בס"ד

מבוא להסתברות וסטטיסטיקה, 88165, פרופ' ג'. שיף

מועד א', סמסטר קיץ, תשע"ג - פתרונות

זמן המבחן: שלוש שעות.
מותר להשתמש בכל חומר עזר ובמחשב כיס.
יש לענות על 6 שאלות (ולא יותר), חשוב לנמק היטב כל תשובה.
כל השאלות שוות במשקל
יש להתחיל את התשובה לכל שאלה בדף נפרד.

1. האחוזים של מבוגרים עם השמנת יתר וההכנסה הממוצעת השנתית בדולרים ב-7 המדינות הגדולות בארה"ב הם כדלקמן:

מדינה	אחוז השמנת יתר	הכנסה ממוצעת
פנסילבניה	29.2	36421
פלורידה	26.6	39272
טקסס	31.0	39493
ניו יורק	28.6	41152
קליפורניה	24.0	43104
אילינוי	28.2	43159
אוהייו	23.9	48821

(כל הנתונים משנת 2010). מצא את קו הרגרסיה של אחוז השמנת היתר כפונקציה של ההכנסה הממוצעת. מצא גם אומדן למקדם המתאם של אחוז השמנת היתר וההכנסה הממוצעת של מדינות גדולות.

למדינות יש אוכלוסיות די שונות -

מדינה	אוכלוסייה	אחוז השמנת יתר	הכנסה ממוצעת
פנסילבניה	11536504	29.2	36421
פלורידה	18801310	26.6	39272
טקסס	25145561	31.0	39493
ניו יורק	12702379	28.6	41152
קליפורניה	37253956	24.0	43104
אילינוי	12830632	28.2	43159
אוהייו	19378102	23.9	48821

איך היית מציע לקחת בחשבון את האוכלוסיות השונות בחישובים שעשית ? האם יש עוד אינפורמציה רלוונטית שהיא חסרה ?

קו הרגרסיה: $y = ax + b$ כאשר a, b פותרים

$$\begin{pmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \sum 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

כאן x מסמן את ההכנסה הממוצעת - נמדוד באלפי דולרים - ו- y מסמן את אחוז השמנת היתר.

$$\begin{pmatrix} 12226 & 291 \\ 291 & 7 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 7928 \\ 192 \end{pmatrix}$$

מקבלים $a \approx -0.477$, $b \approx 47.2$ כלומר כל אלף דולר הכנסה נוספת מורידה את אחוז השמנת היתר ב- 0.48% !

מקדם המתאם:

$$\begin{aligned}
\rho &\approx \frac{\frac{1}{N}\sum x_i y_i - \left(\frac{1}{N}\sum x_i\right)\left(\frac{1}{N}\sum y_i\right)}{\sqrt{\left(\frac{1}{N}\sum x_i^2 - \left(\frac{1}{N}\sum x_i\right)^2\right)\left(\frac{1}{N}\sum y_i^2 - \left(\frac{1}{N}\sum y_i\right)^2\right)}} \\
&= \frac{\left(\sum 1\right)\left(\sum x_i y_i\right) - \left(\sum x_i\right)\left(\sum y_i\right)}{\sqrt{\left(\left(\sum 1\right)\left(\sum x_i^2\right) - \left(\sum x_i\right)^2\right)\left(\left(\sum 1\right)\left(\sum y_i^2\right) - \left(\sum y_i\right)^2\right)}} \\
&= \frac{7 \cdot 7928 - 291 \cdot 192}{\sqrt{\left(7 \cdot 12226 - 291^2\right)\left(7 \cdot 5282 - 192^2\right)}} \\
&= \frac{-313}{\sqrt{656 \cdot 192}} \\
&\approx -0.71
\end{aligned}$$

אם רוצים לקחת בחשבון את כמות האנשים בכל מדינה ומדינה יש לתת יותר משקל למספרים עם יותר אנשים - כלומר במקום הנוסחה למעלה משתמשים ב-

$$\begin{pmatrix} \sum p_i x_i^2 & \sum p_i x_i \\ \sum p_i x_i & \sum p_i \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum p_i x_i y_i \\ \sum p_i y_i \end{pmatrix}$$

אם עושים את זה מוצאים $a \approx -0.556, b \approx 50.2$. כנ"ל מוצאים $\rho \approx -0.70$

החישובים האלה עושים כאילו לכל אחד במדינה יש אותה הכנסה (ואותו סיכוי השמנת יתר). כמובן גם בתוך כל מדינה יש התפלגות והיינו רוצים לקחת את זה בחשבון. עדיף שיהיה לנו אנפורמציה לגבי כל יחיד - לגבי ההכנסה וה- BMI.

---

2. באי בודד יש אוכלוסייה של N דביבונים, מתוכם pN זכרים ו- $(1-p)N$ נקבות. בוחרים באקראי קבוצה של 4 דביבונים.

(א) מה היא ההסתברות לבחור בדיוק 2 זכרים ו-2 נקבות ?

(ב) מצא את הגבול של ההסתברות שמצאת בסעיף (א) כאשר $N \to \infty$.

(ג) איך ניתן למצוא את הגבול שמצאת בסעיף (ב) בלי לחשב קודם את ההסתברות במקרה ש-N הוא סופי ?

(ד) מה הוא הגבול כאשר $N \to \infty$ של תוחלת מספר הזכרים שנבחרו ? מה היא תוחלת מספר הזכרים כאשר N הוא סופי ?

(א) מספר הדרכים לבחור 4 מתוך ה- N:

$$\begin{pmatrix} N \\ 4 \end{pmatrix} = \frac{N(N-1)(N-2)(N-3)}{24}$$

מס' הדרכים לבחור 2 זכרים :

$$\begin{pmatrix} pN \\ 2 \end{pmatrix} = \frac{pN(pN-1)}{2}$$

מס' הדרכים לבחור 2 נקבות :

$$\binom{(1-p)N}{2} = \frac{(1-p)N((1-p)N-1)}{2}$$

לכן הסת' לבחור בדיוק שני זכרים ושני נקבות:

$$\frac{p(1-p)N^2(pN-1)((1-p)N-1)/4}{N(N-1)(N-2)(N-3)/24} = \frac{6p(1-p)N(pN-1)((1-p)N-1)}{(N-1)(N-2)(N-3)}$$

(ב) כאשר $N \to \infty$ זה שואף ל- $6p^2(1-p)^2$.

(ג) $6p^2(1-p)^2$ היא ההסתברות $\mathbf{P}(X=2)$ כאשר $X \sim B(4,p)$. בעצם כאשר N הוא גדול ובוחרים 4 הבחירות הן בלתי תלויות ויש הסת' p לבחור זכר בכל בחירה.

(ד) החלטנו שבגבול $N \to \infty$ מספר הזכרים בתוך ה-4 מתפלג $B(4,p)$. ולכן התוחלת היא $4p$. יוצא שגם כאשר N הוא סופי התוחלת של מספר הזכרים בתוך ה-4 היא $4p$ ללא תלות על N. אם X הוא מספר הזכרים, ניתן לכתוב $X = X_1 + X_2 + X_3 + X_4$ כאשר X_1 שווה 1 אם הראשון זכר ו-0 אחרת, X_2 שווה 1 אם השני זכר ו-0 אחרת וכו'. התוחלת של X היא סכום התוחלות $\mathbf{E}[X_1] + \mathbf{E}[X_2] + \mathbf{E}[X_3] + \mathbf{E}[X_4]$ ו- $\mathbf{E}[X_i] = p$ לכל i.

---

3. בכד יש 10 כדורים ממוספרים מ- 1 עד 10. מטילים קוביה הוגנת, ושולפים כדורים מהכד, עם החזרות, עד שמוציאים כדור עם מספר פחות או שווה למספר המופיע על הקוביה.

(א) מצא את ההסתברות להוציא 3 כדורים מן הכד.

(ב) בהנתן שהוצאו 3 כדורים, מצא את ההסתברות שהמספר על הקוביה הוא 4.

(ג) מה היא התוחלת של מספר הכדורים שמוציאים מן הכד?

הסוד כאן הוא שאם המספר על הקוביה הוא I, אזי מספר הכדורים שיש להוציא מתפלג גאומטרית עם $p = \frac{I}{10}$. יש לזכור שבהתפלגות הגאומטרית עם פרמטר p

$$\mathbf{P}(X=n) = p(1-p)^{n-1}$$

ו-

$$\mathbf{E}(X=n) = \frac{1}{p}$$

(א)

$$\begin{aligned}
\mathbf{P}(X=3) &= \sum_{r=1}^{6} \mathbf{P}(X=3, I=r) \\
&= \sum_{r=1}^{6} \mathbf{P}(X=3|I=r)\mathbf{P}(I=r) \\
&= \frac{1}{6}\sum_{r=1}^{6} \frac{r}{10}\left(1-\frac{r}{10}\right)^2 \\
&= \frac{1}{6}\left(\frac{81}{1000}+\frac{128}{1000}+\frac{147}{1000}+\frac{144}{1000}+\frac{125}{1000}+\frac{96}{1000}\right) \\
&= \frac{721}{6000}
\end{aligned}$$

(ב) על ידי חוק בייס

$$\begin{aligned}\mathbf{P}(I=4|X=3) &= \frac{\mathbf{P}(X=3|I=4)\mathbf{P}(I=4)}{\mathbf{P}(X=3)} \\ &= \frac{\frac{1}{6}\frac{4}{10}\left(1-\frac{4}{10}\right)^2}{\frac{721}{6000}} \\ &= \frac{144}{721}\end{aligned}$$

(ג)

$$\mathbf{E}[X] = \sum_{r=1}^{6}\mathbf{E}[X|I=r]\mathbf{P}(I=r) = \frac{1}{6}\sum_{r=1}^{6}\frac{10}{r} = \frac{10}{6}\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}\right) = \frac{147}{36}$$

---

4. מטילים N קוביות הוגנות. יהא X_1 מספר הפעמים שמקבלים תוצאה "1", X_2 מספר הפעמ-ים שמקבלים תוצאה "2" וכו'.

(א) כתוב את ההסתברות $\mathbf{P}(X_1=i,X_2=j)$.

(ב) על ידי שימוש בזהות

$$(p+q+r)^N = \sum_{i=0}^{N}\sum_{j=0}^{N-i}\frac{N!}{i!j!(N-i-j)!}p^iq^jr^{N-i-j}$$

הוכח ש- $\text{cov}(X_1,X_2) = -\frac{N}{36}$.

(ג) מה הוא $\text{Var}(X_1+X_2+X_3+X_4+X_5+X_6)$?

(ד) על ידי שימוש בתוצאה של הסעיף הקודם, והנוסחה לשונות של סכום של משתנים מקריים, הוכח שוב ש- $\text{cov}(X_i,X_j) = -\frac{N}{36}$.

(א)

$$\mathbf{P}(X_1=i,X_2=j) = \frac{N!}{i!j!(N-i-j)!}\left(\frac{1}{6}\right)^{i+j}\left(\frac{2}{3}\right)^{N-i-j}$$

כאן $i,j\geq 0$, $i+j\leq N$.

(ב) אם גוזרים את הזהות ביחס ל-p מקבלים

$$N(p+q+r)^{N-1} = \sum_{i=0}^{N}\sum_{j=0}^{N-i}\frac{N!}{i!j!(N-i-j)!}ip^{i-1}q^jr^{N-i-j}$$

אם גוזרים את הזהות ביחס ל-q מקבלים

$$N(p+q+r)^{N-1} = \sum_{i=0}^{N}\sum_{j=0}^{N-i}\frac{N!}{i!j!(N-i-j)!}p^ijq^{j-1}r^{N-i-j}$$

ואם גוזרים ביחס לשניהם בזה אחר זה (הסדר לא חשוב) מקבלים

$$N(N-1)(p+q+r)^{N-2}=\sum_{i=0}^{N}\sum_{j=0}^{N-i}\frac{N!}{i!j!(N-i-j)!}ijp^{i-1}q^{j-1}r^{N-i-j}$$

ולכן

$$\mathbf{E}[X]=\sum_{i,j}i\mathbf{P}(X=i,Y=j)=\frac{1}{6}N$$

$$\mathbf{E}[Y]=\sum_{i,j}j\mathbf{P}(X=i,Y=j)=\frac{1}{6}N$$

$$\mathbf{E}[XY]=\sum_{i,j}ij\mathbf{P}(X=i,Y=j)=\frac{1}{36}N(N-1)$$

ו-

$$\mathrm{cov}(X,Y)=\mathbf{E}[XY]-\mathbf{E}[X]\mathbf{E}[Y]=\frac{1}{36}N(N-1)-\frac{1}{36}N^2=-\frac{N}{36}$$

(ג) היות ו- $X_1+X_2+X_3+X_4+X_5+X_6=N$, $\mathrm{Var}(X_1+\ldots+X_6)=0$.

(ד) מהסעיף הקודם יש לנו

$$\begin{aligned}0 &= \mathrm{Var}(X_1+\ldots+X_6)\\ &= \sum_i \mathrm{Var}(X_i)+2\sum_{i,j}\mathrm{cov}(X_i,X_j)\\ &= 6\mathrm{Var}(X_1)+30\mathrm{cov}(X_1,X_2)\end{aligned}$$

בשורה האחרונה ניצלנו את העובדה שהמ"מים $X_1,\ldots,X_6$ הם בעלי אותה התפלגות. עכשיו $X_1\sim B(N,\frac{1}{6})$ ולכן $\mathrm{Var}(X_1)=\frac{5N}{36}$. ולכן

$$\mathrm{cov}(X_1,X_2)=-\frac{1}{5}\mathrm{Var}(X_1)=-\frac{N}{36}$$

---

5. פונקציית הצפיפות המשותפת של הזוג (X,Y) היא

$$f(x,y)=c(1+x+y)\ ,\qquad x\geq 0\ ,\ y\geq 0\ ,\ x+2y\leq 1$$

כאשר c הוא קבוע חיובי.

(א) מצא את הקבוע c.

(ב) מצא את ההסתברות ש- $X>Y$.

(ג) מצא את ההתפלגות השולית של X.

(ד) הסבר את השלבים בחישוב פונקציית הצפיפות של $\max(X,Y)$. בסעיף זה אין צורך לחשב אנטגרלים, אבל למעט זה יש לפרט עד כמה שאפשר את כל השלבים בחישוב.

(א) יש לדרוש

$$c\int_0^{\frac{1}{2}}\left(\int_0^{1-2y}(1+x+y)\,dx\right)\,dy=1$$

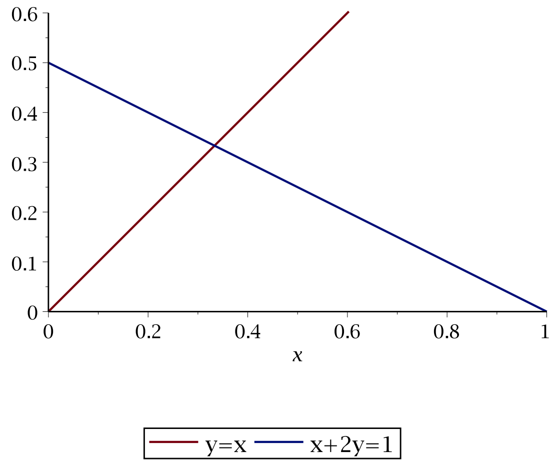
כלומר

$$c\int_0^{\frac{1}{2}}\left[(1+y)x+\frac{1}{2}x^2\right]_0^{1-2y}\,dy=1$$

$$c\int_0^{\frac{1}{2}}\frac{3}{2}(1-2y)\,dy=1$$

$$\frac{3}{2}c\left[y-y^2\right]_0^{\frac{1}{2}}=1 \qquad\Rightarrow\qquad c=\frac{8}{3}$$

(ב) יש לעשות אנטגרל על התחום שבו $x<y$. כלומר המשולש משמאל באיור:



חיתוך שני הישרים ב- $x=y=\frac{1}{3}$. ההסתברות:

$$\begin{aligned}
P &= \frac{8}{3}\int_0^{\frac{1}{3}}\left(\int_x^{\frac{1}{2}(1-x)}(1+x+y)\,dy\right)\,dx\\
&= \frac{8}{3}\int_0^{\frac{1}{3}}\left[(1+x)y+\frac{1}{2}y^2\right]_x^{\frac{1}{2}(1-x)}\,dx\\
&= \frac{8}{3}\int_0^{\frac{1}{3}}\frac{5}{8}-\frac{5}{4}x-\frac{15}{8}x^2\,dx\\
&= \frac{8}{3}\left(\frac{5}{8}\frac{1}{3}-\frac{5}{8}\frac{1}{3^2}-\frac{5}{8}\frac{1}{3^3}\right)\\
&= \frac{25}{81}
\end{aligned}$$

(ג)

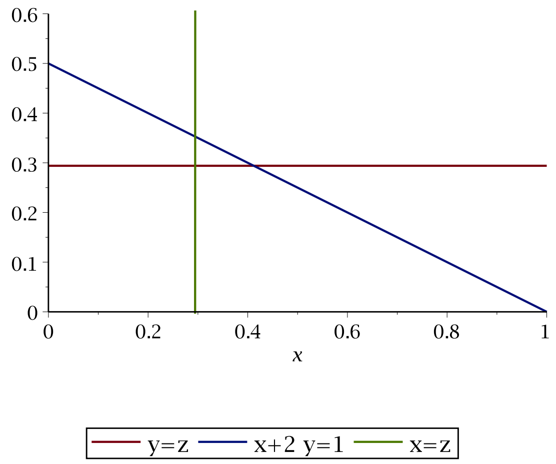
$$f_X(x) = \frac{8}{3}\int_0^{\frac{1}{2}(1-x)}(1+x+y)\,dy$$

$$
\begin{aligned}
&= \frac{8}{3}\left[(1+x)y+\frac{1}{2}y^2\right]_0^{\frac{1}{2}(1-x)} \\
&= \frac{1}{3}(1-x)(5+3x)
\end{aligned}
$$

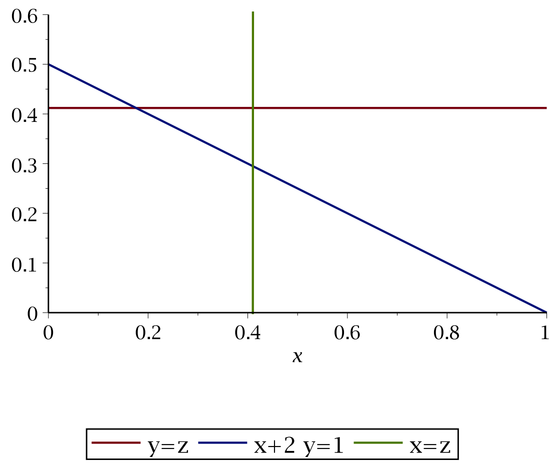
(ד) יש למצוא את פונקציית ההצטברות, כלומר

$$F(z) = \mathbf{P}(\max(X,Y) \le z) = \mathbf{P}(X,Y \le z)$$

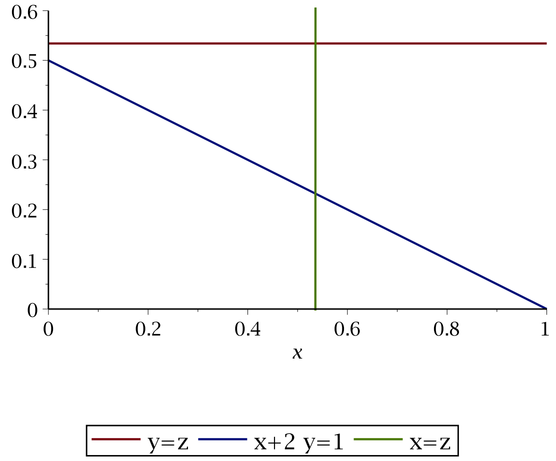
ברור שאם $z \le 0$, $F(z) = 0$. כמו כן אם $z \ge 1$, $F(z) = 1$. יש שלושה מקרים "ביניים" שיש צורך לחשב: אם $0 \le z \le \frac{1}{3}$ אזי צריך לעשות אנטגרל של הצפיפות על ריבוע:



אם $\frac{1}{3} \le z \le \frac{1}{2}$ אזי צריך לעשות אנטגרל של הצפיפות על תחום עם 5 צדים:



ואם $\frac{1}{2} \le z \le 1$ אזי צריך לעשות אנטגרל של הצפיפות על תחום טרפז:



והחישובים מסובכים.

---

6. מטילים קוביה הוגנת 100 פעמים. יהא X_i המספר שמקבלים בהטלה מספר i. מצא קירוב ל-

$$\mathbf{P}\left(10^{44} \le \prod_{i=1}^{100} X_i \le 10^{52}\right)$$

רמז:

$$\log\left(\prod_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} \log X_i$$

וניתן להשתמש במשפט הגבול המרכזי לחישוב הסתברויות לסכום של ממב"תים.
מצא גם את a ואת b כך ש-

$$\mathbf{P}\left(a \le \prod_{i=1}^{100} X_i\right) = \mathbf{P}\left(\prod_{i=1}^{100} X_i \le b\right) = 0.95$$

אם X היא התוצאת של הטלת קוביה

$$\mathbf{E}[\log X] = \frac{1}{6}(\log 1 + \log 2 + \log 3 + \log 4 + \log 5 + \log 6) = \frac{1}{6}\log 720 \approx 0.4762$$

$$\mathbf{E}[(\log X)^2] = \frac{1}{6}((\log 1)^2 + (\log 2)^2 + (\log 3)^2 + (\log 4)^2 + (\log 5)^2 + (\log 6)^2) \approx 0.2958$$

ולכן סטיית התקן של $\log X$ היא בערך 0.2627. השתמשתי כאן ב- $\log$ בבסיס 10.

מזה נובע שהתוחלת של $\sum_{i=1}^{100}\log X_i$ היא בערך 47.62 וסטיית התקן היא בערך 2.627. ועל ידי משפט הגבול המרכזי, ההתפלגות שלה היא בערך נורמלית. ולכן

$$\begin{aligned}\mathbf{P}\left(10^{44} \le \prod_{i=1}^{100} X_i \le 10^{52}\right) &= \mathbf{P}\left(44 \le \sum_{i=1}^{100}\log X_i \le 52\right)\\ &= \mathbf{P}\left(\frac{44-47.62}{2.627} \le \frac{\sum_{i=1}^{100}\log X_i - 47.62}{2.627} \le \frac{52-47.62}{2.627}\right)\end{aligned}$$

$$\begin{aligned} &\approx \Phi(1.67) - \Phi(-1.38) \\ &\approx 0.9525 - (1 - 0.9162) \\ &= 0.8687 \end{aligned}$$

כמו כן:

$$\begin{aligned} \mathbf{P}\left(a \le \prod_{i=1}^{100} X_i\right) &= \mathbf{P}\left(\log a \le \sum_{i=1}^{100} \log X_i\right) \\ &= \mathbf{P}\left(\frac{\log a - 47.62}{2.627} \le \frac{\sum_{i=1}^{100} \log X_i - 47.62}{2.627}\right) \\ &\approx \Phi\left(\frac{47.62 - \log a}{2.627}\right) \end{aligned}$$

זה שווה 0.95 **אם** $\frac{47.62 - \log a}{2.627} \approx 1.65$ **כלומר** $\log a \approx 43.3$. **ולכן** $a \approx 10^{43.3}$.

כמו כן:

$$\begin{aligned} \mathbf{P}\left(\prod_{i=1}^{100} X_i \le b\right) &= \mathbf{P}\left(\sum_{i=1}^{100} \log X_i \le \log b\right) \\ &= \mathbf{P}\left(\frac{\sum_{i=1}^{100} \log X_i - 47.62}{2.627} \le \frac{\log b - 47.62}{2.627}\right) \\ &\approx \Phi\left(\frac{\log b - 47.62}{2.627}\right) \end{aligned}$$

זה שווה 0.95 **אם** $\frac{\log b - 47.62}{2.627} \approx 1.65$ **כלומר** $\log b \approx 52.0$. **ולכן** $b \approx 10^{52.0}$.

---

7. (א) בבית יש שני קיני ג'וקים. בכל יום ג'וק מהקן הראשון עובר לקן השני בהסתברות $\frac{1}{2}$ וג'וק מהקן השני עובר לקן הראשון בהסתברות $\frac{1}{3}$. מצא את ההתפלגות הקבועה של הג'וקים בין שני הקינים.

(ב) מצא את ההסתברות שג'וק שהתחיל בקן הראשון יהיה בקן הראשון אחרי 3 ימים, ואת ההסתברות שג'וק שהתחיל בקן השני יהיה בקן הראשון אחרי 3 ימים.

(ג) בהנתן שבהתחלה יש 100 ג'וקים בקן הראשון ו- 200 בקן השני, מצא קירוב נורמלי להתפלגות של מספר הג'וקים בקן הראשון אחרי 3 ימים.

שרשרת מרקוב עם מטריצת מעבר

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix}$$

(א) יש פתור

$$\begin{pmatrix} s & 1-s \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} s & 1-s \end{pmatrix}$$

מוצאים $s = \frac{2}{5}$: כלומר 40% של הג'וקים בקן ראשון ו- 60% בקן השני.

(ב) מוצאים

$$P^2=\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{3}&\frac{2}{3}\end{pmatrix}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{3}&\frac{2}{3}\end{pmatrix}=\begin{pmatrix}\frac{5}{12}&\frac{7}{12}\\\frac{7}{18}&\frac{11}{18}\end{pmatrix}$$

$$P^3=\begin{pmatrix}\frac{5}{12}&\frac{7}{12}\\\frac{7}{18}&\frac{11}{18}\end{pmatrix}\begin{pmatrix}\frac{1}{2}&\frac{1}{2}\\\frac{1}{3}&\frac{2}{3}\end{pmatrix}=\begin{pmatrix}\frac{29}{72}&\frac{43}{72}\\\frac{43}{108}&\frac{65}{108}\end{pmatrix}$$

ולכן אחרי 3 ימים ג'וק שהתחיל בקן ראשון נמצא בקן ראשון בהסתברות $\frac{29}{72}$ וג'וק שהתחיל בקן שני נמצא בקן ראשון בהסתברות $\frac{43}{108}$

(ג) ולכן - אם מתחילים עם 100 ג'וקים בקן הראשון - אחרי 3 ימים המספר מתוכם הנמצא בקן הראשון מתפלג $B(100,\frac{29}{72})\approx N(\frac{29\cdot100}{72},\sqrt{\frac{29\cdot43\cdot100}{72^2}})$ וכמו כן אם מתחילים עם 200 ג'וקים בקן השני המספר מתוכם הנמצא בקן הראשון אחרי 3 ימים מתפלג $B(200,\frac{43}{108})\approx N(\frac{43\cdot200}{108},\sqrt{\frac{43\cdot65\cdot200}{108^2}})$. ואין תלות ביניהם - לכן סה"כ מספר הג'וקים בקן הראשון אחרי 3 ימים מתפלג בערך

$$N\left(\frac{29\cdot100}{72}+\frac{43\cdot200}{108},\sqrt{\frac{29\cdot43\cdot100}{72^2}+\frac{43\cdot65\cdot200}{108^2}}\right)\approx N(119.9,8.5)$$

---

8. מספרי התרופות החדשות שקבלו אישור ממנהל התרופות בארה"ב (ה- FDA) הם כדלקמן:

שנה	מספר
2001	24
2002	17
2003	21
2004	36
2005	20
2006	22
2007	18
2008	24
2009	26
2010	21
2011	37
2012	30

(א) מצא אומדן לקצב הממוצע השנתי λ של אישור התרופות.

(ב) בדוק את ההשערה שמספר התרופות שאושרו מתפלג פאוסון עם פרמטר λ . יש לעבוד ברמת מובהקות 95%

אחוזוני 95% להתפלגות χ^2_n:

n	אחוזון 95
9	16.9
10	18.3
11	19.7
12	21.0
13	22.4

(ג) ב-8 חודשים הראשונים של שנת 2013 אושרו 16 תרופות חדשות. איך ניתן לקחת את זה בחשבון ?

(א) הממוצע: $24\frac{2}{3}$.

(ב) אם אכן מספר האישורים מתלפג $\mathcal{P}(24\frac{2}{3})$ אזי מצפים ל- $24\frac{2}{3}$ אישורים כל שנה. מוצאים את

$$\chi^2 = \sum_i \frac{(O_i - 24\frac{2}{3})^2}{24\frac{2}{3}} \approx 19.08$$

זה מעל 18.3 . לכן ניתן לדחות את ההשערה של התפלגות פאוסון עם קצב קבוע ברמת מובהקות 95%. יש כאן 10 דרגות חופש כי השתמשנו בממוצע של הנתונים.

(ג) הנתון החדש מתייחס ל-8 חודשים, ולפי ההשערה בתקופה זו מצפים ל- $16\frac{4}{9}$ אישורים. לכן יש להוסיף ל- χ^2 את המספר

$$\frac{(16 - 16\frac{4}{9})^2}{16\frac{4}{9}} \approx 0.012$$

זה נותן $\chi^2 \approx 19.09$, אבל עכשיו יש 11 דרגות חופש - ולכן לא ניתן לדחות ברמת מובהקות 95%. אפשר גם לקחת את הנתון החדש בחשבון בחישוב הממוצע - באים לאותה מסקנה.
