תורת ההסתברות 1

יובל קפלן

סיכום הרצאות פרופ׳ יורי קיפר בקורס ״תורת ההסתברות 1״ (80420)
באוניברסיטה העברית, 8–2007.

תוכן מחברת זו הוקלד ונערך על־ידי יובל קפלן. אין המרצה אחראי לכל טעות שנפלה בו. סודר באמצעות $\LaTeX\,2_\varepsilon$ ב־13 באפריל 2008.

עדכונים ותיקונים יופיעו ב־/http://www.limsoup.net. לתגובות, לתיקונים ובכל עניין אחר, אנא כתבו ל־yuvak@gmx.net.

סיכומים נוספים בסדרה:

2006–7	חשבון אינפיניטסימלי 1	אלגברה לינארית 1
	חשבון אינפיניטסימלי 2	אלגברה לינארית 2
	תורת הקבוצות	
2007–8	מבנים אלגבריים 1	תורת ההסתברות 1

תוכן עניינים

1 תורת ההסתברות

1.1 מרחבי הסתברות

21.1.2008 σ-אלגברה

הגדרה. **σ-אלגברה**[1] על Ω היא קבוצה $\mathcal{F}$ של תת־קבוצות של Ω המקיימת (א) $\Omega \in \mathcal{F}$; (ב) אם $A \in \mathcal{F}$ אז $\Omega \setminus A \in \mathcal{F}$; (ג) לכל סדרה $A_1, A_2, \ldots \in \mathcal{F}$ (סופית או אינסופית), $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (או עד n).

מהתכונות נובע כי $\varnothing \in \mathcal{F}$ ו־$\bigcap_i A_i \in \mathcal{F}$.

טענה 1: לכל משפחה $\{\mathcal{F}_\alpha\}$ של σ-אלגברות, $\bigcap_\alpha \mathcal{F}_\alpha$ גם σ-אלגברה.

הגדרה. ה־σ-אלגברה המיוצרת על־ידי משפחה A, או ה־σ-אלגברה המינימלית המכילה את A, של תתי־קבוצות של Ω, היא $\bigcap\{A$ את המכילות σ-אלגי$\}$.[2]

דוגמה. **σ-אלגברת בורל** היא ה־σ-אלגברה המינימלית שמכילה את כל הקבוצות הפתוחות. (מוגדרת בכל מרחב טופולוגי.)

מידת הסתברות

הגדרה. **הסתברות** P (או **מידת הסתברות**) היא פונקציה אי־שלילית על $\mathcal{F}$ σ-אלגברה על Ω[3] כך ש־$P(\Omega) = 1$ ומתקיימת σ-אדיטיביות – אם $A_1, \ldots \in \mathcal{F}$ זרות, כלומר $A_i \cap A_j = \varnothing$ לכל $i \neq j$, אז $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

מרחב הסתברות

הגדרה. **מרחב הסתברות** הוא שלשה $(\Omega, \mathcal{F}, P)$ כאשר Ω היא **מרחב מדגם** – קבוצה של נקודות ω (**מאורעות אלמנטריים**), $\mathcal{F}$ היא σ-אלגברה של תתי־קבוצות מ־Ω (**מאורעות**) ו־P היא הסתברות.

דוגמה. $([0,1], \mathcal{B}, Leb)$ כאשר $\mathcal{B}$ σ-אלגברת בורל על $[0,1]$; להגדרת Leb, אנו מסתמכים על תוצאה מתורת המידה: קיימת הסתברות יחידה Leb על $\mathcal{B}$ כך ש־$Leb[a,b] = b - a$.[4]

מתקיים $Leb(\{x\}) = 0$ ($\{x\} = \bigcap(x - \frac{1}{n}, x + \frac{1}{n})$).

דוגמה. $\Omega = \{(\alpha_0, \alpha_1, \ldots) \mid \alpha_i \in \{0,1\}\}$; $\mathcal{F}$ היא ה־σ-אלגברה המינימלית המכילה את כל הקבוצות הגליליות $C_{a_0,\ldots,a_n}$ לכל n, כאשר מגדירים

$$C_{a_0,\ldots,a_n} = \{\alpha = (\alpha_0, \alpha_1, \ldots) \mid \forall i = 0, 1, \ldots, n \quad \alpha_i = a_i\}$$

תוצאה מתורת המידה: קיימת ויחידה הסתברות P המוגדרת על $\mathcal{F}$ כך שמתקיים $P(C_{a_0,\ldots,a_n}) = \prod_{i=0}^{n} p_{a_i}$ לכל זוג $p_0, p_1 \geq 0$, $p_0 + p_1 = 1$.[5]

בפרט, עבור $p_0 = p_1 = \frac{1}{2}$, אם נסתכל על הסדרות כעל ייצוג בינארי של מספרים בקטע $[0,1]$,[6] P מתלכדת עם מידת לבג.

[1] נקראת גם **σ-שדה**.

[2] ניתן להגדיר כך כי לפחות σ-אלגברה אחת כזו קיימת – זו שמכילה את כל תת־הקבוצות של Ω.

[3] P לא בהכרח מוגדרת על כל תת־הקבוצות של Ω.

[4] ניתן לדבר על הסתברות קטעים סגורים שכן $[a,b] = \bigcap_n (a - \frac{1}{n}, b + \frac{1}{n})$ ולכן ב־$\mathcal{B}$.

[5] כל אחד מהגורמים במכפלה הוא p_0 או p_1, שכן $a_i = 0$ או $a_i = 1$.

[6] אמנם אין התאמה חד־חד ערכית – למשל, $0.0100\ldots = 0.0011\ldots$ – אבל יש רק מספר בן־מניה של מספרים כאלה ($\mathbb{Q}$), אז אין בעיה.

1.2 הסתברות מותנית

הסתברות מותנית

במרחב הסתברות $(\Omega, \mathcal{F}, P)$, עבור $B \in \mathcal{F}$ כך ש־$P(B) > 0$, נגדיר $P(A|B) = \frac{P(A\cap B)}{P(B)}$ – **ההסתברות של A בהינתן B**.[7]

אם נגדיר $P_B(A) = P(A|B)$ ו־$\mathcal{F}_B = \{C \in \mathcal{F} \mid C \subseteq B\} = \{C \cap B \mid C \in \mathcal{F}\}$, נקבל ש־$(B, \mathcal{F}_B, P_B)$ מרחב הסתברות.

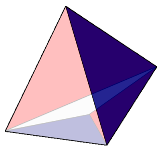
הגדרה. מאורעות A, B **בלתי־תלויים** אם $P(B) = 0$ או אם $P(B) > 0$ ו־$P(A|B) = P(A)$, כלומר $P(A \cap B) = P(A)P(B)$.

נכליל את ההגדרה:

אי־תלות

הגדרה. מאורעות $A_1, \ldots, A_n$ הם **בלתי־תלויים** אם $P(A_1 \cap \ldots \cap A_n) = \prod_{i=1}^n P(A_i)$.
(אינסוף) מאורעות $A_1, A_2, \ldots$ הם בלתי־תלויים אם כל תת־משפחה סופית היא בלתי־תלויה:
$P(A_{i_1} \cap \ldots \cap A_{i_n}) = \prod_{j=1}^n P(A_{i_j})$ ($i_j \neq i_k$).

דוגמה (אי־תלות בזוגות חלשה ממש מאי־תלות). ניקח טטרהדר ונצבע כל אחת משלוש מפאותיו בצבע בודד מבין $\{R, W, B\}$ ואת הבסיס בכל שלושת הצבעים:



בהטלה הוגנת, $P(R) = P(W) = P(B) = \frac{1}{2}$ ו־$P(RW) = P(RB) = P(WB) = \frac{1}{4}$, ולכן המאורעות זרים בזוגות; אבל $P(RWB) = \frac{1}{4}$, ולכן המאורעות אינם זרים.

1.3 נוסחת ההסתברות השלמה

נוסחת ההסתברות השלמה

$A_1, \ldots, A_n, B \in \mathcal{F}$ כאשר A_i זרים, $P(A_i) > 0$ ו־$\bigcup_{i=1}^n A_i = \Omega$.[8] אז

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

1.4 נוסחת בייס

נוסחת בייס

בתנאים כדלעיל ו־$P(B) > 0$:

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

[7] נעיר שלכל $B \subseteq A$ מתקיים $P(A) \leq P(B)$, כי $B = A \cup (B \setminus A)$ ולכן $P(B) = P(A) + P(B \setminus A)$ כאשר $P(B \setminus A) \geq 0$.

[8] למעשה, מספיק לדרוש שמכיל את B.

דוגמה. נתונים n בתי חרושת שמייצרים נורות.[9] אנחנו יודעים כמה נורות כל מפעל מייצר ומה ההסתברות לכך שנורה שמייצר מפעל מסויים תהיה פגומה, ואנו רוצים לדעת מאיזה בית חרושת הגיעה נורה פגומה שקיבלנו (ללא כתובת עליה).

במונחים מתמטיים: נסמן A_i – המאורע שהנורה יוצרה בבית החרושת מספר i, B – המאורע שהנורה פגומה. ידועים לנו $P(B|A_i)$ ו־$P(A_i)$ ואנו מעוניינים ב־$P(A_i|B)$. לשם כך, נוכל להיעזר בנוסחת בייס, שבעצם מאפשרת להפוך את ההתניה.

דוגמה (בעיית ההתרוששות - gambler's ruin). מהמר יכול לקבל בכל משחק 1 ש״ח בהסתברות p ולהפסיד בהסתברות $q = 1 - p$. (למעשה, מדובר בהילוך מקרי פשוט, כאשר עושרו של המהמר מיוצג על־ידי מיקומו על הישר.) המהמר מגיע עם סכום התחלתי k. כאשר מגיע ל־0, יוצא מהמשחק בבושת פנים; כאשר מגיע ל־N, ישר מפסיק לשחק והולך לחנות של הקאדילק.

אנו מעוניינים לחשב את ההסתברות להתרוששות בהינתן סכום התחלתי k. נסמנה p_k.

נסמן ב־B את המאורע בו המהמר התרושש וב־X_n את מצב המהלך המקרי בזמן n.

$$\begin{aligned} p_k &= P(B|X_0 = k) \\ &= P(B|X_0 = k, X_1 = k+1)P(X_1 = k+1|X_0 = k)+ \\ &\quad +P(B|X_0 = k, X_1 = k-1)P(X_1 = k-1|X_0 = k) \\ &= p_{k+1}p + p_{k-1}q \end{aligned}$$

במעבר האחרון הסתמכנו על כך שמתקיימת כאן **תכונת מרקוב**: אין משמעות לידיעת העבר ($X_0 = k$) כאשר יודעים מצב נוכחי ($X_1 = k \pm 1$). נעיר כי תנאי השפה הם $p_0 = 1$, $p_N = 0$. תכונת מרקוב

אנו יודעים כי $p_k = p_k p + p_k(1-p)$ (פירוק הסכום) ו־$p_k = p_{k+1}p + p_{k-1}(1-p)$, ולפיכך, בהנחת $p \neq 0, 1$ (מקרים אלו טריוויאליים), נקבל על־ידי הצבה והעברת אגפים

$$\begin{aligned} 0 &= (p_{k+1} - p_k)p + (p_{k-1} - p_k)(1-p) \\ p_{k+1} - p_k &= \tfrac{1-p}{p}(p_k - p_{k-1}) \end{aligned}$$

ועל־ידי נסיגה, נקבל $p_{k+1} - p_k = (\frac{1-p}{p})^k(p_1 - 1)$.

24.1.2008 על־מנת להתיר את נוסחת הנסיגה, נפריד למקרים. עבור $p \neq \frac{1}{2}$, ניעזר בנוסחת סכום סדרה הנדסית.[10]

$$p_k - p_0 = \sum_{i=0}^{k-1}(p_{i+1} - p_i) = (p_1 - 1)\frac{(\frac{1-p}{p})^k - 1}{(\frac{1-p}{p}) - 1}$$

כלומר $p_k = 1 + (p_1 - 1)\frac{(\frac{1-p}{p})^k - 1}{(\frac{1-p}{p}) - 1}$.

נותר למצוא את p_1. לשם כך, נציב $k = N$:

[9] למעשה, בתי חרושת כאלו די נפוצים; בארץ מייבאים נורות מסין, שמתפוצצות אחרי כמה ימים, אבל במזרח אירופה יש גם מפעלים כאלה.

[10] עבור $p = \frac{1}{2}$, המנה היא $\frac{1-\frac{1}{2}}{\frac{1}{2}} = 1$ והנוסחה איננה מוגדרת.

$$\begin{aligned} 0 &= 1+(p_1-1)\frac{(\frac{1-p}{p})^N-1}{(\frac{1-p}{p})-1} \\ p_1-1 &= -\frac{(\frac{1-p}{p})-1}{(\frac{1-p}{p})^N-1} \end{aligned}$$

ובסך־הכול,

$$p_k = 1-\frac{(\frac{1-p}{p})-1}{(\frac{1-p}{p})^N-1} = \frac{(\frac{1-p}{p})^N-(\frac{1-p}{p})^k}{(\frac{1-p}{p})^N-1}$$

עבור $p=\frac{1}{2}$, נקבל $(1-p)(p_k-p_{k-1}) = p(p_{k+1}-p_k)$ ולכן מתקיים כי $p_{k+1}-p_k = p_k-p_{k-1}$ ואז $p_{k+1}-p_k = p_1-p_0 = p_1-1$. ושוב,

$$p_k-1 = \sum_{i=1}^{k-1}(p_{i+1}-p_i) = k(p_1-1)$$

נציב $k=N$ ונקבל $0=1+N(p_1-1)$ ולכן $p_1-1=-\frac{1}{N}$, ולכן כאשר $p=\frac{1}{2}$,

$$p_k = 1+k(p_1-1) = 1-\frac{k}{N}$$

2 משתנים מקריים

2.1 הגדרה

משתנה מקרי

הגדרה. יהי $(\Omega,\mathcal{F},P)$ מרחב הסתברות. **משתנה מקרי** X זאת העתקה $X:\Omega\to\mathbb{R}$ כך שלכל $a\in\mathbb{R}$, $\{\omega \mid X(\omega)\le a\}\in\mathcal{F}$.

טענה 2: לכל B קבוצת בורל[11] מתקיים $X^{-1}(B)=\{\omega \mid X(\omega)\in B\}\in\mathcal{F}$.[12] (פונקציה המקיימת תכונה זו נקראת **פונקציה מדידה**.)

פונקציית התפלגות

הגדרה. **פונקציית ההתפלגות** של משתנה מקרי X היא

$$F_X(x) = P\{X\le x\} = P\{\omega \mid X(\omega)\le x\}$$

תכונות F_X:[13]

1. $F_X(x)$ מונוטונית עולה;
2. $\lim_{x\to\infty}F_X(x)=1$.

   הוכחה. נובע מ־σ־אדיטיביות: $\bigcup_n\{\omega \mid X(\omega)\le n\}=\Omega$.
3. $\lim_{x\to-\infty}F_X(x)=0$.

   הוכחה. $\bigcap_n\{\omega \mid X(\omega)\le n\}=\varnothing$.
4. F_X רציפה מימין.

[11] B תיקרא **קבוצת בורל** אם היא איבר ב־σ־אלגברת בורל – כאן, על $\mathbb{R}$.
[12] בהגדרה דרשנו זאת עבור קבוצות מהצורה $B=(-\infty,a]$, שהן בפרט קבוצות בורל.
[13] בהוכחות אנו מסתמכים על כך שאם $A_n\nearrow A$, כלומר $A_{n+1}\supseteq A_n$ ו־$\bigcup A_n=A$, או $A_n\searrow A$ מאורעות, אז $\lim_{n\to\infty}P(A_n)=P(A)$ נובע מ־σ־אדיטיביות – בדוק!

הוכחה.

$$\begin{aligned}\lim_{y\to x^+} F_X(y) &= \lim_{n\to\infty} F_X(x+\tfrac{1}{n}) \\ &= P\{\bigcap_{n=1}^{\infty}\{X\le x+\tfrac{1}{n}\}\} \\ &= P\{X\le x\}=F_X(x)\end{aligned}$$

הגדרה. כל פונקציה ממשית F שמקיימת את התכונות לעיל נקראת **פונקציית התפלגות** (בלי התייחסות למשתנה מקרי מסויים).

משפט 3: לכל פונקציית התפלגות F קיים מרחב הסתברות $(\Omega,\mathcal{F},P)$ ומשתנה מקרי X כך ש־F היא פונקציית ההתפלגות של X.

הוכחה. נובע משתי טענות:

למה 1.3: תהי F פונקציית התפלגות. אזי קיימת הסתברות P מוגדרת על σ־אלגברת בורל על $\mathbb{R}$ כך ש־$P((a,b])=F(b)-F(a)$.

הוכחה. בתורת המידה.

למה 2.3: קיים משתנה מקרי על $(\mathbb{R},\mathcal{B},P)$[14] כך ש־$F$ היא פונקציית ההתפלגות שלו.

הוכחה. נגדיר $X(x)=x$. אז $P\{x \mid X(x)\le b\}=P(-\infty,b]=F(b)$ (שהרי $F(-\infty)=0$).

2.2 אי־תלות

28.1.2008 **הגדרה.** $X_1,\ldots,X_n$ משתנים מקריים נקראים **בלתי־תלויים** אם לכל $a_1,\ldots,a_n$ מתקיים $P\{X_1\le a_1,\ldots,X_n\le a_n\}=\prod_{i=1}^n P\{X_i\le a_i\}$. אי־תלות

טענה 4: $X_1,\ldots,X_n$ בלתי־תלויים אס"ם לכל אוסף קבוצות בורל $B_1,\ldots,B_n\subseteq\mathbb{R}$ מתקיים $P\{X_1\in B_1,\ldots,X_n\in B_n\}=\prod_{i=1}^n P\{X_i\in B_i\}$. (כתרגיל.)

הגדרה. נאמר שסדרה של אינסוף משתנים מקריים $X_1,X_2,\ldots$ הם **בלתי־תלויים** אם הם בלתי־תלויים לכל סדרה סופית.

אי־תלות של סדרת משתנים מקריים $X_1,\ldots,X_n$ חזקה ממש מאי־תלות בזוגות.[15]

דוגמה. נבנה משתנים מקריים בלתי־תלויים: יהיו X ו־Y משתנים מקריים עם התפלגות נתונה על $(\Omega,\mathcal{F},P)$. נסתכל על $\Omega\times\Omega$ כעל מרחב המדגם ונגדיר $\tilde{X}(\omega_1,\omega_2)=X(\omega_1)$, $\tilde{Y}(\omega_1,\omega_2)=Y(\omega_2)$. נגדיר $\tilde{P}(A_1\times A_2)=P(A_1)P(A_2)$ ($A_1,A_2\in\mathcal{F}$).[16] נגדיר $\mathcal{F}\times\mathcal{F}$ להיות ה־σ־אלגברה המינימלית המכילה את כל המכפלות $A_1\times A_2$ עבור $A_1,A_2\in\mathcal{F}$.

[14] כאן $\mathcal{B}$ σ־אלגברת בורל, P מהלמה הקודמת.

[15] אפשר להראות זאת על־ידי דוגמת הטטרהדר מקודם, עם הסבה למשתנים מקריים המתפלגים לפי ברנולי (ראה הגדרה להלן).

[16] על־פי תוצאה מתורת המידה, קיימת הסתברות $\tilde{P}$ המוגדרת על $\mathcal{F}\times\mathcal{F}$ המקיימת את התנאי לעיל, והיא יחידה.

$\tilde{X}$ ו־$\tilde{Y}$ בלתי־תלויים:

$$\begin{aligned}\tilde{P}\{\tilde{X} \le a_1, \tilde{Y} \le a_2\} &= \tilde{P}\{\{\omega_1 \mid X(\omega_1) \le a_1\} \times \{\omega_2 \mid Y(\omega_2) \le a_2\}\} \\ &= P\{X \le a_1\}P\{Y \le a_2\} \\ &= \tilde{P}\{\tilde{X} \le a_1\}\tilde{P}\{\tilde{Y} \le a_2\}\end{aligned}$$

משפט 5: תהי $g : \mathbb{R}^n \to \mathbb{R}$ פונקציה רציפה[17] ויהיו $X_1, \ldots, X_n$ משתנים מקריים. אזי $g(X_1, \ldots, X_n)$ הוא משתנה מקרי.

הוכחה. יהי $(\Omega, \mathcal{F}, P)$ מרחב ההסתברות. צריך לבדוק ש־$\{g(X_1, \ldots, X_n) \in B\} \in \mathcal{F}$, כאשר B קבוצה פתוחה.[18] תהי B קבוצה פתוחה. צריך להוכיח

$$\{\omega \mid g(X_1(\omega), \ldots, X_n(\omega)) \in B\} = \{\omega \mid (X_1(\omega), \ldots, X_n(\omega)) \in g^{-1}B\} \in \mathcal{F}$$

$g^{-1}B$ קבוצה פתוחה, לכן מספיק להוכיח שלכל קבוצה פתוחה G, $\{(X_1, \ldots, X_n) \in G\} \in \mathcal{F}$.

נדבר קודם על $G = A = (a_1, b_1) \times \ldots (a_n, b_n)$ קופסה:

$$\{(X_1, \ldots, X_n) \in A\} = \{X_i \in (a_i, b_i) \quad \forall i = 1, \ldots, n\} = \bigcap_{i=1}^{n} \{X_i \in (a_i, b_i)\} \in \mathcal{F}$$

על־פי תוצאה מתורת המידה, כל קבוצה פתוחה היא איחוד בן־מניה של קופסאות. קופסאות נמצאות ב־$\mathcal{F}$, לכן כל איחוד שלהן נמצא ב־$\mathcal{F}$, ובפרט – קבוצות פתוחות נמצאות ב־$\mathcal{F}$.

2.3 התפלגויות בדידות

התפלגות בדידה

הגדרה. נאמר שלמשתנה מקרי X יש **התפלגות בדידה** אם קיימת סדרה של מספרים שונים $x_1, x_2, \ldots$ כך ש־$\sum_{i=1}^{\infty} P\{X = x_i\} = P\{X \in \bigcup_{i=1}^{\infty} x_i\} = 1$. (כלומר, יש מספר בן מנייה של תוצאות אפשריות.)

אפשר לאפיין התפלגות על־ידי פונקציית משקל: $f_X(x_i) = P\{X = x_i\}$ ו־$f_X(x) = 0$ לכל $x \notin \bigcup_{i=1}^{\infty} x_i$; $F_X(x) = P\{X \le x\} = \sum_{i=x_i \le x} f_X(x_i)$.

2.3.1 התפלגות ברנולי

התפלגות ברנולי

משתנה מקרי X הוא בעל **התפלגות ברנולי** אם $f_X(1) = p$, $f_X(0) = q = 1 - p$. כלומר, X מקבל את הערך 1 בהסתברות p ואת הערך 0 בהסתברות $1 - p$.

משתנה מקרי המתפלג לפי ברנולי מתאר הצלחה או כישלון בניסוי בודד.

2.3.2 התפלגות בינומית

התפלגות בינומית

משתנה מקרי X הוא בעל **התפלגות בינומית** אם $f_X(k) = P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}$ עבור $k = 0, 1, \ldots, n$. (ברור שמתקיים $\sum_{k=0}^{n} f_X(k) = 1$.)

[17] למעשה, מספיק לדרוש ש־g בורל, כלומר המקור של כל קבוצת בורל הוא קבוצת בורל. איפיון ברוח הגדרה זו לפונקציה רציפה: המקור של כל קבוצה פתוחה היא קבוצה פתוחה.

[18] זו דרישה מספיקה, כי כל קבוצת בורל ניתן לייצג על־ידי קבוצות פתוחות.

טענה 6: יהיו $X_1,\dots,X_n$ משתנים מקריים בלתי־תלויים ברנולי. אז $X_1+X_2+\dots+X_n$ הוא משתנה מקרי בינומי.

הוכחה. $P\{X_1=\alpha_1,\dots,X_n=\alpha_n\}=\prod_{i=1}^{n}P\{X_i=\alpha_i\}=p^{\sum\alpha_i}q^{n-\sum\alpha_i}$. אנו מעוניינים במקרים בהם $\sum_{i=1}^{n}\alpha_i=k$: כלומר, k α_i־ים מתוך n יקבלו את הערך 1, והשאר – 0. יש $\binom{n}{k}$ דרכים לבחור כך; אז $P\{X_1+\dots+X_n=k\}=\binom{n}{k}p^k(1-p)^{n-k}$.

2.3.3 התפלגות בינומית שלילית

משתנה מקרי X הוא בעל **התפלגות בינומית שלילית** אם $f_X(k)=\binom{n+k-1}{k}(1-p)^kp^n$ עבור $k=0,1,2,\dots$. התפלגות בינומית שלילית

זו התפלגות: נכתוב טור טיילור ל־$(1-q)^{-n}$ ($q=1-p$, $0\le q<1$): הנגזרת ה־k (ב־$q=0$) היא $\frac{n(n+1)\cdots(n+k-1)}{k!}$, אז $p^{-n}=(1-q)^{-n}=\sum_{k=0}^{\infty}\binom{n+k-1}{k}q^k$. נכפול את שני האגפים ב־$p^n$ ונקבל $\sum_{k=0}^{\infty}\binom{n+k-1}{k}q^kp^n=1$. כעת, מכיוון שהסכום הוא 1, נוכל באופן פורמאלי לבנות מכך התפלגות אם נגדיר $F_X(x)=\sum_{x_i\le x}f_X(x_i)$.

משתנה מקרי בעל התפלגות בינומית שלילית מייצג את ההסתברות לכך שלפני n הצלחות (כאשר הצלחה מתקבלת בהסתברות p) יהיו k אי־הצלחות (בהסתברות $1-p$): בסוף יש הצלחה אחת, ובשאר $n+k-1$ המקומות יש להציב k אי־הצלחות ו־$n-1$ הצלחות באופן כלשהו.

2.3.4 התפלגות גיאומטרית

התפלגות גיאומטרית היא מקרה פרטי של התפלגות בינומית שלילית, כאשר $n=1$ (שאז $\sum_{k=0}^{\infty}q^k=(1-q)^{-1}$): $P\{X=k\}=f_X(k)=(1-p)^{k-1}p$ עבור $k=1,2,\dots$: זו ההסתברות ל־$k-1$ אי־הצלחות עד ההצלחה הראשונה. התפלגות גיאומטרית

נגדיר Y להיות מספר אי ההצלחות עד להצלחה מספר n. אפשר לייצג $Y=X_1+\dots+X_n$, כאשר $\{X_i\}$ משתנים מקריים בלתי־תלויים שווי־התפלגות המפולגים כמספר אי־ההצלחות עד להצלחה הראשונה (דהיינו, גיאומטרית). Y הוא בעל התפלגות בינומית שלילית.

2.3.5 התפלגות פואסון

משתנה מקרי X הוא בעל **התפלגות פואסון** אם $P\{X=k\}=f_X(k)=e^{-\lambda}\frac{\lambda^k}{k!}$. ברור שהסכום הוא 1, שכן $\sum\frac{\lambda^k}{k!}=e^{\lambda}$. התפלגות פואסון

4.2.2008 **משפט 7 (פואסון):** יהיו X_n משתנים מקריים בינומיים, $P\{X_n=k\}=\binom{n}{k}p_n^k(1-p_n)^{n-k}$ כאשר $np_n\overset{n\to\infty}{\to}\lambda$. אז $P\{X_n=k\}\overset{n\to\infty}{\to}\frac{\lambda^k}{k!}e^{-\lambda}$.

הוכחה. בתרגול.

אם λ שיחות טלפון מגיעות בקטע $[0,t]$, ההסתברות לשיחה בקטע $[\frac{k}{n}t,\frac{(k+1)}{n}t]$ היא בערך $\lambda\frac{t}{n}$. שיחות שמגיעות בקטעים זרים הן בלתי־תלויות. ההסתברות שמגיעה יותר משיחה אחת

בקטע $[\frac{k}{n}t, \frac{k+1}{n}t]$ היא $o(\frac{t}{n})$. ההסתברות ל־k שיחות בקטע $[0,t]$ היא $P\{X_1+\ldots+X_n=k\}$ כאשר $X_i=1$ כשיש שיחה בקטע i (בהסתברות $\lambda\frac{t}{n}$), $X_i=0$ אחרת – כלומר, X_i בעלי התפלגות ברנולי, ולכן $X_1+\ldots+X_n$ משתנה מקרי בינומי. אז

$$P\{X_1+\ldots+X_n=k\}=\binom{n}{k}(\lambda\frac{t}{n})^k(1-\lambda\frac{t}{n})^{n-k}\to\frac{(\lambda t)^k}{k!}e^{-\lambda t}$$

(כאשר $p_n=\lambda\frac{t}{n}$, $np_n\to\lambda t$; למעשה, $p_n=\lambda\frac{t}{n}+o(\frac{t}{n})$, ואז $np_n\to\lambda t$, אבל זה מספיק כדי שיתקיים המשפט.)

2.4 התפלגויות רציפות

התפלגות רציפה

הגדרה. נאמר שלמשתנה מקרי X יש **התפלגות רציפה** אם קיימת פונקציה f אי־שלילית עם לא יותר ממספר סופי של נקודות אי־רציפות כך שמתקיים $F_X(x)=P\{X\le x\}=\int_{-\infty}^x f(y)dy$.

צפיפות ההתפלגות

f נקראת **צפיפות ההתפלגות**.

2.4.1 התפלגות אחידה

התפלגות אחידה

ההתפלגות האחידה על קטע $[a,b]$, $b>a$ היא

$$f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0 & x\notin[a,b]\end{cases}$$

כשאומרים ‘‘ניקח נקודה באקראי בקטע $[a,b]$’’ מתכוונים לבחירת נקודה עם התפלגות אחידה.

2.4.2 התפלגות נורמלית (Gauss)

התפלגות נורמלית

a, σ פרמטרים; **ההתפלגות הנורמלית** היא $f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. עבור $a=0$, $\sigma=1$, **התפלגות נורמלית סטנדרטית.**

נחליף משתנה: $y=\frac{x-a}{\sigma}\mapsfrom x$. אז $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{y^2}{2}}dy\overset{!}{=}1$.

כדי להוכיח זאת, נחשב

$$(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\frac{x^2}{2}}dx)(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\frac{x^2}{2}}dx)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{\frac{x^2+y^2}{2}}dxdy$$

באמצעות מעבר לקואורדינטות פולאריות ($x=r\cos\varphi$, $y=r\sin\varphi$, ואז $x^2+y^2=r^2$).

2.4.3 התפלגות קושי

התפלגות קושי

התפלגות קושי היא $f(x)=\frac{1}{\pi(1+x^2)}$. מתקיים, כנדרש,

$$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dx}{1+x^2}=\frac{1}{\pi}\arctan x\mid_{-\infty}^{\infty}=\frac{1}{\pi}(\frac{\pi}{2}+\frac{\pi}{2})=1$$

2.4.4 התפלגות מעריכית

התפלגות מעריכית

ההתפלגות המעריכית היא
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$
עם פרמטר $\lambda > 0$. מתקבל $\int_0^\infty f(x)dx = 1$.

ההתפלגות מייצגת את הזמן T עד לשיחת הטלפון הראשונה: ראינו קודם שההסתברות ל־k שיחות טלפון בקטע $[0,t]$ היא $P\{X = k\} = \frac{(\lambda t)^k}{k!}e^{-\lambda t}$, ועבור $k = 0$ (כלומר, אין שיחות בקטע $[0,t]$) נקבל $P\{X = 0\} = e^{-\lambda t}$; ואילו כאן, $P\{T \leq t\} = \int_0^t \lambda e^{-\lambda x}dx = 1 - e^{-\lambda t}$.

2.4.5 התפלגות Γ

התפלגות Γ

התפלגות Γ היא
$$f(x) = \begin{cases} \lambda^{\alpha-1}x^{\alpha}e^{-\lambda x}\frac{1}{\Gamma(\alpha)} & x \geq 0 \\ 0 & x < 0\} \end{cases}$$
כאשר $\Gamma(\alpha) = \int_0^\infty \lambda^\alpha x^{\alpha-1}e^{-\lambda x}dx = \int_0^\infty y^{\alpha-1}e^{-y}dy$ (עם הצבה $y = \lambda x$). (מתקיים $\Gamma(n) = (n-1)!$).

כאשר $\alpha = 1$, זו ההתפלגות האקספוננציאלית.

2.4.6 התפלגות χ^2

התפלגות χ^2

התפלגות χ^2 היא ההתפלגות של $X_1^2 + \ldots + X_n^2$, כאשר X_i בלתי־תלויים עם ההתפלגות הנורמלית הסטנדרטית. זה מקרה פרטי של התפלגות Γ; נראה זאת בהמשך.

2.5 וקטורים מקריים

וקטור מקרי

הגדרה. $(X_1, X_2, \ldots, X_n)$ הוא **וקטור מקרי**, בעל **התפלגות n־מימדית**:
$$F_{(X_1,\ldots,X_n)}(x_1,\ldots,x_n) = P\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$$

במקרה של התפלגות בדידה, $\sum_i f_X(\bar{a}_i) = 1$, $\bar{a}_i \in \mathbb{R}^n$.

דוגמה. במקרה הדו־מימדי, התפלגות אחידה בעיגול או במלבן; למשל, מלבן עם שטח S. אפשר להגדיר $f(x_1,x_2) = \frac{1}{S}$ כאשר $(x_1,x_2) \in R$, $f(x_1,x_2) = 0$ אחרת.

דוגמה. בהינתן B ו־C משתנים מקריים בלתי־תלויים עם התפלגות אחידה על $[0,1]$, חשבו את $P\{x^2 + 2Bx + C > 0 \quad \forall x\}$.

2.6 פרדוקס Bertrand

בעיה: בהינתן מעגל מרדיוס 1, נבחר מיתר באקראי ונסמן את אורכו ב־X. $P\{X > \sqrt{3}\} = ?$

דרך 1: נבחר את מרכז המיתר באקראי; זה מגדיר את המיתר באופן יחיד (עד־כדי סיבוב המעגל). על־פי חישוב טריגונומטרי, נקבל שאם המרכז נמצא במעגל מרדיוס $\frac{1}{2}$, אורכו גדול מ־ או שווה ל־$\sqrt{3}$; לכן נקבל $P\{X > \sqrt{3}\} = \frac{(\frac{1}{2})^2\pi}{1^2\pi} = \frac{1}{4}$.

דרך 2: תהי Z נקודה כלשהי על המעגל; נבחר את קצהו השני של המיתר באופן אקראי. על־פי חישוב טריגונומטרי, נקבל שאורך הקשת שבחירת נקודות ממנה תביא למיתר שאורכו כנדרש הוא שליש מהיקף המעגל, ולכן $P\{X > \sqrt{3}\} = \frac{1}{3}$.

מסקנה: יש חשיבות לאופן הגדרת משתנים מקריים בחישוב ההסתברות.

3 תוחלת, שונות ותיאום

3.1 תוחלת

תוחלת

הגדרה. תוחלת (תוחלת מתמטית; expectation) מוגדרת בנפרד עבור משתנים מקריים בדידים ורציפים:[19]

יהי X משתנה מקרי עם התפלגות בדידה ופונקציית משקל $f_X(x)$ כך ש־$\sum_{i=1}^{\infty} f_X(x_i) = 1$, x_i נקודות שונות. נניח ש־$\sum f_X(x_i)|x_i| < \infty$. התוחלת של X: $\mathbb{E}X = \sum_{i=1}^{\infty} f_X(x_i)x_i$.

יהי X עם התפלגות רציפה עם פונקציית צפיפות $f_X(x)$ כך ש־$\int_{-\infty}^{\infty} f_X(x)|x|dx < \infty$. אז $\mathbb{E}X = \int_{-\infty}^{\infty} f_X(x)xdx$.

בשני המקרים, אנו דורשים התכנסות בהחלט[20] בכדי שבמקרה הרציף לא נהיה תלויים בסדר הסכימה, ובמקרה הרציף, כדי שלא נהיה תלויים באופן לקיחת הגבולות בחישוב האינטגרל הלא־אמיתי.

משפט 8: יהיו X ו־Y משתנים מקריים כך ש־$\mathbb{E}X$, $\mathbb{E}Y$ קיימים. אז $\mathbb{E}(X+Y)$ קיימת ו־$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$, ואם λ קבוע, $\mathbb{E}(\lambda X)$ קיימת ו־$\mathbb{E}(\lambda X) = \lambda \mathbb{E}X$. 7.2.2008

הוכחה (משתנים מקריים בדידים). נניח ש־X מקבל ערכים $x_1, x_2, \ldots$ ו־Y מקבל ערכים $y_1, y_2, \ldots$. אז $X+Y$ מקבל ערכים $\{x_i + y_j\}$. z_i הן נקודות שונות בקבוצה $\{x_i + y_j \quad \forall i,j\}$. $\mathbb{E}(X+Y)$ קיימת, כי

[19] לו ידענו תורת המידה, יכולנו לאחד את ההגדרות.
[20] אין צורך להכניס את $f_X(x_i)$ לערך מוחלט, כיוון שזו פונקציה אי־שלילית.

$$\begin{aligned}\sum_{i=1}^{\infty}|z_i|P\{X+Y=z_i\} &= \sum_{k=1}^{\infty}\sum_{i,j\,:\,x_i+y_j=z_k}|x_i+y_j|P\{X=x_i,Y=y_j\}\\ &\le \sum_{k=1}^{\infty}\sum_{i,j\,:\,x_i+y_j=z_k}(|x_i|+|y_j|)P\{X=x_i,Y=y_j\}\\ &= \sum_{i,j}(|x_i|+|y_j|)P\{X=x_i,Y=y_j\}\\ &= \sum_{i,j}|x_i|P\{X=x_i,Y=y_j\}+\\ &\quad+\sum_{i,j}|y_j|P\{X=x_i,Y=y_j\}\\ &= \sum_i|x_i|P\{X=x_i\}+\sum_j|y_j|P\{Y=y_j\}\\ &= \mathbb{E}|X|+\mathbb{E}|Y|<\infty\end{aligned}$$

כי $\sum_j P\{X=x_i,Y=y_j\}=P\{X=x_i\}$, וזאת כי $\{Y=y_j\}$ זרים ולכן $\bigcup_j\{X=x_i\}\cap\{Y=y_j\}=\{X=x_i\}$.

כעת, כשאנו יודעים שיש התכנסות בהחלט, ניתן "למחוק" את סימני הערך המוחלט ויתקבל שוויון.

הוכחת ההומוגניות – כתרגיל (גם למקרה של משתנים מקריים רציפים.)

טענה 9: יהי X משתנה מקרי ו־g פונקציית בורל[21] כך ש־$\mathbb{E}g(X)$ קיים.[22] אז (1) אם X משתנה מקרי בדיד אז $\mathbb{E}g(X)=\sum_{i=1}^{\infty}g(x_i)f_X(x_i)$, כאשר $\{x_i\}$ הם הערכים השונים ש־X מקבל; (2) אם X משתנה מקרי רציף (בעל התפלגות רציפה), $\mathbb{E}g(X)=\int_{-\infty}^{\infty}g(x)f_X(x)dx$.

הוכחה. (1) יהיו z_k הערכים השונים בקבוצה $\{g(x_i)\}$. אז

$$\begin{aligned}\mathbb{E}g(X) &= \sum_{k=1}^{\infty}z_kP\{g(X)=z_k\}\\ &= \sum_{k=1}^{\infty}\sum_{i\,:\,g(x_i)=z_k}g(x_i)P\{X=x_i\}\\ &= \sum_{x_i}g(x_i)P\{X=x_i\}=\sum_{x_i}g(x_i)f_X(x_i)\end{aligned}$$

(2) קצת סיפור להוכיח, אז לא נעשה זאת.

טענה 10: יהיו X ו־Y משתנים מקריים בלתי תלויים כך ש־$\mathbb{E}X$ ו־$\mathbb{E}Y$ קיימים. אזי $\mathbb{E}XY$ קיימת ו־$\mathbb{E}XY=\mathbb{E}X\mathbb{E}Y$.

הוכחה (התפלגות בדידה). X מקבל ערכים $\{x_i\}$ שונים, Y מקבל ערכים $\{y_j\}$ שונים; z_k האיברים השונים בקבוצה $\{x_iy_j\}$.

$$\begin{aligned}\mathbb{E}XY &= \sum_{k=1}^{\infty}z_kP\{XY=z_k\}\\ &= \sum_{k=1}^{\infty}\sum_{i,j\,:\,x_iy_j=z_k}x_iy_jP\{X=x_i,Y=y_j\}\\ &= \sum_{k=1}\sum_{i,j\,:\,x_i,y_j=z_k}x_iy_jP\{x=x_i\}P\{Y=y_j\}\\ &= \sum_{i,j}x_iy_jP\{X=x_i\}P\{Y=y_j\}\\ &= (\sum_i x_iP\{X=x_i\})(\sum_j y_jP\{Y=y_j\})=\mathbb{E}X\mathbb{E}Y\end{aligned}$$

[21] כזכור, g נקראת פונקציית בורל אם לכל קבוצה A בורל $g^{-1}(A)$ היא קבוצת בורל.

[22] $g(X)$ משתנה מקרי, שכן $\{g(X(\omega))<x\}\in\mathcal{F}$ שקול ל־$\{X(\omega)\in g^{-1}(-\infty,x)\}\in\mathcal{F}$, כלומר $X^{-1}(g^{-1}(-\infty,x))\in\mathcal{F}$.

3.1.1 תוחלות של התפלגויות בדידות

ברנולי

$X=1$ בהסתברות p, $X=0$ אחרת; $\mathbb{E}X=p$.

בינומי

$\mathbb{E}(X_1+\ldots+X_n)=np$.

גיאומטרי

$$\begin{aligned}\sum_{k=1}^{\infty}kp(1-p)^{k-1} &= p\sum_{k=1}^{\infty}(-(1-p)^k)' \\ &= -p(\sum_{k=1}^{\infty}(1-p)^k)' \\ &= -p(\tfrac{1-p}{p})'=\tfrac{1}{p}\end{aligned}$$

מציאת $\mathbb{E}X$ כאשר X פואסון ($\mathbb{E}X=\lambda$) או בינומי שלילי – כתרגיל.

3.1.2 תוחלות של התפלגויות רציפות

אקספוננציאלי

$\int_{-\infty}^{\infty}x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}$

נורמלי (a,σ)

$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}xe^{\frac{-(x-a)^2}{2\sigma}}dx=a$.

Γ

בהתפלגות Γ יש פונקציית צפיפות $f_X(x)=\frac{1}{\Gamma(\alpha)}x^{\alpha-1}\lambda^{\alpha}e^{-\lambda x}$. 11.2.2008

הצפיפות כש־x שלילי היא 0, לכן נוכל לחשב את האינטגרל רק מ־0.

$$\begin{aligned}\int_0^{\infty}xf_X(x)dx &= \frac{1}{\lambda\Gamma(\alpha)}\int_0^{\infty}x^{\alpha}\lambda^{\alpha+1}e^{-\lambda x}dx \\ &= \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)}\cdot\frac{1}{\Gamma(\alpha+1)}\int_0^{\infty}x^{\alpha}\lambda^{\alpha+1}e^{-\lambda x}dx \\ &= \frac{\alpha}{\lambda}\end{aligned}$$

שכן $\frac{1}{\Gamma(\alpha+1)}\int_0^{\infty}x^{\alpha}\lambda^{\alpha+1}e^{-\lambda x}dx=1$ – התפלגות Γ עם פרמטר $\alpha+1$, ו־$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)$.

קושי

$f_X(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$ התוחלת לא קיימת:

$\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|x|}{1+x^2}dx=\infty$, כי $\int_0^A\frac{x}{1+x^2}dx=\frac{1}{2}\int_0^A\frac{d(x^2+1)}{x^2+1}=\frac{1}{2}\ln(A^2+1)\overset{A\to\infty}{\to}\infty$

ולפי ההגדרה, תוחלת קיימת אם $\mathbb{E}|X|<\infty$.

נורמלית

פונקציית הצפיפות היא $f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}$. נחשב:

$$\begin{aligned}\mathbb{E}X &= \tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-\frac{(x-a)^2}{2\sigma^2}}dx \\ &= \tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)e^{-\frac{(x-a)^2}{2\sigma^2}}dx+ \\ &\quad +a\tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-a)^2}{2\sigma^2}}dx\end{aligned}$$

אנו יודעים כי $\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=1$; נחליף משתנה ונקבל

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}ye^{-\frac{y^2}{2\sigma^2}}dy=0$$

אך $\int_{-\infty}^{0}-=\int_0^{\infty}$, לכן בסך־הכול $\mathbb{E}X=a$.

3.2 שונות

הגדרה. **השונות** של משתנה מקרי היא $\operatorname{Var}X=\mathbb{E}(X-\mathbb{E}X)^2$. שונות

דוגמה (שונות ההתפלגות הנורמלית). ניעזר בעובדה שאם $g(x)=(x-a)^2$, מתקיים

$$\mathbb{E}g(X)=\begin{cases}\sum_i g(x_i)f_X(x_i) & X \text{ בדיד} \\ \int_{-\infty}^{\infty}g(x)f_X(x)dx & X \text{ רציף}\end{cases}$$

ונחשב:

$$\begin{aligned}\tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-a)^2e^{-\frac{(x-a)^2}{2\sigma^2}}dx &= \tfrac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}y^2e^{-\frac{y^2}{2\sigma^2}}dy \\ &= \tfrac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\tfrac{y}{\sigma})^2e^{-\frac12(\frac{y}{\sigma^2})^2}d(\tfrac{y}{\sigma} \\ &= \sigma^2\tfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2e^{\frac{z^2}{2}}dz\end{aligned}$$

נבצע אינטגרציה בחלקים:

$$\tfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}z^2e^{-\frac{z^2}{2}}dz=\begin{vmatrix}z & ze^{-\frac{z^2}{2}}dz \\ 1 & -e^{-\frac{z^2}{2}}\end{vmatrix}=\tfrac{1}{\sqrt{2\pi}}\int e^{-\frac{z^2}{2}}dz=1.$$

בסך־הכול, $\mathbb{E}(X-a)^2=\sigma^2$.

הגדרה. $\sigma=\sqrt{\operatorname{Var}X}$ היא **סטיית התקן** (standard variation).[23]

3.3 תיאום

הגדרה. יהיו X ו־Y משתנים מקריים. $\operatorname{cov}(X,Y)=\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$ הוא **התיאום** (covariance). אם $\operatorname{cov}(X,Y)=0$, אומרים ש־$X$ ו־Y **לא־מתואמים זה לזה**. תיאום

באופן כללי, $|ab|\le\frac12(a^2+b^2)$, לכן אם $\operatorname{Var}X$, $\operatorname{Var}Y$ קיימים גם $\operatorname{cov}(X,Y)$ קיימת:

$$|(X-\mathbb{E}X)(Y-\mathbb{E}Y)|\le\frac12((X-\mathbb{E}X)^2+(Y-\mathbb{E}Y)^2)$$

[23] לא כל־כך משתשים בזה במתמטיקה.

אם X ו־Y בלתי־תלויים, אז

$$\begin{aligned}\operatorname{cov}(X,Y) &= \mathbb{E}((X-\mathbb{E}X)(Y-\mathbb{E}Y))\\ &= \mathbb{E}(XY - X\mathbb{E}Y - Y\mathbb{E}X + \mathbb{E}X\mathbb{E}Y)\\ &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y - \mathbb{E}X\mathbb{E}Y + \mathbb{E}X\mathbb{E}Y\\ &= \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y = 0\end{aligned}$$

(כמובן, נוסחה זו – $\operatorname{cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X\mathbb{E}Y$ – נכונה באופן כללי.)

מסקנה 11: משתנים מקריים בלתי־תלויים הם לא־מתואמים.

שאלה: האם משתנים מקריים לא־מתואמים הם תמיד בלתי־תלויים?

דוגמה נגדית: X ו־Y ברנולי בלתי־תלויים שווי־התפלגות כך ש־$P\{X=1\} = P\{Y=1\} = \frac{1}{2}$. נגדיר $U = X+Y$, $V = |X-Y|$. כדי לחשב את $\operatorname{cov} UV = \mathbb{E}UV - (\mathbb{E}U)(\mathbb{E}V)$, נמצא $\mathbb{E}U = \frac{1}{4}\cdot 2 + \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 0 = 1$, $\mathbb{E}V = \frac{1}{2}\cdot 1 = \frac{1}{2}$, $\mathbb{E}UV = \frac{1}{2}\cdot 1 = \frac{1}{2}$: אז $\operatorname{cov} UV = 0$, כלומר U ו־V לא מתואמים. עם זאת, $P\{U=0, V=0\} = \frac{1}{4}$ שונה מ־$P\{U=0\}P\{V=0\} = \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{8}$, ולכן הם לא בלתי־תלויים.

למה 12: $X_1,\dots,X_n$ משתנים מקריים לא־מתואמים. אז $\operatorname{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \operatorname{Var} X_i$. 14.2.2008

הוכחה. נסמן $\mu_i = \mathbb{E}X_i$.

$$\begin{aligned}\operatorname{Var}(X_1+\ldots+X_n) &= \mathbb{E}(X_1+\ldots+X_n - (\mu_1+\ldots+\mu_n))^2\\ &= \mathbb{E}(\textstyle\sum_{i=1}^n (X_i-\mu_i))^2\\ &= \mathbb{E}(\textstyle\sum_{i=1}^n (X_i-\mu_i)^2 + \sum_{i\neq j}(X_i-\mu_i)(X_j-\mu_j))\\ &= \textstyle\sum_{i=1}^n \operatorname{Var} X_i + \sum_{i\neq j}\operatorname{cov}(X_i,X_j)\end{aligned}$$

אך $\operatorname{cov}(X_i,X_j) = 0$, ומתקבל הדרוש.

4 אי־שוויונים

משפט 13 (אי־שוויון בסיסי): תהי $h \ge 0$ פונקציה אי־שלילית בורל (מספיק רציפה) ו־X משתנה מקרי כך ש־$\mathbb{E}h(X)$ קיים. אז לכל $a > 0$, $P\{h(X) \ge a\} \le \frac{\mathbb{E}h(X)}{a}$. 11.2.2008

הוכחה. נגדיר מאורע $A = \{h(x) \ge a\}$ ופונקציה מציינת $\mathbf{1}_A$.[24] ברור ש־$\mathbb{E}\mathbf{1}_A = P(A)$.

אפשר לכתוב $h(X) - a\mathbf{1}_A \ge 0$ תמיד, אז $\mathbb{E}(h(X) - a\mathbf{1}_A) \ge 0$ (שהרי זוהי תכונה נוספת של התוחלת: לכל משתנה מקרי $X \ge 0$, $\mathbb{E}X \ge 0$ אם קיימת, וזה ברור כי $\mathbb{E}X = \sum x_i f_X(x_i)$ כאשר $x_i \ge 0$, ובאופן דומה במקרה הרציף). כלומר, $\mathbb{E}h(X) - a\mathbb{E}\mathbf{1}_A \ge 0$, ומכאן מתקבל האי־שוויון $\mathbb{E}h(X) \ge aP(A)$.

מאי־שוויון זה נובעים אי־השוויונים הבאים:

[24] indicator; לעתים גם מסומן χ_A. מתקיים $\mathbf{1}_A(\omega) = 1$ עבור $\omega \in A$, $\mathbf{1}_A(\omega) = 0$ אחרת.

משפט 14 (אי־שוויון מרקוב): $P\{|X|\geq a\}\leq\frac{\mathbb{E}|X|}{a}$

הוכחה. ניקח $h(X)=|X|$.

משפט 15 (אי־שוויון צ׳בישב): $P\{|X-\mathbb{E}X|\geq a\}\leq\frac{\operatorname{Var}X}{a^2}$

הוכחה. עבור $h(X)=X^2$ נקבל $P\{|X|\geq a\}=P\{X^2\geq a^2\}\leq\frac{\mathbb{E}X^2}{a^2}$. במקרה ש־$h(X)=|X-\mathbb{E}X|^2$, נקבל את הנדרש.

דוגמה (הערכה). יהי X משתנה מקרי כך ש־$\mathbb{E}X=33$, $\operatorname{Var}X=16$. נרצה להעריך את $P\{23\leq X\leq 43\}$:

$$\begin{aligned}P\{23\leq X\leq 43\} &= P\{-10\leq X-33\leq 10\}\\ &= P\{|X-33|\leq 10\}\\ &= 1-P\{|X-33|>10\}\\ &\geq 1-\tfrac{16}{10^2}\end{aligned}$$

לפי אי־שוויון צ׳בישב.

משפט 16 (אי־שוויון קושי־שוורץ): X ו־Y משתנים מקריים כך ש־$\mathbb{E}X^2$ ו־$\mathbb{E}Y^2$ קיימים. אז $\mathbb{E}|XY|\leq(\mathbb{E}X^2\mathbb{E}Y^2)^{\frac{1}{2}}$.

הוכחה. $(X-yY)^2\geq 0$ (y קבוע). אז $0\leq\mathbb{E}(X-yY)^2=\mathbb{E}X^2-2y\mathbb{E}XY+y^2\mathbb{E}Y^2$. זהו פולינום ריבועי ב־$y$. פולינום ריבועי זה אי־שלילי לכל y אם״ם $\mathbb{E}X^2\mathbb{E}Y^2-(\mathbb{E}XY)^2\geq 0$.[25]

5 התכנסות משתנים מקריים

14.2.2008 **הגדרה.** יהיו X ו־$X_1,X_2,\ldots$ משתנים מקריים. נאמר ש־$X_n\xrightarrow{p}X$ (**מתכנס בהסתברות**) אם לכל $\varepsilon>0$, $P\{|X_n-X|>\varepsilon\}\overset{n\to\infty}{\longrightarrow}0$.

הגדרה. יהיו X ו־$X_1,X_2,\ldots$ משתנים מקריים. נאמר ש־$X_n\to X$ **כמעט תמיד/כמעט בכל מקום** אם קיים מאורע N כך ש־$P(N)=0$ ו־$X_n(\omega)\overset{n\to\infty}{\longrightarrow}X(\omega)$ לכל $\omega\in\Omega\setminus N$.

הגדרה. אם $X_1,X_2,\ldots$ משתנים מקריים עם פונקציות התפלגות $F_{X_1},F_{X_2},\ldots$ ו־F_X פונקציית התפלגות של משתנה מקרי X כך ש־$F_{X_n}\to F_X$ בכל נקודת רציפות של F_X, אז X_n **מתכנס ל־X בהתפלגות** ($X_n\xrightarrow{d}X$ או $X_n\Rightarrow X$). התכנסות בהתפלגות

משפט 17: $X_n\to X$ כמעט תמיד אם״ם $\lim_{n\to\infty}P\{\bigcup_{k\geq n}\{|X_k-X|>\varepsilon\}\}=0$ לכל $\varepsilon>0$. בפרט, התכנסות כמעט תמיד יותר חזקה (או לא יותר חלשה) מהתכנסות בהסתברות. כלומר, התכנסות כמעט תמיד גוררת התכנסות בהסתברות.[26] כ״ת $\Longleftarrow$ בהסתברות

[25] אנו מחשבים מתי $b^2-4ac\leq 0$ עבור פולינום ay^2+by+c שעבורו $a=\mathbb{E}Y^2$, $b=-2\mathbb{E}XY$ ו־$C=1$.

[26] זה נכון כי $P\{\bigcup_{k\geq n}\{|X_k-X|>\varepsilon\}\geq P\{|X_n-X|>\varepsilon\}$.

הוכחה. $X_n \to X$ כמעט תמיד, לכן קיימת קבוצה $N \in \mathcal{F}$ עם $P(N) = 0$ כך שלכל $\varepsilon > 0$ ו־$\omega \in \Omega \setminus N$ קיים $n_\varepsilon(\omega)$ כך ש־$\sup_{k \geq n_\varepsilon(\omega)} |X_k(\omega) - X(\omega)| \leq \varepsilon$. לפיכך $P\{\bigcup_{n=1}^\infty \bigcap_{k\geq n}\{|X_k(\omega) - X(\omega)| \leq \varepsilon\}\} = 1$. נתבונן במאורע המשלים: $P\{\bigcap_{n=1}^\infty \bigcup_{k\geq n}\{\omega \mid |X_k(\omega) - X(\omega)| > \varepsilon\}\} = 0$.

נסמן $A_n = \bigcup_{k\geq n}\{\omega \mid |X_k(\omega) - X(\omega) > \varepsilon\}$; A_n יורדות כאשר $n \to \infty$, ולפי תכונות ההסתברות (מ־σ־אדיטיביות), גם $P(A_n) \searrow P(\bigcap_n A_n)$. לכן נקבל בסך־הכול $\lim_{n\to\infty} P\{\bigcup_{k\geq n}\{\omega \mid |X_k(\omega) - X(\omega)| > \varepsilon\}\} = 0$.

כ״ת $\not\Leftarrow$ בהסתברות

דוגמה (התכנסות בהסתברות ולא כמעט תמיד). $\Omega = [0,1]$, $\mathcal{F}$ σ־אלגברת בורל, $P = Leb$ (כלומר, $[a,b] \subset [0,1]$ אז $P[a,b] = b - a$). $\omega \in [0,1]$. צריך כעת לבנות משתנים מקריים – פונקציות.

נייצג כל טבעי כ־$n = 2^k + m$, כאשר $m = 0, 1, \ldots, 2^k - 1$. נגדיר באמצעות ייצוג זה

$$X_n(\omega) = \begin{cases} 1 & \omega \in [m2^{-k}, (m+1)2^{-k}] \\ 0 & \omega \in [0,1] \setminus [m2^{-k}, (m+1)2^{-k}] \end{cases}$$

מתקיים $X_n \xrightarrow{p} 0$, כי $P\{|X_n| > \varepsilon\} = 2^{-k} \to 0$. מצד שני, אין קבוצה עליה כל המשתנים המקריים מתאפסים.[27]

בהסתברות $\Longleftarrow$ כ״ת לתת־סדרה 3.4.2008

משפט 18: אם $X_n \xrightarrow{p} X$ אז קיימת תת־סדרה X_{n_i} כך ש־$X_{n_i} \to X$ כמעט תמיד.

הוכחה. לכל $\varepsilon > 0$ מתקיים $P\{|X_n - X| > \varepsilon\} \to 0$, לכן נוכל לכל m לבחור $n(m)$ כך ש־$P\{|X_{n(m)} - X| > \frac{1}{m}\} \leq 2^{-m}$. נסמן $Y_m = X_{n(m)}$. אז $P\{|Y_m - X| > \frac{1}{m}\} \leq 2^{-m}$. אם נסמן $A_m = \{|Y_m - X| > \frac{1}{m}\}$, נקבל $\sum_m P(A_m) = \sum_{m=1}^\infty 2^{-m} = 1 < \infty$. מבורל־קנטלי (שנוכיח בהמשך), $P(\bigcap_{n=1}^\infty \bigcup_{m>n} A_m) = 0$,[28] כלומר קיימת $\tilde{\Omega} \subseteq \Omega$ כך ש־$P(\tilde{\Omega}) = 1$ ולכל $\omega \in \tilde{\Omega}$ קיים $M(\omega)$ כך שאם $m \geq M(\omega)$, $\omega \notin A_m$. כלומר, ל־$m \geq M(\omega)$, $|Y_m(\omega) - X(\omega)| \leq \frac{1}{m}$. לכן $Y_n(\omega) \to X(\omega)$ כמעט תמיד.

בהסתברות $\Longleftarrow$ בהתפלגות

משפט 19: אם $X_n \xrightarrow{p} X$ אז $X_n \overset{d}{\Rightarrow} X$.

הוכחה. $X_n \xrightarrow{p} X$ אז $P\{|X_n - X| > \varepsilon\} \to 0$. נקבל

$$\begin{aligned} F_n(x) = P\{X_n \leq X\} &= P\{X_n \leq x, X \leq x + \varepsilon\} + P\{X_n \leq x, X > x + \varepsilon\} \\ &\leq P\{X \leq x + \varepsilon\} + P\{|X_n - X| > \varepsilon\} \\ &= F(x + \varepsilon) + P\{|X_n - X| > \varepsilon \\ &\to F(x + \varepsilon) \end{aligned}$$

האי־שוויון נשמר גם כשעוברים לגבול עליון: $\limsup F_n(x) \leq F(x + \varepsilon)$. באופן דומה,

$$\begin{aligned} F(x - \varepsilon) &= P\{X \leq x - \varepsilon\} \\ &= P\{X \leq x - \varepsilon\}, X_n \leq x\} + P\{X \leq x - \varepsilon, X_n > x\} \\ &\leq F_n(x) + P\{|X_n - X| > \varepsilon\} \end{aligned}$$

[27] נקבע ω; הוא שייך לסדרה אינסופית של קטעים מהסוג הזה, שעליהם $X_{n_k}(\omega) = 1$. לכן $X_{n_k} \not\to 0$.

[28] זהו הגבול העליון: קבוצת ה־ω־ות שמופיעים באינסוף מבין A_n.

(כי כאשר מתעלמים מהתנאי $X \le x-\varepsilon$ ההסתברות גדלה). אז $F(x-\varepsilon) \le \liminf F_n(x)$.
בנקודת רציפות, נעבור לגבול $\varepsilon \to 0$: נקבל $\overline{\lim}_{n\to\infty} F_n(x) \le F(x) \le \underline{\lim}_{n\to\infty} F_n(x)$
ולכן $F(x) = \lim_{n\to\infty} F_n(x)$.

בהתפלגות $\not\Longleftarrow$ בהסתברות, גם לא לת"ס

דוגמה (התכנסות בהתפלגות אך לא בהסתברות). ניקח $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים שווי־התפלגות ברנולי, $P\{X_i = 1\} = \frac{1}{2}$. ברור, אם כן, $X_n \Rightarrow X$ עבור X משתנה מקרי ברנולי.

נניח שקיימת תת־סדרה $X_{n_i} \xrightarrow{p} X$. תת־סדרה זו זהה לסדרה כולה, אז נדבר עליה. מתקיים $P\{|X_n - X| > \delta\} \to 0$ לכל δ, ולכן גם $P\{|X_{n+1} - X| > \delta\} \to 0$. מאי־שוויון המשולש נקבל $P\{|X_{n+1} - X_n| > 2\delta\} \to 0$. אבל משתנים אלה בלתי־תלויים ולכן $P\{X_{n+1} = 1, X_n = 0\} = P\{X_{n+1} = 0, X_n = 1\} = \frac{1}{4}$, ומכאן נקבל סתירה עבור $\delta < \frac{1}{2}$, כי $P\{|X_{n+1} - X_n| = 1\} = \frac{1}{2}$ (ולכן $P\{|X_{n+1} - X_n| > 2\delta\} \to \lambda > \frac{1}{2}$).

בהתפלגות לקבוע $\Longleftarrow$ בהסתברות

משפט 20: אם $X_n \overset{d}{\Rightarrow} a$ (a קבוע), אז $X_n \xrightarrow{p} a$.[29]

הוכחה. פונקציית ההתפלגות של a היא $F(x) = P\{a \le x\}$, שרציפה בכל נקודה חוץ מב־a. אז מהתכנסות בהתפלגות, נקבל מצד אחד $1 = F(a+\varepsilon) = \lim_{n\to\infty} P\{X_n \le a+\varepsilon\}$ ומצד שני $0 = F(a-\varepsilon) = \lim_{n\to\infty} P\{X_n \le a-\varepsilon\}$. ניקח הפרש: $P\{a-\varepsilon < X_n \le a+\varepsilon\} \to 1$, ולכן $P\{|X-a| > \varepsilon\} \to 0$.

6 חוק המספרים הגדולים

6.1 בצורה החלשה

14.2.2008

משפט 21: יהיו $X_1, X_2, \ldots$ משתנים מקריים לא־מתואמים כך ש־$\mathbb{E}X_i = \mu_i$, $\operatorname{Var} X_i = \sigma_i^2$ קיימים, ונניח ש־$\frac{\sigma_1^2+\ldots+\sigma_n^2}{n^2} \to 0$. אז לכל $\varepsilon > 0$, $P\{|\frac{X_1+\ldots+X_n}{n} - \frac{\mu_1+\ldots+\mu_n}{n}| > \varepsilon\} \overset{n\to\infty}{\to} 0$.

הוכחה. נשתמש באי־שוויון צ'בישב[30]:

$$\begin{aligned} P\{|\frac{X_1+\ldots+X_n-(\mu_1+\ldots+\mu_n)}{n}| > \varepsilon\} &= P\{|X_1+\ldots+X_n-(\mu_1+\ldots+\mu_n)| > n\varepsilon\} \\ &\le \frac{\operatorname{Var}(X_1+\ldots+X_n)}{\varepsilon^2 n^2} \\ &= \frac{\sum_{i=1}^n \sigma_i^2}{\varepsilon^2 n^2} \overset{n\to\infty}{\to} 0 \end{aligned}$$

משפט 22 (חוק המספרים הגדולים בצורה החלשה): יהיו $X_1, X_2, \ldots$ משתנים מקריים לא־מתואמים כך ש־$\mathbb{E}X_i = \mu$ ומתקיים $\frac{\sigma_1^2+\ldots+\sigma_n^2}{n^2} \to 0$ (למשל אם $\sigma_i = \sigma$). אז

$$P\{|\frac{X_1+\ldots+X_n}{n} - \mu| > \varepsilon\} \to 0$$

[29] ניתן להראות בכיוון ההפוך שגם התכנסות בהסתברות לקבוע לא מספיקה להתכנסות כמעט תמיד.
[30] $P\{|X-\mu| > \varepsilon\} \le \frac{\operatorname{Var} X}{\varepsilon^2}$

6.2 שימושים סטטיסטיים

6.2.1 מציאת פרמטר של התפלגות

אנו מניחים (או יודעים) ש־$\xi_1, \xi_2, \ldots \xi_n$ משתנים מקריים בלתי־תלויים שווי־התפלגות ברנולי, אבל איננו יודעים מהו $\theta = P\{\xi_1 = 1\}$; המטרה היא למצוא את θ. 18.2.2008

נעבוד במרחב ההסתברות $(\Omega, \mathcal{F}, P_\theta)$ עבור $\theta \in [0,1]$, כאשר Ω הוא מרחב ה־n־יות $\omega = (x_1, \ldots, x_n)$ בהן $x_i = 0, 1$ ו־$P_\theta(\omega) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$.

אנו רוצים לבנות **מעריך** $T_n(\omega)$ – פונקציה על Ω, שמטרתה לתת הערכה ל־θ בהינתן תוצאות $x_1, \ldots, x_n$ של n ניסויים על־ידי הצבת $\omega = (x_1, \ldots, x_n)$ ב־$T_n(\omega)$.

אפשר לקחת מעריך $T_n(\omega) = \frac{x_1+\ldots+x_n}{n}$, או $T_n(\omega) = \frac{\xi_1(\omega)+\ldots+\xi_n(\omega)}{n}$. לפי חוק המספרים הגדולים בצורה החלשה, $P_\theta\{|\frac{\xi_1(\omega)+\ldots+\xi_n(\omega)}{n}| > \varepsilon\} \overset{n\to\infty}{\to} 0$. כלומר, $T_n \overset{p}{\to} \theta$ – המעריך **עקבי** (consistent). בנוסף, $\mathbb{E}T_n = \frac{\theta+\ldots+\theta}{n} = \theta$ – המעריך **מאוזן** (unbiased).[31]

לפי אי־שוויון צ׳בישב, $P_\theta\{|T_n(\omega) - \theta| \geq \delta\} \leq \frac{\mathbb{E}(T_n-\theta)^2}{\delta^2}$. נחשב: התוחלת של כל ξ_i היא θ, לכן

$$\mathbb{E}(T_n - \theta)^2 = \frac{\mathbb{E}(\sum_{i=1}^n(\xi_i - \theta))^2}{n^2} = \frac{\operatorname{Var}\sum_{i=1}^n \xi_i}{n^2} = \frac{\sum_{i=1}^n \operatorname{Var}\xi_i}{n^2}$$

ומכיוון שכולם שווי־התפלגות, נקבל $\mathbb{E}(T_n-\theta)^2 = \frac{\operatorname{Var}\xi_1}{n}$.

נחשב את השונות: $\operatorname{Var}\xi_1 = \mathbb{E}(\xi_1 - \theta)^2 = \mathbb{E}\xi_1^2 - \theta^2 = \theta - \theta^2 = \theta(1-\theta)$. אז בסך־הכול $P_\theta\{|T_n(\omega) - \theta| \geq \delta\} \leq \frac{\mathbb{E}(T_n-\theta)^2}{\delta^2} = \frac{\theta(1-\theta)}{n\delta^2}$.

נסמן $\frac{\theta(1-\theta)}{n\delta^2} = (\frac{1}{\delta}\sqrt{\frac{\theta(1-\theta)}{n}})^2$, ואז $\frac{1}{\lambda} = \frac{1}{\delta}\sqrt{\frac{\theta(1-\theta)}{n}}$. אז נקבל $\delta = \lambda\sqrt{\frac{\theta(1-\theta)}{n}}$. $P_\theta\{|T_n - \theta| > \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\} \leq \frac{1}{\lambda^2}$. אם כן, $P_\theta\{|T_n = \theta| \leq \lambda\sqrt{\frac{\theta(1-\theta)}{n}}\} \geq 1 - \frac{1}{\lambda^2}$. אז בהסתברות גדולה מ־ או שווה ל־ $1 - \frac{1}{\lambda^2}$, נקבל $T_n - \lambda\sqrt{\frac{\theta(1-\theta)}{n}} \leq \theta \leq T_n + \lambda\sqrt{\frac{\theta(1-\theta)}{n}}$.

אנו עדיין רוצים "לסלק" את θ; נשים לב שלכל $\theta \in [0,1]$, $\theta(1-\theta) \leq \frac{1}{4}$. נעיר כי אי־השוויון נכון לכל λ, כי לפי אי־שוויון צ׳בישב הוא נכון לכל λ. לכן נקבל $T_n - \frac{\lambda}{2\sqrt{n}} \leq \theta \leq T_n + \frac{\lambda}{2\sqrt{n}}$.

נבחר למשל $\lambda = 3$: אז ההסתברות שנקבל היא $1 - \frac{1}{9} = \frac{8}{9} = 0.8888$. אם לא מסתפקים בהסתברות זו, צריך לבחור λ גדול יותר. נקבל $\theta \in [T_n - \frac{3}{2\sqrt{n}}, T_n + \frac{3}{2\sqrt{n}}]$. קטע זה נקרא **קטע סֶמֶךְ** (confidence interval) עם רמת סְמָךְ (reliability) 0.8888.

6.2.2 קירוב פולינומיאלי

משפט 23 (משפט ויירשטראס): תהא f פונקציה רציפה על $[0,1]$ ונגדיר

$$B_n(x) = \sum_{k=0}^n f(\frac{k}{n})\binom{n}{k}x^k(1-x)^{n-k}$$

ל־$x \in [0,1]$ – **פולינום ברנשטיין**. אז $B_n(x) \overset{n\to\infty}{\to} f(x)$ במידה שווה על $[0,1]$.

[31] יכולנו לקחת מעריך אחר: למשל, $\tilde{T}_n(\omega) = \frac{a_1\xi_1(\omega)+\ldots+a_n\xi_n(\omega)}{n}$ כאשר $a_1, \ldots, a_n \geq 0$, $\sum_{i=1}^n a_i = n$.

הוכחה. יהיו $\xi_1,\dots,\xi_n$ משתנים מקריים ברנולי שווי־התפלגות כך ש־$P\{\xi_1=1\}=x$. אזי $S_n=\xi_1+\ldots+\xi_n$ הוא משתנה מקרי בינומי, $P\{S_n=k\}=\binom{n}{k}x^k(1-x)^{n-k}$. אז $B_n(x)=\mathbb{E}f(\frac{S_n}{n})=\sum_{k=0}^n f(\frac{k}{n})P\{S_n=k\}$.

$$|B_n(x)-f(x)|=|\sum_{k=0}^n f(\frac{k}{n})P\{S_n=k\}-f(x)|$$

ניתן לכתוב $f(x)=\sum_{k=0}^n f(x)P\{S_n=k\}$, אז

$$\begin{aligned} &= |\sum_{k=0}^n (f(\tfrac{k}{n})-f(x))P\{S_n=k\}| \\ &\le \sum_{k=0}^n |f(\tfrac{k}{n})-f(x)|P\{S_n=k\} \end{aligned}$$

נתון $\varepsilon>0$ שרירותי; נבחר $\delta>0$ כך ש־$|x-y|\le\delta$ אז $|f(x)-f(y)|\le\varepsilon$ (מרציפות במידה שווה על $[0,1]$). נסמן $M=\sup_{y\in[0,1]}f(y)$. נכתוב

$$\begin{aligned} &= \sum_{k\,:\,|\frac{k}{n}-x|\le\delta}|f(\tfrac{k}{n})-f(x)|P\{S_n=k\}+ \\ &\quad +\sum_{k\,:\,|\frac{k}{n}-x|>\delta}|f(\tfrac{k}{n})-f(x)|P\{S_n=k\} \\ &\le \varepsilon+2M\sum_{k\,:\,|\frac{k}{n}-x|>\delta}P\{S_n=k\} \\ &= \varepsilon+2MP\{|\tfrac{S_n}{n}-x|>\delta\} \end{aligned}$$

לפי חוק המספרים הגדולים בצורה החלשה, הסתברות זו שואפת ל־0 כש־$n\to\infty$, כי $\mathbb{E}\xi_1=x$. נבחר N כך שלכל $n\ge N$ $P\{|\frac{S_n}{n}-x|>\delta\}\le\frac{\varepsilon}{2M}$.

לפי אי־שוויון צ׳בישב, $P\{|\frac{S_n}{n}-x|>\delta\}\le\frac{\mathrm{Var}\,S_n}{\delta^2n^2}=\frac{\mathrm{Var}\,\xi_1}{\delta^2n}=\frac{x(1-x)}{\delta^2n}\le\frac{1}{4\delta^2n}\overset{n\to\infty}{\to}0$, בלי תלות ב־$x$.[32] כעת אפשר לבחור N כך שלכל $n\ge N$ מתקיים $\frac{1}{4\delta^2n}\le\frac{\varepsilon}{2M}$, כלומר $n\ge\frac{2M}{4\delta^2\varepsilon}$ ללא תלות ב־x. לכן ההתכנסות היא במידה שווה.

ניתן להרחיב את השאלה: במקום התפלגות ברנולי ובינום, נדבר על התפלגות פואסון ונגדיר $R_n(x)=e^{-nx}\sum_{k=0}^n f(\frac{k}{n})\frac{(nx)^k}{k!}$ עבור $x\in[0,1]$.

למה 24: יהיו X ו־Y משתנים מקריים בלתי־תלויים עם התפלגות פואסון עם פרמטר λ ו־μ, בהתאמה; אז ל־$X+Y$ יש התפלגות פואסון עם פרמטר $\lambda+\mu$.

הוכחה.

$$\begin{aligned} P\{X+Y=n\} &= \sum_{k=0}^n P\{X=k,Y=n-k\} \\ &= \sum_{k=0}^n e^{-\lambda}\frac{\lambda^k}{k!}e^{-\mu}\frac{\mu^{n-k}}{(n-k)!} \\ &= e^{-(\lambda+\mu)}\frac{1}{n!}\sum_{k=0}^n\frac{n!}{k!(n-k)!}\lambda^k\mu^{n-k} \\ &= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!} \end{aligned}$$

באינדוקציה, אם $X_1,\dots,X_n$ משתנים מקריים בלתי־תלויים פואסון עם פרמטרים $\lambda_1,\dots,\lambda_n$, אז $X_1+\ldots+X_n$ משתנה מקרי פואסון עם פרמטר $\lambda_1+\ldots+\lambda_n$.

יהיו $X_1,X_2,\ldots$ משתנים מקריים בלתי־תלויים פואסון עם פרמטר x. אז לפי הלמה $S_n=X_1+\ldots+X_n$ פואסון עם פרמטר nx. מכאן, $R_n(x)=\sum_{k=0}^n f(\frac{k}{n})P\{S_n=k\}$.[33]

[32] זאת כי $x(1-x)\le\frac{1}{4}$ לכל $x\in[0,1]$: המקסימום מתקבל עבור $x=\frac{1}{2}$.

[33] זו לא $\mathbb{E}f(\frac{S_n}{n})$, כי אנו מתעלמים מערכים שליליים. אבל אפשר לכתוב שזהו $\mathbb{E}f(\frac{S_n}{n})\mathbf{1}_{[0,n]}(S_n)$.

טענה 25: אם f פונקציה רציפה על $[0,1]$ ו־$x \in (0,1)$, אז $f(x) \overset{n\to\infty}{\to} R_n(x)$.[34]

21.2.2008

הוכחה. מתקיים

$$\begin{aligned} |f(x) - e^{-nx}\sum_{k=0}^{n} f(\tfrac{k}{n})\tfrac{(nx)^k}{k!}| &= |f(x)e^{-nx}\sum_{k=0}^{\infty}\tfrac{(nx)^k}{k!} - \\ &-e^{-nx}\sum_{k=0}^{n} f(\tfrac{k}{n})\tfrac{(nx)^k}{k!}| \\ &\le |f(x)|e^{-nx}\sum_{k=n+1}^{\infty}\tfrac{(nx)^k}{k!} + \\ &+e^{-nx}\sum_{k=0}^{n}|f(x)-f(\tfrac{k}{n})|\tfrac{(nx)^k}{k!} \end{aligned}$$

מהרציפות, לכל $\varepsilon > 0$ קיים $\delta > 0$ כך שאם $|x-y| \le \delta$, $x, y \in [0,1]$ אז $|f(x)-f(y)| < \varepsilon$.
$0 < x < 1$, אז נבחר δ כך ש־$x + \delta < 1$. בנוסף, נסמן $\sup_{y\in[0,1]}|f(y)| = M < \infty$.

$$\begin{aligned} &\le |f(x)|e^{-nx}\sum_{k=n+1}^{\infty}\tfrac{(nx)^k}{k!} + \\ &+e^{-nx}\sum_{k\le n\,:\,|\frac{k}{n}-x|\le\delta}|f(x)-f(\tfrac{k}{n})|\tfrac{(nx)^k}{k!} + \\ &+e^{-nx}\sum_{k\le n\,:\,|\frac{k}{n}-x|>\delta}|f(x)-f(\tfrac{k}{n})|\tfrac{(nx)^k}{k!} \\ &\le Me^{-nx}\sum_{k-nx=n(1-x)+1}\tfrac{(nx)^k}{k!} + \varepsilon + 2MP\{|\tfrac{S_n}{n} - x| > \delta\} \end{aligned}$$

כאשר $S_n = X_1 + \ldots + X_n$ עבור $\{X_i\}$ משתנים מקריים בלתי־תלויים שווי־התפלגות פואסון עם פרמטר x. נחלק את $k - nx = n(1-x) + 1$ ב־n ונקבל $\frac{k}{n} - x = (1-x) + \frac{1}{n} > \delta$ לפי בחירת δ; אז

$$\begin{aligned} &\le \varepsilon + 3MP\{|\tfrac{S_n}{n} - x| > \delta\} \\ &\le \varepsilon + 3M\frac{\operatorname{Var}\frac{S_n}{n}}{\delta^2} \\ &= \varepsilon + 3M\frac{\operatorname{Var}S_n}{n^2\delta^2} \\ &= \varepsilon + 3M\frac{\operatorname{Var}X_1}{n\delta^2} \end{aligned}$$

ומתקיים $\operatorname{Var}X_1 = x$, כי $\operatorname{Var}X_1 = \mathbb{E}X_1^2 - (\mathbb{E}X_1)^2 = \mathbb{E}X_1^2 - x^2$ ובנוסף $e^{-x}\sum_{k=0}^{\infty}k^2\frac{x^k}{k!} = x^2 + x$.

6.2.3 שיטת מונטה־קרלו

רוצים לחשב $\int_0^1 f(x)dx$, $f(x)$ רציפה ו־$0 \le f(x) \le 1$.[35]

יהיו $X_1, Y_1, X_2, Y_2, \ldots$ משתנים מקריים בלתי־תלויים שווי־התפלגות עם התפלגות אחידה על הקטע $[0,1]$. נגדיר $Z_i = \begin{cases} 1 & f(X_i) > Y_i \\ 0 & f(X_i) \le Y_i \end{cases}$. אז $\mathbb{E}Z_i = P\{f(X_i) > Y_i\}$.

$$P\{(X_i, Y_i) \in\} = P\{X_i \in [a,b]\}P\{Y_i \in [a,c]\} = areaR$$

$\{Z_i\}$ משתנים מקריים בלתי־תלויים ברנולי, $P\{Z_i = 1\} = \int_0^1 f(x)dx$.

שיטה: נקרב את $\int_0^1 f(x)dx$ על־ידי $\frac{1}{n}\sum_{i=1}^{n} Z_i \xrightarrow{p} \mathbb{E}Z_i = \int_0^1 f(x)dx$.

[34] ההתכנסות במידה שווה היא על כל קטע סגור ב־$[0,1)$, אבל לא על כל $[0,1]$: אם $x = 1$, נוכיח בעתיד שההתכנסות היא לא ל־$f(1)$ אלא ל־$\frac{f(1)}{2}$.

[35] כדי להגיע למצב כזה, אפשר להוסיף או להוריד קבוע ולחלק ב־sup של הפונקציה.

6.3 בצורה החזקה

25.2.2008 **משפט 26 (החוק החזק של המספרים הגדולים):** $(X_n)_{n=1}^{\infty}$ משתנים מקריים בלתי־תלויים ושווי־התפלגות. נניח ש־$\mathbb{E}(|X_1|^n) < \infty$[36] וש־$\mathbb{E}(X_1) = \mu$.[37] נסמן לכל n טבעי $S_n = \sum_{k=1}^{n} X_k$.
אזי $P(\lim_{n\to\infty} \frac{1}{n} \cdot S_n = \mu) = 1$.

ניסוח אחר: לכל ε, $P($קיים N כך שלכל $n > N$ $|\frac{S_n}{n} - \mu| < \varepsilon) = 1$.[38]

תרגיל: מצאו סדרה של משתנים מקריים $\{Z_n\}_{n=1}^{\infty}$ כך שלכל ε $\lim_{k\to\infty} P(|Z_n| < \varepsilon) = 1$ אבל קיים ε כך ש־$P(\exists N : \forall n > N \quad |Z_n| < \varepsilon) = 0$.

דוגמה. יהיו $\{A_n\}_{n=1}^{\infty}$ מאורעות בלתי־תלויים המקיימים $P(A_n) = \frac{1}{n}$ ו־$Z_n = 1$ אם A_n קורה, $Z_n = 0$ אחרת.

למה 27 (בורל־קנטלי): (א) יהיו $\{A_n\}_{n=1}^{\infty}$ מאורעות ונניח $\sum_{n=1}^{\infty} P(A_n) < \infty$. אז מתקיים $P(\exists N : \forall n > N \quad \neg A_n) = 1$.[39]

(ב) תהי $\{A_n\}_{n=1}^{\infty}$ סדרת מאורעות בלתי־תלויים ונניח $\sum_{n=1}^{\infty} P(A_n) = \infty$. אז מתקיים $P(\exists N : \forall n > N \quad \neg A_n) = 0$.

הוכחה. (א) יהי $\bar{A}$ המאורע שקיים N כך שלכל $n > N$, A_n לא מתקיים. אנחנו רוצים לחשב את ההסתברות של $\bar{A}$. נגדיר $\bar{A}_N$ להיות המאורע שלכל $n > N$, A_n לא מתקיים, ואז $\bar{A} = \bigcup_{n=1}^{\infty} \bar{A}_N$, ובפרט $P(\bar{A}) \geq P(\bar{A}_N)$ לכל N.

$\bar{A}_N = \bigcap_{n=N+1}^{\infty} A_n^C = [\bigcup_{n=N+1}^{\infty} A_n]^C$ ואם כן,

$P(\bar{A}_N) = 1 - P(\bigcup_{n=N+1}^{\infty} A_n) \geq 1 - \sum_{n=N+1}^{\infty} P(A_n) \to 1$, כי זהו זנב של טור מתכנס. לכן $P(\bar{A}) \geq \limsup_{N\to\infty} P(\bar{A}_N) = 1$.

(ב) נשתמש באותם הסימונים. נשים לב ש־$\bar{A}_N \subseteq \bar{A}_{N+1}$. לכן $P(\bar{A}) = \lim_{N\to\infty} P(\bar{A}_N)$. לכן מספיק להראות שתחת ההנחות, $P(\bar{A}_N) = 0$ לכל N.

$\bar{A}_N = \bigcap_{n=N+1}^{\infty} A_n^C$

A_n בלתי־תלויים, לכן גם A_n^C בלתי־תלויים. אז

$P(\bar{A}_N) = \lim_{k\to\infty} P(\bigcap_{n=N+1}^{k} A_n^C) = \lim_{k\to\infty} \prod_{n=N+1}^{k} P(A_n^C)$.[40]

נחלק לשני מקרים: אם קיים $n > N$ כך ש־$P(A_n) = 1$, אז $P(A_n^C) = 0$ ולכל $k > n$, $\prod_{n=N+1}^{k} P(A_n^C) = 0$ ואז $P(\bar{A}_N) = 0$.

אחרת, הסיכויים תמיד חיוביים ונוכל לכתוב $\prod_{n=N+1}^{k} P(A_n^C) = e^{\sum_{n=N+1}^{k} \log P(A_n^C)}$.
נזכור ש־$\log(x) \leq x - 1$ לכל x, ולכן $\log(P(A_n^C)) \leq P(A_n^C) - 1 = -P(A_n)$. לכן $P(\bigcap_{n=N+1}^{k} A_n^C) \leq e^{-\sum_{n=N+1}^{k} P(A_n)}$. אבל $\lim_{k\to\infty} \sum_{n=N+1}^{k} P(A_n) = \infty$, ולכן $\lim_{k\to\infty} P(\bigcap_{n=N+1}^{\infty} A_n^C) = 0$ וזהו.

[36] ניתן להחליף דרישה זו בדרישה ש־$\mathbb{E}(|X_1|)$ קיימת, אך אז ההוכחה תהיה קשה בהרבה.
[37] תוחלת זו קיימת כי הרי הנחנו ש־$\mathbb{E}(|X|^n)$ קיימת.
[38] תזכורת – החוק החלש: לכל ε, $\lim_{n\to\infty} P(|\frac{1}{n} S_n - \mu| < \varepsilon) = 1$.
[39] במילים: ההסתברות לכך שקיים N כך שלכל $n > N$ המאורע A_n לא מתקיים היא 1.
[40] הגדרת אי־תלות מדברת על תת־אוסף סופי, לכן עלינו להשתמש בגבול.

למה 28: יהיו $\{Y_n\}_{n=1}^{\infty}$ משתנים מקריים בלתי־תלויים ושווי־התפלגות המקיימים $\mathbb{E}(Y_1^4) < \infty$ ו־$\mathbb{E}(Y_1) = 0$. אז לכל n, $\mathbb{E}[(\sum_{k=1}^n Y_k)^4] \leq 3n^2\mathbb{E}(Y_1^4)$.

הוכחה. ראשית נשים לב ש־$\mathbb{E}(Y_1^4) \geq [\mathbb{E}(Y_1^2)]^2$.

$(\sum_{k=1}^n Y_k)^4 = \sum_{1\leq i_1,i_2,i_3,i_4\leq n} Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}$. אם כן,

$\mathbb{E}[(\sum_{k=1}^n Y_k)^4] = \sum_{1\leq i_1,i_2,i_3,i_4\leq n} \mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4})$

נחשב את $\mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4})$. נחלק למספר מקרים:

1. נניח ש־$i_1 \notin \{i_2,i_3,i_4\}$. אז $\mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}) = \mathbb{E}(Y_{i_1})\mathbb{E}(Y_{i_2}Y_{i_3}Y_{i_4}) = 0$, כי Y_{i_1} בלתי־תלוי ב־$Y_{i_2}Y_{i_3}Y_{i_4}$ ו־$\mathbb{E}(Y_{i_1}) = 0$. דבר דומה מתקיים עבור $i_2 \notin \{i_1,i_3,i_4\}$ וכו'. לכן $\mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}) \neq 0$ רק אם
2. $i_1 = i_2 = i_3 = i_4$, שאז $\mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}) = \mathbb{E}(Y_{i_1}^4)$, או אם
3. הם מחולקים לשתי קבוצות בגודל 2, שאז $\mathbb{E}(Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}) = [\mathbb{E}(X_1^2)]^2 \leq \mathbb{E}(Y_1)^4$.

ישנן n רביעיות מסוג (2), וישנן $3\binom{n}{2}$ רביעיות מסוג (3). לכן

$\mathbb{E}[(\sum_{k=1}^n Y_k)^4] \leq \mathbb{E}(Y_1^4)(n+3\binom{n}{2}) \leq 3n^2\mathbb{E}(Y_1^4)$.

למה 29: תהי $\{Z_n\}_{n=1}^{\infty}$ סדרה של משתנים מקריים ו־Z משתנה מקרי. אז אם לכל ε, 28.2.2008
$P(_{|Z_n - Z| < \varepsilon, n > N \text{ כך שלכל } N \text{ קיים}}) = 1$, אז $P(_{|Z_n - Z| < \varepsilon, n > N \text{ כך שלכל } N \text{ קיים } \varepsilon \text{ לכל}}) = 1$.

הוכחה. לכל ε יהי $A_\varepsilon = \{_{|Z_n - Z| < \varepsilon, n > N \text{ כך שלכל } N \text{ קיים}}\}$. נשים לב שאם $\varepsilon_1 < \varepsilon_2$ נקבל $A_{\varepsilon_1} \subseteq A_{\varepsilon_2}$.

נתון שלכל ε, $P(A_\varepsilon) = 1$. אנו רוצים להראות $P(\bigcap_{\varepsilon>0} A_\varepsilon) = 1$.

מה הקושי? ידוע שבהינתן $\aleph_0$ מאורעות שהסתברות כל אחד מהם היא 1 גם הסתברות חיתוכם היא 1. אך לאוסף שאינו בן מניה של מאורעות, טענה זו אינה בהכרח נכונה.

לכל $k = 1,2,\ldots$, $P(A_{\frac{1}{k}}) = 1$ ולכן $P(\bigcap_{k=1}^{\infty} A_{\frac{1}{k}}) = 1$, שכן זהו חיתוך בן־מניה.

לכל ε קיים k_0 כך ש־$\frac{1}{k_0} < \varepsilon$. לכן $A_{\frac{1}{k_0}} \subseteq A_\varepsilon$. לכן $\bigcap_{k=1}^{\infty} A_{\frac{1}{k}} \subseteq A_{\frac{1}{k_0}} \subseteq A_\varepsilon$, לכן $\bigcap_{k=1}^{\infty} A_{\frac{1}{k}} \subseteq \bigcap_{\varepsilon>0} A_\varepsilon$.

לכן $P(\bigcap_{\varepsilon>0} A_\varepsilon) \geq P(\bigcap_{k=1}^{\infty} A_{\frac{1}{k}}) = 1$, ולכן $P(\bigcap_{\varepsilon>0} A_\varepsilon) = 1$.

השתמשנו בעובדה הבאה: אם $P(B_k) = 1$ לכל $k = 1,2,\ldots$, אז $P(\bigcap_{k=1}^{\infty} B_k) = 1$. ממה זה נובע? אחת האקסיומות של תורת ההסתברות אומרת שאם $\{A_n\}_{n=1}^{\infty}$ סדרה של מאורעות זרים בזוגות אז $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

נשמיט את ההנחה ש־A_n זרים בזוגות. במקרה זה, נגדיר $\bar{A}_n = A_n \setminus \bigcup_{k=1}^{n-1} A_k$, ואז $P(\bar{A}_n) \leq P(A_n)$ ו־$\{\bar{A}_n\}$ זרים בזוגות. $P(\bigcup_{n=1}^{\infty} A_n) = P(\bigcup_{n=1}^{\infty} \bar{A}_n) = \sum_{n=1}^{\infty} P(\bar{A}_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

אז $P(\bigcap_{k=1}^{\infty} B_k) = 1 - P((\bigcap_{k=1}^{\infty} B_k)^C) = 1 - P(\bigcup_{k=1}^{\infty} B_k^C) \geq 1 - \sum_{k=1}^{\infty} P(B_k^C) = 1$.

דוגמה. מאורעות $\{B_\alpha\}_{0\le\alpha\le1}$ המקיימים $P(B_\alpha)=1$ לכל α אבל $P(\bigcap B_\alpha)\lneq 1$:[41] יהי X משתנה מקרי מפולג אחיד בקטע $[0,1]$ (דהיינו, לכל $0\le a<b\le1$ מתקיים כי $P(a\le X\le b)=b-a$). נגדיר $B_\alpha=\{X\neq\alpha\}$. אז לכל α, $P(B_\alpha)=1$, בעוד ש־$P(\bigcap_{0\le\alpha\le1}B_\alpha)=P(X\notin[0,1])=0$.

מסקנה: יש להיזהר באיחודים ובחיתוכים שאינם בני־מניה.

25.2.2008 נחזור להוכחת החוק החזק של המספרים הגדולים:

הוכחה. נראה שלכל ε, $P(\text{קיים } N \text{ כך שלכל } n>N,\ |\frac{1}{n}S_n-\mu|<\varepsilon)$: נגדיר $A_n=\{|\frac{1}{n}S_n-\mu|\ge\varepsilon\}$. עלינו להראות ש־$1=P(\text{קיים } N \text{ כך שלכל } n>N,\ A_n \text{ לא קורה})$. לכן, לפי בורל־קנטלי, מספיק להראות ש־$\sum_{n=1}^\infty P(A_n)<\infty$. נעריך את $P(A_n)$:
$P(A_n)=P(|\frac{1}{n}S_n-\mu|\ge\varepsilon)=P((\frac{1}{n}S_n-\mu)^4\ge\varepsilon^4)\le\frac{\mathbb{E}[(\frac{1}{n}S_n-\mu)^4]}{\varepsilon^4}$ לפי אי־שוויון מרקוב. נגדיר $Y_n=X_n-\mu$. אז $\{Y_n\}_{n=1}^\infty$ בלתי־תלויים, ו־$\mathbb{E}(Y_1)=\mathbb{E}(x_1)-\mu=0$ ו־$\mathbb{E}(Y_1^r)<\infty$.[42]

כמו כן, $\frac{1}{n}\sum_{k=1}^n Y_k=\frac{1}{n}\sum_{k=1}^n(X_k-\mu)=\frac{1}{n}(\sum_{k=1}^n X_k-n\mu)=\frac{1}{n}S_n-\mu$. לכן

$$\begin{aligned}\mathbb{E}[(\tfrac{1}{n}S_n-\mu)^4] &= \mathbb{E}[(\tfrac{1}{n}\textstyle\sum_{k=1}^n Y_k)^4]\\ &= \tfrac{1}{n^4}\mathbb{E}[(\textstyle\sum_{k=1}^n Y_k)^4]\\ &\le \frac{3n^2\mathbb{E}(Y_1^4)}{n^4}\\ &= \frac{3\mathbb{E}(Y_1^4)}{n^2}\end{aligned}$$

לכן $P(A_n)\le\frac{3\mathbb{E}(Y_1^4)}{\varepsilon^4}\cdot\frac{1}{n^2}$, ולכן $\sum_{n=1}^\infty P(A_n)<\infty$.

28.2.2008 **משפט 30 (הכיוון ההפוך של החוק החזק של המספרים הגדולים):** $\{X_n\}_{n=1}^\infty$ משתנים מקריים בלתי־תלויים ושווי־התפלגות, ויהי $\mu\in\mathbb{R}$ ונניח שמתקיים $P(\frac{1}{n}\sum_{i=1}^n X_k\overset{n\to\infty}{\to}\mu)=1$, אז $\mathbb{E}(X-1)=\mu$, ובפרט ל־X_1 יש תוחלת.[43]

הוכחה. נראה שיש תוחלת.[44]

למה 1.30: אם ל־X_1 אין תוחלת, אז $\sum_{n=1}^\infty P(|X_1|>n)=\infty$.

הוכחה. תרגיל.

[41] כלומר, הקושי עם חיתוך שאינו בן־מניה אמיתי.

[42] $Y_1^r=(x_1-\mu)^4=\sum_{i=0}^4\binom{4}{i}x_1^i(-\mu)^{4-i}=c_1+c_2x+c_3x^2+c_4x^3+c_5x^4$. לכן מתקיים $\mathbb{E}(Y_1^4)=c_1+c_2\mathbb{E}(X_1)+\ldots+c_5\mathbb{E}(X_1^4)$, ולכן $\mathbb{E}(Y_1^4)<\infty$.

למה 1.29: אם $\mathbb{E}(|X|^n)<\infty$, אז לכל $k<n$, $\mathbb{E}(|X|^k)<\infty$.

הוכחה. יהיו $W=1$ אם $|X|\le1$, $W=0$ אחרת, $V=0$ אם $|X|\le1$, $V=1$ אחרת. אז $|X|=V|X|+W|X|$ ו־$|X|^k|=V|X|^k+W|X|^k$. אם כן,
$$\mathbb{E}(|X|^k)=\mathbb{E}(V|X|^k)+\mathbb{E}(W|X|^k)\le\mathbb{E}(|X|^n)+1<\infty$$

[43] לא בחומר הקורס.

[44] ברגע שיש, ברור שההתכנסות היא לדבר הנכון.

לכן $\sum_{n=1}^{\infty} P(|X_n| > n) = \infty$ והמאורעות $\{|X_n| > n\}$ בלתי־תלויים. אז מבורל־קנטלי, נובע שבסיכוי 1, אינסוף מהמאורעות $|X_n| > n$ קורים. מחשבון פשוט נקבל שאם $|X_n| > n$ אז $|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n-1}\sum_{k=1}^{n-1} X_k| > \frac{1}{2}$. אבל אם זה קורה אינסוף פעמים, אין התכנסות.

7 פונקציות יוצרות

פונקציה יוצרת

הגדרה. $G_a(s) = \sum_{k=0}^{\infty} a_k s^k$ נקראת **פונקציה יוצרת** של סדרת מספרים $a = (a_0, a_1, \ldots)$. 3.3.2007

נזכיר שלטור חזקות יש רדיוס r בו הוא מתכנס (כלומר, לכל $|s| < r$, $\sum_{k=0}^{\infty} a_k s^k < \infty$).
אם $r > 0$, ניתן לגזור את הטור איבר־איבר לכל $|s| < r$, ואז $G_a(0) = a_0$, $G'_a(0) = a_1$, ובאופן כללי $G_a^{(n)}(0) = a_n n!$ - כלומר, $a_n = \frac{G_a^{(n)}(0)}{n!}$: מקדמי הטור מוגדרים באופן יחיד על־ידי נגזרות הפונקציה היוצרת.

מסקנה 31: אם קיים $r > 0$ כך שלכל $|s| < r$ מתקיים $G_a(s) = G_b(s)$ לשתי סדרות $a = (a_0, a_1, \ldots)$ ו־$b = (b_0, b_1, \ldots)$, אזי $a = b$.

עד כה, לא בהכרח דיברנו על הסתברות. בהסתברות נתעניין במשתנה מקרי X המקבל ערכים טבעיים (או 0), ונדבר על הפונקציה היוצרת $G_X(s) = \sum_{k=0}^{\infty} p_k s^k$, כאשר $p_k = P\{X = k\}$.[45]
רדיוס ההתכנסות r מקיים $r \geq 1$: אם $|s| < 1$, כל המקדמים $p_k \leq 1$ ולכן הטור חסום מלמעלה על־ידי טור גיאומטרי מתכנס.

משפט 32 (אבל): נניח שלטור חזקות $G_a(s) = \sum_{k=0}^{\infty} a_k s^k$ המקדמים אי־שליליים ושהטור מתכנס לכל $|s| < 1$. אזי $\lim_{s\to 1} G_a(s) = G_a(1) (\in \mathbb{R} \cup \{\infty\})$.[46]

נניח $G_X(s) = \sum_{k=0}^{\infty} p_k s^k$. לכל $|s| < 1$ קיימת $G'(s) = \sum_{k=1}^{\infty} k p_k s^{k-1}$. נציב $s = 1$ ונקבל, לפי משפט אבל, $G'_X(1) = \lim_{s\to 1} G'_X(s) = \sum_{k=1}^{\infty} k p_k = \mathbb{E}(X)$. כלומר, מצאנו נוסחה לחישוב התוחלת.

נתבונן בנגזרת השנייה: $G''_X(s) = \sum_{k=2}^{\infty} k(k-1) p_k s^{k-2}$, ואם נציב $s = 1$ נקבל

$$\begin{aligned}
G''_X(1) = \lim_{s\to 1} G''_X(s) &= \sum_{k=2}^{\infty} k(k-1) p_k \\
&= \mathbb{E}(X(X-1)) = \mathbb{E}X^2 - \mathbb{E}X \\
&= \mathbb{E}X^2 - (\mathbb{E}X)^2 + (\mathbb{E}X)^2 - \mathbb{E}X \\
&= \operatorname{Var} X + (\mathbb{E}X)^2 - \mathbb{E}X
\end{aligned}$$

כלומר, $G''_X(1) = \operatorname{Var} X + (G'_X(1))^2 - G'_X(1)$, ולפיכך $\operatorname{Var} X = G''_X(1) - (G'_X(1))^2 + G'_X(1)$ - מצאנו נוסחה לחישוב השונות.

עם זאת, נשים לב שאנו מוגבלים לעיסוק במשתנים מקריים שמקבלים ערכים טבעיים בלבד.

[45] דהיינו, בונים את הטור מסדרת ההסתברויות.
[46] כלומר, G_a רציפה ב־1.

דוגמה. התפלגות בינומית שלילית: X – מספר ניסויי ברנולי שנכשלו עד להצלחה מספר r: $P\{X=k\} = \binom{k-1}{r-1}p^r q^{k-1}$ (כאשר $q=1-p$, $k \geq r$). זה נקרא בינומי שלילי כי $\sum_{k=1}^{\infty} \binom{k-1}{r-1} q^k = (1-q)^{-r}$. אם $r=1$, המשתנה המקרי נקרא גיאומטרי.

נרצה לחשב את הפונקציה היוצרת עבור X משתנה מקרי בינומי שלילי, Y משתנה מקרי עם התפלגות גיאומטרית ($P\{Y=k\} = pq^{k-1}$). $\sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} = (1-q)^{-r}$, ולכן

$$\begin{aligned} G_X(s) &= \sum_{k=r}^{\infty} p_k s^k \\ &= p^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} q^{k-r} s^k \\ &= (ps)^r \sum_{k=r}^{\infty} \binom{k-1}{r-1} (qs)^{k-r} \\ &= (ps)^r (1-qs)^{-r} \\ &= \left(\frac{ps}{1-qs}\right)^r \end{aligned}$$

מכאן, אם נציב $r=1$, $G_Y(s) = \frac{ps}{1-qs}$. אם כן, $G_X(s) = (G_Y(s))^r$. נוסחה זו איננה מקרית, ובעזרת טענה שנוכיח מיד, ניתן להסיק מכאן ש־X הוא סכום של r משתנים מקריים בלתי־תלויים שווי־התפלגות עם התפלגות גיאומטרית (כמו Y).

הגדרה. קונבולוציה (convolution) של שתי סדרות $a=(a_0,a_1,\ldots)$, $b=(b_0,b_1,\ldots)$ היא סדרה $a*b=c=(c_0,c_1,\ldots)$ המוגדרת על־ידי $c_n = \sum_{i=1}^{n} a_i b_{n-i}$. קונבולוציה

הפונקציה היוצרת של c:

$$\begin{aligned} G_c(s) = \sum_{n=0}^{\infty} c_n s^n &= \sum_{n=0}^{\infty} \left(\sum_{i=0}^{n} a_i b_{n-i}\right) s^{i+(n-i)} \\ &= \sum_{n=0}^{\infty} \sum_{i=0}^{n} a_i s^i b_{n-i} s^{n-i} \\ &= \sum_{i=1}^{\infty} \sum_{n=i}^{\infty} a_i s^i b_{n-i} s^{n-i} \\ &= \sum_{i=0}^{\infty} a_i s^i \sum_{k=0}^{\infty} b_k s^k = G_a(s) \cdot G_b(s) \end{aligned}$$

(בשלב האחרון, $k=n-i$.) כלומר, אם $c=a*b$ אז $G_c(s) = G_a(s) \cdot G_b(s)$.

נניח ש־X ו־Y משתנים מקריים בלתי־תלויים המקבלים ערכים טבעיים כך שמתקיים $P\{X=k\}=p_k$ ו־$P\{Y=k\}=q_k$.[47] נסתכל על $Z=X+Y$:

$$r_n = P\{Z=n\} = P\{X+Y=n\} = \sum_{k=0}^{n} P\{X=k, Y=n-k\} = \sum_{k=0}^{n} P\{X=k\}P\{Y=n-k\} = \sum_{k=0}^{n} p_k q_{n-k}$$

כלומר, $r=p*q$, ולכן $G_{X+Y}(s) = G_Z(s) = G_X(s) \cdot G_Y(s)$. כלומר, אם X ו־Y בלתי־תלויים, הפונקציה היוצרת של הסכום $X+Y$ היא מכפלת הפונקציות היוצרות של X ו־Y.

מכאן, באינדוקציה, אם $X_1,\ldots,X_n$ משתנים מקריים בלתי־תלויים המקבלים ערכים טבעיים, אז $G_{X_1+\ldots+X_n}(s) = \prod_{i=1}^{n} G_{X_i}(s)$.

נזחור למה שעשינו קודם: יהיו $X_1,\ldots,X_r$ משתנים מקריים בלתי־תלויים עם התפלגות גיאומטרית. אז $G_{X_1+\ldots+X_n}(s) = \left(\frac{ps}{1-qs}\right)^r$. לכן $X_1+\ldots+X_r$ הוא משתנה מקרי בינומי שלילי. כלומר, למשתנה מקרי $X = X_1+\ldots+X_n$ יש אותה פונקציית התפלגות (או אותה

[47] כאן אין קשר בין p ל־q.

פונקצית משקל) של סכום של r משתנים מקריים בלתי־תלויים שווי התפלגות עם התפלגות גיאומטרית.

נניח $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים שווי־התפלגות המקבלים ערכים טבעיים עם פונקציה יוצרת G_X, ויהי N משתנה מקרי אחר בלתי־תלוי בכל X_i המקבל ערכים טבעיים עם פונקציה יוצרת G_N. נגדיר $S = X_1 + X_2 + \ldots + X_N$ (אם $N = 0$, $S = 0$). כלומר, לכל $\omega \in \Omega$, $S(\omega) = X_1(\omega) + X_2(\omega) + \ldots + X_{N(\omega)}(\omega)$.

טענה 33: $G_S(t) = G_N(G_X(t))$.

הוכחה.

$$\begin{aligned} G_S(t) &= \textstyle\sum_{n=0}^\infty t^n P\{S = n\} \\ &= \textstyle\sum_{n=0}^\infty t^n P\{X_1 + \ldots + X_N = n\} \\ &= \textstyle\sum_{n=0}^\infty t^n \sum_{k=0}^\infty P\{X_1 + \ldots + X_k = n \mid N = k\} P\{N = k\} \end{aligned}$$

אבל N בלתי־תלוי ב־X_i, ובאופן כללי אם A ו־B בלתי־תלויים אז $P(A \mid B) = P(A)$. אז

$$\begin{aligned} &= \textstyle\sum_{n=0}^\infty t^n \sum_{k=0}^\infty P\{X_1 + \ldots + X_k = n\} P\{N = k\} \\ &= \textstyle\sum_{k=0}^\infty P\{N = k\} \sum_{n=0}^\infty t^n P\{X_1 + \ldots + X_k = n\} \end{aligned}$$

(מותר להחליף את הסכימה כי כל הטורים מתכנסים.)

$$\begin{aligned} &= \textstyle\sum_{k=0}^\infty P\{N = k\} G_{X_1 + \ldots + X_n}(t) \\ &= \textstyle\sum_{k=0}^\infty P\{N = k\} (G_X(t))^k \\ &= G_N(G_X(t)) \end{aligned}$$

שהרי $G_N(u) = \sum_{k=0}^\infty P\{N = k\} u^k$.

7.1 תהליכי הסתעפות

יש אוכלוסייה בדור n שמספרה Z_n ובדור $n+1$ כל אב (אם) מספר i מוליד X_i בנים (בנות) כך 6.3.2008
ש־$X_1, \ldots, X_{Z_n}$ משתנים מקריים בלתי־תלויים שווי־התפלגות עם פונקציה יוצרת G.

Z_0 בזמן 0

$$Z_1 = X_1 + \ldots + X_{Z_0}$$
$$Z_2 = \tilde{X}_1 + \ldots + \tilde{X}_{Z_1}$$

כל ה־X־ים הם בלתי־תלויים ויש להם אותה התפלגות עם פונקציה יוצרת G.

$$G_N(s) = \textstyle\sum_{i=0}^\infty p_i s^i = \sum_{i=0}^\infty e^{-\lambda} \frac{\lambda^i}{i!} s^i = e^{-\lambda} \sum_{i=0}^\infty \frac{(\lambda s)^i}{i!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)}$$

$G_S(u) = G_N(G_X(u)) = e^{\lambda(1-p+pu-1)} = e^{\lambda p(u-1)}$ – S הוא משתנה מקרי פואסון עם פרמטר $p\lambda$.

טענה 34: נניח $Z_0 = 1$. $G_{Z_n}(s) = G(G(\ldots(G(s))\ldots))$ n־פעמים.

הוכחה. באינדוקציה. $n = 1$: $G_{Z_1}(s) = G(s)$.

נניח שזה נכון לכל $n = 1, \ldots, k$ ונוכיח ל־$n = k+1$.

$G_{Z_{k+1}}(s) = G_{Z_k}(G(s))$ לפי הנוסחה (שכן $Z_{k+1} = X_1 + \ldots + X_{Z_k}$, $\{X_i\}$ בלתי־תלויים ושווי־התפלגות עם G כפונקציה יוצרת). לפי הנחת האינדוקציה, $G_{Z_k}(u) =$

$G(G(\ldots(G(u))\ldots))$ k פעמים, אז $G_{Z_k+1}(s) = G(G(\ldots(G(s))\ldots))$, כנדרש.

טענה 35: $\mu = \mathbb{E}X_1$, $\mu_n \overset{\text{def}}{=} \mathbb{E}Z_n = \mu^n$, $\sigma^2 \overset{\text{def}}{=} \operatorname{Var} Z_n = \sigma^2\mu^{n-1}\frac{\mu^n-1}{\mu-1}$ עבור $\mu \neq 1$, $=\sigma^2 n$ עבור $\mu = 1$ ‏($\sigma^2 = \operatorname{Var} X_1$).

הוכחה. $\lim_{s\uparrow 1} G_n'(s) = (G_n'(1)) = \mathbb{E}Z_n$.[48]

$G_n(s) = G_{n-1}(G(s))$;[49] נגזור ונקבל $G_n'(s) = G_{n-1}'(G(s))G'(s)$. נעבור לגבול $s \uparrow 1$: $\mathbb{E}Z_n = \mathbb{E}Z_{n-1}\mu$.

$\lim_{s\uparrow 1} G(s) = 1$, אז $\lim_{s\uparrow 1} G_{n-1}'(G(s)) = \lim_{u\to 1} G_{n--1}'(u) = \mathbb{E}Z_{n-1}$. לכן $\mathbb{E}Z_n = \mu^n$.

הוכחנו $\operatorname{Var} Y = G_Y''(1) + G_Y'(1) - (G_Y'(1))^2$.

10.3.2008 [חסר בינתיים]

7.2 חזרות של מהלך אקראי פשוט

13.3.2008 $S_0 = 0$, $S_n = \sum_{k=1}^{\infty} X_k$ כאשר $\{X_k\}$ משתנים מקריים בלתי־תלויים שווי־התפלגות, $X_k = 1$ בהסתברות p, $X_k = -1$ אחרת. אנו מעוניינים לתאר את חזרת S_n ל־0.

נגדיר $p_0(n) = P\{S_n = 0\}$. מתקיים $p_0(0) = 1$.

נגדיר גם $f_0(n) = P\{S_1 \neq 0, S_2 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0\}$. גם כאן $f_0(0) = 0$.

אלה בעצם ההסתברויות של סדרות המאורעות הנתונות על־ידי $A_n = \{S_n = 0\}$ ו־$B_k = \{S_1 \neq 0, \ldots, S_{k-1} \neq 0, S_n = 0\}$.

$\sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} f_0(n)$ היא ההסתברות לכך שזמן החזרה T_0 לאפס הוא סופי. מכיוון שהמאורעות B_n זרים, הסתברות זו שווה ל־$P(\bigcup_{n=1}^{\infty} B_n)$. אם $P(\bigcup B_n) = 1$, המהלך נקרא **חוזר**; אחרת, המהלך **איננו חוזר**.

נגדיר $P(s) = \sum_{n=0}^{\infty} s^n p_0(n)$ ו־$F(s) = \sum_{n=1}^{\infty} s^n f_0(n)$ עבור $0 \leq s < 1$.[50]

משפט 36: (א)[51] $P(s) = 1 + P(s)F(s)$; (ב) $P(s) = (1-4pqs^2)^{-\frac{1}{2}}$; (ג) $F(s) = 1-(1-4pqs^2)^{\frac{1}{2}}$.

הוכחה. $A_n = \bigcup_{k=1}^{n}(B_k \cap A_n)$. זהו איחוד זר, לכן מנוסחת ההסתברות השלמה נקבל $P(A_n) = \sum_{k=1}^{n} P(A_n \mid B_k)P(B_k)$. לכן $p_0(n) = \sum_{k=1}^{n} P(A_n \mid B_k) f_0(k)$.

A_n מתקיים בתנאי ש־B_k התקיים אם ורק אם $X_{k+1} + \ldots + X_n = 0$, לכן מתקיים $P\{X_{k+1} + \ldots + X_n = 0\} = P\{S_{n-k} = 0\} = p_0(n-k)$. מכאן נקבל את הנוסחה $p_0(n) = \sum_{k=1}^{n} p_0(n-k) f_0(k)$. לכן

---

[48] למעשה, לא תמיד ניתן לדבר על $G_n'(1)$ ממש, שכן הפונקציה לא בהכרח מוגדרת מעבר ל־1; כל הנגזרות הן גבולות $G^{(n)}(1) = \lim_{s\uparrow 1} G^{(n)}(s)$.

[49] מותר לדבר על הרכבה שכזו, שכן $\sum_{i=0}^{\infty} p_i s^i$ $G_Y(s)$ עבור $s \geq 0$ ופונקציה זו עולה כאשר $s \uparrow 1$; $G_Y() = 1$, אז $G_Y(s) < 1$ ב־$0 \leq s < 1$.

[50] ב־$s = 1$, הטורים אולי מתבדרים.

[51] יותר כללי מלמהלך הזה, אבל לא נרחיב למה.

$$\begin{aligned}P(s)-1 &= \sum_{n=1}^{\infty}\sum_{k=1}^{n} p_0(n-k)s^{n-k}f_0(k)s^k\\ &= \sum_{k=1}^{\infty}\sum_{n=k}^{\infty} f_0(k)s^k p_0(n-k)s^{n-k}\\ &= (\sum_{k=1}^{\infty} f_0(k)s^k)(\sum_{m=0}^{\infty} p_0(m)s^m)\\ &= F(s)P(s)\end{aligned}$$

(ב) רוצים לחשב $p_0(n)=P\{S_n=0\}$. אם $S_n=0$, אז n חייב להיות זוגי: $p_0(n)=0$ לכל n אי־זוגי. ניקח $n=2l$; יש $\binom{2l}{l}$ אפשרויות ללכת l פעמים ימינה ושמאלה. $p_0(2l)=\binom{2l}{l}(pq)^l$.

$P(s)=\sum_{l=0}^{\infty}\binom{2l}{l}(pq)^l s^{2l}$

תרגיל מבית־ספר: $\binom{2l}{l}=\frac{(2l)!}{(l!)^2}=\frac{2^l l!(1\cdot3\cdot\ldots\cdot(2l-1))}{(l!)^2}=\frac{2^l(1\cdot3\cdot\ldots\cdot(2l-1))}{l!}$

אז נוכל לכתוב $=\sum_{l=0}^{\infty}\frac{1\cdot3\cdot\ldots\cdot(2l-1)}{l!}(2pqs^2)^l$

תרגיל מאינפי: $(1-2x)^{-\frac{1}{2}}=\sum_{l=0}^{\infty}\frac{1\cdot3\cdot\ldots\cdot(2l-1)}{l!}x^l$ **(בדוק!)**. $pq\le\frac{1}{4}$ תמיד, וטור טיילור הזה ודאי מתכנס בתחום הזה. לכן נוכל להציב $x=2pqs^2$ ולקבל $(1-4pqs^2)^{-\frac{1}{2}}=$.

(ג) $F(s)=\frac{P(s)-1}{P(s)}=1-\frac{1}{P(s)}=1-(1-4pqs^2)^{\frac{1}{2}}$.

כעת, נחשב:

$F(1)=\lim_{s\to1}(1-(1-4pqs^2)^{\frac{1}{2}})=1-(1-4pq)^{\frac{1}{2}}$. $F(1)=\sum_{n=1}^{\infty}f_0(n)$ היא הסתברות החזרה, כלומר ההסתברות לכך שהמהלך חוזר בזמן סופי. נבדוק מתי היא 1:

$(1-4pq)^{\frac{1}{2}}=(1-4p(1-p))^{\frac{1}{2}}=(1-4p+4p^2)^{\frac{1}{2}}=((1-2p)^2)^{\frac{1}{2}}=((q-p)^2)^{\frac{1}{2}}=|q-p|$

אם כן, $F(1)=1-|q-p|=1$ אם $p=q=\frac{1}{2}$, אחרת $F(1)<1$.

8 שרשראות מרקוב

שרשרת מרקוב

הגדרה. סדרה של משתנים מקריים $X_1,X_2,\ldots$ מגדירה **שרשרת מרקוב** אם 17.3.2008

$$P(X_{i+1}=a\mid X_1,X_2,\ldots,X_i)=P(X_{i+1}=a\mid X_i)$$

כלומר, השרשרת "זוכרת" רק את המשתנה המקרי הקודם.

דוגמה. סדרה של משתנים בלתי־תלויים היא דוגמה לשרשרת מרקוב.

8.1 מטריצות מרקוב

מטריצת מרקוב

הגדרה. נתונה קבוצה של s מצבים $\{1,2,\ldots,s\}$. **מטריצת מרקוב** היא מטריצה P הנתונה על־ידי $P_{ij}\ge0$, כאשר $1\le i,j\le s$ ו־$\sum_{j=1}^{s}P_{ij}=1$; **הסתברויות המעבר** P_{ij} מייצגות את ההסתברות לעבור ממצב i למצב j.[52]

דוגמה. $s=3$. בכל תא מופיעה ההסתברות שנעבור מהמצב המתואר בשורה למצב המתואר בעמודה:

[52] לכל i, סכום ה־P_{ij} הוא 1, לכן הם מגדירים התפלגות.

	נאה	מעונן	גשום
גשום	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
מעונן	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
נאה	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$

הסכום בכל שורה הוא 1.

סדרה X_1, X_2, X_3 מתאימה למטריצה P_{ij} אם מתקיים $P(X_2 = j \mid X_1 = i) = P_{ij}$ ו־$P(X_3 = j \mid X_1 = k, X_2 = i) = P_{ij}$. נניח $q_i = P(X_1 = i)$ כש־$\sum_{i=1}^{s} q_i = 1$ ו־$q_i \geq 0$. ניתן לומר מה ההתפלגות המלאה של המשתנים ומהו מרחב המדגם: $\Omega = \{1, 2, \ldots, s\}^3$, ולכל $(i, j, k) \in \Omega$, $P(X_1 = i, X_2 = j, X_3 = k) = q_i p_{ij} p_{jk}$.

עבור n משתנים, מרחב המדגם יהיה $\Omega = \{1, 2, \ldots, s\}^n$, ועבור $(i_1, \ldots, i_n) \in \Omega$, $P(X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n) = q_{i_1} p_{i_1 i_2} \cdot \ldots \cdot_{i_{n-1} i_n}$.

כעת, נרצה לחשב מה ההסתברות $P(X_3 = k \mid X_1 = i)$ (מאורע k קורה היום אם קרה מאורע i שלשום):

$$\begin{aligned}
P(X_3 = k \mid X_1 = i) &= \sum_j P(X_3 = k \mid X_1 = i, X_2 = j) P(X_2 = j, X_1 = i) \\
\frac{P(X_3=k, X_1=i)}{P(X_1=i)} &= \sum_j \frac{P(X_3=k, X_2=j, X_1=i}{P(X_1=i)} \\
&= \sum_j \frac{q_i P_{ij} P_{jk}}{q_i} \\
&= \sum_j P_{ij} P_{jk}
\end{aligned}$$

זהו האיבר ה־ik במכפלת המטריצה P עם עצמה. כלומר, $P(X_3 = j \mid X_1 = i) = (P^2)_{ij}$.

נכליל:

למה 37: בשרשרת מרקוב $\{X_t\}$ עם הסתי׳ מעבר P_{ij}, $P(X_t = j \mid X_1 = i) = (P^{t-1})_{ij}$.

דוגמה (הילוך מקרי). מרחב המצבים הוא $S = \{0, 1, 2, \ldots, 100\}$. מכל נקודה הולכים ימינה בהסתברות $\frac{1}{2}$ ושמאלה בהסתברות $\frac{1}{2}$, ואם נמצאים בקצה, חוזרים פנימה בהסתברות 1: כלומר, $P_{i,i+1} = P_{i,i-1} = \frac{1}{2}$, $P_{0,1} = P_{100,99} = 1$.

8.2 התנהגות לאורך זמן

עבור t גדול, מה נוכל לומר על, למשל, $P(X_t = i \mid X_1 = 0)$?

הגדרה. עבור מטריצת מרקוב P_{ij} אומרים ש**ניתן לעבור ממצב** i_0 **למצב** j_0 אם קיימת סדרת מצבים $i_1, \ldots, i_k$ כך ש־$P_{i_0 i_1} > 0$, $P_{i_1 i_2} > 0$, $\ldots$, $P_{i_{k-1} i_k} > 0$, $P_{i_k j_0} > 0$. כלומר, בשרשרת מרקוב $X_1, X_2, \ldots$ לפי מטריצה זו מתקיים $P(X_{k+2} = j \mid X_1 = i) > 0$ – או, לפי הלמה, ש־$(P^{k+1})_{ij} > 0$.

הגדרה. מטריצת מקרוב נקראת **ארגודית**[53] אם ניתן לעבור מכל מצב לכל מצב. ארגודיות

[53] לעתים קוראים למטריצה כזו **אי־פריקה**.

אם חושבים על קבוצת המצבים $S=\{1,\ldots,s\}$ כקדקודי גרף מכוון ומסמנים קשת מ־i ל־j רק אם $P_{ij}>0$, ניתן לומר שהמטריצה ארגודית אס"ם הגרף המכוון קשיר־חזק.

דוגמה. בדוגמת ההילוך המקרי מקודם, יש קשירות חזקה: מכל נקודה אפשר להגיע לכל נקודה.

רגולריות

הגדרה. מטריצת מרקוב P היא **רגולרית** אם קיים k כך ש־$(P^k)_{ij}>0$ לכל $1\le i,j\le s$.[54]

דוגמה. בדוגמת ההילוך המקרי, תנאי זה איננו מתקיים:

$$\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}=\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

וכן הלאה.

משפט 38: עבור מטריצה רגולרית P_{ij} מתקיים $\lim_{n\to\infty}(P^n)_{ij}=\pi_j$ לכל $1\le i\le s$, כאשר π_j הוא וקטור הסתברויות המקיים $\sum_{i=1}^s \pi_i P_{ij}=\pi_j$. בניסוח הסתברותי: עבור שרשרת מרקוב רגולרית, $\lim_{n\to\infty} P(X_n=j\mid X_1=i)=\pi_j$.

התפלגות סטציונרית

ההתפלגות π נקראת **ההתפלגות הסטציונרית** עבור P.

נסמן ב־$\mathbf{1}$ את וקטור העמודה המורכב מאחדות. אם $P\ge 0$ מרטיצת מרקוב, אזי $P\cdot\mathbf{1}=\mathbf{1}$. לכן $\mathbf{1}$ הוא וקטור עצמי עם ערך עצמי 1. מכאן, קיים גם וקטור שורה עצמי π המקיים $\pi P=\pi$.

דוגמה. בדוגמת ההילוך המקרי, $\pi_i=\frac{1}{100}$ לכל $1\le i<100$ ו־$\pi_0=\pi_{100}=\frac{1}{200}$. (נשים לב שקיום π כזה לא תלוי בהיות המטריצה רגולרית.)

ניגש להוכחת המשפט:

הוכחה. נתחיל במקרה הפרטי בו $P_{ij}>0$ לכל $1\le i,j\le s$. נסמן $c=\min_{0\le i,j\le s}P_{ij}$. נסתכל על פעולת P על וקטורי עמודות $0\le u_1,\ldots,u_s$ ונראה שלכל וקטור u, $\lim_{n\to\infty}P^n u=\lambda\mathbf{1}$, כאשר λ תלוי ב־u. נסמן $T(u)=u_{max}-u_{min}$ (הקואורדינטה המקסימלית והמינימלית ב־u) ונחשב את $T(Pu)$:

$$(Pu)_i=\sum_j P_{ij}u_j\le c\cdot u_{min}+(1-c)u_{max}$$

מכאן, $(Pu)_{max}\le c\cdot u_{min}+(1-c)u_{max}$ וגם $(Pu)_{min}\ge c\cdot u_{max}+(1-c)u_{min}$, ולכן

$$T(Pu)=(Pu)_{max}-(Pu)_{min}\le-(1-2c)u_{min}+(1-2c)u_{max}=(1-2c)T(u)$$

באינדוקציה, $T(P^n u)=(1-2c)^n T(u)$. אם המטריצה גדולה מ־2×2, $1-2c<1$, לכן $T(P^n u)\overset{n\to\infty}{\to}0$ בקצב מעריכי, ולכן $\lim_{n\to\infty}P^n u=\lambda\mathbf{1}$.

$P^n(P)$ היא הכפלת P^n ב־s וקטורי עמודה שכל אחד מהם ייתן וקטור עמודה קבוע, לכן נקבל שורות זהות. נסמן שורה זו ב־π. סכום השורות של P^n הוא תמיד 1, לכן גם הגבול מקיים זאת: $\sum_{i=1}^s \pi_i=1$.

[54] יש לשים לב שאין מדובר ברגולריות במובן שמוגדר באלגברה לינארית.

$$(P) = \begin{pmatrix} -\pi P- \\ \vdots \\ -\pi P- \end{pmatrix} = (P)\begin{pmatrix} -\pi- \\ \vdots \\ -\pi- \end{pmatrix} = (\lim P^n)(P)$$. $\lim P^n = \lim P^{n+1}$, ולכן $\pi P = \pi$.

במקרה הכללי, נעזרים במקרה הראשון עבור P^k [55] כדי לראות שקיים c עבורו מתקבל $T(P^{kn}u) = (1-2c)^n T(u)$.

נניח ש־P מטריצת מרקוב ארגודית. כדי לבדוק אם היא רגולרית, מסתכלים בגרף המכוון שתיארנו על מסלולים סגורים פשוטים. מסמנים ב־d את המחלק המשותף המקסימלי של ארכי כל המסלולים הללו. P רגולרית אס"ם $d = 1$.[56] נוכיח מקרה פרטי:

20.3.2008 **טענה 39:** אם קיימים שני מסלולים סגורים פשוטים בגרף המכוון שמוגדר על־ידי P עם ארכים l_1 ו־l_2 המקיימים $(l_1, l_2) = 1$ ואפשר לעבור מכל מצב לכל מצב, אזי המטריצה P רגולרית.

הוכחה. $(P^k)_{ij} > 0$ לכל i ו־j, אז לכל i ו־j יש מסלול בגרף מ־i ל־j באורך k.

למה 1.39: אם $(l_1, l_2) = 1$, כל מספר הגדול מ־l_1 ומ־l_2 ניתן לכתיבה כסכום $al_1 + bl_2$ עבור $a, b \geq 0$.

הוכחה. בלי הגבלת הכלליות, $l_1 < l_2$. כאשר עוברים על כל הכפולות של l_1 מקבלים את כל השאריות מודולו l_2; ניתן להשתמש בהן ולקבל כל מספר מספיק גדול כנדרש.

קיים N כך שבין כל i ו־j יש מסלול באורך קטן מ־ או שווה ל־N, כי מספר הזוגות סופי ואפשר למצוא מסלול מכל i לכל j. נגדיר $k = 10N + l_1 l_2$: ניתן למצוא $a \geq 0$ ו־$b \geq 0$ שיפתרו את המשוואה $m_1(i) + al_1 + m + bl_2 + m_2(j) = k$, ולכן לכל i ו־j יש מסלול באורך k המחבר את i ל־j: כלומר, $(P^k)_{ij} > 0$.

דוגמה (ערבוב חפיסת קלפים). סידור הקלפים $1, 2, \ldots, 52$ הוא תמורה $\rho \in S_{52}$: בוחרים מקום בין 1 ל־52; מוציאים את הקלף העליון ומחזירים אותו למקומו החדש; מבצעים מכפלה בחבורת התמורות כאשר בוחרים באקראי אחת מ־52 התמורות הללו.

יש מסלול שמביא לחילוף $(i \quad 1)$ לכל $2 \leq i \leq 52$, וחילופים אלה יוצרים את כל התמורות. יש מסלול מהתמורה $(s \ \ldots \ 2 \quad 1)$ לכל תמורה, ולכן מטריצת מרקוב המוגדרת על־ידי פעולה זו היא רגולרית.

מכל ρ יוצאות 51 קשתות ולכל תמורה מגיעות 51 קשתות, אז בשרשרת מרקוב יש בכל עמודה מספר קבוע של איברים השונים מ־0. מכאן, π עבור מטריצה זו נותן משקל $\frac{1}{52!}$ לכל אחת מהתמורות.

מסקנה: אם חוזרים על פעולת הערבוב n פעמים ל־n גדול מספיק, סידור הקלפים יהיה בקירוב אחיד על כל $52!$ הסידורים האפשריים. (כדי לחשב את טיב הקירוב, כלומר מתי $51^n \approx 52!$, נוכל להיעזר בנוסחת סטירלינג.)

[55] k כך ש־$P^k > 0$ – קיים כזה, מרגולריות P.

[56] כלומר, מספיק למצוא קבוצה קטנה של מסלולים שמביאה למחלק משותף מקסימלי 1.

9 פונקציות יוצרות מומנטים

24.3.2008 מומנט

הגדרה. עבור משתנה מקרי X, $\mathbb{E}X^k$ נקרא **המומנט ה־k של** X.

פונקציה יוצרת מומנטים

הגדרה. יהי X משתנה מקרי. **הפונקציה יוצרת המומנטים** של X היא

$$M_X(t) = \mathbb{E}e^{tX} = \begin{cases} \sum e^{tx_k} P(X = x_k) & \text{בדיד ומקבל ערכים } x_k \ X \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \text{רציף עם צפיפות } f \ X \end{cases}$$

נשים לב ש־$\frac{dM_X(t)}{dt}\ |_{t=0} = \mathbb{E}X$, ... , $\frac{d^kM_X(t)}{dt^k}\ |_{t=0} = \mathbb{E}X^k$: $M_X^{(k)}(0) =$ כלומר, מנגזרות M_X ב־$t = 0$ ניתן לקבל את המומנטים.

בניגוד לפונקציה יוצרת, פונקציה יוצרת מומנטים מוגדרת גם למשתנים מקריים רציפים.

$M_X(t)$ שימושית כאשר קיים $a > 0$ כך ש־$M_X(t)$ קיים לכל $|t| < a$: נשים לב ש־$M_X(t) = \mathbb{E}e^{tX} = \mathbb{E}(\sum_{k=0}^{\infty} \frac{t^kX^k}{k!}) = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathbb{E}(X^k)$; זהו טור חזקות, שרדיוס ההתכנסות שלו יכול להיות 0 בלבד (ויש משתנים מקריים שהפונקציה יוצרת המומנטים שלהם קיימת רק ב־$t = 0$).

אם X מקבל ערכים טבעיים, $M_X(t) = \mathbb{E}e^{tX} = \sum_{k=0}^{\infty} e^{tk} f_X(k) = G_X(e^t)$, כאשר G_X היא הפונקציה היוצרת של X.

9.1 תכונות

משפט 40: יהיו X ו־Y משתנים מקריים בלתי־תלויים כך ש־$M_X(t)$ ו־$M_Y(t)$ קיימים לכל $|t| < a$ ($a > 0$). אז לכל $|t| < a$, $M_{X+Y}(t) = M_X(t)M_Y(t)$.

הוכחה. (נובע מכך שתוחלת סכום משתנים מקריים בלתי־תלויים היא מכפלת התוחלות.)

משפט 41: יהי X משתנה מקרי כך ש־$M_X(t)$ קיים לכל $|t| < a$ ($a > 0$) ונניח שעבור משתנה מקרי Y מתקיים לכל $|t| < b \leq a$ ($b > 0$) $M_X(t) = M_Y(t)$. אזי פונקציית ההתפלגות F_X של X שווה לפונקציית ההתפלגות F_Y של Y.[57][58]

אם X ו־Y משתנים מקריים ולכל $k \in \mathbb{N}$ מתקיים $\mathbb{E}X^k = \mathbb{E}Y^k$, פונקציות ההתפלגות של X ו־Y לא בהכרח שוות (ניתן לבנות דוגמה נגדית) – אלא אם $M_X(t)$ ו־$M_Y(t)$ קיימות בסביבה של 0, כלומר רדיוס ההתכנסות חיובי.[59]

משפט 42 (הרציפות): יהיו $X_1, X_2, \dots$ משתנים מקריים עם פונקציות יוצרות מומנטים $M_{X_1}(t), M_{X_2}(t), \dots$ שקיימות לכל $|t| < a$ ($a > 0$). נניח $\lim_{n\to\infty} M_{X_n}(t) = M(t)$ לכל $|t| < b \leq a$ ($b > 0$), ונניח ש־$M(t)$ היא פונקציה יוצרת מומנטים של משתנה מקרי X. אז פונקציות ההתפלגות $F_{X_1}, F_{X_2}, \dots$ מתכנסות לפונקציית ההתפלגות F_X של X בכל נקודת רציפות של F_X.

[57] נקבל מכך גם ש־$b = a$.

[58] על־אף שמשפט זה איננו טריוויאלי, לא נוכיח אותו.

[59] מכאן מתקבל תנאי מספיק לשוויון בין התפלגויות X ו־Y במקרה זה.

9.2 פונקציות יוצרות מומנטים של משתנים מוכרים

9.2.1 משתנה מקרי נורמלי

יהי $X \sim N(0,1)$ (תוחלת 0 ושונות 1). אזי

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2-2tx+t^2)}e^{\frac{1}{2}t^2}dx \\ &= e^{\frac{t^2}{2}}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2}d(x-t)\right) \\ &= e^{\frac{t^2}{2}} \end{aligned}$$

אם $X \sim N(\mu,\sigma^2)$, נגדיר $Y = \frac{X-\mu}{\sigma}$; זהו משתנה מקרי נורמלי סטנדרטי ($X = \sigma Y + \mu$).
לפי החישוב למעלה, $M_Y(t) = e^{\frac{t^2}{2}}$, ואז

$$\begin{aligned} M_X(t) = \mathbb{E}e^{tx} &= \mathbb{E}e^{t\sigma Y + t\mu} \\ &= e^{t\mu}\mathbb{E}e^{(t\sigma)Y} \\ &= e^{t\mu}M_Y(t\sigma) \\ &= e^{t\mu}e^{\frac{1}{2}\sigma^2t^2} = e^{t\mu+\frac{1}{2}\sigma^2t^2} \end{aligned}$$

אם $X \sim N(\mu_1,\sigma_1^2)$ ו־$Y \sim N(\mu_2,\sigma_2^2)$ משתנים מקריים בלתי־תלויים, פונקציית ההתפלגות של $Z = X+Y$ היא

$$\begin{aligned} M_Z(t) &= M_{X+Y}(t) \\ &= M_X(t)M_Y(t) \\ &= e^{\mu_1t+\frac{1}{2}\sigma_1^2t^2}e^{\mu_2t+\frac{1}{2}\sigma_2^2t^2} \\ &= e^{(\mu_1+\mu_2)t+\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2} \end{aligned}$$

זוהי פונקציה יוצרת מומנטים של משתנה מקרי נורמלי עם ממוצע $\mu_1+\mu_2$ ושונות $\sigma_1^2+\sigma_2^2$.
פונקציה יוצרת מומנטים מגדירה באופן יחיד את ההתפלגות, לכן $Z \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

9.2.2 התפלגות Γ

$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}$ עבור $x \geq 0$, $f(x) = 0$ אחרת.
יהי X משתנה מקרי עם התפלגות Γ עם פרמטרים λ ו־α. אז

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} e^{tx}\frac{\lambda^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}dx \\ &= \frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^{\infty} x^{\alpha-1}e^{(t-\lambda)x}dx \end{aligned}$$

עבור $t > \lambda$, אינטגרל זה קיים; על־ידי החלפת משתנה $u = (t-\lambda)x$, נקבל בסוף $M_X(t) = (\frac{\lambda}{\lambda-t})^\alpha$ לכל $t < \lambda$.

9.2.3 התפלגות אקספוננציאלית

התפלגות אקספוננציאלית מתקבלת עבור $\alpha = 1$ (פונקציית צפיפות $f(x) = \lambda e^{-\lambda x}$). אם $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי־תלויים שווי־התפלגות עם פרמטר λ, מהחישוב לעיל

נקבל $M_{X_i}(t) = \frac{\lambda}{\lambda-t}$ ו־$M_{X_1+\ldots+X_n}(t) = (\frac{\lambda}{\lambda-t})^n$.

9.2.4 התפלגות χ^2

יהיו $X_1, X_2, \ldots, X_n$ משתנים מקריים בלתי־תלויים נורמליים סטנדרטיים ($X_i \sim N(0,1)$).
נסתכל על $Z = X_1^2 + \ldots + X_n^2$: ההתפלגות שלו נקראת התפלגות χ^2 עם n דרגות חופש. זוהי התפלגות Γ עם פרמטרים $\lambda = \frac{1}{2}$ ו־$\alpha = \frac{n}{2}$:

$$\begin{aligned} M_Z(t) = (M_{X_i^2}(t))^n &= (\int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx)^n \\ &= (\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2(\frac{1}{1-2t})}} dx)^n \\ &= (\frac{1}{1-2t})^{\frac{n}{2}} \\ &= (\frac{\frac{1}{2}}{\frac{1}{2}-t})^{\frac{n}{2}} \end{aligned}$$

באופן כללי, $\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma$, וכאן $\sigma^2 = \frac{1}{1-2t}$; התוצאה $(\frac{\frac{1}{2}}{\frac{1}{2}-t})^{\frac{n}{2}}$ מתארת התפלגות Γ עם פרמטרים $\lambda = \frac{1}{2}$ ו־$\alpha = \frac{n}{2}$, כפי שרצינו להראות.

10 משפט הגבול המרכזי

משפט 43 (הגבול המרכזי): יהיו $X_1, X_2, \ldots$ משתנים מקריים בלתי־תלויים שווי־התפלגות עם $\mathbb{E}X_1 = \mu$, $\mathrm{Var}\, X_1 = \sigma^2$. נגדיר $Y_n = \frac{\sum_{i=1}^n (X_i - \mu)}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$. אז לכל x, 27.3.2008

$$P\{Y_n \le x\} \overset{n\to\infty}{\to} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du = \Phi(x)$$

כלומר $Y_n \Rightarrow Z \sim N(0,1)$.

הוכחה (לא בדיוק). נוכיח רק למקרה ש־$M_{X_1}(t) = \mathbb{E}e^{tX_1}$ קיימת לכל $t \in (-a,a)$.

$$\begin{aligned} M_{Y_n}(t) &= (M_{\frac{X_1-\mu}{\sigma\sqrt{n}}}(t))^n = (\mathbb{E}e^{t\frac{X_1-\mu}{\sigma\sqrt{n}}})^n \\ &= \left(1 + \mathbb{E}(t\frac{X_1-\mu}{\sigma\sqrt{n}}) + \frac{1}{2}\mathbb{E}(t^2\frac{(X_1-\mu)^2}{\sigma^2 n} + o(\frac{1}{n})\right)^n \end{aligned}$$

אבל $\mathbb{E}(X_1 - \mu) = 0$ ואז $\mathbb{E}(X_1-\mu)^2 = \sigma^2$, ו־$\frac{o(\frac{1}{n})}{\frac{1}{n}} \to 0$ ולכן $o(\frac{1}{n}) < \frac{1}{n}$. לכן

$$\begin{aligned} &= (1 + 0 + \frac{1}{2} \cdot \frac{t^2}{n} + o(\frac{1}{n}))^n \\ &\to e^{\frac{t^2}{2}} = \mathbb{E}(e^{tZ}) \end{aligned}$$

עבור $Z \sim N(0,1)$. לפי משפט הרציפות ומשפט היחידות, לכל x $\mathbb{E}_{Y_n}(x) \overset{n\to\infty}{\to} \Phi(x)$.

דוגמה. יהי $X \sim N(0,1)$. נתבונן בסדרה $X, -X, X, -X, \ldots$. כאן יש התכנסות בהתפלגות אך לא התכנסות אחרת.

דוגמה (קירוב נורמלי). X הוא מספר הפעמים שמתקבל פלי בהטלת מטבע כשר בין 40 הטלות. נרצה לחשב את $P\{X = 21\}$. הערך המדוייק הוא $P\{X = 21\} = \binom{40}{21}(\frac{1}{2})^{40}$. ניתן גם למצוא רק קירוב. התוחלת של X היא 20 והשונות היא 10.

$$\begin{aligned} P\{X=21\}=P\{20.5<X\le 21.5\} &= P\{0.5<X-20\le 1.5\} \\ &= P\{\tfrac{0.5}{\sqrt{10}}<\tfrac{X-20}{\sqrt{10}}\le\tfrac{1.5}{\sqrt{10}}\} \\ &\approx \Phi(\tfrac{1.5}{\sqrt{10}})-\Phi(\tfrac{0.5}{\sqrt{10}}) \end{aligned}$$

(החלוקה היא ב־$\sqrt{10}$ כי $\sqrt{10}=\frac{1}{2}\sqrt{40}=\sigma\sqrt{n}$. בנוסף, השתמשנו בכך שאם $P\{Y_n\le b\}\to\Phi(b)$ ו־$P\{Y_n\le a\}\to\Phi(a)$ אזי $P\{a<Y_n\le b\}\to\Phi(b)-\Phi(a)$.)

דוגמה (חביבה לבחינות). הוכחנו כי $e^{-n\lambda}\sum_{k=0}^n f(\frac{k}{n})\frac{(n\lambda)^k}{k!}\to f(\lambda)$ עבור $0<\lambda<1$ כאשר f רציפה וחסומה; עכשיו נראה באמצעות חוק המספרים הגדולים בצורה החלשה שעבור $\lambda=1$, ביטוי זה שואף ל־$\frac{1}{2}f(1)$.

הוכחה. ראשית נוכיח $e^{-n}\sum_{k=0}^n \frac{n^k}{k!}\overset{n\to\infty}{\to}\frac{1}{2}$: ביטוי זה הוא, עבור X_i פואסון עם פרמטר 1,

$$\begin{aligned} P\{X_1+\ldots+X_n\le n\} &= P\{\textstyle\sum_{i=1}^n(X_i-1)\le 0\} \\ &= P\{\frac{\sum_{i=1}^n(X_i-1)}{\sqrt{n}}\le 0\} \\ &\to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^0 e^{-\frac{u^2}{2}}du=\frac{1}{2} \end{aligned}$$

(הפונקציה $h(u)=e^{-\frac{u^2}{2}}$ זוגית, לכן $\int_{-\infty}^0 h(u)du=\int_0^\infty h(u)du=\frac{1}{2}$.)

כעת,

$$\begin{aligned} e^{-n}\sum_{k=0}^n f(\tfrac{k}{n})\frac{n^k}{k!} &= e^{-n}\sum_{k=0}^n(f(\tfrac{k}{n}-f(1))\frac{n^k}{k!}+ \\ &\quad +f(1)e^{-n}\sum_{k=0}^n\frac{n^k}{k!} \end{aligned}$$

יודעים ש־$\frac{1}{2}\to e^{-n}\sum_{k=0}^n\frac{n^k}{k!}$, לכן נותר להראות $e^{-n}\sum_{k=0}^n(f(\frac{k}{n})-f(1))\frac{n^k}{k!}\to 0$.

יהי $\varepsilon>0$; נבחר $\delta>0$ כך שאם $|x-1|<\delta$ אז $|f(x)-f(1)|\le\varepsilon$. אז

$$\begin{aligned} e^{-n}\sum_{k=0}^n(f(\tfrac{k}{n})-f(1))\frac{n^k}{k!} &\le e^{-n}\sum_{k:1-\frac{k}{n}>\delta}|f(\tfrac{k}{n})-f(1)|\frac{n^k}{k!}+ \\ &\quad +e^{-n}\sum_{k:1-\frac{k}{n}\le\delta,k\le n}|f(\tfrac{k}{n})-f(1)|\frac{n^k}{k!} \\ &\le 2Me^{-n}\sum_{k:1-\frac{k}{n}>\delta}\frac{n^k}{k!}+\varepsilon \end{aligned}$$

כאשר $M=\sup_{x\in[0,1]}f(x)$.

נשים לב ש־$\sum_{k:1-\frac{k}{n}>\delta}e^{-n}\frac{n^k}{k!}=P\{1-\frac{X_1+\ldots+X_n}{n}>\delta\}$,[60] ולפי החוק החלש של המספרים הגדולים, $P\{1-\frac{X_1+\ldots+X_n}{n}>\delta\}\to 0$, כנדרש.

11 תהליך פואסון

31.3.2008 **הגדרה.** ***תהליך פואסון*** זאת משפחת משתנים מקריים $N=\{N(t),t\ge 0\}$ המקבלים ערכים טבעיים כך ש־ תהליך פואסון

א. $N(t)$ משפחה עולה, כלומר אם $t>s$ אז $N(t)\ge N(s)$, ו־$N(0)=0$;

[60] הסכום "אוסף" את ההסתברות לקבל k־ים המקיימים את התנאי, כאשר k הוא הערך של המשתנה המקרי $X_1+\ldots+X_n\sim P(n)$ (שהרי $X_i\sim P(1)$).

ב. אם $s < t$ אז $N(t) - N(s)$ בלתי־תלוי בכל $N(u)$ עבור $u \le s$;

ג. $$P\{N(t+h) - N(t) = k\} = \begin{cases} \lambda h + o(h) & k = 1 \\ 1 - \lambda h + o(h) & k = 0 \end{cases}$$

לכן נובע שאם $k > 1$ אז $o(h)$, ולמעשה $\sum_{k>1} P\{N(t+h) - N(t) = k\} = o(h)$: סכום ההסתברויות הוא 1, ועבור $k = 0, 1$ מקבלים $1 + o(h)$. לכן ההסתברות לכל $k > 1$ צריכה גם להיות $o(h)$.

משמעות t היא זמן.

משפט 44: $N(t)$ הוא משתנה מקרי פואסון עם פרמטר λt, כלומר $P\{N(t) = k\} = e^{-\lambda t}\frac{(\lambda t)^k}{k!}$.

הוכחה. נסמן $P_k(t) = P\{N(t) = k\}$.

$$\begin{aligned} P_k(t+h) &= P\{N(t+h) = k\} \\ &= \textstyle\sum_{m=0}^{k} P\{N(t+h) = k \mid N(t) = m\} P\{N(t) = m\} \\ &= \textstyle\sum_{m=0}^{k} P\{N(t+h) - N(t) = k - m \mid N(t) = m\} P\{N(t) = m\} \\ &\overset{\text{ב'}}{=} \textstyle\sum_{m=0}^{k} P\{N(t+h) - N(t) = k - m\} P\{N(t) = m\} \\ &= \textstyle\sum_{m=0}^{k-2} P\{N(t+h) - N(t) = k - m\} P\{N(t) = m\} + \\ &\quad + P\{N(t+h) - N(t) = 1\} P\{N(t) = k-1\} + \\ &\quad + P\{N(t+h) - N(t) = 0\} P\{N(t) = k\} \\ &= o(h) + (\lambda h + o(h)) P_{k-1}(t) + (1 - \lambda h + o(h)) P_k(t) \end{aligned}$$

לפיכך, $P_k(t+h) = \lambda h P_{k-1}(t) + P_k(t) - \lambda h P_k(t) + o(h)$. נחלק ב־$h$ ונקבל $\frac{P_k(t+h) - P_k(t)}{h} = \lambda P_{k-1}(t) - \lambda P_k(t) + \frac{o(h)}{h}$. כאשר $h \to 0$, $P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t)$.

זה נכון עבור $k \ge 1$. עבור $k = 0$ נקבל

$$\begin{aligned} P_0(t+h) &= P\{N(t+h) = 0\} \\ &= P\{N(t+h) - N(t) = 0 \mid N(t) = 0\} P\{N(t) = 0\} \\ &= P\{N(t+h) - N(t) = 0\} P\{N(t) = 0\} \\ &= (1 - \lambda h + o(h)) P_0(t) \end{aligned}$$

לפיכך, $P_0(t+h) = (1 - \lambda h + o(h)) P_0(t)$.

נחלק ב־h ונקבל $\frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \frac{o(h)}{h}$. כאשר $h \to 0$, $P_0'(t) = -\lambda P_0(t)$.

קיבלנו מערכת משוואות: $P_k'(t) = \lambda P_{k-1}(t) - \lambda P_k(t)$ ו־$P_0'(t) = -\lambda P_0(t)$, כלומר $P_0(t) = e^{-\lambda t}$, כאשר תנאי ההתחלה הוא $P_0(0) = 1$.

נגדיר $q_k(t) = e^{\lambda t} P_k(t)$. אז

$$\begin{aligned} q_k'(t) &= \lambda e^{\lambda t} P_k(t) + e^{\lambda t} P_k'(t) \\ &= \lambda q_k(t) + e^{\lambda t}(\lambda P_{k-1}(t) + \lambda P_k(t)) \\ &= \lambda q_k(t) + \lambda q_{k-1}(t) - \lambda q_k(t) \\ &= \lambda q_{k-1}(t) \end{aligned}$$

קיבלנו $q_0(t)=1$, $q_k'(t)=\lambda q_{k-1}(t)$. מכאן, $q_1'(t)=\lambda$ ולכן $q_1(t)=\lambda t$.[61] נמשיך ונקבל $q_2'(t)=\lambda(\lambda t)$ ולכן $q_2(t)=\frac{\lambda^2 t^2)}{2}=\frac{(\lambda t)^2}{2}$ וכו', עד $q_k(t)=\frac{(\lambda t)^k}{k!}$. לכן $P_k(t)=e^{-\lambda t}\frac{(\lambda t)^k}{k!}$.

דוגמה: הטלפון של סימאון פואסון

נסמן $T_k=\inf\{t\ :\ N(t)=k\}$ – זמן ההגעה של שיחת טלפון מספר k.

משפט 45: נגדיר $X_k=T_k-T_{k-1}$ עבור $k=1,2,\ldots$. אז $X_1,X_2,\ldots$ משתנים מקריים בלתי־תלויים מעריכיים: $P\{X_1\in[a,b]\}=\int_a^b \lambda e^{-\lambda t}dt$, $P\{X_1\le t\}=1-e^{-\lambda t}$ כאשר $a\ge 0$.

יהיו $X_1,X_2,\ldots$ משתנים מקריים בלתי־תלויים שווי־התפלגות עם התפלגות אקספוננציאלית עם פרמטר λ, ונגדיר $T_0=0$, $T_k=X_1+X_2+\ldots+X_k$.

נגדיר $N(t)=\max\{k\ge 0\mid T_k\le t\}$.

משפט 46: $N(t)$ משתנה פואסון עם פרמטר λt, כלומר $P\{N(t)=k\}=e^{-\lambda t}\frac{(\lambda t)^k}{k!}$.

הוכחה. מצאנו קודם של־T_k יש התפלגות Γ: $P\{T_k\le t\}=\int_0^t\frac{\lambda^k s^{k-1}}{(k-1)!}e^{-\lambda s}ds$. (כאשר $k\to\infty$, ביטוי זה שואף ל־0, לכן קיים מקסימום; זה מתקיים כי סדרת המאורעות יורדת: $\{X_1+\ldots+X_{k+1}\le t\}\supset\{X_1+\ldots+X_k\le t\}$.)

נשים לב כי $\{T_k\le t\}=\{N(t)\ge k\}$. באמצעות זאת, נקבל

$$\begin{aligned}P\{N(t)=k\} &= P\{N(t)\ge k\}-P\{N(t)\ge k+1\}\\ &= \{T_k\le t\}-\{T_{k+1}\le t\}\\ &= \int_0^t\frac{\lambda^k s^{k-1}}{(k-1)!}e^{-\lambda s}ds-\int_0^t\frac{\lambda^{k+1}s^k}{k!}e^{-\lambda s}ds\end{aligned}$$

נחשב ונקבל

$$\begin{aligned}\int_0^t\frac{\lambda^{k+1}s^k}{k!}e^{-\lambda s}ds &= \frac{\lambda^{k+1}}{k!}\int_0^t s^k e^{-\lambda s}ds\\ (\text{אינטגרציה בחלקים}\to) &= \frac{\lambda^{k+1}}{k!}\left|\begin{array}{cc} s^k & e^{-\lambda s}ds\\ ks^{k-1} & -\frac{1}{\lambda}e^{-\lambda s}\end{array}\right|\\ &= \frac{\lambda^{k+1}}{k!}\left(-\frac{1}{k}s^k e^{-\lambda s}\right)\Big|_0^t+\frac{\lambda^{k+1}}{k!}\cdot\frac{k}{\lambda}\int_0^t s^{k-1}e^{-\lambda s}ds\\ &= -\frac{\lambda^k}{k!}t^k e^{-\lambda t}+\frac{\lambda^k}{(k-1)!}\int_0^t s^{k-1}e^{-\lambda s}ds\end{aligned}$$

לכן נקבל $P\{N(t)=k\}=\frac{\lambda^k}{k!}t^k e^{-\lambda t}=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$, כנדרש.

[61] קבוע האינטגרציה הוא 0 מכיוון ש־$P_k(0)=0$ עבור $k>1$.