k$ $k \leq \min(n, N)$	$\sum_k e^{tk} p(k)$	$\frac{n \cdot N}{N+M}$	$n \cdot \frac{N}{N+M} \cdot \frac{N}{N+M} \cdot \frac{M}{N+M} \cdot \frac{M}{N+M} - \left(\frac{n \cdot N}{N+M}\right)^2$	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	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha \cdot \beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	$\alpha = \beta = 1 \rightarrow \text{Uniform}$
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	