



מרצה : ד"ר גלית יום-טוב

מתרגל : אריק סנדרוביץ'

תאריך הבחינה : 11.10.2015

מס' סטודנט _____

שם _____

הנדסת מערכות שירות – 096324

מועד ב' – סמסטר אביב 2015

פתרון

- המבחן נמשך **שלוש** שעות. באחריותך להחזירו במועד.
- **המבחן בחומר סגור.**
- **מותר להשתמש במחשבון מדעי רגיל.**
- המבחן כולל **23** 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 20 נקודות	תרגילי בית
שאלה 2	_____ מתוך 25 נקודות	בחינות קודמות/הרצאות/תרגולים
שאלה 3	_____ מתוך 35 נקודות	יישום
שאלה 4	_____ מתוך 20 נקודות	תיאוריה

סה"כ _____ מתוך 100 נקודות

בהצלחה !

שאלה 1 - תרגילי בית (20 נקודות).

חלק 1 – אמידת סבלנות (10 נקודות)

להלן ארבעה מודלים לתיאור סבלנות של לקוחות:

1. מודל אקספוננציאלי סטנדרטי.

הזמן שלקוח מוכן להמתין הוא מהתפלגות אקספוננציאלית עם פרמטר θ .

2. מודל אקספוננציאלי עם נטישות מיידיות.

לקוח המתקשר ומסתבר לו שהוא נדרש להמתין בתור, נוטש באופן מיידי בהסתברות p (באופן מעשי זמן הנטישה ייקח מספר שניות). במידה והלקוח לא נוטש מיידי, סבלנותו תהיה אקספוננציאלית עם פרמטר θ .

3. סבלנות בת שתי פאזות.

סבלנות $X + Y^d$. כאשר X הינו משתנה מקרי (או קבוע) המתאים לזמן בו הלקוח "מרגיש בסדר" ולא חושב על נטישה, המ"מ Y מתאים לזמן בו הלקוח מתחיל להיות קצר רוח, והתפלגות Y הינה אקספוננציאלית עם פרמטר θ . לבסוף, X ו Y בלתי תלויים.

4. התנהגות אדפטיבית.

סבלנות הלקוחות משתנה במשך שעות היום (לדוגמא בשעות אשר הלקוחות מצפים להמתין זמן רב תהיה להם סבלנות ארוכה).

עליכם למדל את סבלנותם של לקוחות ממענה קולי טלפוני מסוים. כדי לבחור מודל מתאים לסוגים שונים של לקוחות, שורטטו ארבעת הגרפים הבאים:

- כל גרף מייצג סוג לקוחות מסוים (PS, IN, NE, NW).
- כל גרף מתאר פיזור נקודות, כאשר כל נקודה מחושבת כממוצע במשך רבע שעה.
- ערכי נקודה בגרף הם: ממוצע זמן ההמתנה של הלקוחות שהמתינו בתור, ברבע שעה מסוימת, מול אחוז הנוטשים של לקוחות אשר המתינו בתור באותה רבע שעה.
- בכל גרף משורטט קו ישר אשר נוצר על ידי רגרסיה ליניארית בין שני המשתנים.

Figure 1 NE customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

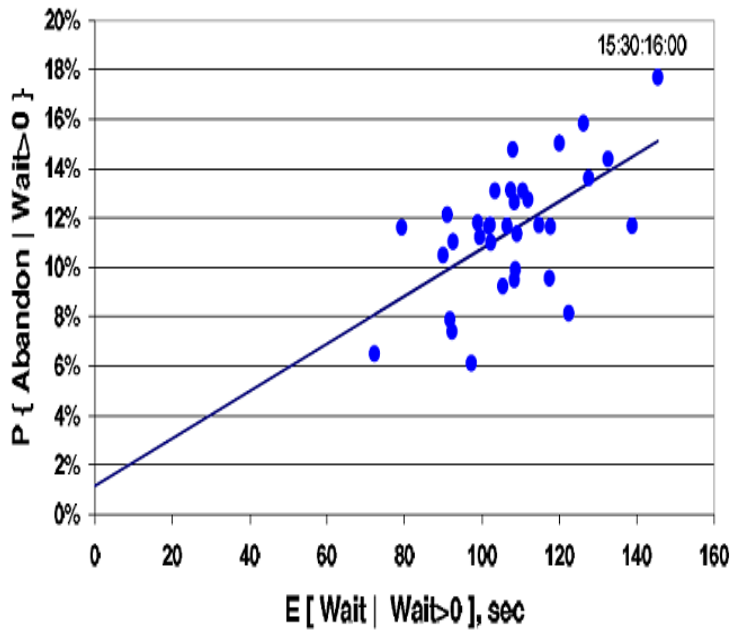


Figure 2 NW customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

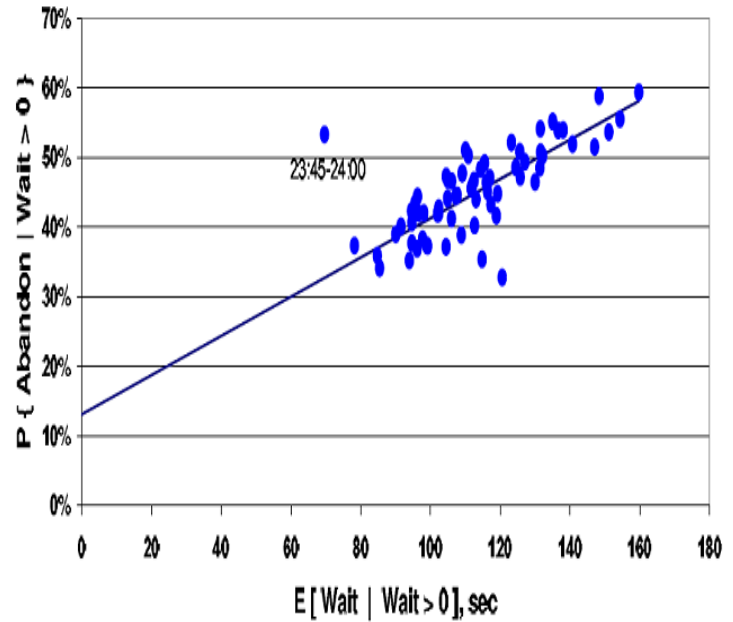


Figure 3 PS customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.

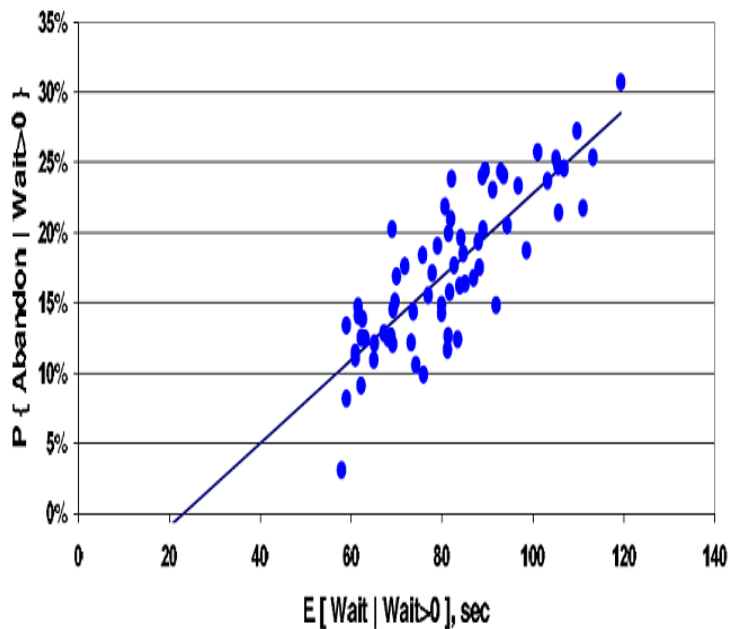
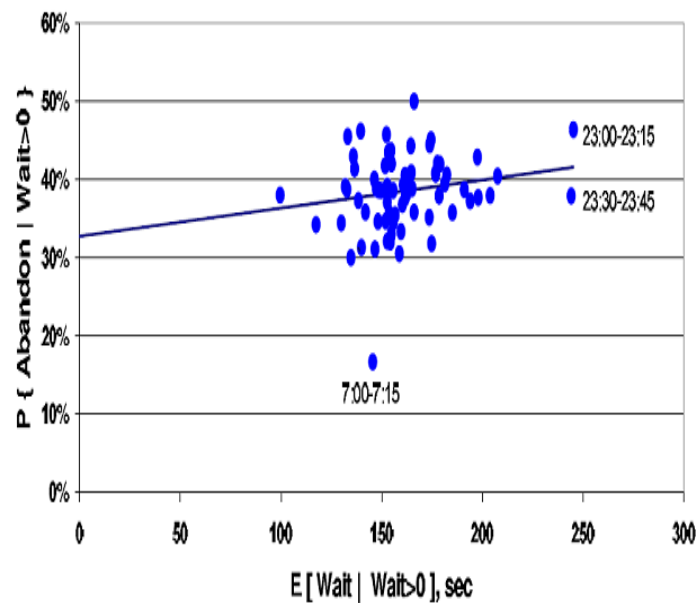


Figure 4 IN customers.
 $P\{\text{Abandon} \mid \text{Wait} > 0\}$ vs. $E[\text{Wait} \mid \text{Wait} > 0]$.



1.1.1 הסבירו בקצרה מדוע הנתונים שנבחרו עבור מידול הסבלנות (אותם נתונים אשר מופיעים בארבעת הגרפים), הם נתונים המדברים על לקוחות שהמתינו בתור בלבד, ולא על כלל הלקוחות. (במילים אחרות, מדוע לא נכללו הלקוחות ששורתו מיידית?) **(2 נקודות)**

תשובה:

כאשר אנו אומדים סבלנות לקוחות אשר לא המתינו בתור, אלו לא תורמים מידע לגבי סבלנותם.

על סמך הגרפים הנ"ל, התאימו מודל סבלנות מתאים (מארבעת המודלים המוצגים בשאלה) לכל אחד מארבעת סוגי הלקוחות. הסבירו את תשובתכם בכל סעיף.

1.1.2 לקוחות NE (2 נקודות)

תשובה:

לקוחות NE

המודל המתאים הוא מודל אקספוננציאלי סטנדרטי. מאחר ובמודל אקספוננציאלי מתקיים הקשר $P(\text{Abandon} | \text{Wait} > 0) = \theta \cdot E(\text{Wait} | \text{Wait} > 0)$, עבור גרף בו קיים קשר ליניארי העובר בראשית הצירים נבחר את המודל האקספוננציאלי הסטנדרטי (הגרף "כמעט" עובר דרך הראשית).

1.1.3 לקוחות NW (2 נקודות)

תשובה:

לקוחות NW

מאחר וישנה נקודת חיתוך עם ציר Y בערך ב- $P(\text{Abandon} | \text{Wait} > 0) \approx 0.12$, המודל המתאים יהיה מודל אקספוננציאלי עם נטישות מיידיות, בו ישנה הסתברות חיובית לנטוש ללא זמן המתנה משמעותי ולאחר מכן אנו מקבלים קו ישר, כלומר לקוחות אשר החליטו לא לנטוש ישר הם בעלי סבלנות אקספוננציאלית.

1.1.4 לקוחות PS (2 נקודות)

תשובה:

לקוחות PS

מאחר והחיתוך עם ציר ה- X הינו בערך ב- $E(\text{Wait} | \text{Wait} > 0) \approx 21\text{sec}$, אנו נבחר את מודל שתי הפאזות כאשר X הינו מ"מ דטרמיניסטי השווה ל 21 שניות ולאחר מכן אנו עוברים לפאזה השנייה Y אשר יוצרת קו ישר ולכן מתאימה להתפלגות אקספוננציאלית.

1.1.5 לקוחות IN (2 נקודות)

תשובה:

לקוחות IN

מאחר ואנו מקבלים קו ישר בעל שיפוע קטן מאד אנו למדים כי ההסתברות לנטישה של לקוחות אלו כמעט ולא תלויה בזמן ההמתנה שלהם ולכן המודל המתאים ביותר יהיה המודל האדפטיבי.

חלק 2 – ניתוח נתוני דוח ACD (10 נקודות)

להלן טבלה המסכמת חלק מנתוני דוח ACD ממוקד שרות איטלקי.

	ASA	Avg Aban Time	ACD Calls	Aban Calls	Avail Time	Staffed Time	Avg ACW time
Totals	29.99	126.24	6719	929	63:27:00	1088:50:00	68.17
8:00	475.00	0.00	1	0	5:25:00	13:32:00	51.49
8:30	8.73	15.55	113	11	3:06:00	23:48:00	55.72
9:00	4.86	0.00	239	0	7:39:00	41:51:00	56.11
9:30	17.08	16.46	391	39	1:05:00	44:48:00	68.89
10:00	25.03	10.71	432	28	1:21:00	53:06:00	71.54
10:30	1.16	0.00	392	0	5:12:00	54:34:00	58.49
11:00	0.15	4.50	357	2	6:48:00	53:56:00	63.82
11:30	0.00	0.00	350	0	9:12:00	60:02:00	76.06
12:00	5.59	13.17	334	12	3:00:00	58:29:00	70.38
12:30	11.90	9.84	337	19	1:05:00	58:47:00	61.8
13:00	7.40	9.40	298	15	1:57:00	33:35:00	74.26
13:30	11.43	13.09	289	23	1:27:00	33:15:00	63.1
14:00	6.35	15.09	323	22	1:27:00	55:13:00	70.76
14:30	6.40	13.10	345	21	1:51:00	57:48:00	63.56
15:00	22.76	18.05	432	22	1:28:00	58:17:00	58.89

15:30	2.41	4.57	574	7	2:35:00	57:55:00	48.89
16:00	1.08	5.00	533	1	4:38:00	58:06:00	48.57
16:30	50.28	36.57	406	121	1:00:00	59:17:00	74.7
17:00	163.99	50.97	431	317	0:00:00	53:53:00	54.6
17:30	426.70	214.25	128	216	0:00:00	31:37:00	80.27
18:00	584.69	737.27	13	49	0:00:00	26:08:00	157.15

להלן פירוט השדות בדוח :

- ASA – Average Speed of Answer (sec)
זמן המתנה ממוצע בתור של לקוחות שקיבלו שירות.
- Avg Aban Time – Average Abandon Time (sec)
זמן המתנה ממוצע בתור של לקוחות שנטשו.
- ACD Calls
מספר הלקוחות שקיבלו שירות
- Aban Calls
מספר הלקוחות שנטשו.
- Avail Time – Available Time
סך כל זמן הבטלה של השרתים.
- Staffed Time
סך כל זמן ה- Log in של השרתים (זמן בין התחברות לעמדת העבודה ועד יציאה מעמדת העבודה. בזמן זה השרת יכול לענות לשיחות, לטפל בניירת, לבצע פעילויות אחרות או להיות בטל).
- Avg ACW Time- Average after call work time
זמן עבודה ממוצע בשיחה אחרי שהסתיימה

- 1.2.1. תארו כיצד יש לחשב את סך ההגעות, אחוז הנוטשים ותוחלת זמן ההמתנה באינטרוול של חצי שעה.

(3 נקודות)

תשובה:

$$\text{Average waiting time} = \frac{\text{Avg Speed Ans} \times \text{ACD Calls} + \text{Avg Aban Time} \times \text{Aban Calls}}{\text{Number of arrivals}},$$

where

$$\text{Number of arrivals} = \text{ACD Calls} + \text{Aban Calls}.$$

$$\text{Fraction of abandoning customers} = \frac{\text{Aban Calls}}{\text{Number of arrivals}}.$$

- 1.2.2. תארו בפירוט כיצד תאמדו את ממוצע זמן הסבלנות בכל אינטרוול? כיצד תחשבו את ממוצע זמן הסבלנות היומי? (3 נקודות)

תשובה:

$$\text{Average patience} = \frac{\text{Average wait}}{P\{\text{Ab}\}}$$

לכל אינטרוול, ואז נשקלל על פי מופעים לרמת היום.

1.2.3. תארו כיצד ניתן לחשב את משך השירות הממוצע ואת האיוש הממוצע בכל אינטרוול באמצעות נתוני הדו"ח. (4 נקודות)

תשובה:

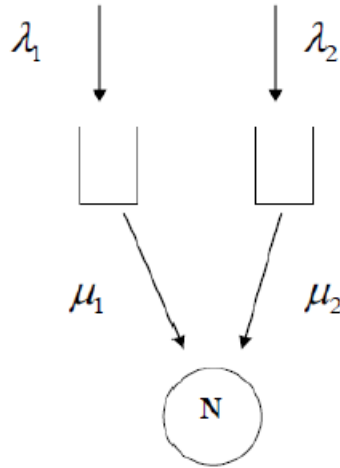
$$\text{Average service time} = \frac{\text{Staffed Time} - \text{Avail Time}}{\text{ACD Calls}} .$$

$$\text{Staffing} = \frac{\text{ACD time} + \text{ACW time}}{(1/2 \text{ hour}) \times \text{utilization}}$$

שאלה 2 – בחינות קודמות/הרצאות/תרגולים (25 נקודות).

חלק 1 - מודלי נוזלים ותכן של מערכות שירות (10 נקודות).

נתונה המערכת הבאה, אשר מופיעה בתרגול מספר 4 :



המערכת משרתת שני סוגים של לקוחות, לקוחות רגילים (Regular) ולקוחות VIP. במערכת N שרתים מסוג זה. נסמן את קצבי ההגעות המשתנים בזמן: $\lambda_1(t)$ עבור לקוחות רגילים ו- $\lambda_2(t)$ עבור לקוחות VIP. קצבי השירות קבועים בזמן ומסומנים ב- μ_1 עבור לקוחות רגילים המשרתים אצל שרת מסוג 1, μ_2 עבור לקוחות VIP. נסמן ב- $Q_1(t)$ ו- $Q_2(t)$ את כמות הלקוחות מסוג רגיל ו-VIP במערכת בזמן t (בהתאמה), והניחו כי המערכת ריקה בזמן $t=0$.

2.1.1. האם אפשר לכתוב את המשוואות שמתארות את מספר הלקוחות במערכת ללא הנחות נוספות? אם כן, כתוב את המשוואות. אם לא, הסבר. (3 נקודות)

תשובה:

לא ניתן לכתוב את משוואת הנוזלים ללא הנחה על תעדוף בין סוגי הלקוחות.

2.1.2. דני, מנהל התפעול של המערכת מציע להפריד את קבוצת השרתים לשתי קבוצות: קבוצה אחת שתעסוק בלקוחות רגילים בלבד, וקבוצה שניה שתעסוק בלקוחות VIP בלבד. לאיזה חסרון של המערכת המקורית הפרדה זו תיתן מענה? הסבר. (3 נקודות)

תשובה:

דני מציע לעבור ממערכת V לשתי מערכות I.

במערכת המקורית כל המוקדנים הם בעלי הכשרה זהה. במערכת אותה מציע דני, יהיו מוקדנים שיהיו כשירים לתת מענה ללקוחות VIP מה שיעלה את רמת השירות של לקוחות אלו. בנוסף, הפרדה זו תקצר את זמן ההכשרה ולכן תוריד את עלויות העובדים.

תשובה לא נכונה: לקוחות VIP יקבלו מענה מהיר יותר, או שניתן יהיה להקצות קיבולת גבוהה ללקוחות VIP. במערכת המקורית תחת הנחת PREEMPTION – לקוחות VIP יהיו למעשה במערכת דומה לזו שמציע דני מבחינה תפעולית. עם זאת הם יקבלו שירות פחות טוב משרתים שהם GENERALISTS.

2.1.3 מהו החסרון הבולט במערכת שמציע דני? הציעו חלופה להצעה של דני שתיתן מענה לחסרון זה. (4

נקודות)

תשובה:

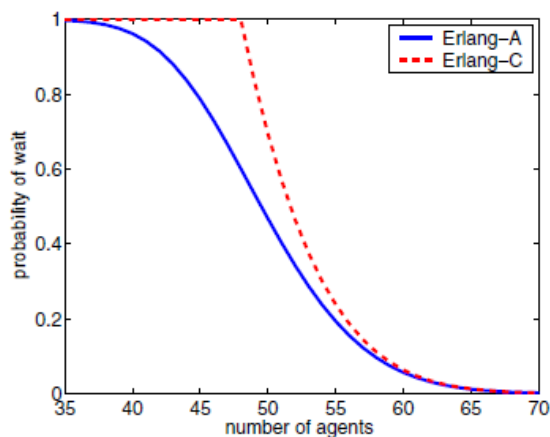
החסרון הוא ששרתי VIP או REGULAR עשויים להיות מורעבים בחלקים מסויימים של הזמן (תלוי בקצב המופע ובזמני העיבוד). כדי להתגבר על חסרון זה אפשר למשל להפעיל מערכת תמיכה בין השרתים, כך שלמשל שרתי VIP יתנו תמיכה גם ללקוחות רגילים כאשר אין לקוחות VIP בתור. למעשה נעבור משני תורי I לתור N.

להלן שני תרשימים של השתנות שני מדדי שירות כפונקציה של מספר השרתים עבור שני המודלים - ERLANG-A ו-ERLANG-C:

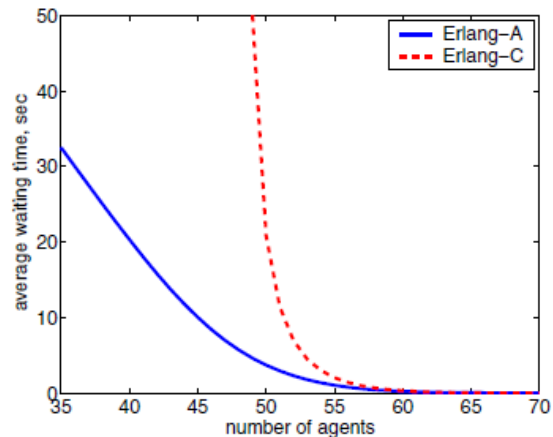
Erlang-A vs. Erlang-C

48 calls per min, 1 min average service time,
2 min average patience

probability of wait
vs. number of agents



average wait
vs. number of agents



2.2.1 התרשימים שמתאימים ל- ERLANG-C (הקיום המקווקים) מציגים התנהגות אסימפטוטית, ולעומת זאת ERLANG-A (הקו החלק) לא מראה התנהגות כזו. מה היא הסיבה לכך? הסבר באופן מדויק ככל הניתן.
(3 נקודות)

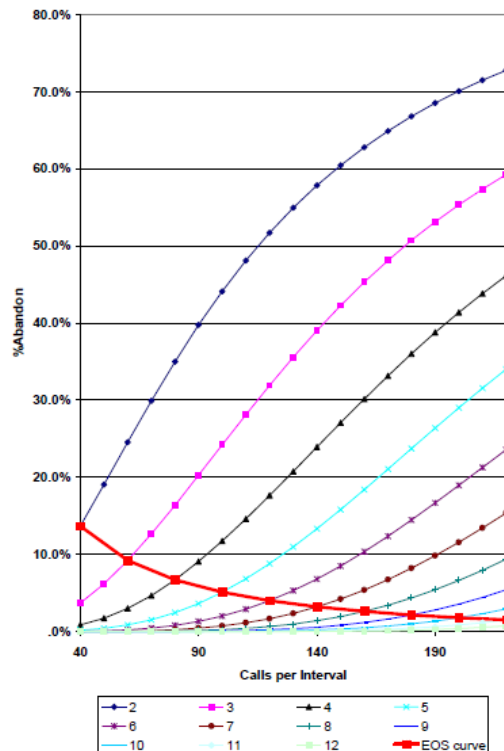
תשובה:

מודל Erlang-A תמיד יציב (לכל מספר מוקדנים כולל 0).

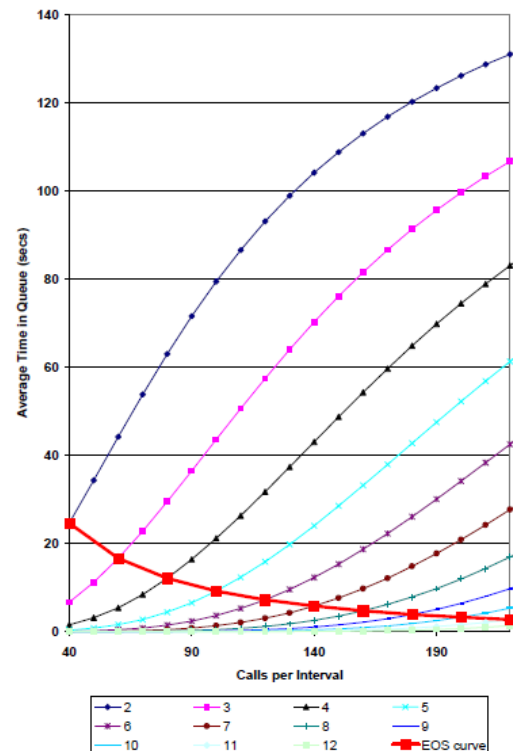
מודל Erlang-C לא יציב עבור מספר מוקדנים קטן מהעומס המוצע – לכן יש לו אסימפטוטה ב-48 מוקדנים.

להלן שני תרשימים של שני מדדי שירות למודל ERLANG-A, עבור הנתונים המופיעים מתחת לתרשימים:

Probability to abandon



Average wait



Example: $1/\mu = 2$ minutes, $1/\theta = 3$ minutes;
 λ varies from 40 to 230 calls per hour, in steps of 10;
 n varies from 2 to 12.

כל סדרה בתרשים תואמת מספר שונה של מוקדנים שהוצבו לטפל בביקוש הנכנס.

2.2.2 מדוע יש דמיון בין שני הגרפים? הסבר בטיעון מתמטי. (3 נקודות)

תשובה:

מאחר והקשר:

$$P\{ab\} = \theta EW$$

מתקיים במדויק עבור מודל Erlang-A יש קורלציה של 1 בין שני הגרפים (כלומר, יש להם אותו מבנה).

2.2.3 בהנחה שקצב ההגעה הינו 140 לקוחות לשעה, כמה מוקדנים צריך, על סמך הגרפים הנ"ל, כדי שזמן ההמתנה הממוצע יהיה פחות מ 30 שניות ואחוז הנוטשים לא יעלה על 9% ? (3 נקודות)

תשובה:

בכדי לעמוד בשתי רמות השירות הנ"ל צריך 6 מוקדנים.

2.2.4 העקומה היורדת המכונה בתרשימים EOS curve מספרת על יתרון לגודל. הסבר מדוע? השתמש במושג "עומס מוצע". (3 נקודות)

תשובה:

במערכות שירות גדולות עם מספר רב של שרתים קיימת תופעה שנקראת יתרון לגודל – Economies of Scale בתחומים תפעוליים מסויימים.

בגרף לעיל רואים כי העומס המוצע פר שרת נשאר קבוע על קו ה-EOS, עם זאת ככל שהמערכת גדולה יותר ביצועי המערכת משתפרים. ההסתברות למצא שרת פנוי עולה ככל שגודל המערכת עולה, מה שמביא לעליה במדדי הביצוע מקודת מבטו של הלקוח.

זה קורה רק בתחום ה-QED (הביטא בגרפים היא 0).

להלן דוגמה שמציגה מדדי ביצוע של מודל Erlang-A עם הגידול במערכת בתחום התפעולי מכוון יעילות (Efficiency-Driven) (כלומר, כאשר משתמשים בכלל האיוש: $\gamma = 1/6$, $n \approx n - \gamma R$).

$$E[S] = 6 \text{ min}, E[\tau] = 9 \text{ min}$$

λ/hr	n	Occupancy	$P\{Ab\}$	$E[W_q]$	$P\{W_q > 0\}$
12	1	73.4%	38.8%	3:29	73.4%
48	4	89.8%	25.2%	2:16	75.6%
192	16	97.5%	18.7%	1:41	85.4%
768	64	99.8%	16.8%	1:31	97.2%
3,072	256	100.0%	16.7%	1:30	100.0%
↓	↓	↓	↓	↓	↓
∞	∞	100%	16.7%	1:30	100%

2.2.5 האם התופעה של היתרון לגודל קיימת במקרה זה? הסבר. (3 נקודות)

תשובה:

בתחום זה אין קיום של יתרון לגודל, כיוון שלא כל מדדי הביצוע משתפרים עם העליה בגודל. לדוגמה ההסתברות להמתנה נעשית יותר גרועה.

שאלה 3 – יישום (37 נקודות)

להלן קטע ממאמר שפורסם בעיתון הארץ, בנושא הרכבת הקלה בתל אביב אין צורך לקרוא את הכתבה מאחר וההסברים והנתונים כלולים הגוף השאלה על סעיפיה השונים.



ביום ראשון יחלו שש שנים של חפירות שצפויות לשנות את פניו של גוש דן. ברשויות ובחברה המבצעת מבקשים להרגיע, אבל יש מי שבכל זאת מתכוננים לגרוע מכל



אילן ליאור |
11:47 27.07.2015

80 שנה חלפו מאז עלה לראשונה רעיון הקמתה של **הרכבת הקלה בתל אביב**. שרי תחבורה וראשי ערים התחלפו, תוכניות הוגשו, שונו, אושרו והוקפאו, תקציבים הוזרמו ואפילו אבני פינה הונחו, אבל הרכבת נותרה חלום רחוק. בשבוע הבא - 28 שנים אחרי שקו פתח תקוה-תל אביב נבחר להיות החלוץ במערכת הסעת ההמונים בגוש דן ו-12 שנים אחרי פרסום המכרז הראשון לקו האדום של הרכבת - התוכניות צפויות להפוך למציאות. עבודות **הרכבת הקלה**, התחתית בחלקה, יחלו אז במלוא המרץ, והתנועה בגוש דן תעבור מהפכה. אם לא יהיו עיכובים בלתי צפויים, הרכבת הקלה תחצה את תל אביב בעוד שש שנים - אך הדרך ליעד ארוכה ורצופת מכשולים.

"אנחנו נעשה את המקסימום להקל, אבל זה הולך להיות קשה", מודה עוזי יצחקי, מנכ"ל משרד התחבורה. "זו עבודה שהיתה צריכה להתבצע באמצע המאה הקודמת.

בעקבות עבודות התשתית של הרכבת הקלה בתל-אביב, החליט משרד התחבורה להפעיל שירותי הסעה (שאטלים) להקל על הכניסה למרכז העיר: אלו למעשה אוטובוסים שיצאו מחניונים מרכזיים המפוזרים בפאתי העיר תל אביב, ויפזרו את הנוסעים בין יעדים מרכזיים בעיר תוך שהם מורידים נוסעים (ללא העלאת נוסעים חדשים).

3.1 (13 נקודות) אנו מעוניינים בניתוח ארוך טווח של הפעלת שירותי ההסעה של כל יום בשבוע (אופק התכנון הוא אינסוף ימים). קצב הגעת הנוסעים החזוי לחניון "גני יהושע" הינו 1200 נוסעים לשעה. משרד התחבורה הגדיר כי נוסע לא יחכה לשאטל מעל 5 דקות בממוצע.

3.1.1 (3 נקודות) מהו מספר הנוסעים הממוצע שיחכה בתחנה? הסבר ופרט את ההנחות שהנחת לצורך החישוב.

תשובה:

תנאי שימוש בליטל: קיום גבול עבור התהליך הארוך-טווח.

על פי חוק ליטל:

$$L_p = \lambda_p W_p \leq 20 \cdot 5 = 100$$

3.1.2 (3 נקודות) כעת הניחו כי משך ההמתנה הממוצע של נוסע בתחנה הינו 5 דקות בדיוק. נסמן את התהליך המונה את מספר הנוסעים בתחנה ב- $L(t)$. הניחו כי תהליך המופע של אוטובוסים לתחנה הוא תהליך פואסון הומוגני בזמן והוא בלתי תלוי במופעי הנוסעים. כמה נוסעים יחכו בתחנה, בממוצע, ברגע המופע של אוטובוס? הסבירו.

תשובה:

ברגע מופע של אוטובוס יחכו בתחנה 100 נוסעים. אנו יודעים זאת על סמך PASTA והפתרון של סעיף 3.1.1. האוטובוסים דוגמים את המערכת בתהליך פואסון ולכן רואים את התהליך $L(t)$ שהוא תהליך חידוש (מתקיימים התנאים ל-PASTA) בממוצע הארוך-טווח שלו.

3.1.3. (3 נקודות) כעת הניחו כי גם הנוסעים מופיעים בתהליך פואסון לתחנה. מהו זמן ההמתנה הממוצע של נוסע המופיע לתחנה, כאשר תהליך המופע של האוטובוסים לתחנה הוא תהליך פואסון עם קצב λ לשעה? נמק. ניתן להשתמש בעמוד העזר המצורף בעמוד הבא.

תשובה:

התשובה היא: $\frac{1}{\lambda}$

נימוק: אמידה מוטת – מקדם ההשתנות של זמן מופעי בין אוטובוסים הוא 1 (תהליך פואסון) ואז כשמציבים בנוסחאות שמצורפות מתקבל:

$$\tau = \frac{1}{2} \cdot \frac{1}{\lambda} \cdot 2 = \frac{1}{\lambda}$$

3.1.4. (4 נקודות) הוחלט כי רוצים לצמצם את זמן המתנה הממוצע של נוסע שמגיע לתחנה באקראי (מופע פואסון) אך ללא הוספת אוטובוסים נוספים. כלומר הניחו כי ממוצע הזמן הבין-מופעי של אוטובוסים ישאר זהה לזמן משאלה 3.1.3. כיצד ניתן להשיג מטרה זו? מה יהיה זמן ההמתנה הממוצע של נוסע?

תשובה:

ניתן להוריד את זמני ההמתנה של לקוחות על ידי הורדת השונות של תהליך המופע של האוטובוסים. אם האוטובוסים יגיעו לפי תהליך דטרמניסטי (בדיוק בזמן), זמן ההמתנה ירד לחצי ממה שהיה קודם. על פי הנוסחה:

$$\begin{aligned} \Rightarrow \tau &= \frac{1}{T} \cdot (\text{area under } X, \text{ over } [0, T]) \\ &\approx \frac{1}{T} \cdot \left(\frac{1}{2}(T_1^b)^2 + \frac{1}{2}(T_2^b)^2 + \dots + \frac{1}{2}(T_{R_b(T)}^b)^2 \right) \\ &= \frac{R_b(T)}{T} \cdot \frac{1}{2} \cdot \frac{T_1^2 + \dots + T_{R_b(T)}^2}{R_b(T)} \xrightarrow{T \rightarrow \infty} \frac{1}{E(T_1^b)} \cdot \frac{1}{2} \cdot E(T_1^b)^2, \text{ by SLLN} \\ &= \underbrace{\frac{1}{2}E(T_1^b)}_{\text{"Deterministic" answer}} \underbrace{[1 + c^2(T_1^b)]}_{\text{Bias, due to variability}}, \quad c = \frac{\sigma}{E} \text{ coefficient of variation.} \end{aligned}$$

הערך הנמוך ביותר לזמן ההמתנה מתקבל כאשר התפלגות הזמן הבין-מופעי של אוטובוסים יהיה דטרמניסטי (כי הפונקציה מונוטונית עולה כתלות במקדם ההשתנות (עבור ממוצע זהה, כפי שנתון בשאלה). עבור זמן דטרמניסטי הנוסע ימתין חצי מהזמן בסעיף 3.1.3 ולכן התשובה היא:

$$\tau = \frac{1}{2} \cdot \frac{1}{\lambda}$$

Application of PASTA: Biased Sampling

A *renewal process* is a counting process with iid interarrivals.

Descriptions: $R = \{R(t), t \geq 0\}$ or $\{T_1, T_2, \dots\}$ iid, or $\{S_1, S_2, \dots\}$
 Example: Poisson exponential Erlang

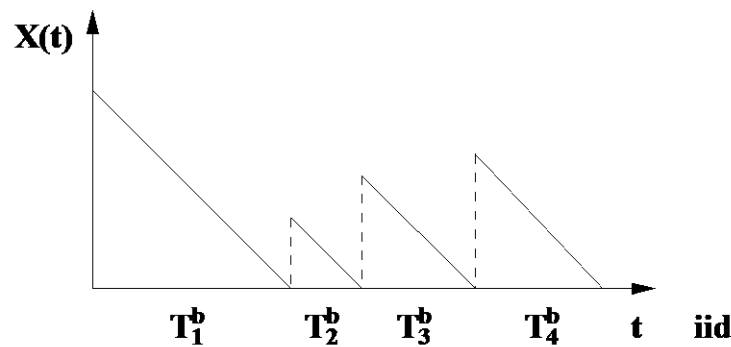
Story: Buses arrive to a bus stop according to a renewal process $R_b = \{R_b(t), t \geq 0\}$.

T_i^b — times between arrivals of the buses.

Passengers arrive to the bus stop in a completely random fashion (Poisson).

S_i^p — arrival times of the passengers.

Question: How long, on average, do they wait? Plan service-level.



$A = \{A(t), t \geq 0\}$ = Poisson arrivals of passengers.

$X = \{X(t), t \geq 0\}$ = state = *Virtual waiting time*.

$$\text{PASTA: } \lim_{T \uparrow \infty} \frac{1}{T} \int_0^T X(t) dt = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=1}^N X(S_n^p) = \bar{\tau}$$

$$\begin{aligned} \Rightarrow \tau &= \frac{1}{T} \cdot (\text{area under } X, \text{ over } [0, T]) \\ &\approx \frac{1}{T} \cdot \left(\frac{1}{2}(T_1^b)^2 + \frac{1}{2}(T_2^b)^2 + \dots + \frac{1}{2}(T_{R_b(T)}^b)^2 \right) \\ &= \frac{R_b(T)}{T} \cdot \frac{1}{2} \cdot \frac{T_1^2 + \dots + T_{R_b(T)}^2}{R_b(T)} \xrightarrow{T \rightarrow \infty} \frac{1}{E(T_1^b)} \cdot \frac{1}{2} \cdot E(T_1^b)^2, \text{ by SLLN} \\ &= \underbrace{\frac{1}{2}E(T_1^b)}_{\text{"Deterministic" answer}} \underbrace{[1 + c^2(T_1^b)]}_{\text{Bias, due to variability}}, \quad c = \frac{\sigma}{E} \text{ coefficient of variation.} \end{aligned}$$

Check Poisson bus arrivals to derive Paradox:

$$\frac{1}{2} (\text{"stochastic" answer}) = \frac{1}{2} (\text{"deterministic" answer}).$$

3.2 (15 נקודות) באחד החניונים מספר השאטלים לא מספיק כדי לעמוד בביקוש ולכן הוחלט להפעיל שירות מוניות ספיישל בתשלום מוזל אשר יתמוך בעומס. רמת השירות בשירות זה מוגדרת כדלהלן: **ממוצע המתנה של לכל היותר 5 דקות עבור נוסעים שנאלצים להמתין**.

יש לתכנן את מספר המוניות שישרתו את הנוסעים. קצב המופע של הנוסעים לשירות המוניות הינו 120 לשעה (הנח מופע פואסון). תחילה, מתכנני המערכת משערכים את ממוצע זמן השירות (כלומר זמן שלוקח למונית לצאת מהחניון ולחזור לאיסוף הבא) ב-20 דקות נסיעה. הניחו כי נוסעים לא נוטשים ומחכים עד להגעת המונית הבאה וכי כל מונית יכולה לקחת נוסע אחד לכל היותר.

3.2.1. (3 נקודות) מהו מספר המוניות המינימאלי שיבטיח את יציבות המערכת?

תשובה:

מכיוון שאין נטישות, תנאי היציבות הוא:

$$N \geq R$$

$$N \geq \frac{\lambda}{\mu} = \frac{120}{3} = 40$$

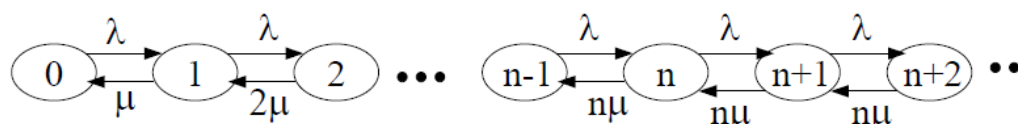
כלומר, יש צורך ב-40 מוניות לפחות.

נמדל את המערכת במודל המרקובי Erlang-C. כלומר, קצב המופע של הנוסעים הוא פואסון עם קצב 120 לשעה, משך השירות מתפלג מעריכית עם תוחלת משך שירות של 20 דקות, ומספר המוניות הינו קבוע – נסמנו ב-N. שים לב כי זהו מודל מצב-יציב.

3.2.2. (3 נקודות) שרטטו את דיאגרמת קצבי המעבר של מודל זה וכתבו את קצבי המעבר לכלל מצבי המערכת.

תשובה:

דיאגרמה של EC:



$$\lambda_j \equiv \lambda, \quad j \geq 0,$$

$$\mu_j = (j \wedge n)\mu, \quad j \geq 1.$$

תשובה:

Steady-state distribution:

$$\begin{aligned}\pi_i &= \frac{R^i}{i!} \pi_0, & i \leq n, \\ &= \frac{n^n \rho^i}{n!} \pi_0, & i \geq n, \\ \pi_0 &= \left[\sum_{j=0}^{n-1} \frac{R^j}{j!} + \frac{R^n}{n!(1-\rho)} \right]^{-1},\end{aligned}$$

where $R = \lambda/\mu$ is the offered load.

3.2.4. (3 נקודות) נגדיר את $E(2, n)$ כהסתברות להמתין למונית. השתמש במשוואות המצב היציב וב- $E(2, n)$ על מנת לרשום את **התפלגות מספר האנשים בתור**.

תשובה:

Number-in-queue:

$$P\{L_q = i\} = E_{2,n} \cdot (1 - \rho)\rho^i, \quad i > 0,$$

or

$$L_q = \begin{cases} 0 & \text{wp } 1 - E_{2,n} \\ \text{Geom}(1 - \rho) & \text{wp } E_{2,n} \end{cases}$$

כאשר $E(2, n)$ הינו:

$$P\{W_q > 0\} \triangleq E_{2,n} = \sum_{i \geq n} \pi_i = \frac{R^n}{n!} \frac{1}{1 - \rho} \cdot \pi_0.$$

3.2.5. (3 נקודות) מהו מספר המוניות הדרוש, כך שזמן המתנה הממוצע (ללקוחות ממתנינים) לא יעלה על 5 דקות. יש לחשב במדויק.

תשובה:

מספר האנשים בתור בהינתן המתנה הוא מ"מ גאומטרי:

$$L_q \sim \text{Geon}(1-\rho)$$

$$EL_q = \frac{1}{1-\rho} = \frac{1}{1-\frac{R}{n}}$$

$$R = \lambda / \mu = 40$$

$$EL_q = \frac{n}{n-40}$$

$$EW_q = \frac{EL_q}{\lambda} \leq \frac{1}{12}$$

$$EL_q = \frac{n}{n-40} \leq 10$$

$$9n \geq 400$$

$$n \geq 44.4$$

$$n_{\min} = 45$$

3.3. (9 נקודות) לאחר ביצוע ניסויים של נסיעות המונית דרך המסלול שנקבע מראש, התברר כי ממוצע השירות הינו 20 דקות וסטיית התקן הינה 30 דקות.

3.3.1. (3 נקודות) איזו הנחה מהנחות מודל Erlang-C ממצא זה מפר? נמק.

תשובה:

נתונים אלו מצביעים על כך שכפי הנראה משך השירות אינו אקספוננציאלי (למשתנה מקרי אקספוננציאלי מקדם השתנות = 1).

בעקבות הממצאים הוחלט לעבור למודל כללי יותר של המערכת: מודל תורים M/G/N. נשתמש בקירוב G/G/N שנלמד בהרצאות:

Kingman's Exponential Law:

$$\frac{W_q}{E(S)} \approx \begin{cases} \exp\left(\text{mean} = \frac{1}{n} \cdot \frac{1}{1-\rho} \cdot \frac{C^2(A)+C^2(S)}{2}\right) & , \text{wp } E_{2,n}, \\ 0 & , \text{otherwise.} \end{cases}$$

In particular, a popular measure for service-level, used to determine the number-of-servers n , is:

$$P\{W_q > x \cdot E(S)\} \approx E_{2,n} \cdot \exp\left(-x \cdot \frac{2n(1-\rho)}{C^2(A) + C^2(S)}\right), \quad x > 0.$$

Heavy-Traffic Approximation:

$$E(W_q) \approx E(S) \cdot \frac{1}{n} \cdot \frac{E_{2,n}}{1-\rho} \cdot \frac{C^2(A) + C^2(S)}{2}.$$

or equivalently,

$$E(W_q) \approx E(W_{q,M/M/n}) \cdot \frac{C^2(A) + C^2(S)}{2}.$$

3.3.2. (3 נקודות) הניחו כי מספר המוניות נשאר כפי שנקבע על ידי מודל Erlang-C. מהו ממוצע זמן ההמתנה של נוסע על פי הקירוב? כתבו חישוב מדויק.

תשובה:

מקדם השתנות המופעים הוא 1, בעוד שלשירותים הוא 1.5. מציבים בנוסחה ומכפילים ב-5 (שזה משך ההמתנה הישן).

התוצאה: 8.125 דקות

3.3.3. (3 נקודות) הסבירו מדוע המדד $P(W_q \geq xE[s])$ חשוב במערכות שירות.

תשובה:

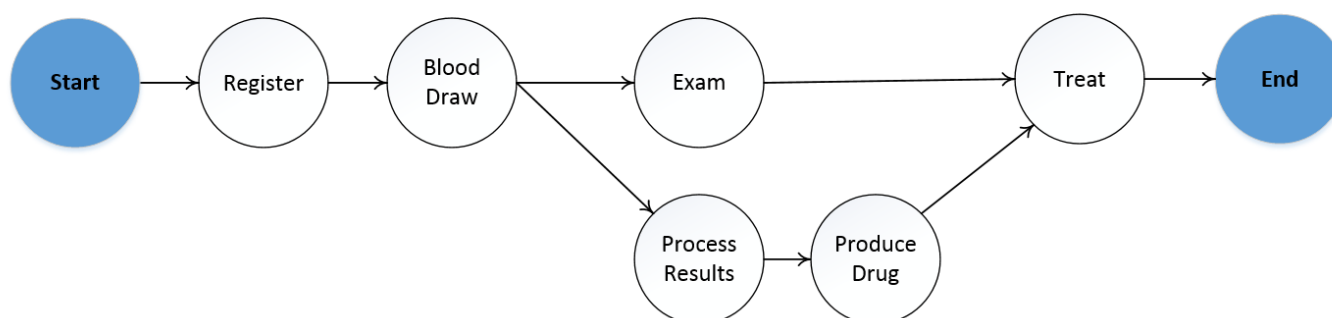
נוסחה זו מביאה בחשבון שני אלמנטים הקושרים זמני המתנה לשביעות רצון של לקוחות:

1. רמת שירות סבירה נמדדת הרבה פעמים ביחס לזמן שירות ממוצע, אם השירות קצר אנשים מצפים להמתין פחות זמן.

2. מדידת אחוזון לעומת מדידת ממוצע. אנשים זוכרים את החוויות הגרועות שלהם, ולכן חשוב לדעת מה אחוז האנשים שחוו חוויה לא נעימה ולא רק מה היה הזמן הממוצע שכולם יחד חוו.

שאלה 4 – תיאוריה (18 נקודות)

להלן רשת שירות (Service Network) המהווה מודל לתהליך השירות בבית חולים לטיפול סרטן:



חולים עוברים את הפעולות: (1) קבלה (Register), (2) נטילת דם (Blood Draw), (3) בדיקת רופא (Exam), (4) עיבוד בדיקת הדם (Process Result), (5) הכנת תרופה לטיפול (Produce Drug) ו-(6) טיפול (Treat). על פי סדר הקדימויות המתואר ברשת השירות, הפיצול לאחר Blood Draw הינו פיצול מסוג AND שבו היישות של המטופל מתפצלת לשני הנתבים יחד, והפעולה Treat לא תתחיל עד לאיחוד בין שני הנתבים. נסמן ב- T_i את משך הזמן שמטופל שווה בפעולה ה- i כאשר i מסמן את מספר הפעולה ($i=1, \dots, 6$). הניחו כי משך פעולה תלוי רק בסוג הפעולה ולא במטופל ספציפי.

4.1 (3 נקודות) הניחו כי המשכים, T_i , הינם דטרמיניסטיים וכי אין אילוצי משאב (המשאבים תמיד זמינים לביצוע הפעולות). נסמן ב- T_D את משך הביצוע של טיפול במטופל יחיד. כתבו את הביטוי ל- T_D כפונקציה של T_i .

תשובה:

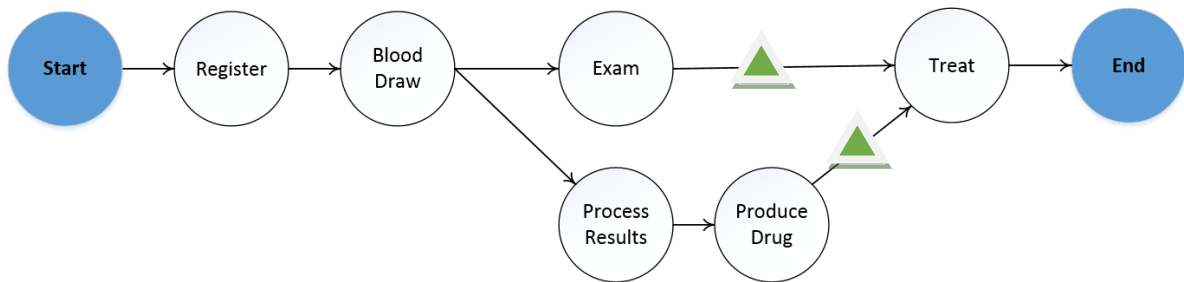
$$T_D = T_1 + T_2 + \max\{T_3, T_4 + T_5\} + T_6$$

4.2 (6 נקודות) כעת הניחו כי משכי הפעולות סטוכסטיים ובלתי תלויים האחד בשני.

4.2.1 (3 נקודות) האם גם במצב בו יש מספר בלתי מוגבל של משאבים יהיו קיימים תורים במערכת? אם כן, השתמשו בסימונים המקובלים עבור תורים אלו וסמנו את המיקום בהם נוצרים תורים אלו בתרשים שלהלן והסבירו מדוע נוצרים תורים אלו. אם לא, הסבירו מדוע אין תורים במערכת.

תשובה:

כן, סימון:



אלו הם תורי סנכרון בין סיום ייצור התרופה לבדיקת הרופא.

לא, הסבר:

המקרה היחיד בו לא יהיו תורי סינכרון הוא אם משכי הפעולות דטרמניסטיים ו- $T_3 = T_4 + T_5$.

4.2.2 (3 נקודות) נסמן את משכי הפעולות ב- X_i . הניחו כי $E[X_i] = T_i$ לכל i . נסמן ב- T_S את משך

הטיפול הכולל של מטופל יחיד (המשיכו להניח כי אין אילוצי משאבים). הוכח/הפוך: עבור הרשת

המתוארת בשאלה מתקיים $E[T_S] \geq T_D$.

תשובה:

$$E[T_S] = EX_1 + EX_2 + EX_6 + E[\max\{X_3, X_4 + X_5\}] \geq T_1 + T_2 + T_6 + \max\{EX_3, EX_4 + EX_5\} = T_D$$

המעבר * בשל אי שוויון ינסן.

4.3 (6 נקודות) כעת הניחו כי כל פעולה מתאימה לתחנת שירות עם מספר משאבים סופי. כמו כן, הניחו כי אנו בתרחיש דינאמי, כלומר מטופלים רבים עוברים ברשת. לשימושכם סימולטור מדויק ככל הניתן של המערכת (המבוסס על נתוני אמת).

4.3.1 (3 נקודות) כיצד תחשבו את הנתיב הקריטי בתרחיש הדינאמי-סטוכסטי בעזרת הסימולטור?

תשובה:

נריץ את הסימולטור N פעמים (N מספיק גדול בשביל מובהקות סטטיסטית). בכל הרצה נתעד מה היה הנתיב הקריטי לסיום הטיפול. נחזיר את ההתפלגות של הנתיב הקריטי, כלומר את ההסתברות שכל נתיב היה הנתיב הארוך ביותר.

4.3.2 (3 נקודות) לאחר הרצת הסימולטור יצא כי תחנת הרופאים אשר מבצעת את הפעולה Exam, היא התחנה בעלת הנצילות משאב הגבוהה ביותר ובה משכי ההמתנה למשאב הארוכים ביותר. הוכח/הפוך את הטענה: במערכת יציבה קיצור משך השהיה של מטופל בתחנת הרופאים יביא בהכרח להקטנת משך הטיפול הכולל של מטופל.

הוכחה/הפרכה:

הטענה אינה נכונה.

ניתן דוגמה נגדית. זמן ההמתנה בתחנת הרופאים סופי (מערכת יציבה). ייתכן מצב בו לכל מספר משאבים בתחנת הרופאים יתקיים: $T_3 + W_3 \leq T_4 + T_5$. במצב כזה, גם עם אינסוף משאבים בתחנת הרופאים, נקבל את אותו משך שהיה של מטופלים במערכת.

4.4 (3 נקודות) הצע שתי חלופות אפשריות להקטנת משך השהיה הכולל של המטופלים. לכל חלופה, התייחס באיזה סוג המתנה היא מטפלת.

תשובה:

1. הגדלת מספר המשאבים – טיפול בתורי משאב.
2. קיצור משכי עיבוד – טיפול בתורי סנכרון.

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	$X \sim \text{Mult}(n, p)$	$p(x) = \frac{n!}{x_1! \dots x_k!} \prod_{i=1}^k p_i^{x_i}$ $x = (x_1, \dots, x_k)$	$0 \leq p_i; \sum_{i=1}^k p_i = 1; n \geq 1$ $p = (p_1, \dots, p_k)$	$0 \leq x_i; \sum_{i=1}^k x_i = n$		$E(X_i) = np_i$ $i = 1, \dots, k$	$\text{cov}(X_i, X_j) = -np_i p_j$ $(i \neq j)$	$X_i \sim \text{Bin}(n, p_i)$
Negative Binomial	$X \sim NB(n, p)$	$p(k) = \binom{k-1}{n-1} p^n q^{k-n}$	$0 \leq p = 1 - q \leq 1; n \geq 1$	$k = n, n+1, \dots$	$\left[\frac{pe^{qt}}{1-qe^{qt}} \right]^n$	$\frac{np}{p}$	$\frac{npq}{p^2}$	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	$X \sim P(\lambda)$	$p(k) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda \geq 0$	$k = 0, 1, \dots$	$e^{\lambda(e^t - 1)}$	λ	λ	
Hypergeometric	$X \sim H(n, M, n)$	$p(k) = \frac{\binom{N}{k} \binom{M-k}{n-k}}{\binom{N+M}{n}}$	$N, M > 0$ $1 \leq n \leq N+M$	$\max(0, n-M) \leq k \leq \min(n, N)$	$\sum_{k=\max(0, n-M)}^{\min(n, N)} e^{kt} p(k)$	$\frac{nM}{N+M}$	$n \cdot \frac{M}{N+M} \cdot \frac{N-M}{N+M-1}$ $= \frac{NM}{N+M-1}$	N - special elements M - ordinary ones n - sample size $N+M$ - total
		Dirichy 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	$X \sim N(\mu, \sigma^2)$	$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{(t-\mu)^2}{2\sigma^2} \right\}$	$-\infty < \mu < \infty; \sigma > 0$	$-\infty < t < \infty$	$\exp \left\{ t\mu + \frac{1}{2} t^2 \sigma^2 \right\}$	μ	σ^2	