PROBABILITY AND STATISTICS

הסתברות וסטטיסטיקה

מאת יוג'ין קנציפר

■ דף פתרונות לשאלות בית 5

התפלגויות בדידות מיוחדות

■

שאלה H5.1

א. ישנם 19 מספרים כאלה, אזי ההסתברות הדרושה היא $\frac{14}{100}=0.14$.

ב. ישנם 19 מספרים כאלה: $07,17,27,37,47,57,67,77,87,97,70,71,72,73,74,75,76,78,79$.

אזי התשובה היא $\frac{19}{100}=0.19$.

שאלה H5.2

זמן המתנת האוטובוס מקבל את הערכים $0,1,2,3,4,5,6,7,8,9,10,11,12,13,14$. אזי, ההסתברות הדרושה היא

$$P(X>5)=P(X=6)+P(X=7)+P(X=8)+...+P(X=13)+P(X=14)=\frac{9}{15}=0.6$$

שאלה H5.3

הגדרות הניסוי והמשתנה	סוג ופרמטרים של התפלגות	$E[X]$
ניסוי: בחירה מקרית <u>עם החזרה</u> של שתי ספרות מאוסף {1, 2, 3, 4, 5, 6, 7, 8, 9}. **משתנה:** X = {מספר הספרות המתחלקות ב-3 בין שתיים שנבחרו}	$\text{Bin}\left(2,\frac{1}{3}\right)$	$\frac{2}{3}$
ניסוי: בחירה מקרית של סיפרה אחת מאוסף {1, 2, 3, 4, 5, 6, 7, 8, 9}. **משתנה:** X = {סיפרה שנבחרת אקראית}	$\text{U}_d(1,9)$	5
ניסוי: בחירה מקרית <u>וללא החזרה</u> של שלוש ספרות מאוסף {1, 2, 3, 4, 5, 6, 7, 8, 9}. **משתנה:** X = {מספר ספרות אי זוגיות בין שלוש שנבחרו}	$\text{Hyp}(9,5,3)$	$\frac{5}{3}$
ניסוי: בחירה מקרית <u>עם החזרה</u> של סיפרה מהסדרה {1, 2, 3, 4, 5, 6, 7, 8, 9} **משתנה:** X = {מספר הבחירות עד קבלת סיפרה "5" בפעם השנייה}	$\text{NegBin}\left(\frac{1}{9},2\right)$	18

שאלה H5.4

א. צריך לחשב את ההסתברות שנגיע לכדור האדום בפעם הראשונה בהוצאה הרבעית. במילים אחרות, מדובר על ביצוע סדרת ניסוי ברנולי (הוצאה עם החזרה!) עד ההצלחה הראשונה. נגדיר מישתנה מקרי X = {מספר הוצאות עד ההצלחה הראשונה – גילוי כדור אדום}. מתקיים: $X \sim \mathrm{G}(p)$ עם הפרמטר $p = \frac{2}{10} = \frac{1}{5}$. כתוצאה,

$$P(X=4) = \frac{1}{5}\cdot\left(1-\frac{1}{5}\right)^{4-1} = \frac{64}{625}.$$

ב. $P(X>3) = \left(1-\frac{1}{5}\right)^{3} = \frac{64}{125}$.

שאלה H5.5

נצטרך להטיל את הקובייה יותר מ–5 פעמים אם ב–5 ההטלות הראשונות לא יופיע מספר 6. אזי, מדובר על ביצוע סדרת ניסוי ברנולי עד ההצלחה הראשונה (התפלגות גיאומטרית בעל פרמטר $p = \frac{1}{6}$). כתוצאה,

$$P(X>5) = \left(1-\frac{1}{6}\right)^{5} = \left(\frac{5}{6}\right)^{5}.$$

שאלה H5.6

ניתן לראות כי הדפסת סימן היא ניסוי ברנולי ("הצלחה" – טעות בהדפסה). אזי, מספר הטעויות X בעמוד הדפסה אחד הוא משתנה בינומי, $X \sim \mathrm{Bin}(n=500, p=0.01)$. את ההסתברות המבוקשת אפשר לחשב בקירוב בעזרת התפלגות פואסון מכיוון ש– $n=500$ הוא מספר גדול מאוד ו–$p=0.01$ הוא מספר קטן מאוד אך $\lambda = 500\cdot 0.01 = 5$ הוא מספר בינוני. בדרך זו, מגיעים לתשובה:

$$P(X\le 1) = P(X=0)+P(X=1) \approx e^{-5} + 5e^{-5} \approx 0.04.$$

שאלה H5.7

אם X הוא מספר הנהגים ללא רישיון באוכלוסיה המונה 100 איש, מיד מגיעים למסקנה כי $X \sim \mathrm{Bin}\left(n=100, p=\frac{5}{1000}=\frac{1}{200}\right)$. כאן, $n=100$ הוא מספר גדול, $p=0.005$ הוא מספר קטן אך $\lambda = n\cdot p = \frac{1}{2}$ הוא מספר בינוני. אזי, מותר להשתמש בקירוב פואסון כדי להגיע לתשובה:

$$P(X\ge 1) = 1-P(X=0) \approx 1-e^{-1/2} \approx 0.39.$$

שאלה H5.8

מספר כולל של הצלחות במערכת של 3 ניסוי ברנולי מתפלג בינומית, $X \sim \mathrm{Bin}(3,p)$ כך ש–

$$P(X=k) = C_3^k\, p^k (1-p)^{3-k} \quad \text{עבור} \quad k=0,\dots,3.$$

נתון כי $P(X=2)=12\cdot P(X=3)$. מזה נובע כי $3p^2(1-p)=12p^3$. אזי, $p=\frac{1}{5}$.

שאלה H5.9

נסמן ב— X את רווח השחקן בהטלת קובייה בודדת. פונקציית הסתברות של X ניתנת על ידי הטבלה:

x	-5	-4	-3	-2	-1	$+15$
$P_X(x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

א. תוחלת הרווח היא

$$E[X]=\sum_x x\cdot P_X(x)=\frac{1}{6}\left(-5-4-3-2-1+15\right)=0.$$

שונות הרווח היא

$$\mathrm{var}[X]=E[X^2]-\left(E[X]\right)^2=E[X^2]=\sum_x x^2\cdot P_X(x)=\frac{1}{6}\left(25+16+9+4+1+225\right)\approx 46.7$$

ב. במשחק בודד השחקן יקבל כסף מהקופה בהסתברות $1/6$. מספר כולל של זכיות Y בסדרה של 10 הטלות מתפלג בינומית $Y\sim \mathrm{Bin}\left(10,\frac{1}{6}\right)$ כך ש—

$$P(Y=k)=C_{10}^{k}\left(\frac{1}{6}\right)^{k}\left(\frac{5}{6}\right)^{10-k} \quad \text{עבור} \quad k=0,\ldots,10.$$

אזי,

$$P(Y=3)=C_{10}^{3}\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{7}\approx 0.155.$$

שאלה H5.10

בהגרלה 100 כרטיסים. נסמן ב— m את מספר הפרסים. מספר הכרטיסים ללא פרס הוא $100-m$. ההסתברות להשיג לפחות פרס אחת שווה לאחת פחות ההסתברות לא לזכות באף פרס. אזי,

$$1-\frac{100-m}{100}\cdot\frac{99-m}{99}=\frac{2}{100}.$$

מכאן נובע כי מספר הפרסים $m=1$.

שאלה H5.11

א. $\frac{6\cdot 5\cdot 4}{6^3}$; ב. $\frac{6}{6^3}$; ג. $1-\frac{5^3}{6^3}$.

שאלה H5.12

יהיה X — מספר הכדורים השחורים במדגם של n כדורים. יהיה Y — מספר הכדורים האדומים באותו המדגם. מתקיים:

$$P(X=k)=\frac{C_6^k C_4^{n-k}}{C_{10}^n} \text{ ו— } P(Y=k)=\frac{C_4^k C_6^{n-k}}{C_{10}^n} \text{ עבור } k=0,\dots,n.$$

נתון כי $P(X=n)=5\cdot P(Y=n)$. מזה נובע כי $C_6^n=5\cdot C_4^n$ כך ש—

$$(6-n)(5-n)=6.$$

למשוואה זו ישנם שני פתרונות: $n=3$ ו— $n=8$. בגלל ההגבלה $n\le 4$ אנחנו מגיעים לתשובה

$$n=3.$$

שאלה H5.13

צריך לחשב הסתברות מותנית $P(\text{"קובייה הוגנת"}/\text{"5"})$. שימוש במשפט בייס מביא:

$$P(\text{"קובייה הוגנת"}/\text{"5"})=\frac{P(\text{"5"}/\text{"קובייה הוגנת"})P(\text{"קובייה הוגנת"})}{P(\text{"5"})}.$$

כדי לחשב את ההסתברות $P(\text{"5"})$, נישתמש בנוסחא להסתברות שלמה אשר מביאה:

$$P(\text{"5"})=P(\text{"5"}/\text{"קובייה הוגנת"})P(\text{"קובייה הוגנת"})$$
$$+P(\text{"5"}/\text{"קובייה מוטה"})P(\text{"קובייה מוטה"})$$

כיוון ש— $P(\text{"5"})=\frac{1}{6}\cdot\frac{1}{2}+\frac{5}{36}\cdot\frac{1}{2}=\frac{11}{72}$, מקבלים:

$$P(\text{"קובייה הוגנת"}/\text{"5"})=\frac{\frac{1}{6}\cdot\frac{1}{2}}{\frac{11}{72}}=\frac{6}{11}.$$

שאלה H5.14

ניתן לראות כי $X_n=(-1)^X$ כאשר X הוא מספר החלפות סימן בדרך מ— X_0 ל— X_n. מספר זה הוא משתנה מקרי בעל התפלגות בינומית $X\sim\mathrm{Bin}(n,1-p)$ כך ש—

$$P(X=k)=C_k^n(1-p)^k p^{n-k}.$$

ברור כי הערכים האפשריים של X_n הם $+1$ ו— -1. כתוצאה, יש לחשב את ההסתברויות $P(X_n=+1)$ ו— $P(X_n=-1)$.

חישוב:

$$P(X_n=+1)=P(X=\text{זוגי})=P(X=0)+P(X=2)+P(X=4)+\dots$$

נעשה "טריק" ונחשב $E[X_n]$ (ראה/י שאלת כיתה C5.11):

$$E[X_n] = E\left[(-1)^X\right] = \sum_x (-1)^x P_X(x) = P(X = \text{זוגי}) - P(X = \text{אי זוגי})$$

מצד אחד, בכיתה קיבלנו כי

$$E[X_n] = E\left[(-1)^X\right] = \sum_{k=0}^{n} C_n^k (p-1)^k p^{n-k} = (2p-1)^n$$

כך ש—

$$P(X = \text{זוגי}) - P(X = \text{אי זוגי}) = (2p-1)^n.$$

מצד שני,

$$P(X = \text{זוגי}) + P(X = \text{אי זוגי}) = 1.$$

שתי משוואות אלו מביאות:

$$P(X_n = +1) = P(X = \text{זוגי}) = \frac{1+(2p-1)^n}{2},$$

$$P(X_n = -1) = 1 - P(X_n = +1) = \frac{1-(2p-1)^n}{2}.$$

שתי נוסחאות אלו מהוות את פונקציית ההסתברות של משתנה מקרי X_n.

שאלה H5.15

א. אם נסמן ב— X את מספר עוגות הגבינה אשר זומנו על ידי 5 סועדים, יש לחשב את ההסתברות $P(X \le 1)$. מכיוון ש— $X \sim \text{Hyp}(16,6,5)$, ההסתברות $P(X = k)$ היא

$$P(X = k) = \frac{C_6^k C_{10}^{5-k}}{C_{16}^5}.$$

אזי,

$$P(X \le 1) = P(X = 0) + P(X = 1) = \frac{C_{10}^5}{C_{16}^5} + \frac{C_6^1 C_{10}^4}{C_{16}^5}.$$

ב. יהיה R רווח של בית קפה מעוגות של יום אתמול. מנתוני השאלה עולה כי

$$R = 1.5 \cdot X + 2 \cdot (5 - X) = 10 - \frac{X}{2}.$$

התוחלת הרווח היא

$$E[R] = 10 - \frac{1}{2} E[X].$$

שונות הרווח היא

$$\text{var}[R] = \frac{1}{4} \text{var}[X].$$

מכיוון ש— $X \sim \text{Hyp}(16,6,5)$, אנחנו יכולים לחשב את התוחלתו והשונותו בקלות:

$$E[X] = n\frac{D}{N} = 5 \cdot \frac{6}{16} = \frac{15}{8} \approx 1.88,$$

$$\operatorname{var}[X] = n\frac{D}{N}\left(1-\frac{D}{N}\right)\left(1-\frac{n-1}{N-1}\right) = 5\cdot\frac{6}{16}\cdot\left(1-\frac{6}{16}\right)\cdot\left(1-\frac{4}{15}\right) = \frac{55}{64} \approx 0.86 .$$

נתונים אלה מביאים:

$$E[R] \approx 9.06 ,\ \operatorname{var}[R] \approx 0.215 .$$

שאלה H5.16

נתון כי $X \sim P(\lambda)$ כך ש—

$$P(X=k) = \frac{\lambda^k}{k!}e^{-\lambda} \qquad \text{עבור} \qquad k = 0,1,\ldots$$

א.

$$E[X!] = \sum_x x!\,P_X(x) = \sum_{k=0}^{\infty} k!\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}\lambda^k = \frac{e^{-\lambda}}{1-\lambda} .$$

התשובה קיימת עבור $0 \le \lambda < 1$ בלבד.

ב.

$$E[(X+1)!] = \sum_x (x+1)!\,P_X(x) = \sum_{k=0}^{\infty}(k+1)!\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}\sum_{k=0}^{\infty}(k+1)\lambda^k$$
$$= e^{-\lambda}\frac{\partial}{\partial\lambda}\sum_{k=0}^{\infty}\lambda^{k+1} = e^{-\lambda}\frac{\partial}{\partial\lambda}\left(\frac{\lambda}{1-\lambda}\right) = \frac{e^{-\lambda}}{(1-\lambda)^2} .$$

התשובה קיימת עבור $0 \le \lambda < 1$ בלבד.

שאלה H5.17

התפלגות היפרגיאומטרית שלילית מוגדרת על ידי פונקציית הסתברות

$$P(X_m = k) = C_{k-1}^{m-1}\frac{C_{N-k}^{D-m}}{C_N^D} \qquad \text{עבור} \qquad k = m, m+1, \ldots, N-D+m .$$

נתבונן בפונקציית הסתברות זה עבור m סופי ופרמטרים D ו— N גדולים מאוד ($D>>1$ ו— $N>>1$) כאלה ש— $p=\frac{D}{N}$ מקבל ערך קבוע.

- אין לפשט C_{k-1}^{m-1}.
- באמצעות נוסחת סטירלינג, מקבלים:

$$C_N^D = \frac{N!}{D!(N-D)!} \approx \frac{\sqrt{2\pi}N^{N+\frac{1}{2}}e^{-N}}{\sqrt{2\pi}D^{D+\frac{1}{2}}e^{-D}\sqrt{2\pi}N^{N-D+\frac{1}{2}}(1-p)^{N-D+\frac{1}{2}}e^{-N+D}}$$
$$= \frac{N^D}{\sqrt{2\pi}D^{D+\frac{1}{2}}(1-p)^{N-D+\frac{1}{2}}}$$

כאן, $p=\dfrac{D}{N}$. במהלך החישוב, השתמשנו בקירובים:

$$N!\approx\sqrt{2\pi}\,N^{N+\frac{1}{2}}e^{-N},\quad D!\approx\sqrt{2\pi}D^{D+\frac{1}{2}}e^{-D},$$

$$(N-D)!\approx\sqrt{2\pi}(N-D)^{N-D+\frac{1}{2}}e^{-N+D}=\sqrt{2\pi}N^{N-D+\frac{1}{2}}(1-p)^{N-D+\frac{1}{2}}e^{-N+D}.$$

- באמצעות אותה הנוסחת סטירלינג, נפשט את הגורם האחרון:

$$C_{N-k}^{D-m}=\frac{(N-k)!}{(D-m)!(N-D+m-k)!}.$$

נוסחת סטירלינג מביאה:

$$(D-m)!\approx\sqrt{2\pi}(D-m)^{D-m+\frac{1}{2}}e^{-D+m}\approx\sqrt{2\pi}D^{D-m+\frac{1}{2}}\left(1-\frac{m}{D}\right)^{D-m+\frac{1}{2}}e^{-D+m}$$
$$\approx\sqrt{2\pi}D^{D-m+\frac{1}{2}}e^{-D},$$

$$(N-k)!\approx\sqrt{2\pi}(N-k)^{N-k+\frac{1}{2}}e^{-N+k}=\sqrt{2\pi}N^{N-k+\frac{1}{2}}\left(1-\frac{k}{N}\right)^{N-k+\frac{1}{2}}e^{-N+k}$$
$$\approx\sqrt{2\pi}N^{N-k+\frac{1}{2}}e^{-N},$$

$$(N-D+m-k)!\approx\sqrt{2\pi}\left(N-D+m-k\right)^{N-D+m-k+\frac{1}{2}}e^{-N+D-m+k}$$
$$=\sqrt{2\pi}(N-D)^{N-D+m-k+\frac{1}{2}}\left(1+\frac{m-k}{N-D}\right)^{N-D+m-k+\frac{1}{2}}e^{-N+D-m+k}$$
$$\approx\sqrt{2\pi}(N-D)^{N-D+m-k+\frac{1}{2}}e^{-N+D}=\sqrt{2\pi}N^{N-D+m-k+\frac{1}{2}}(1-p)^{N-D+m-k+\frac{1}{2}}e^{-N+D}.$$

שילוב הנוסחאות מביא את הקירוב הבא:

$$C_{N-k}^{D-m}\approx\frac{\sqrt{2\pi}N^{N-k+\frac{1}{2}}e^{-N}}{\sqrt{2\pi}D^{D-m+\frac{1}{2}}e^{-D}\sqrt{2\pi}N^{N-D+m-k+\frac{1}{2}}(1-p)^{N-D+m-k+\frac{1}{2}}e^{-N+D}}$$
$$=\frac{N^{D-m}}{\sqrt{2\pi}D^{D-m+\frac{1}{2}}(1-p)^{N-D+m-k+\frac{1}{2}}}$$

חישוב זה מאפשר לפתח קירוב עבור היחס $\dfrac{C_{N-k}^{D-m}}{C_N^D}$. מקבלים:

$$\frac{C_{N-k}^{D-m}}{C_N^D} \approx \frac{N^{D-m}}{\sqrt{2\pi}D^{D-m+\frac{1}{2}}(1-p)^{N-D+m-k+\frac{1}{2}}} \frac{\sqrt{2\pi}D^{D+\frac{1}{2}}(1-p)^{N-D+\frac{1}{2}}}{N^D}$$
$$= p^m(1-p)^{k-m}.$$

כתוצאה, אנחנו מגיעים לנוסחא מקורבת:

$$P(X_m = k) = C_{k-1}^{m-1}\frac{C_{N-k}^{D-m}}{C_N^D} \approx C_{k-1}^{m-1}p^m(1-p)^{k-m}.$$

סוף החישוב.

שאלה H5.18

יש להוכיח:

$$\frac{C_D^{m-1}C_{N-D}^{k-m}}{C_N^{k-1}}\frac{D-(m-1)}{N-(k-1)} = C_{k-1}^{m-1}\frac{C_{N-k}^{D-m}}{C_N^D}.$$

חישוב:

$$\frac{C_D^{m-1}C_{N-D}^{k-m}}{C_N^{k-1}}\frac{D-(m-1)}{N-(k-1)}$$
$$= \frac{\frac{D!}{(m-1)!(D-m+1)!}\cdot\frac{(N-D)!}{(N-D+m-k)!(k-m)!}}{\frac{N!}{(N-k+1)!(k-1)!}}\cdot\frac{D-(m-1)}{N-(k-1)}$$
$$= \underbrace{\frac{D!(N-D)!}{N!}}_{1/C_N^D}\cdot\underbrace{\frac{(k-1)!}{(m-1)!(k-m)!}}_{C_{k-1}^{m-1}}\cdot\underbrace{\frac{(N-k+1)!}{(N-k+1)}}_{(N-k)!}\cdot\underbrace{\frac{(D-m+1)}{(D-m+1)!}}_{1/(D-m)!}\cdot\frac{1}{(N-D+m-k)!}$$
$$= C_{k-1}^{m-1}\cdot\frac{1}{C_N^D}\cdot\underbrace{\frac{(N-k)!}{(N-D+m-k)!(D-m)!}}_{1/C_{N-k}^{D-m}} = C_{k-1}^{m-1}\frac{C_{N-k}^{D-m}}{C_N^D}.$$

סוף הוכחה.

שאלה H5.19

- מצב הכד לפני הניסוי: D כדורים אדומים ו–$(N-D)$ כדורים שחורים.
- מוציאים באקראי וללא החזרה n כדורים. בתוכם ישנם מספר לא ידוע של כדורים אדומים –אותו נסמן ב– X. מתנאי השאלה עולה כי $X \sim \mathrm{Hyp}(N,D,n)$. כתוצאה, ההסתברות למצוא בדיוק k כדורים אדומים בין n הכדורים שהוצאו היא

$$P(X = k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}.$$

- מצב הכד לאחר הוצאת n כדורים: $D-k$ כדורים אדומים ו—$(N-D-(n-k))$ כדורים שחורים.

- לאחר החלפת הצבעים, עומדים לרשותנו $n-k$ כדורים אדומים ו—k כדורים שחורים. אותם אנחנו מחזירים לכד אשר מכיל $D-k$ כדורים אדומים ו—$(N-D-(n-k))$ כדורים שחורים.

- לאחר שהחזרנו n כדורים אלה חזרה לכד, הוא מכיל:
  - $D-k+n-k=D+n-2k$ כדורים אדומים
  - $N-D-(n-k)+k=N-D-n+2k$ כדורים שחורים.
  - סך הכל, מספר הכדורים בכד הוא $D+n-2k+N-D-n+2k=N$ כהלכה.

- בשלב הזה מוצאים באקראי כדור אחד. הוא יהיה כדור אדום בהסתברות
$$P(R/k)=\frac{D+n-2k}{N}.$$

זוהי **הסתברות מותנית**: בהחלט, ביטוי זה נכון **אם ידוע** כי בין n הכדורים שהוצאנו מהכד ישנם בדיוק k כדורים אדומים.

אזי, על פי נוסחא להסתברות שלמה, ההסתברות להוציא כדור אדום בשלב האחרון של הניסוי היא

$$P_R=\sum_{k=0}^{n}P(R/k)\cdot P(X=k)=\sum_{k=0}^{n}\frac{D+n-2k}{N}\cdot P(X=k)$$
$$=\frac{D+n}{N}\sum_{k=0}^{n}P(X=k)-\frac{2}{N}\sum_{k=0}^{n}k\cdot P(X=k).$$

את שני הסכומים אפשר לחשב בקלות כי

- הסכום הראשון מהווה נרמול של פונקצית הסתברות:
$$\sum_{k=0}^{n}P(X=k)=1$$

- הסכום השני מהווה תוחלת של משתנה היפרגיאומטרי $X\sim \mathrm{Hyp}(N,D,n)$. כתוצאה,
$$\sum_{k=0}^{n}k\cdot P(X=k)=E[X]=n\frac{D}{N}.$$

שילוב הנוסחאות מביא את התשובה להסתברות להוציא כדור אדום:

$$P_R=\frac{D+n}{N}-2n\frac{D}{N^2}=\frac{ND+n(N-2D)}{N^2}.$$

בשלב האחרון, כדאי לבדוק שהתשובה עבור P_R נמצאת בגבולות הנכונות. זאת אומרת, יש לבדוק כי $0\le P_R\le 1$. לשם כך, צריך לפתור את אי השוויון

$$0\le\frac{ND+n(N-2D)}{N^2}\le 1.$$

(עבודה עצמית).

שאלה H5.20

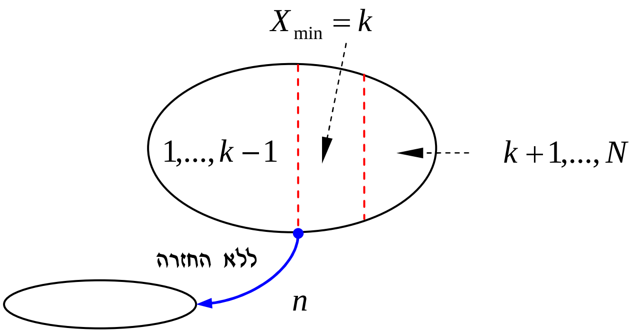
א. $\frac{C_{15}^{5}}{C_{30}^{5}}$.

ב. $\frac{C_{14}^{4}}{C_{30}^{5}}=5\frac{C_{1}^{1}}{C_{30}^{1}}\frac{C_{14}^{4}}{C_{29}^{4}}$.

שאלה H5.21

א. ברור כי הערכים האפשריים של $X_{\min}$ נמצאים בתווך $1\le x_{\min}\le N-n+1$.

ב. כדי למצוא את התוחלת $E[X_{\min}]$ עלינו לדעת את פונקצית ההסתברות של $X_{\min}$. אותה אפשר לחשב באמצעות גישה קלאסית להסתברות:



$P(X_{\min}=k)=\frac{|X_{\min}=k|}{|\Omega|}$.

כאן, גודל של מרחב המדגם הוא $|\Omega|=C_N^n$.

גודל המאורע $|X_{\min}=k|$ הוא מספר האופציות להגיע ל— $X_{\min}=k$:

$$|X_{\min}=k|=C_{k-1}^{0}C_{1}^{1}C_{N-k}^{n-1}=C_{N-k}^{n-1}.$$

כתוצאה, $P(X_{\min}=k)=\frac{C_{N-k}^{n-1}}{C_N^n}$.

<u>חישוב התוחלת:</u>

$$E[X_{\min}]=\sum_{x_{\min}}x_{\min}P_X(x_{\min})=\sum_{k=1}^{N-n+1}kP_X(X_{\min}=k)=\frac{1}{C_N^n}\sum_{k=1}^{N-n+1}kC_{N-k}^{n-1}.$$

נוח להריץ את הסכום על $j=N+1-k$ כך ש—

$$E[X_{\min}]=\frac{1}{C_N^n}\sum_{k=1}^{N-n+1}kC_{N-k}^{n-1}=\frac{1}{C_N^n}\sum_{j=n}^{N}(N-j+1)C_{j-1}^{n-1}=\frac{1}{C_N^n}\left[(N+1)\sum_{j=n}^{N}C_{j-1}^{n-1}-\sum_{j=n}^{N}jC_{j-1}^{n-1}\right]$$

- על מנת לחשב את הסכום $S(n,N)=\sum_{j=n}^{N}jC_{j-1}^{n-1}$, נשתמש בנוסחא שאותה הוכחנו בשאלה H3.1–(ב):

$$jC_{j-1}^{n-1}=nC_j^n$$

כדי לרשום: $S(n,N)=\sum_{j=n}^{N}jC_{j-1}^{n-1}=n\sum_{j=n}^{N}C_j^n$. עכשיו, הנוסחא משאלה H3.3 מביאה:

$$.S(n,N) = n\underbrace{\sum_{j=n}^{N} C_j^n}_{C_{N+1}^{n+1}} = nC_{N+1}^{n+1}$$

כתוצאה,

$$.S(n,N) = \sum_{j=n}^{N} jC_{j-1}^{n-1} = nC_{N+1}^{n+1}$$

- על מנת לחשב את הסכום השני $T(n,N) = \sum_{j=n}^{N} C_{j-1}^{n-1}$ נכתוב $T(n,N) = \sum_{k=n-1}^{N-1} C_k^{n-1}$ ונשתמש בנוסחא משאלה H3.3 כדי להגיע ל—

$$.T(n,N) = \sum_{k=n-1}^{N-1} C_k^{n-1} = C_N^n$$

כתוצאה,

$$E[X_{\min}] = \frac{1}{C_N^n}\left[(N+1)\underbrace{\sum_{j=n}^{N} C_{j-1}^{n-1}}_{T(n,N)=C_N^n} - \underbrace{\sum_{j=n}^{N} jC_{j-1}^{n-1}}_{S(n,N)=nC_{N+1}^{n+1}}\right] = \frac{1}{C_N^n}\left[(N+1)C_N^n - nC_{N+1}^{n+1}\right]$$

$$= (N+1) - n\frac{C_{N+1}^{n+1}}{C_N^n} = (N+1) - n\frac{(N+1)!\,n!\,(N-n)!}{(n+1)!\,(N-n)!\,N!}$$

$$= (N+1) - n\frac{N+1}{n+1} = \frac{N+1}{n+1}.$$