מבוא להסתברות: סיכום ודף נוסחאות

גרסה 1.1, נובמבר 2010

ברק שושני
baraksh@gmail.com | http://baraksh.co.il/

1 קומבינטוריקה

1.1 עקרון הספירה

אם בשלב הראשון של ניסוי כלשהו יש n תוצאות אפשריות, ולכל אחת מתוצאות אלה יש m תוצאות אפשריות בשלב השני, אז יש בסה"כ nm תוצאות אפשריות לניסוי.

1.2 סידור n עצמים בשורה: פרמוטציות

$$n! = n \cdot (n-1) \cdots 3 \cdot 2 \cdot 1$$

יש $n!$ דרכים שונות לסדר n עצמים בשורה: n אפשרויות לעצם הראשון, $n-1$ אפשרויות לשני וכך הלאה עד שלעצם האחרון נשארת רק אפשרות אחת.
יש $(n-1)!$ דרכים לסדר n עצמים במעגל: אם נסדר קודם את העצמים בשורה ואז נצמיד את הראשון והאחרון נקבל מעגל. אך לאותו מעגל ניתן להגיע ב־n צורות שונות (למשל עבור $n=3$: $123, 231, 312$ כולם מייצגים את אותו המעגל) ולכן יש לחלק ב־n.
ניתן לקבל קירוב טוב לעצרת (**קירוב סטירלינג**) כאשר $n \to \infty$ באמצעות הנוסחה $n! \sim \left(\frac{n}{\mathrm{e}}\right)^n \sqrt{2\pi n}$.

1.3 בחירת k עצמים מתוך n ללא החזרה ועם חשיבות לסדר

$$n^{\underline{k}} = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

יש $n^{\underline{k}}$ דרכים שונות לבחור k עצמים מתוך קבוצה של n ולסדר אותם בשורה: n אפשרויות לעצם הראשון, $n-1$ אפשרויות לשני וכך הלאה עד שלעצם האחרון נשארות $n-k+1$ אפשרויות. נשים לב כי $n^{\underline{n}} = n!$.

1.4 בחירת k עצמים מתוך n ללא החזרה וללא חשיבות לסדר: מקדמים בינומיים

$$\binom{n}{k} = \frac{n^{\underline{k}}}{k!} = \frac{n!}{(n-k)!k!}$$

יש $\binom{n}{k}$ דרכים שונות לבחור k עצמים מתוך n עצמים כאשר אין חשיבות לסדר: ראשית נבחר את העצמים ב־$n^{\underline{k}}$ דרכים. אך כל קבוצה של k עצמים מופיעה $k!$ פעמים (כמספר הפרמוטציות שלה), לכן יש לחלק ב־$k!$. כמו כן נשים לב כי יש $\binom{n}{k}$ תת־קבוצות שונות בגודל k של קבוצה בגודל n.
הבינום של ניוטון:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

תכונות של האינדקס התחתון:

$$\binom{n}{0} = \binom{n}{n} = 1, \qquad \binom{n}{k} = \binom{n}{n-k}$$

סכומים על האינדקס התחתון:

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0, \qquad \sum_{k=0}^{n} \binom{n}{k} = 2^n$$

$$\sum_{k=0}^{n} k \binom{n}{k} = 2^{n-1} n$$

$$\sum_{k=0}^{n} k^2 \binom{n}{k} = 2^{n-2} n (n+1)$$

סכום על האינדקס העליון:

$$\sum_{k=0}^{n} \binom{k}{m} = \binom{n+1}{m+1}$$

מספר זהויות שימושיות נוספות:

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

$$\binom{n}{k} = \frac{n}{k} \binom{n-1}{k-1}$$

$$\binom{n}{k}\binom{k}{m} = \binom{n}{m}\binom{n-m}{k-m}$$

$$\sum_{k=0}^{n} \binom{m+k}{k} = \binom{m+n+1}{n}$$

קיימים גם **מקדמים מולטינומיים**:

$$\binom{n}{k_1, \ldots, k_m} = \frac{n!}{k_1! \cdots k_m!}$$

אם $k_1 + \cdots + k_m = n$ אז יש $\binom{n}{k_1, \ldots, k_m}$ חלוקות שונות של קבוצה בעלת n עצמים ל־m קבוצות זרות בעלות $k_1, \ldots, k_m$ עצמים.

1.5 בחירת k עצמים מתוך n עם החזרה ועם חשיבות לסדר

יש n^k דרכים שונות לבחור k עצמים מתוך קבוצה בעלת n עצמים עם החזרה: n אפשרויות לעצם הראשון, n אפשרויות לשני וכך הלאה. ניסוח אחר: אם בכל שלב בניסוי יש n תוצאות אפשריות, אז לאחר k שלבים יש n^k תוצאות אפשריות בסה"כ.

1.6 בחירת k עצמים מתוך n עם החזרה וללא חשיבות לסדר

$$\left(\!\!\binom{n}{k}\!\!\right) = \binom{k+n-1}{k} = \binom{k+n-1}{n-1}$$

הוכחה (בקטן כי היא ארוכה): לכל אחד מ־n העצמים בקבוצה נשייך מספר בין 0 ל־k הקובע את מספר הפעמים שאותו עצם נבחר, כאשר סכום המספרים המשויכים חייב להסתכם ל־k. צורה נוחה למצוא את מספר האפשרויות לעשות זאת היא לסדר k כוכבים בשורה ולהפריד ביניהם באמצעות $n-1$ חוצצים, כך שנוצרות n קבוצות (רצפים של כוכבים) אשר עשויות להיות בגודל 0 עד k. לדוגמה, עבור $k=5$ ו־$n=4$, מיקום אפשרי של חוצצים הוא:

$$\star || \star\star\star | \star$$

משמעות הסידור: ניקח את העצם הראשון בקבוצה פעם אחת, את העצם השני אפס פעמים, את העצם השלישי שלוש פעמים ואת העצם הרביעי פעם אחת - בסה"כ 5 עצמים, כנדרש. כעת, מספר האפשרויות לבחור מיקום ל־$n-1$ החוצצים מתוך $k+n-1$ מיקומים אפשריים (כולל הקצוות) הוא $\binom{k+n-1}{n-1}$.

1.7 סידור k עצמים ב־n תאים כך שבכל תא יש לפחות עצם אחד

$$\binom{k-1}{n-1}$$

כמו בסעיף הקודם, נסדר k כוכבים בשורה ונפריד ביניהם באמצעות $n-1$ חוצצים, כאשר הפעם בין כל שני כוכבים יכול להיות רק חוצץ אחד וכל חוצץ חייב להיות בין שני כוכבים - בסה"כ $k-1$ אפשרויות למיקום החוצצים.

2 הסתברות

2.1 אקסיומות ונוסחאות בסיסיות

יהי S מרחב מדגם ויהי $A \in S$ מאורע, אז:

$$P\left(S\right)=1,\qquad P\left(A\right)\in\left[0,1\right],\qquad P\left(\overline{A}\right)=1-P\left(A\right)$$

כאשר $\overline{A}$ הוא המשלים של A. אם S מרחב סופי וסימטרי אז:

$$P\left(A\right)=\frac{\left|A\right|}{\left|S\right|}$$

2.2 ההסתברות של איחוד מאורעות

יהיו $A_i \in S$, $1\le i\le m$, מאורעות זרים בזוגות, אז:

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P\left(A_i\right)$$

אם A_i אינם זרים בזוגות אז (**עקרון ההכלה וההפרדה**):

$$\begin{aligned}P\left(\bigcup_{i=1}^{n}A_i\right) &= \sum_{i=1}^{n}P\left(A_i\right)-\sum_{i<j}P\left(A_i\cap A_j\right)+\dots\\ &\dots+\left(-1\right)^{n+1}P\left(\bigcap_{i=1}^{n}A_i\right)\end{aligned}$$

כאשר $\left|A\right|$ מציין את מספר האיברים בקבוצה A. **אי־שוויון בול** קובע כי:

$$P\left(\bigcup_{i=1}^{n}A_i\right)\le\sum_{i=1}^{n}P\left(A_i\right)$$

2.3 הסתברות מותנית

יהיו A,B מאורעות. אם $P\left(B\right)>0$ אז ההסתברות המותנית של A בהינתן B היא:

$$P\left(A|B\right)=\frac{P\left(A\cap B\right)}{P\left(B\right)}$$

מכאן:

$$P\left(A\cap B\right)=P\left(A|B\right)P\left(B\right)$$

ובאופן כללי יותר:

$$\begin{aligned}P\left(\bigcap_{i=1}^{n}A_i\right) &= P\left(A_1\right)P\left(A_2|A_1\right)P\left(A_3|A_2\cap A_1\right)\cdots\\ &\cdots P\left(A_n|A_{n-1}\cap\cdots\cap A_1\right)\end{aligned}$$

נוסחת ההסתברות השלמה - אם $B=\biguplus_i B_i$ (איחוד זר) אז:

$$P\left(A\right)=\sum_{i=1}^{n}P\left(A|B_i\right)P\left(B_i\right)$$

נוסחת בייס:

$$P\left(A|B\right)=\frac{P\left(B|A\right)P\left(A\right)}{P\left(B\right)}$$

נוסחת בייס + נוסחת ההסתברות השלמה:

$$P\left(A|B\right)=\frac{P\left(B|A\right)P\left(A\right)}{P\left(B|A\right)P\left(A\right)+P\left(B|\overline{A}\right)P\left(\overline{A}\right)}$$

או במקרה הכללי, אם $A=\biguplus_i A_i$:

$$P\left(A_i|B\right)=\frac{P\left(B|A_i\right)P\left(A_i\right)}{\sum_{j=1}^{n}P\left(B|A_j\right)P\left(A_j\right)}$$

2.4 מאורעות בלתי־תלויים

המאורעות A,B הם בלתי־תלויים אם ורק אם:

$$P\left(A\cap B\right)=P\left(A\right)P\left(B\right)$$

או באופן שקול:

$$P\left(A|B\right)=P\left(A\right),\qquad P\left(B|A\right)=P\left(B\right)$$

כלומר הידע ש־B קרה לא משנה את ההסתברות ש־A יקרה ולהפך.
המאורעות A_i, $1\le i\le n$, הם בלתי־תלויים אם ורק אם **לכל תת־קבוצה** $A_{i_1},\dots,A_{i_r}$ שלהם מתקיים:

$$P\left(\bigcap_{i=1}^{r}A_i\right)=\prod_{i=1}^{r}P\left(A_i\right)$$

3 משתנה מקרי

הגדרה: משתנה מקרי $X:\Omega\to\mathbb{R}$ הוא פונקציה ממרחב המדגם Ω לשדה המספרים הממשיים $\mathbb{R}$. מסמנים:

$$P\left(X=k\right)=P\left(\left\{\omega\in\Omega|X\left(\omega\right)=k\right\}\right)$$

כלומר ההסתברות ש־X מקבל ערך מסוים (או טווח של ערכים) היא ההסתברות של קבוצת כל המאורעות ω במרחב המדגם כך ש־$X\left(\omega\right)$ מקבל את הערך או טווח הערכים הנדרש.

3.1 תוחלת

יהי X משתנה מקרי. התוחלת של X היא:

$$\mathrm{E}\left(X\right)=\sum_{x}x\cdot P\left(X=x\right)$$

כאשר הסכום הוא על כל הערכים האפשריים של X. מכאן, התוחלת היא למעשה ממוצע משוקלל של כל ערכי X. באופן כללי הביטוי $\mathrm{E}\left(X^n\right)$ נקרא "**המומנט** ה־n של X"; התוחלת היא המומנט הראשון. לכל פונקציה $f:\mathbb{R}\to\mathbb{R}$ מתקיים:

$$\mathrm{E}\left(f\left(X\right)\right)=\sum_{x}f\left(x\right)P\left(X=x\right)$$

כאשר הסכימה היא על כל הערכיים האפשריים של X (ולא של $f\left(X\right)$). התוחלת היא לינארית:

$$\mathrm{E}\left(aX+b\right)=a\mathrm{E}\left(X\right)+b$$

אם מתקיים $P\left(a\le X\le b\right)$ אז בהכרח גם $a\le\mathrm{E}\left(X\right)\le b$. יתרה מזאת, אם $X\ge Y$ אז בהכרח $\mathrm{E}\left(X\right)\ge\mathrm{E}\left(Y\right)$.

3.2 שונות

יהי X משתנה מקרי. השונות של X היא:

$$\mathrm{Var}\left(X\right)=\mathrm{E}\left(\left[X-\mathrm{E}\left(X\right)\right]^2\right)=\mathrm{E}\left(X^2\right)-\left[\mathrm{E}\left(X\right)\right]^2$$

השונות מעלה קבועים בריבוע, ואינה מושפעת מהזזות:

$$\mathrm{Var}\left(aX+b\right)=a^2\mathrm{Var}\left(X\right)$$

סטיית התקן מוגדרת להיות השורש של השונות:

$$\sigma_X=\sqrt{\mathrm{Var}\left(X\right)}$$

3.3 פונקציה יוצרת מומנטים

יהי X משתנה מקרי. הפונקציה יוצרת המומנטים היא:

$$M_X\left(t\right)=\mathrm{E}\left(\mathrm{e}^{tX}\right)=\sum_x \mathrm{e}^{tx}P\left(X=x\right)$$

מתקיים:

$$M_X\left(0\right)=1,\qquad M_X'\left(0\right)=\mathrm{E}\left(X\right),\qquad M_X''\left(0\right)=\mathrm{E}\left(X^2\right)$$

וכך הלאה, כאשר הגזירה היא לפי t. לכן השונות היא:

$$\mathrm{Var}\left(X\right)=M_X''\left(0\right)-\left[M_X'\left(0\right)\right]^2$$

אם X ו־Y בלתי־תלויים אז:

$$M_{X+Y}\left(t\right)=M_X\left(t\right)M_Y\left(t\right)$$

הפונקציה יוצרת המומנטים קובעת את ההתפלגות באופן יחיד, לכן ניתן להשתמש בנוסחה זה למציאת ההתפלגות של **סכום משתנים מקריים**. דרך נוספת היא באמצעות **נוסחת הקונבולוציה**:

$$P\left(X+Y=m\right)=\sum_k P\left(X=k\right)P\left(Y=m-k|X=k\right)$$

3.4 שכיח

השכיח הוא הערך הנפוץ ביותר בהתפלגות, כלומר הערך של המשתנה המקרי שמתקבל בהסתברות הגבוהה ביותר. ערך זה אינו בהכרח יחיד (לדוגמה: התפלגות אחידה).

4 מספר משתנים מקריים

שני משתנים מקריים X ו־Y הם בלתי־תלויים אם ורק אם לכל $a,b\in\mathbb{R}$ המאורעות $\{X\leq a\}$ ו־$\{Y\leq b\}$ הם בלתי־תלויים. אי־תלות של מספר משתנים מקריים מוגדרת בדומה לאי־תלות של מספר מאורעות.

4.1 תוחלת

מתקיים:

$$\mathrm{E}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{E}\left(X_i\right)$$

לכל סט של משתנים מקריים (המוגדרים על אותו מרחב מדגם), **גם אם אינם בלתי־תלויים**. מכאן, אם ניתן לרשום:

$$X=\sum_{i=1}^n X_i$$

כאשר X_i הם **אינדיקטורים** שמקבלים את הערך 1 בהסתברות p_i ו־0 אחרת, אז נקבל:

$$\mathrm{E}\left(X\right)=\sum_{i=1}^n p_i$$

4.2 שונות משותפת

יהיו X,Y משתנים מקריים. אז השונות המשותפת שלהם היא:

$$\begin{aligned}\mathrm{Cov}\left(X,Y\right)&=\mathrm{E}\left(\left[X-\mathrm{E}\left(X\right)\right]\left[Y-\mathrm{E}\left(Y\right)\right]\right)\\&=\mathrm{E}\left(XY\right)-\mathrm{E}\left(X\right)\mathrm{E}\left(Y\right)\end{aligned}$$

קל לראות כי:

$$\begin{gathered}\mathrm{Cov}\left(X,X\right)=\mathrm{Var}\left(X\right)\\\mathrm{Cov}\left(X,a\right)=0\\\mathrm{Cov}\left(X,Y\right)=\mathrm{Cov}\left(Y,X\right)\\\mathrm{Cov}\left(aX,bY\right)=ab\mathrm{Cov}\left(X,Y\right)\\\mathrm{Cov}\left(X+Y,Z\right)=\mathrm{Cov}\left(X,Z\right)+\mathrm{Cov}\left(Y,Z\right)\\\mathrm{Cov}\left(\sum_{i=1}^n X_i,\sum_{j=1}^n Y_j\right)=\sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}\left(X_i,Y_j\right)\end{gathered}$$

אם המשתנים X,Y הם בלתי־תלויים, אז:

$$\mathrm{Cov}\left(X,Y\right)=0$$

אבל ההפך אינו בהכרח נכון. כמו כן מתקיים הקשר:

$$\mathrm{Var}\left(X+Y\right)=\mathrm{Var}\left(X\right)+\mathrm{Var}\left(Y\right)+2\,\mathrm{Cov}\left(X,Y\right)$$

ובאופן כללי:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{Var}\left(X_i\right)+2\sum_{i<j}\mathrm{Cov}\left(X_i,X_j\right)$$

בפרט, אם המשתנים X_i בלתי־תלויים:

$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n \mathrm{Var}\left(X_i\right)$$

מקדם המתאם של X,Y הוא:

$$\rho_{X,Y}=\frac{\mathrm{Cov}\left(X,Y\right)}{\sigma_X\sigma_Y}$$

מתקיים תמיד $-1\leq\rho_{X,Y}\leq 1$, כאשר 1 משמעותו שמתקיים קשר לינארי חיובי בין המשתנים, -1 משמעותו שקיים קשר לינארי שלילי ו־0 משמעותו שאין קשר לינארי.

4.3 תוחלת מותנית

יהיו X,Y משתנים מקריים. התוחלת של X בהינתן ש־$Y=y$ היא:

$$\mathrm{E}\left(X|Y=y\right)=\sum_x x\cdot P\left(X=x|Y=y\right)$$

כאשר הסכום הוא על כל הערכים האפשריים של X. חשוב לשים לב כי $\mathrm{E}\left(X|Y=y\right)$ הוא מספר. ניתן גם להגדיר את הפונקציה $\mathrm{E}\left(X|Y\right)$, שהיא משתנה מקרי בפני עצמה. זהות חשובה (**נוסחת התוחלת השלמה**) היא:

$$\mathrm{E}\left(X\right)=\mathrm{E}\left(\mathrm{E}\left(X|Y\right)\right)$$

כלומר:

$$\mathrm{E}\left(X\right)=\sum_y \mathrm{E}\left(X|Y=y\right)P\left(Y=y\right)$$

כמו כן מתקיים:

$$\mathrm{E}\left(g\left(X\right)Y|X\right)=g\left(X\right)\mathrm{E}\left(Y|X\right)$$

מכאן עם $g\left(X\right)=X$ ושימוש בנוסחת התוחלת השלמה נקבל:

$$\mathrm{E}\left(XY\right)=\mathrm{E}\left(X\cdot\mathrm{E}\left(Y|X\right)\right)$$

לכן אם X,Y בלתי־תלויים:

$$\mathrm{E}\left(XY\right)=\mathrm{E}\left(X\right)\mathrm{E}\left(Y\right)$$

4.4 שונות מותנית וסכום משתנים באורך מקרי

יהיו X,Y משתנים מקריים. השונות של X בהינתן Y היא:

$$\begin{aligned}\mathrm{Var}\left(X|Y\right) &= \mathrm{E}\left(\left[X-\mathrm{E}\left(X|Y\right)\right]^2\middle|Y\right)\\ &= \mathrm{E}\left(X^2\middle|Y\right)-\left[\mathrm{E}\left(X|Y\right)\right]^2\end{aligned}$$

מכאן (**נוסחת השונות השלמה**):

$$\mathrm{Var}\left(X\right)=\mathrm{E}\left[\mathrm{Var}\left(X|Y\right)\right]+\mathrm{Var}\left[\mathrm{E}\left(X|Y\right)\right]$$

יהי N משתנה מקרי המקבל ערכים שלמים אי־שליליים ויהיו X_i משתנים מקריים בלתי־תלויים, שווי־התפלגות ובלתי־תלויים ב־N, אז מתקיים:

$$\mathrm{E}\left(\sum_{i=1}^{N}X_i\right)=\mathrm{E}\left(N\right)\mathrm{E}\left(X_i\right)$$

$$\mathrm{Var}\left(\sum_{i=1}^{N}X_i\right)=\mathrm{E}\left(N\right)\mathrm{Var}\left(X_i\right)+\left[\mathrm{E}\left(X_i\right)\right]^2\mathrm{Var}\left(N\right)$$

5 התפלגויות

5.1 התפלגות אחידה

X הוא מספר שלם בין a ל־b הנבחר באקראי.

$$X\sim\mathrm{Uni}\left(a,\ldots,b\right)$$

יש בסה"כ $n=b-a+1$ אפשרויות וכל אחת בעלת הסתברות שווה, לכן ההתפלגות, התוחלת, השונות והפונקציה יוצרת המומנטים הן:

$$P\left(X=k\right)=\frac{1}{b-a+1}=\frac{1}{n}$$

$$\mathrm{E}\left(X\right)=\frac{a+b}{2}$$

$$\mathrm{Var}\left(X\right)=\frac{\left(b-a+1\right)^2-1}{12}=\frac{n^2-1}{12}$$

$$M_X\left(t\right)=\frac{\mathrm{e}^{at}-\mathrm{e}^{(b+1)t}}{n\left(1-\mathrm{e}^{t}\right)}$$

5.2 התפלגות בינומית

X הוא מספר ההצלחות ב־n ניסויים בלתי־תלויים אשר סיכוי ההצלחה בכל אחד מהם הוא p.

$$X\sim\mathrm{Bin}\left(n,p\right)$$

נבחר k מתוך n הניסויים. ניסויים אלה הצליחו, בהסתברות p, ו־$n-k$ הניסויים הנותרים נכשלו, בהסתברות $1-p$. לכן ההתפלגות היא:

$$P\left(X=k\right)=\binom{n}{k}p^k\left(1-p\right)^{n-k}$$

התוחלת והשונות הן:

$$\mathrm{E}\left(X\right)=np,\qquad \mathrm{Var}\left(X\right)=np\left(1-p\right)$$

השכיח הוא:

$$\mathrm{Mode}\left(X\right)=\lfloor(n+1)p\rfloor\ \text{ or }\ \lfloor(n+1)p\rfloor-1$$

והפונקציה יוצרת המומנטים היא:

$$M_X\left(t\right)=\left(1-p+p\mathrm{e}^t\right)^n$$

אם $X\sim\mathrm{Bin}\left(n,p\right)$ ו־$Y\sim\mathrm{Bin}\left(m,p\right)$ הם בלתי־תלויים, אז מתקיים:

$$X+Y\sim\mathrm{Bin}\left(n+m,p\right)$$

5.3 התפלגות בינומית שלילית וגאומטרית

התפלגות בינומית שלילית:

X הוא מספר הניסויים עד להצלחה ה־n (כולל), כשהניסויים בלתי־תלויים והסיכוי להצלחה בכל אחד מהם הוא p.

$$X\sim\mathrm{NB}\left(n,p\right)$$

נבחר $n-1$ מתוך $k-1$ הניסויים הראשונים. ניסויים אלה הצליחו, בהסתברות p^{n-1}, וה־$k-n$ הנותרים נכשלו, בהסתברות $1-p$. לאחר מכן עלינו להכפיל את התוצאה ב־p משום שהניסוי האחרון הצליח. לכן ההתפלגות היא:

$$P\left(X=k\right)=\binom{k-1}{n-1}p^n\left(1-p\right)^{k-n}$$

התוחלת והשונות הן:

$$\mathrm{E}\left(X\right)=\frac{n}{p},\qquad \mathrm{Var}\left(X\right)=n\frac{1-p}{p^2}$$

והפונקציה יוצרת המומנטים היא:

$$M_X\left(t\right)=\left(\frac{p\mathrm{e}^t}{1-\left(1-p\right)\mathrm{e}^t}\right)^n$$

אם $X\sim\mathrm{NB}\left(n,p\right)$ ו־$Y\sim\mathrm{NB}\left(m,p\right)$ הם בלתי־תלויים, אז מתקיים:

$$X+Y\sim\mathrm{NB}\left(n+m,p\right)$$

התפלגות גאומטרית:

זהו מקרה פרטי של התפלגות בינומית שלילית כאשר $n=1$.

$$X\sim\mathrm{Geo}\left(p\right)\equiv\mathrm{NB}\left(1,p\right)$$

ההתפלגות היא:

$$P\left(X=k\right)=p\left(1-p\right)^{k-1}$$

$$P\left(X\le k\right)=1-\left(1-p\right)^k$$

התוחלת והשונות הן:

$$\mathrm{E}\left(X\right)=\frac{1}{p},\qquad \mathrm{Var}\left(X\right)=\frac{1-p}{p^2}$$

והשכיח הוא 1, כי ההסתברות יורדת ככל ש־k עולה. הפונקציה יוצרת המומנטים היא:

$$M_X\left(t\right)=\frac{p\mathrm{e}^t}{1-\left(1-p\right)\mathrm{e}^t}$$

אם $X_i\sim\mathrm{Geo}\left(p\right)$ עבור $1\le i\le n$ הם בלתי־תלויים, אז מתקיים:

$$\sum_{i=1}^{n}X_i\sim\mathrm{NB}\left(n,p\right)$$

ההתפלגות הגאומטרית היא ההתפלגות הבדידה **היחידה** שהיא "חסרת זיכרון", כלומר, בכל שלב בניסוי יש אותו סיכוי להצלחה ללא תלות בכמות או בתוצאות הניסויים הקודמים (למשל, למטבע שמוטל אין "זיכרון" של ההטלות הקודמות). ניתן לתאר תכונה זו באמצעות הנוסחה:

$$P\left(X>m+n|X>m\right)=P\left(X>n\right)$$

כלומר, לכל $n,m\in\mathbb{N}$, אם ידוע לנו שכבר נערכו m ניסויים כושלים, ההסתברות ש־n ניסויים נוספים ייכשלו שווה להסתברות ש־n ניסויים ייכשלו אם נתחיל את התהליך מהתחלה. מכאן, אם מצאנו התפלגות המקיימת:

$$P\left(X=k+1\right)=cP\left(X=k\right)$$

כאשר c קבוע כלשהו, התפלגות זו בהכרח תהיה גאומטרית (הקבוע c הוא פשוט ההסתברות לכישלון בניסוי).

5.4 התפלגות היפרגאומטרית

X הוא מספר הפריטים "המיוחדים" שנמצאו בדגימה (ללא החזרה) בגודל n מתוך אוכלוסייה בגודל N הכוללת D פרטים מיוחדים.

$$X \sim \mathrm{HG}\left(N, D, n\right)$$

נבחר k פריטים מיוחדים מתוך D המיוחדים שבאוכלוסייה, ונבחר את $n-k$ הפריטים הלא־מיוחדים שנשארו מתוך $N-D$ הלא־מיוחדים באוכלוסייה. נחלק את מספר האפשרויות בגודל מרחב המדגם, שהוא n פריטים כלשהם שנדגמו מאוכלוסייה בגודל N. לכן ההתפלגות היא:

$$P\left(X=k\right) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

התוחלת והשונות הן:

$$\mathrm{E}\left(X\right) = n\frac{D}{N}, \qquad \mathrm{Var}\left(X\right) = n\frac{D}{N}\left(1-\frac{D}{N}\right)\frac{N-n}{N-1}$$

והשכיח הוא $\left\lfloor \frac{(n+1)(D+1)}{N+2} \right\rfloor$.

5.5 התפלגות פואסונית

X הוא מספר ההצלחות בתהליך בינומי כאשר $p \to 0$, $n \to \infty$, אך המכפלה $\lambda = np$ קבועה. לכן ההתפלגות משמשת לקירוב של תהליכים בהם יש הסתברות נמוכה להצלחה ומספר רב של ניסיונות, לדוגמה מספר טעויות ההדפסה בספר או מספר האנשים שמגיעים לגיל 100 באוכלוסייה.

$$X \sim \mathrm{Pois}\left(\lambda\right)$$

ההתפלגות, התוחלת והשונות הן:

$$P\left(X=k\right) = \frac{\mathrm{e}^{-\lambda}\lambda^k}{k!}, \qquad \mathrm{E}\left(X\right) = \mathrm{Var}\left(X\right) = \lambda$$

והשכיח הוא $\lfloor\lambda\rfloor$, או $\lambda-1$ אם $\lambda\in\mathbb{Z}$. הפונקציה יוצרת המומנטים היא:

$$M_X\left(t\right) = \mathrm{e}^{\lambda\left(\mathrm{e}^t-1\right)}$$

אם $X \sim \mathrm{Pois}\left(\lambda\right)$ ו־$Y \sim \mathrm{Pois}\left(\mu\right)$ הם בלתי־תלויים, אז מתקיים:

$$X+Y \sim \mathrm{Pois}\left(\lambda+\mu\right)$$

6 אי־שוויונים, קירובים ומשפטי גבול

6.1 אי־שוויון מרקוב

יהי X משתנה מקרי אי־שלילי, אז לכל $a>0$ מתקיים:

$$P\left(X \ge a\right) \le \frac{\mathrm{E}\left(X\right)}{a}$$

ניתן לבצע הזזה של המשתנה, כלומר $P\left(X-b \ge a-b\right)$, כדי לשפר את החסם - כל עוד $X-b$ אי־שלילי. כמו כן, במקרה של מספרים שלמים $P\left(X>a\right) = P\left(X \ge a+1\right)$.

6.2 אי־שוויון צ'בישב

יהי X משתנה מקרי בעל שונות $\sigma^2 = \mathrm{Var}\left(X\right)$, אז לכל $k>0$ מתקיים:

$$P\left(\left|X-\mathrm{E}\left(X\right)\right| \ge k\right) \le \frac{\sigma^2}{k^2}$$

או באופן שקול:

$$P\left(\left|X-\mathrm{E}\left(X\right)\right| \ge k\sigma\right) \le \frac{1}{k^2}$$

קיימת גם **גרסה חד־צדדית** של שוויון זה:

$$P\left(X-\mathrm{E}\left(X\right) \ge k\right) \le \frac{\sigma^2}{\sigma^2+k^2}$$

או באופן שקול:

$$P\left(X-\mathrm{E}\left(X\right) \ge k\sigma\right) \le \frac{1}{1+k^2}$$

אבחנה חשובה - אם המשתנה המקרי הוא **סימטרי סביב התוחלת**, כלומר $P\left(X-\mathrm{E}\left(X\right)\right) = P\left(\mathrm{E}\left(X\right)-X\right)$, אז ניתן לרשום:

$$P\left(X-\mathrm{E}\left(X\right) \ge k\right) = \frac{1}{2}P\left(\left|X-\mathrm{E}\left(X\right)\right| \ge k\right)$$

6.3 אי־שוויון צ'רנוף

יהי X משתנה מקרי ותהי:

$$M_X\left(t\right) = \mathrm{E}\left(\mathrm{e}^{tX}\right)$$

הפונקציה יוצרת המומנטים שלו. אז מתקיים:

$$P\left(X \ge a\right) \le \mathrm{e}^{-ta}M_X\left(t\right), \qquad t>0$$

$$P\left(X \le a\right) \le \mathrm{e}^{-ta}M_X\left(t\right), \qquad t<0$$

באמצעות מציאת ה־t (החיובי או השלילי לפי המקרה) שנותן את הערך הנמוך ביותר לפונקציה $\mathrm{e}^{-ta}M_X\left(t\right)$, נוכל למצוא את החסם הטוב ביותר.

6.4 אי־שוויון ינסן

הגדרה: פונקציה $f\left(x\right)$ נקראת **קמורה** אם $f''\left(x\right) \ge 0$ לכל x, **וקעורה** אם $f''\left(x\right) \le 0$. אם $f\left(x\right)$ קמורה אז מתקיים:

$$\mathrm{E}\left[f\left(X\right)\right] \ge f\left[\mathrm{E}\left(X\right)\right]$$

6.5 קירוב לינארי

יהיו X,Y משתנים מקריים. התחזית הלינארית האופטימלית, $\hat{Y}$, של Y לפי X ניתנת באמצעות הנוסחה:

$$\frac{\hat{Y}-\mathrm{E}\left(Y\right)}{\sigma_Y} = \rho_{X,Y}\frac{X-\mathrm{E}\left(X\right)}{\sigma_X}$$

כאשר:

$$\rho_{X,Y} = \frac{\mathrm{Cov}\left(X,Y\right)}{\sigma_X\sigma_Y}$$

הוא מקדם המתאם (ע"ע). ביטוי מפורש ל־$\hat{Y}$ הוא:

$$\hat{Y} = aX+b$$

כאשר:

$$a = \frac{\mathrm{Cov}\left(X,Y\right)}{\mathrm{Var}\left(X\right)}, \qquad b = \mathrm{E}\left(Y\right) - \frac{\mathrm{Cov}\left(X,Y\right)}{\mathrm{Var}\left(X\right)}\mathrm{E}\left(X\right)$$

השגיאה הריבועית הממוצעת (MSE) של התחזית $\hat{Y}$ היא:

$$\mathrm{MSE}\left(\hat{Y}\right) = \mathrm{E}\left[\left(\hat{Y}-Y\right)^2\right]$$

והיא מספקת מדד למידת הדיוק של הקירוב הלינארי. אם $\mathrm{MSE}\left(\hat{Y}\right) = 0$ הקירוב הוא מדויק.

6.6 חוק המספרים הגדולים

יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים ושווי־התפלגות בעלי תוחלת μ ושונות σ^2. אז מתקיים:

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(X_i\right) = \frac{1}{n}\cdot n\mu = \mu$$

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}\left(X_i\right) = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

לכן לכל $\varepsilon > 0$ מתקיים, לפי אי־שוויון צ'בישב:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

ובגבול $n \to \infty$ נקבל:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$$

מכאן, ככל שנגדיל את מספר הניסויים n, הממוצע $\frac{1}{n}\sum_{i=1}^{n} X_i$ של תוצאות הניסויים יתקרב לתוחלת μ.

6.7 משפט הגבול המרכזי

יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים ושווי־התפלגות בעלי תוחלת μ ושונות σ^2. אז לכל $a \in \mathbb{R}$ מתקיים:

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq a\right) = \Phi\left(a\right)$$

או:

$$\lim_{n\to\infty} P\left(\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \leq a\right) = \Phi\left(a\right)$$

כאשר:

$$\Phi\left(a\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} \mathrm{e}^{-\frac{1}{2}x^2}\mathrm{d}x$$

היא ההתפלגות הנורמלית הסטנדרטית, שהיא התפלגות רציפה המקיימת $\Phi\left(\infty\right) = 1$. יהי $A_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ הממוצע של תוצאות n הניסויים הראשונים. קל לראות כי למשתנה המקרי:

$$Y_n = \frac{A_n - \mu}{\sigma/\sqrt{n}}$$

יש תוחלת 0 ושונות 1, ומתקיים:

$$\lim_{n\to\infty} P\left(Y_n \leq a\right) = \Phi\left(a\right)$$

מכאן, ככל שנגדיל את מספר הניסויים n, התפלגות הממוצע "המתוקנן" Y_n תתקרב להתפלגות הנורמלית הסטנדרטית.
אם ברשותנו טבלת התפלגות נורמלית בעלת ערכים חיוביים בלבד, נוכל להשתמש בנוסחה:

$$\Phi\left(-a\right) = 1 - \Phi\left(a\right)$$

כמו כן נשים לב כי קיים קירוב גם לאי־שוויון בכיוון ההפוך (ההסתברות המשלימה):

$$\lim_{n\to\infty} P\left(Y_n \geq a\right) = 1 - \Phi\left(a\right)$$

ולאי שוויון דו־צדדי ($|Y_n| \leq a$):

$$\lim_{n\to\infty} P\left(-a \leq Y_n \leq a\right) = \Phi\left(a\right) - \Phi\left(-a\right) = 2\Phi\left(a\right) - 1$$

עבור קירוב של משתנה מקרי בינומי שלילי, נציג אותו כסכום של משתנים גאומטריים; עבור קירוב של משתנה מקרי פואסוני, נציג אותו כסכום של משתנים פואסוניים עם $\lambda = 1$; עבור קירוב של משתנה מקרי בינומי, נציג אותו כסכום של אינדיקטורים (משתנים בינומיים עם $n = 1$). במקרה האחרון הקירוב נחשב טוב אם $np\left(1-p\right) \geq 10$ בערך.
כמו כן, אם המשתנה המקרי שלנו מקבל ערכים שלמים בלבד (זה בדר"כ המצב), ואנו רוצים לבדוק למשל מתי $P\left(X > b\right)$, אז ניתן להפוך את הקירוב לקצת יותר טוב אם משנים את b למספר החצי־שלם הקרוב ביותר ל־b כך ש־b נכלל בתחום, למשל עבור $P\left(\sum X_i \geq 100\right)$ אפשר לחשב $P\left(\sum X_i \geq 99.5\right)$.

7 שונות

7.1 סכומים נפוצים

$$\sum_{k=0}^{n} k = \frac{n\left(n+1\right)}{2}$$

$$\sum_{k=m}^{n} k = \frac{\left(n+m\right)\left(n+1-m\right)}{2}$$

$$\sum_{k=0}^{n} k^2 = \frac{n\left(n+1\right)\left(2n+1\right)}{6}$$

$$\sum_{k=m}^{n} k^2 = \frac{\left(n+1-m\right)\left(2n^2+2m^2+2nm+n-m\right)}{6}$$

$$\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q}, \qquad \sum_{k=m}^{n} q^k = \frac{q^m - q^{n+1}}{1-q}$$

$$\sum_{k=0}^{\infty} kq^k = \frac{q}{\left(1-q\right)^2}, \qquad \sum_{k=0}^{\infty} k^2q^k = \frac{q\left(1+q\right)}{\left(1-q\right)^3}$$

$$\sum_{k=1}^{n} k\cdot k! = \left(n+1\right)! - 1$$

7.2 טיפים לפתרון בעיות

- אם מצאנו הסתברויות של מאורעות שהאיחוד הזר שלהם הוא כל מרחב המדגם, כדאי לבדוק שסכום ההסתברויות הוא אכן 1.
- אם צריך למצוא את ההסתברות ש"לפחות איבר אחד מתוך קבוצה כלשהי מקיים תנאי כלשהו" אז לרוב עדיף למצוא את אחד מינוס ההסתברות המשלימה, כלומר ההסתברות "לא קיים אף איבר בקבוצה המקיים את התנאי". אחרת יש סכנה לספירות כפולות.
- לפונקציה $p\left(1-p\right)$ יש נקודת מקסימום ב־$p = \frac{1}{2}$, וערכה שם הוא $\frac{1}{4}$.
- למציאת התוחלת של מכפלת שני משתנים מקריים $\mathrm{E}\left(XY\right)$ ניתן להתנות באחד מהם ולהשתמש בנוסחת התוחלת השלמה.
- אם קיבלנו התפלגות שהיא סימטרית סביב ערך c (כלומר $P\left(X=c+k\right) = P\left(X=c-k\right)$) אז $\mathrm{E}\left(X\right) = c$.
- כדי להראות ששני משתנים מקריים X, Y הם תלויים מספיק למצוא k מסוים שעבורו $P\left(X|Y=k\right) \neq P\left(X\right)$.
- אם צריך למצוא את היחס בין שני ערכים (למשל הסתברויות), לפעמים חלק מהביטויים מצטמצמים ואין צורך לחשב אותם. לדוגמה: בחישוב יחס בין הסתברויות מותנות המכנה עשוי להצטמצם.