יסודות תורת ההסתברות

אוהד נ. פלדהיים
אורי גוראל גורביץ׳

מכון איינשטיין למתמטיקה
האוניברסיטה העברית בירושלים

12 בנובמבר 2018

פרק 5

התוחלת

ה"יתרון" בתורת הסיכויים הוא מכפלת הסכום שאנו מייחלים לו בהסתברות לזכות בו; זה הוא הסכום שראוי לחלקו כאשר אנחנו לא מעוניינים ליטול את הסיכון אלא לחלק [את הסכום המיוחל] לפי ההסתברויות. חלוקה זו היא החלוקה ההוגנת היחידה אם מתעלמים מנסיבות משונות; מפני שנתחים שווים של הסתברות מובילים לנתחים שווים מהסכום המיוחל. אנחנו נכנה את המושג הזה בשם "תקווה מתמטית".

– פְּיֶיר סִימוֹן לַפְּלַס, הצגתו הראשונה של מושג התוחלת ב**תיאוריה אנליטית של התפלגות**, 1814.

בפרק זה נגדיר את מושג ה**תוחלת**. מושג מתמטי שיתאר תוצאה ממוצעת צפויה של ניסוי. נשים לב שזו הפעם הראשונה בה נתעניין בניתוח הערכים שמקבל משתנה מקרי, ולא רק במאורעות שמיוצגים על ידם. בפרק הבא נקשור את התוחלת עם הממוצע של תוצאות ניסויים חוזרים בלתי-תלויים, ואולם עוד בטרם נבסס קשר זה, יתגלו לפנינו שימושים נוספים של התוחלת עת נקבל באמצעותה הערכות טובות יותר להתפלגותם של משתנים מקריים. בפרקים הבאים נכליל את התוחלת ואת שימושיה בהציגנו את מושגי ה**שונות** וה**מומנט**.

5.1 תוחלת

תּוֹחֶלֶת מְמֻשָּׁכָה מַחֲלָה לֵב וְעֵץ חַיִּים תַּאֲוָה בָאָה.

– משלי י"ג 12

את הגדרתה של התוחלת המובאת להלן אפשר לראות כ"ממוצע משוקלל של מרחב המדגם לפי פונקציית ההסתברות הנקודתית".

הגדרה 5.1 (תוחלת)**.** יהי X משתנה מקרי המוגדר על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$.

ה**תוחלת (expectation)** של X, מוגדרת על ידי

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\})$$

במידה וטור זה מתכנס בהחלט במובן הרחב (ר' 1.7). אחרת נאמר שלמשתנה X אין תוחלת.

טענה 5.2 (הגדרה שקולה לתוחלת). יהי X משתנה מקרי המוגדר על מרחב הסתברות בדידה. נזכר כי $\mathrm{Supp}(X) = \{x \in \mathbb{R} \ : \ \mathbb{P}(X = x) > 0\}$ ונשים לב שזו היא קבוצה בדידה, כלומר כזו אשר ניתן לסכום עליה. התוחלת של X ניתנת לחישוב על ידי

$$\mathbb{E}(X) = \sum_{s \in \mathbb{R}} s \cdot \mathbb{P}(X = s) = \sum_{s \in \mathbb{R}_+} s \cdot \mathbb{P}(X = s) - \sum_{s \in \mathbb{R}_-} -s \cdot \mathbb{P}(X = s)$$

במקרה שלפחות אחד הטורים באגף ימין מתכנס למספר סופי. אחרת נאמר כי ל-X אין תוחלת. בפרט, התוחלת של משתנה מקרי תלויה רק בהתפלגותו.

הוכחה.

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{s \in \mathrm{Supp}(X)} \sum_{\omega \in X^{-1}(s)} X(\omega)\mathbb{P}(\{\omega\}) = \sum_{s \in \mathbb{R}} s \cdot \mathbb{P}(X = s)$$

■

מטענה 5.2 אנו למדים כי תוחלת היא תכונה של התפלגות וכי היא אינה תלויה במשתנה המקרי ובמרחב ההסתברות עליו הוא מוגדר.

דוגמא 5.3 (תוחלת משתנים מקריים מוכרים).

(א) תוחלת משתנה מקרי **ברנולי** $X \sim \mathrm{Ber}(p)$ היא

$$\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = \mathbf{p}.$$

לפיכך, התוחלת של $\mathbb{1}(A)$ (האינדיקטור של המאורע A) היא $\mathbb{P}(A)$.

(ב) תוחלת משתנה מקרי **אחיד** על $[N]$ היא

$$\mathbb{E}(X) = \sum_{n \in [N]} n\,\mathbb{P}(X = n) = \frac{N(N+1)}{2N} = \frac{\mathbf{N+1}}{\mathbf{2}}.$$

(ג) תוחלת משתנה מקרי **בינומי** $X \sim \mathrm{Bin}(N, p)$ היא

לראות איך זה מתנהג באמת

$$\begin{aligned}
\mathbb{E}(X) = \sum_{n=0}^{N} n \binom{N}{n} p^n (1-p)^{N-n} \quad &= \sum_{n=0}^{N} n \frac{N!}{(N-n)!n!} p^n (1-p)^{N-n} \\
= \sum_{n=1}^{N} \frac{N!}{(N-n)!(n-1)!} p^n (1-p)^{N-n} &\overset{[m=n-1]}{=} \sum_{m=0}^{N-1} \frac{N!}{(N-m-1)!m!} p^{m+1} (1-p)^{N-m-1} \\
= Np \sum_{m=0}^{N-1} \frac{(N-1)!}{(N-1-m)!m!} p^m (1-p)^{N-1-m} &\overset{\text{בינום}}{=} Np\Big(p + (1-p)\Big)^{N-1} = \mathbf{Np}.
\end{aligned}$$

(ד) תוחלת משתנה מקרי **פואסוני** $X \sim \mathrm{Po}(\lambda)$ היא

$$\mathbb{E}(X) = \sum_{n=0}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} = \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \overset{[m=n-1]}{=} \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

נשים לב שזו גם תוחלתו של משתנה $Y \sim \mathrm{Bin}(N, \lambda/N)$ כפי שניתן היה לצפות מטענה ‼ (אע"פ שנחוצים תנאים נוספים לשימור של תוחלת תחת גבול).

(ה) תוחלת משתנה מקרי **גיאומטרי** $X \sim \text{Geo}(p)$ היא

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} np(1-p)^{n-1}$$

כדי לחשב טור זה נזכר בנוסחה לסכום של טור הנדסי עבור $|x| < 1$:

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

נגזור את שני הצדדים ונקבל:

$$\sum_{n=1}^{\infty} nx^{n-1} = \frac{1}{(1-x)^2}$$

כאשר הגזירה מוצדקת כיון שטור הנגזרות מתכנס במידה שווה בסביבת x. נציב כעת $x = 1-p$ ונקבל:

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} np(1-p)^{n-1} = p\frac{1}{p^2} = \mathbf{\frac{1}{p}}$$

טענה 5.4 (תכונות התוחלת). יהיו X, Y מ"מ בדידים בעלי תוחלת סופית המוגדרים באותו מרחב הסתברות, אזי

(א) אם $\mathbb{P}(X \geq 0) = 1$ אז $\mathbb{E}(X) \geq 0$. אם בנוסף $\mathbb{P}(X > 0) > 0$ אז $\mathbb{E}(X) > 0$

(ב) **ליניאריות התוחלת:** $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ לכל $a, b \in \mathbb{R}$.

(ג) **מונוטוניות התוחלת:** אם $\mathbb{P}(X \geq Y) = 1$ אז $\mathbb{E}(X) \geq \mathbb{E}(Y)$.

הוכחה למ"מ בדיד. **סעיף א.**

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) \geq 0$$

מכיון שכל המחוברים אי-שליליים - שהרי אם $X(\omega) < 0$ אז לפי ההנחה שלנו $\mathbb{P}(\{\omega\}) = 0$ ואם לפחות אחד מהם חיובי, אז התוחלת חיובית.

סעיף ב. נחשב לפי הגדרה:

$$\mathbb{E}(aX + bY) = \sum_{\omega \in \Omega} (aX(\omega) + bY(\omega))\mathbb{P}(\{\omega\})$$
$$= a\sum_{\omega \in \Omega} X(\omega)\mathbb{P}(\{\omega\}) + b\sum_{\omega \in \Omega} Y(\omega)\mathbb{P}(\{\omega\}) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

סעיף ג. נרשום $X = (X - Y) + Y$ ונחשב לפי הסעיפים הקודמים:

$$\mathbb{E}(X) \overset{\text{סעיף ב}}{=} \mathbb{E}(X - Y) + \mathbb{E}(Y) \overset{\text{סעיף א}}{\geq} \mathbb{E}(Y).$$

■

דוגמא 5.5 (תוחלת מ"מ בינומי באמצעות ליניאריות התוחלת).

יהיו $p \in (0,1)$ ו-$N \in \mathbb{N}$ ויהי $X \sim \text{Bin}(N, p)$ מ"מ. בדוגמא 5.3ג חישבנו את תוחלתו של X באופן ישיר. כעת נחשבה ביתר קלות באמצעות ליניאריות התוחלת. יהיו $\{X_n\}_{n \in [N]}$ מ"מ ב"ת המתפלגים ברנולי עם סיכוי הצלחה p ונסמן את סכומם ב-S. לפי טענה 4.42 S ו-X שווי התפלגות ולכן הם שווי תוחלת. לפי ליניאריות התוחלת, טענה 5.4ב, מתקיים

דוג׳:

בעיה: בכד אחד יש כדורים a אדומים, b שחורים
שולפים בלי החזרה k כדורים אדומים מה תוחלת מספר כדורים שהוצאתי ?

פתרון: ההסתברות של כל k כדורים להיות אדום זהה ולכן נסמן $X_i = \{$ הכדור ה-i אדום $\}$

אינדיקטור

$$X_i = \begin{cases} 1 & \text{כדור אדום} \\ 0 & \text{אחרת} \end{cases}$$

נשים לב כי כאן המשתנים תלויים זה בזה אך זה לא משנה לתוחלת

$$E\left(\sum_{x_i=1}^{k} x_i\right) = \sum_{x_i=1}^{k} E(x_i) = k\cdot E(x_i) = k\cdot\left(\frac{a}{a+b}\cdot 1 + \cancel{\frac{1-k}{a+b}\cdot 0}\right) = \frac{ka}{a+b}$$

סיכוי של כדור אדום

הערות:

(1) נשים לב שהתוחלת של אינדיקטור לא מבדילה בין משתנים תלויים וב״ת

(2) התשובה לשאלה הייתה נשארת זהה גם אם היינו מחזירים את הכדורים לכד

$$\mathbb{E}\left(\sum_{n=1}^{N} X_i\right) = \sum_{n=1}^{N} \mathbb{E}(X_i) = \sum_{n=1}^{N} p = Np.$$

כאשר השוויון האמצעי נובע מדוגמא 5.3א.

טענה 5.6 (נוסחא לתוחלת של מ"מ טבעי באמצעות פונקציית ההתפלגות השיורית). יהי X מ"מ המקיים $\mathrm{Supp}(X) \subset \mathbb{N}$. אזי,

$$\mathbb{E}(X) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n)$$

הוכחה. נחשב תוחלת לפי ההגדרה השקולה וננצל את העובדה שהמחוברים בטור אי-שליליים כדי להחליף סדר סכימה.

$$E(X) = \sum_{n=1}^{\infty} n\mathbb{P}(X=n) = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mathbb{P}(X=n) = \sum_{\substack{k,n\in\mathbb{N}\\ k\leq n}} \mathbb{P}(X=n) = \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} \mathbb{P}(X=n) = \sum_{n\in\mathbb{N}} \mathbb{P}(X \geq n)$$

■

דוגמא 5.7 (תוחלת מ"מ גיאומטרי באמצעות פונקציית התפלגות שיורית). יהי $X \sim \mathrm{Geo}(p)$ עבור $p \in (0,1)$. בדוגמא 5.3ה חישבנו את תוחלתו של X תוך שימוש בחשבון אינפיניטיסימלי. כעת נחשבה ביתר קלות באמצעות טענה 5.6. נרשום

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n) = \sum_{n=1}^{\infty} (1-p)^{n-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

אחד הכלים העיקריים שעמדו לרשותנו ליצירה ועיבוד של משתנים מקריים הוא היכולת להפעיל עליהם פונקציות ואופרטורים. כעת נתעניין בשאלה כיצד לחשב את תוחלתו של משתנה מקרי שנוצר על ידי הפעלה של פונקציה על משתנה מקרי אחר.

טענה 5.8 (תוחלת של פונקציה של מ"מ). יהי X מ"מ בדיד ותהי $f : \mathbb{R} \to \mathbb{R}$ פונקציה. אזי המ"מ $Y = f(X)$ מקיים

$$\mathbb{E}(Y) = \sum_{x\in\mathbb{R}} f(x) p_X(x)$$

בתנאי שטור זה מתכנס בהחלט.

אם צריך לחשב את התוחלת של X
בעזרת המשתנה Y לפעם נוחות

הוכחה. נחשב לפי הגדרה.

$$\mathbb{E}(Y) = \sum_{\omega\in\Omega} f(X(\omega))\mathbb{P}(\{\omega\}) = \sum_{x\in\mathrm{Supp}(X)} \sum_{\omega\in X^{-1}(x)} f(X(\omega))\mathbb{P}(\{\omega\})$$
$$= \sum_{x\in\mathrm{Supp}(X)} f(x) \sum_{\omega\in X^{-1}(x)} \mathbb{P}(\{\omega\}) = \sum_{x\in\mathbb{R}} f(x) p_X(x).$$

נשים לב שלפי הגדרה, התוחלת קיימת אם ורק אם הטורים מתכנסים בהחלט ולכן החלפת סדר הסכימה מותרת. ■

✎ **בעיה 5.1** (תוחלת ופונקציות של יותר ממשתנה אחד). יהיו $X_1,\dots,X_k$ מ"מ בדידים ותהי $f:\mathbb{R}^k\to\mathbb{R}$ פונקציה. יש להוכיח כי המ"מ $Y=f(X_1,\dots,X_k)$ מקיים

$$\mathbb{E}\big(Y\big)=\sum_{\substack{(x_1,\dots,x_k)\\ x_i\in\mathrm{Supp}(X_i)}} f(x_1,\dots,x_k)p_{X_1,\dots,X_k}(x_1,\dots,x_k)$$

(נזכר כי $p_{X_1,\dots,X_k}$ היא פונקציית ההתפלגות המשותפת האטומית שהוגדרה בהגדרה 4.21)

טענה 5.9 (תוחלת של מכפלת מ"מ בלתי-תלויים). יהיו X,Y מ"מ ב"ת בדידים בעלי תוחלת סופית על מרחב הסתברות, אז התוחלת של XY קיימת ומקיימת,

$$\mathbb{E}(XY)=\mathbb{E}(X)\mathbb{E}(Y)$$

הוכחה.

$$\mathbb{E}(XY)=\sum_{x\in\mathbb{R}}\sum_{y\in\mathbb{R}}xyp_{X,Y}(x,y)=\sum_{x\in\mathbb{R})}\sum_{y\in\mathbb{R}}xyp_X(x)p_Y(y)$$
$$=\sum_{x\in\mathbb{R}}xp_X(x)\sum_{y\in\mathbb{R}}yp_Y(y)=\mathbb{E}(X)\mathbb{E}(Y).$$

ת.פ שווים כאן אינטגרלים
והם כופלים של יחיבניים

למעשה ההכללה הבאה מתקיימת.

טענה 5.10 (תוחלת של מכפלת מ"מ בלתי-תלויים). יהיו X,Y מ"מ בדידים על מרחב הסתברות. אזי X ו-Y בלתי-תלויים אם ורק אם לכל שתי פונקציות $f,g:\mathbb{R}\to\mathbb{R}$ מתקיים $\mathbb{E}(f(X)g(Y))=\mathbb{E}(f(X))\mathbb{E}(g(Y))$.

הוכחה. נניח שלכל $f,g:\mathbb{R}\to\mathbb{R}$ מתקיים $\mathbb{E}(f(X)g(Y))=\mathbb{E}(f(X))\mathbb{E}(g(Y))$. אזי לכל $x\in\mathrm{Supp}(X)$ ולכל $y\in\mathrm{Supp}(Y)$ מתקיים

$$\mathbb{P}(X=x,Y=y)=\mathbb{E}(\mathbb{1}(X=x)\mathbb{1}(Y=y))=\mathbb{E}(\mathbb{1}(X=x)\mathbb{E}(\mathbb{1}(Y=y))=\mathbb{P}(X=x)\mathbb{P}(Y=y)$$

ולכן, לפי אבחנה 4.31 המ"מ ב"ת. הכיוון השני, לפי טענה 4.34, נובע מטענה 5.9. ■

דוגמא 5.11 (מכפלת שתי קוביות).

מהי תוחלת מכפלת שתי קוביות בלתי תלויות?

תשובה: אם X_1 ו-X_2 הם משתנים מקריים בלתי תלויים המתפלגים אחיד ב-$[6]$ הרי שתוחלת מכפלתם הינה

$$\mathbb{E}(X_1X_2)=\mathbb{E}(X_1)\mathbb{E}(X_2)=\left(\frac{7}{2}\right)^2=12\frac{1}{4}.$$

5.2 תוחלת תחת התניה

בפרק הקודם הגדרנו את התפלגותו של משתנה מקרי בהנתן מאורע A בעל הסתברות חיובית באמצעות החלפת פונקציית ההסתברות ב-$\mathbb{P}_A$, פונקציית ההסתברות המותנית בA (ראה הגדרה 3.2). כעת נוכל לחשב את התוחלת של המשתנה המקרי בהינתן מאורע A לפי

$$\mathbb{E}(X \mid A) = \sum_{\omega \in \Omega} X(\omega)\mathbb{P}_A(\{\omega\}) = \sum_{s \in \mathrm{Supp}(X)} s\mathbb{P}(X = s \mid A).$$

נוסחא זו מאפשרת לנו להגדיר אנלוג לנוסחת ההסתברות השלמה עבור התוחלת:

טענה 5.12 (נוסחאת התוחלת השלמה). תהי $(A_n)_{n \in \mathbb{N}}$ חלוקה של מרחב הסתברות $(\Omega, \mathcal{F}, \mathbb{P})$. מ"מ X בעל תוחלת סופית מקיים

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{E}(X \mid A_i)\mathbb{P}(A_i).$$

כאשר אנו מפרשים איברים באגף ימין כשווים ל-0 אם $\mathbb{P}(A_i) = 0$.

הוכחה. נרשום לפי הגדרה $\mathbb{E}(X) = \sum_{s \in \mathbb{R}} s\,\mathbb{P}(X = s)$. נפעיל את נוסחת ההסתברות השלמה ונקבל

$$\mathbb{E}(X) = \sum_{s \in \mathrm{Supp}(X)} \sum_{i=1}^{\infty} s\,\mathbb{P}(X = s|A_i)\mathbb{P}(A_i) \overset{\text{התכנסות בהחלט}}{=} \sum_{i=1}^{\infty} \mathbb{P}(A_i) \sum_{s \in \mathrm{Supp}(X)} s\,\mathbb{P}(X = s \mid A_i)$$
$$= \sum_{i=1}^{\infty} \mathbb{E}(X \mid A_i)\mathbb{P}(A_i)$$

■

דוגמא 5.13 (תוחלת תחת התניה).

בקופסה שני מטבעות. הראשון נופל על עץ בהסתברות p_1 השני נופל על עץ בהסתברות p_2. בוחרים באקראי מטבע ומטילים אותו n פעמים. מה תוחלת מספר העצים שהתקבלו?

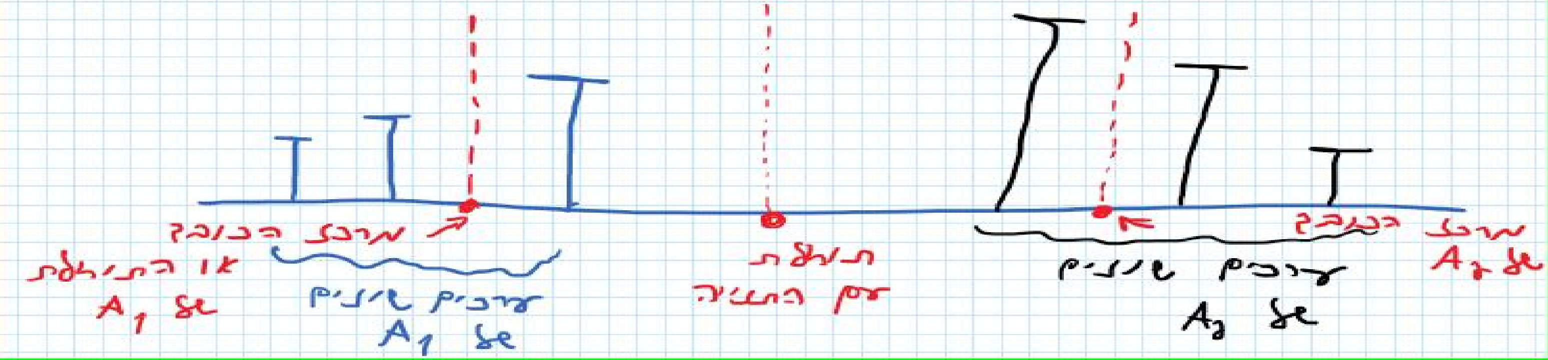
תשובה: בהנתן שנבחר המטבע הראשון התוחלת היא np_1. בהנתן שנבחר המטבע השני התוחלת היא np_2. התוחלת של מספר העצים היא לפיכך $\frac{1}{2}np_1 + \frac{1}{2}np_2 = n\frac{p_1+p_2}{2}$.

5.3 חסם התוחלת (אי-שוויון מרקוב)

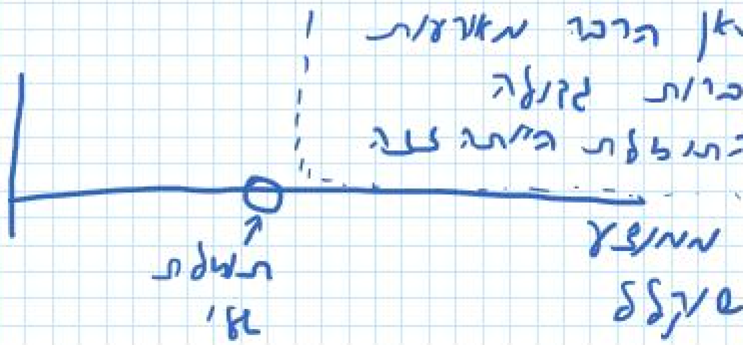
אבסורד שגור הוא: "זו כיתה כה מוכשרת עד שלכל התלמידים בה ציונים מעל הממוצע". ובאמת - אחת מתכונותיו של הממוצע, ואפילו של ממוצע משוקלל היא שהוא תמיד קטן מהערך המירבי. תופעה זו ניתן להכליל עבור מספרים חיוביים ולקבוע שלא יתכן שיותר ממחצית הילדים בכיתה יקבלו ציון שהוא כפליים מהממוצע. באמצעות אבחנה זו נגיע לאי-שוויון הבא שהוצע בשנת 1853 בידי ז'ול בְּיֶנֶמֶה (Ir□n□e-Jules Bienaym□), הוכח בידי פאפנוטי צ'בישב (Pafnuty Chebyshev) ב-1867, ונקרא על שם תלמידו אנדרי מרקוב (Andrei Markov).

ניתן לחשוב על תוחלת כממוצע חשבוני משוקלל במשמעה

ככה זה נראה ויזואלית:



אינטואיציה של א"ש מרקוב:



אין כאן הרבה מאסה
עם הסתברות גבוהה
אחרת התוחלת הייתה גבוה

בדיוק כמו ממוצע
חשבוני משוקלל

הוכחה של א"ש מרקוב

אני מע' התוחלת

$$E(x) = E(x \cdot \mathbb{1}_{\{x \geq a\}}) + E(x \cdot \mathbb{1}_{\{x \leq a\}}) \geq E(x \cdot \mathbb{1}_{\{x \geq a\}}) = a \cdot P(x > a)$$

נחלק את המשוואה ב a וסיימנו.

שימוש ליניאריות בתוחלת:

בעיה: תהיינה $\{A_k\}_{k\in[N]}$ קב׳ כאשר $|A_k|=n$ קב׳ כלשהם לא ידוע מה הם האיברים בקב׳

הוכח ש: $N<2^{n-1}$ אזי קיימת קב׳ B כך שלכל A_i מתקיים כי $B\cap A_i \neq A_i, \phi$

פתרון: נכתוב $A=\bigcup\{A_k\}_{k\in[N]}$ ולכל $a\in A_k$ נגדיר $X_a\sim Ber(1/2)$ ב״ת כלומר האם a בקב׳ A_k

או לא באופן אקראי

נסמן $B=\{a \mid X_a=1\}$ מאורע

$$E_i^0=\{X_a=0 \ \ \forall a\in A_i\}$$ מייצג את כל a שלא נמצא בקב׳

$$E_i^1=\{X_a=1 \ \ \forall a\in A_i\}$$

מייצג את כל a שנמצא בקב׳

נשים לב כי $\mathbb{P}(E_i^0)=\mathbb{P}(E_i^1)=\frac{1}{2^n}$

$$E=E_i^0\cup E_i^1$$

מאורע רע

E מוגדר להיות המאורע הרע או ש a לא נמצא באף קב׳ או ש a נמצא בכל הקבוצות.

נשים לב כי בוודאות מתקיים כי

אינדיקטור ← $\mathbb{1}_E \underset{\text{יש הפרש}}{\le} \underbrace{\mathbb{1}_{E_i^0}+\mathbb{1}_{E_j^1}}$

מחזיר 0,1 אם a שייך לקב׳ כלשהי או a לא שייך לאף קב׳ [רק $max=1$]

אני רץ על כל הקב׳ ושואל האם a נמצא
או לא נשים לב שאני רץ עם אינדקסים שונים
j,i ויכול להיות שנמצאים j,i שמייצגים קב׳ שונות
כך ש a לא נמצא ב E_i / a נמצא ב E_j ואקבל $\geq$

נפעיל תוחלת על המשוואה הנ״ל ואקבל:

$$E(\mathbb{1}_E)\le\left(\sum \mathbb{1}_{E_i^0}+\sum \mathbb{1}_{E_i^1}\right)=N\cdot 2\cdot\frac{1}{2^n}<1$$

מכיוון שה תוחלת של המאורע הרע קטנה מ-1 בוודאות קיימת קב׳ B שעונה על תנאי השאלה ומכאן הוכח

משפט 5.14 (אי-שוויון מרקוב)**.** יהי X מ"מ אי-שלילי. אזי לכל $a > 0$ מתקיים

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

הוכחה. בהנתן X, נגדיר משתנה מקרי חדש $Y = a\mathbb{1}(X \geq a)$ ונשים לב שמתקיים $X \geq Y$. לפי מונוטוניות התוחלת, $\mathbb{E}(X) \geq \mathbb{E}(Y)$. על כן, $\mathbb{E}(X) \geq \mathbb{E}(Y) = a\mathbb{P}(X \geq a)$ ואם נחלק ב-a נקבל את האי-שוויון המבוקש. ∎

במבט ראשון אי-שוויון מרקוב נראה כהערכה גרועה למדי של ההסתברות של המאורע $\{X \geq a\}$, ואולם תכונת הליניאריות של התוחלת מאפשרת לנו לחשב באופן מדוייק גם תוחלת של סכום מ"מ תלויים. מן הבחינה הזו תוחלתם של מ"מ נגישה יותר מאשר פונקציית ההתפלגות שלהם ולכן גם אומדן גס שכזה יכול להיות טוב ופשוט יותר לחישוב מאשר נסיונות ישירים לחסום את הסתברות המאורע. להלן כמה דוגמאות.

דוגמא 5.15 (התוחלת ופרדוקס יום ההולדת)**.**

נשוב ונתאר את תסריט דוגמא 2.13 באופן מופשט. יהיו $\{X_n\}_{n\in[N]}$ מ"מ ב"ת אחידים על $[M]$. נבקש לדעת מה ההסתברות שלפחות $k > 1$ מהמ"מ קיבלו את אותו הערך.

חסם מרקוב

תשובה: נסמן משתנים מציינים $Y_{i_1,\ldots,i_k} = \mathbb{1}(\{X_{i_1} = X_{i_2} = \cdots = X_{i_k}\})$ עבור $i_1, \ldots, i_k \in [N]$ שונים זה מזה. נסמן את אוסף המשתנים הללו ב-$\mathcal{Y}$. נשים לב שאנו מעוניינים לחסום את המאורע $\mathbb{P}(\sum_{Y\in\mathcal{Y}} \geq 1)$. לכל $i_1, \ldots, i_k \in [N]$ המשתנה $Y_{i_1,\ldots,i_k}$ הוא משתנה ברנולי עם סיכוי הצלחה M^{1-k} ולכן, כפי שחישבנו בדוגמא 5.3א, $\mathbb{E}(Y_{i_1,\ldots,i_k}) = M^{1-k}$. נחשב לפי א"ש מרקוב

$$\mathbb{P}\left(\sum_{Y\in\mathcal{Y}} Y \geq 1\right) \overset{\text{מרקוב}}{\leq} \mathbb{E}\left(\sum_{Y\in\mathcal{Y}} Y\right) \overset{\text{טענה 5.4}}{=} \sum_{Y\in\mathcal{Y}} \mathbb{E}(Y) = \binom{N}{k} M^{1-k} \leq \frac{N^k}{M^{k-1}k!}.$$

נשים לב שהחסם שקיבלנו מתלכד עם החסם מדוגמא 2.18 שהושג באמצעות א"ש בול הדבר אינו מקרי כפי שמתגלה בבעיה 9.5 להלן.

דוגמא 5.16 (התוחלת ובעיית האספן)**.**

נבחן תיאור חדש של בעיית האספן (דוגמא 4.32), הפעם מנקודת מבט של אספן האומד את מספר ביצי ההפתעה שעליו לרכוש בטרם יקבל בובה חדשה.

אספן רוכש ביצי הפתעה. כל ביצה מכילה אחד מ-n סוגים של בובות אהובות אשר נבחרים באקראי ובאופן ב"ת. ראשית רוכש האספן ביצה אחת. לאחר מכן, הוא רוכש ביצים עד שיקבל בובה השונה מהבובה שהופיעה בביצה הראשונה. בשלב הבא הוא רוכש ביצים עד שיקבל בובה ששונה משני סוגי הבובות שכבר ברשותו וכן הלאה עד שהוא משיג עותק של כל אחת מהבובות. נחשב את תוחלת מספר הבובות שעליו לרכוש, ונשתמש בא"ש מרקוב להעריך כמה בובות עליו לרכוש כדי שבהסתברות של לפחות 0.5 יהיה לו עותק יחיד מכל בובה.

תשובה: היות שבשלב ה-i, הסיכוי בכל רכישה שהביצה שנקנתה תכיל בובה חדשה הוא $\frac{n-i+1}{n}$ באופן שאינו תלוי בכמות הבובות שכבר רכשנו באותו שלב, הרי שלאור תכונת חוסר הזיכרון – ניתן לתאר את כמות הביצים שנרכשו בשלב ה-i כמ"מ גיאומטרי $X_i \sim \mathrm{Geo}(\frac{n-i+1}{n})$. אנו מתבקשים אפוא לחשב את $\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)$. ניעזר בליניאריות התוחלת ובחשבון מדוגמא 5.3א ונקבל

כשרוכש ה-I הוא משיג
בובה חדשה עבור
$n-i+1$

פתרון פרדוקס יום ההולדת:

נגדיר מ"מ $y_{i_1 \ldots i_k} = \mathbb{1}(x_{i_1} \ldots x_{i_k})$

x מייצג אנשים y מ"מ שהוא אינדיקטור שמקבל איש ומקבל 1 אם האנשים נולדו באותו יום בשנה ו 0

אם אחרת דוג' $y_{1,3,5} = \mathbb{1}(x_1, x_3, x_5)$ האם x_1, x_3, x_5 נולדו באותו יום?

אנו מעוניינים לחסום את ההסתברות של המאורע $P\left(\sum_{y \in} y \geq 1\right)$

סיכוי שk אנשים נולדו באותו יום

נשים לב כי כל $y_{i \ldots k}$ הוא משתנה ברנולי עם סיכוי $\frac{1}{m^k}$ כי $\frac{1}{m}$ זה הסיכוי שמישהו נולד בתאריך מסוים

ו k מייצג את כמות האנשים שמ"מ y בודק בעצמו:

א"ש מרקוב

$$P\left(\sum_{y \in \mathcal{Y}} y \geq 1\right) \leq \frac{E\left(\sum_{y \in \mathcal{Y}} y\right)}{1} = \sum_{y \in \mathcal{Y}} E(y) = \frac{1}{m^{k-1}} \binom{N}{k} \leq \frac{N^k}{m^{k-1} k!}$$

תכונות לינ' של התוחלת

סיכוי שמישהו נולד מתוך m ימים

כל הצירופים האפשריים של ימי ההולדת של האנשים בלא חשיבות לסדר

מכיוון שאני סופר אנשים מ $x_1, \ldots, x_k$

הסיכוי שאדם אחד נולד בתאריך הנוכחי שלו הוא 1

חסם כלשהו מאינפי

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{טענה 5.4ב}}{=} \sum_{i=1}^{n} \mathbb{E}(X_i) \overset{\text{דוגמה 5.3א}}{=} \sum_{i=1}^{n} \frac{n}{n-i+1} \overset{[j=n-i+1]}{=} n \sum_{j=1}^{n} \frac{1}{j} \le n(\log n + 1)$$

ולפי א"ש מרקוב נקבל

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \ge 2n(\log n + 1)\right) \le \frac{1}{2}$$

ולכן די ב-$2n(\log n + 1)$ ביצים בכדי שהסיכוי לא לקבל ביצה אחת מכל סוג יהיה לכל היותר חצי. נשים לב שתוצאה זו מתלכדת עם החסם שקיבלנו בדוגמא 4.32.

✎ **בעיה 5.2.** להוכיח את א"ש בול (משפט 2.17) באמצעות א"ש מרקוב על משתנים מציינים.

✎ **בעיה 5.3.** להוכיח את עקרון ההכלה וההדרה (טענה 2.22) באמצעות תוחלות. הדרכה: יש להגדיר את המשתנה המקרי $X = \prod_{i=1}^{n} (\mathbb{1}(\cup_{k=1}^{n} A_k) - \mathbb{1}(A_i))$ ולהוכיח שהוא שווה תמיד לאפס. כעת ניתן לפתוח את הסוגריים ולחשב את תוחלת X בעזרת לינאריות התוחלת.

כשם שלא יתכן שכל התלמידים קיבלו ציון גבוה מהממוצע, כן לא יתכן שכולם קיבלו ציון נמוך ממנו. למשל, לא יתכן שהתלמיד בעל הציון הגבוה ביותר בכיתה עם ממוצע ציונים של 90 קיבל ציון 75.

טענה 5.17 (אי-שוויון מרקוב הפוך). יהי X מ"מ בעל תוחלת סופית.

$$\mathbb{P}\big(X \ge \mathbb{E}(X)\big) > 0$$

הוכחה למקרה הבדיד. נגדיר משתנה מקרי $Y = \mathbb{E}(X) - X$. לא יתכן ש-$\mathbb{P}(Y \le 0) = 0$ שכן אז $\mathbb{P}(Y > 0) = 1$ ולכן לפי טענה 5.4 נקבל ש-$\mathbb{E}(Y) > 0$ בסתירה ללינאריות התוחלת. ■

נשים לב שבאופן כללי טענה 5.17 לא נותנת לנו כל הערכה להסתברות המאורע $X \ge \mathbb{E}(X)$. למשל, יתכן שכל התלמידים בכיתה קיבלו 80, למעט תלמיד אחד שקיבל 90. במקרה זה, תוחלת הציון של תלמיד מקרי תהיה גבוהה במעט מ-80 ולרוב הגדול של התלמידים יהיה ציון נמוך יותר.

*5.4 תוחלת מותנית כמשתנה מקרי

נוכל לשאול "כיצד משתנה התוחלת לאחר שנחשף מידע על המשתנה המקרי". רעיון מתקדם זה מגולם בסוג חדש של משתנה מקרי המכונה **תוחלת מותנית**. אומנם ניתן להבין מושג זה גם מבלי לעסוק להשתמש במושג החלוקה המתואר בפרק 4.5, אך לקורא מומלץ להעמיק בהבנת מושג זה בטרם יקרא את המשך הפרק, שכן הדבר צפוי להקל על הבנתו.

בפרק הקודם ראינו כי בהנתן משתנים מקריים X ו-Y, ניתן לחשב, עבור כל $x \in \mathrm{Supp}(X)$ את התוחלת של Y בהנתן המאורע $\{X = x\}$. נגדיר איפוא

הגדרה 5.18 (תוחלת מותנית). יהיו X, Y משתנים מקריים המוגדרים על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$. **התוחלת המותנית (conditional expectation)** של Y בהינתן X, שנסמנה $\mathbb{E}(Y \mid X)$ היא המשתנה המקרי Z המוגדר על ידי

$$\mathbb{E}(Y \mid X)(\omega) = Z(\omega) := \begin{cases} \mathbb{E}(Y \mid X = X(\omega)) & \mathbb{P}(\omega) > 0 \\ 0 & \mathbb{P}(\omega) = 0 \end{cases}$$

$E[X|Y]$

אם נחשוב על התוחלת $\mathbb{E}(Y)$ בתור התחזית שלנו לערך של Y, הרי שהתוחלת המותנית, $\mathbb{E}(Y \mid X)$ היא התחזית שלנו לערך של Y בהנתן המידע הגלום ב-X. במונחי פרק 4.5, נוכל לומר כי מאחר שהתוחלת המותנית הוגדרה כפונקציה של X, הרי שלפי טענה 4.55 החלוקה של התוחלת המותנית גסה יותר מהחלוקה של X. מבחינה טכנית – התוחלת המותנית מתאימה לכל $\omega \in X^{-1}(a)$ את הערך הממוצע של Y על פני מחלקה זו.

שימו לב: $E[X|Y]$ הוא מ"מ ולא מספר

כעת נוכל לכתוב את טענה 5.12 בצורה הנאה

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y \mid X)).$$

$E(Y) = \sum_{s \in R} P(X=s) \cdot E(Y|X=s)$

נוס' תוחלת שלמה

דוגמא 5.19 (תוחלת מותנית).

יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי תלויים בעלי התפלגות ברנולי עם פרמטר p ונסמן את סכומם $X = \sum_{i=1}^{n} X_i$. מה התוחלת המותנית של X_1 בהנתן X?

תשובה: נחשב את ההתפלגות המותנית של X_1 בהנתן המאורע $X = k$ עבור $k \in [n]$. לשם כך, עלינו לחשב את ההסתברות של המאורע $X = k$ וגם $X_1 = 0$. למשתנה המקרי $X - X_1$ התפלגות בינומית עם פרמטרים $n - 1$ ו-p והוא בלתי תלוי ב-X_1 ולכן

$$\mathbb{P}(X_1 = 1 \mid X = k) = \frac{\mathbb{P}(X_1 = 1\,,\ X = k)}{\mathbb{P}(X = k)} = \frac{\mathbb{P}(X_1 = 1\,,\ X - X_1 = k - 1)}{\mathbb{P}(X = k)}$$
$$= \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(X - X_1 = k)}{\mathbb{P}(X = k)} = \frac{p\binom{n-1}{k-1}(1-p)^{n-k}p^{k-1}}{\binom{n}{k}(1-p)^{n-k}p^k} = \frac{k}{n}$$

באופן דומה נקבל $\mathbb{P}(X_1 = 0 \mid X = k) = \frac{n-k}{n}$ ולכן $\mathbb{E}(X_1 \mid X = k) = \frac{k}{n}$. נציב את X ונקבל $\mathbb{E}(X_1 \mid X) = \frac{X}{n}$. כעת נוכל לבדוק שאכן מתקיים $\mathbb{E}(\mathbb{E}(X_1 \mid X)) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{np}{n} = p = \mathbb{E}(X_1)$.

✎ **בעיה 5.4.** מה התוחלת המותנית של X בהנתן X_1 בדוגמא 5.19?

✎ **בעיה 5.5.** יהי X משתנה מקרי בעל תוחלת הנתמך על $\mathbb{N}$. יש להראות כי $X \sim \text{Geo}(p)$ עבור $p \in (0,1)$ כלשהו אם ורק אם לכל $s \in \mathbb{N}_0$ מתקיים

$$\mathbb{E}(X \mid X > s) = \mathbb{E}(X) + s.$$

טענה 5.20 (תכונות של התוחלת המותנית). יהיו X, Y, Z משתנים מקריים בדידים, תהי $f : \mathbb{R} \to \mathbb{R}$ ונסמן $W = f(Z)$. התכונות הבאות מתקיימות.

(א) **ליניאריות:** $\mathbb{E}[X + Y \mid Z] = \mathbb{E}[X \mid Z] + \mathbb{E}[Y \mid Z]$.

ע"י עובדות בירושה מ תכונות התוחלת

(ב) **תכונת ההרכבה:** $\mathbb{E}[\mathbb{E}[X \mid Z] \mid W] = \mathbb{E}[X \mid W]$.

(ג) **מונוטוניות:** אם $x \overset{a.s}{\geq} y$ אז $E[x|z] \geq E[y|z]$

(ד) **הוצאת החלק הידוע:** $\mathbb{E}[XW \mid Z] \overset{\text{a.s.}}{=} W\mathbb{E}[X \mid Z]$.

(ה) **השמטת החלק שאינו תלוי:** אם Z ב"ת ב-X אז $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ כמעט תמיד.

סיכוי של הנגזרה

✎ **בעיה 5.6.** להוכיח את טענה 5.20.

למעשה ניתן לראות כי ערכיו של X כלל לא שיחקו תפקיד בהגדרת $\mathbb{E}(Y \mid X)$ ולכן למעשה נוכל לחשוב על התוחלת המותנית במונחים של התניה בחלוקה המושרית על ידי X. צורה אחרת לחשוב על התוחלת המותנית היא בתור המשתנה המקרי המקרב את Y באופן הטוב ביותר מבין כל המשתנים המקריים שמשרים את החלוקה הנתונה על ידי X. נפתח רעיון זה בפרק הפניה.

הסבר

כשם שהתנינו במשתנה מקרי יחיד - כך נוכל גם להתנות בוקטור מקרי.

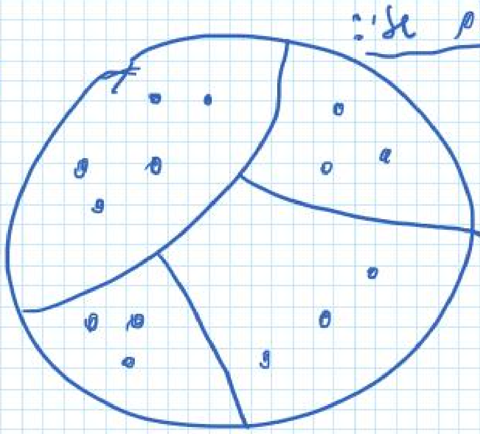
הגדרה 5.21 (תוחלת מותנית). יהיו $X_1, \ldots, X_N, Y$ משתנים מקריים המוגדרים על מרחב הסתברות בדידה $(\Omega, \mathcal{F}, \mathbb{P})$. **התוחלת המותנית** של Y בהינתן $X_1, \ldots, X_N$, שנסמנה $\mathbb{E}(Y \mid X_1, \ldots, X_N)$ היא המשתנה המקרי Z המוגדר על ידי

$$\mathbb{E}(Y \mid X_1, \ldots, X_N)(\omega) = Z(\omega) := \begin{cases} \mathbb{E}(Y \mid X_1 = X_1(\omega), \ldots, X_N = X_N(\omega)) & \mathbb{P}(\omega) > 0 \\ 0 & \mathbb{P}(\omega) = 0 \end{cases}$$

נשים לב שמונחי חלוקות, זו למעשה "התניה בחלוקה שנוצרת מ-$X_1, \ldots, X_n$".

הסבר ל: $E(E(x|z)|w) = E(x|w)$

z מחלק את העולם שלי:



לבצע תוחלת לכל אזור מחלקים בהינתן W

זה בדיוק כמו לעשות תוחלת לכל העולם שלי בהינתן W

הסבר ל הוצאת החלק הידוע: בדיוק כמו בממוצע חשבוני משוכלל

$$\frac{3a_1 + 3a_2 + \ldots + 3a_n}{n} = 3 \cdot \frac{a_1 + \ldots + a_n}{3}$$

הוכחה פורמלית להוצאת החלק הידוע:

$$E[XW|Z](w) = E[XW|Z \overset{*}{=} Z(w)] = E[X\underbrace{f(Z(w))}_{W}\,|\,Z=Z(w)] = W(w)\cdot E[X|Z=Z(w)]$$

$w\in\Omega$

$$= W\cdot E[X|Z](w)$$

*הערה: לא להיבהל ממה שכתוב כאן בסוגר $Z:\Omega\to\mathbb{R}$

$Z(w) =$ מס' כלשהו

סיכום של הוצאת החלק שאינו תלוי:

תכונה אינטואיציה: אם אני רוצה לדעת את התוחלת של הגובה הילדים בארץ אז

$E[$גובה | צבע אהוב של הילד$] = E[$גובה$]$

לא קשור לגובה

אם נסתכל על כלל ההתפלגות המותנית כאשר X,Z משתנים ב"ת אז

$$P(X|Z) = \frac{P(X\cap Z)}{P(Z)} = \frac{P(X)P(Z)}{P(Z)} = P(X)$$

הסבר:

למשל $E[X|Z]$ אני משערך של X אני מידע של X

$E[X_i|Z]$

$x_{i=1}$

x_n

$x_1, \ldots, x_n \in X$

ומחשב את הממוצע שלו בסביבות של $|\Omega)$

מידע על Z

בעיות הרחבה והעשרה

✎ **בעיה 5.7.** עשרה שופטים בתחרות שחיה צורנית מעניקים לקבוצה הקנדית ציונים מקריים המתפלגים אחיד ב-$[10]$. יש לחשב את תוחלת הציון המינימלי ואת תוחלת הציון המקסימלי שקיבלה הקבוצה. (רמז: ניתן להעזר בטענה 5.6)

✎ **בעיה 5.8.** n אנשים מנהלים טורניר של הטלת קוביות. כל זוג אנשים מטילים כל אחד קוביה עד אשר אחד מהם מטיל תוצאה גבוהה יותר. טורניר נקרא טרנזיטיבי אם בכל מקרה שבו שחקן א' ניצח את שחקן ב' ושחקן ב' ניצח את שחקן ג' ניצח שחקן א' את שחקן ג'. יש להשתמש בא"ש מרקוב בכדי לקבל חסם עליון גס להסתברות שהטורניר שתואר יהיה טרנזיטיבי.

✎ **בעיה 5.9.** מניחים שמונה צריחים על לוח שח-מט במקומות אקראיים. יש להוכיח כי ההסתברות שאף שני צריחים לא יאיימו זה על זה היא לכל היותר $1/7$.

✎ **בעיה 5.10.** מעוניינים לסדר n אנשים במעגל. כל אדם רשאי להציג k דרישות לזוגות אנשים שאינו מעוניין להמצא בדיוק ביניהם. יש להראות שכל עוד $k < \binom{n}{2}$ ישנה דרך אפשרית לסדר את האנשים. (רמז: מושיבים את האנשים באקראי)

✎ **בעיה 5.11** (התפלגות היפר-גיאומטרית)**.** נתון שבהטלת n מטבעות התקבלו m תוצאות של עץ. מה תוחלת מספר העצים ב-k ההטלות הראשונות?