סיכום משפטים חשובים בהסתברות 2:

כפיר גולדפרב 2021

אי-שוויוניים:

אי-שוויון מרקוב:

יהי X משתנה מקרי המקבל ערכים אי-שליליים בעל תוחלת סופית אזי לכל $\mathbb{R} \ni t > 0$ מתקיים:

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

אי-שוויון צ'בישב:

יהי X משתנה מקרי בעל שונות סופית אז לכל $\mathbb{R} \ni t > 0$ מתקיים:

$$P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$$

אי-שוויון צ'רנוף:

יהיו $X_1, \dots, X_n$ משתנים מקרים בלתי תלויים המחזירים ערכים בטווח $[0,1]$ וגם $X = \sum_{i=1}^{n} X_i$, אזי:

1. לכל $t > 0$ מתקיים: $P(X \geq E(X) + t) \leq e^{-\frac{2t^2}{n}}$
2. לכל $t > 0$ מתקיים: $P(X \leq E(X) - t) \leq e^{-\frac{2t^2}{n}}$
3. לכל $t > 0$ מתקיים: $P(|X - E(X)| \geq t) \leq 2e^{-\frac{2t^2}{n}}$
4. לכל $\varepsilon > 0$ מתקיים: $P\big(X \leq (1-\varepsilon)E(X)\big) \leq e^{-\frac{\varepsilon^2 E(X)}{3}}$
5. לכל $0 < \varepsilon \leq 3/2$ מתקיים: $P\big(X \geq (1+\varepsilon)E(X)\big) \leq e^{-\frac{\varepsilon^2 E(X)}{3}}$

אי-שוויון צ'רנוף-הופדינג (מקרה פרטי של אי-שוויון צ'רנוף):

יהיו $X_1, \dots, X_n$ משתנים מקרים בלתי תלויים כך שלכל $1 \leq i \leq n$ מתקיים: $P(X_i = 1) = P(X_i = -1) = \frac{1}{2}$
ויהי: $X = \sum_{i=1}^{n} X_i$, אזי לכל $\mathbb{R} \ni t > 0$ מתקיים:

$$P(X \geq t) \leq e^{-\frac{t^2}{2n}} \text{ וגם } P(X \leq -t) \leq e^{-\frac{t^2}{2n}}$$

משפטי גבולות:

החוק החלש של המספרים הגדולים:

תהי $\{X_n\}_{i=1}^{\infty}$ סדרה של משתנים מקרים בלתי תלויים שווי התפלגות, בעלי תוחלת סופית μ,

אזי לכל $\mathbb{R} \ni \varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = 0$$

החוק החזק של המספרים הגדולים:

תהי $\{X_n\}_{i=0}^{\infty}$ סדרה של משתנים מקרים בלתי תלויים שווי התפלגות, בעלי תוחלת סופית μ, ושונות סופית,
אזי:

$$P\left(\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu\right) = 1$$

התכנסות בהסתברות:

נאמר ש-$\{X_n\}_{i=0}^{\infty}$ סדרה של משתנים מקרים <u>מתכנסת בהסתברות</u> למשתנה מקרי X (סימון: $X_n \xrightarrow{P} X$), אזי לכל $\mathbb{R} \ni \varepsilon > 0$ מתקיים:

$$\lim_{n\to\infty} P\left(\underbrace{|X_n - X| \geq \varepsilon}_{\{\omega\in\Omega:|X_n(\omega)-X(\omega)|\geq\varepsilon\}}\right) = 0$$

דוגמה:

לפי החוק החלש והתכנסות בהסתברות מתקיים:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

התכנסות בהסתברות כמעט תמיד בהסתברות 1:

נאמר ש-$\{X_n\}_{i=0}^{\infty}$ סדרה של משתנים מקרים <u>מתכנסת בהסתברות של כמעט תמיד בהסתברות 1</u> למשתנה מקרי X (סימון: $X_n \xrightarrow{a.s.} X$) אם:

$$P\left(\underbrace{\lim_{n\to\infty} X_n = X}_{\left\{\omega\in\Omega:\lim_{n\to\infty} X_n(\omega)=X(\omega)\right\}}\right) = 1$$

דוגמה:

לפי החוק החזר והתכנסות בהסתברות כמעט תמיד בהסתברות 1 מתקיים:

$$\frac{X_1 + \cdots + X_n}{n} \xrightarrow{a.s.} \mu$$

טענה:

התכנסות בהסתברות כמעט תמיד בהסתברות 1 גוררת התכנסות בהסתברות אבל ההיפך לא תמיד נכון.

השיטה ההסתברותית:

פתרון בעיות קומבינטוריות שמנסים לפתור בעיות קיום, כדי להראות שמשהו קיים – להראות שההסתברות שלו גדולה מ-0 אחרת שווה ל-0.

מספר רמזי: $\boldsymbol{R(k,l)}$

מהו הגרף עם מספר הקודקודים n המינימאלי כך שיש בו בוודאות קליקה בגודל k או קבוצה בלתי תלויה בגודל l.

<u>קליקה</u> = תת גרף שלם (שכל הקודקודים בו מחוברים לכולם).

<u>קבוצה בלתי תלויה</u> = תת גרף ריק (שאף קודקוד לא מחובר לאף קודקוד בקבוצה).

טענה:

אם $\binom{n}{t}\cdot 2^{1-\binom{t}{2}}$ אז $R(t,t)>n$.

כלומר n הוא חסם תחתון עובר מספר רמזי כאשר $k=l=t$.

רעיון ההוכחה:

נבנה גרף עם n קודקודים באופן אקראי, כלומר לכל צלע אופציונאלית בגרף נטיל מטבע הוגן, אם יצא פלי נשים את הצלע בגרף, אם יצא עץ לא נשים את הצלע בגרף, ולכן כל צלע בגרף - קיימת בהסתברות של $\frac{1}{2}$.

על מנת להוכיח טענה זו נשתמש בשיטה ההסתברותית, נבחר גרף אקראי, ונראה שההסתברות שהגרף לא מקיים את הטענה שווה לאפס כלומר בהכרח הטענה תמיד מתקיימת.

הוכחה:

לאחר שבנינו גרף עם n קודקודים באופן אקראי ולכל צלע נטיל מטבע הוגן, נגדיר $S_1,S_2,\dots,S_{\binom{n}{t}}$ כל תת הקבוצות של $\{1,\dots,n\}$ מגודל t, לכל $1\le i\le\binom{n}{t}$ יהי A_i המאורע $G[S_i]$ הוא קליק או קבוצה בלתי תלויה, לכל $1\le i\le\binom{n}{t}$ מתקיים:

$$P(A_i)=\left(\frac{1}{2}\right)^{\binom{n}{t}}+\left(\frac{1}{2}\right)^{\binom{n}{t}}=2\left(\frac{1}{2}\right)^{\binom{n}{t}}=2^{1-\binom{n}{t}}$$

נגדיר G טוב – אם G <u>אינו</u> מכיל קליק בגודל t ואינו מכיל קבוצה בלתי תלויה בגודל t, אחרת G רע.

ולכן:

$$P\left(G\ \text{רע}\right)=P\left(\bigcup_{i=1}^{\binom{n}{t}}A_i\right)\le\sum_{i=1}^{\binom{n}{t}}P(A_i)=\binom{n}{t}\cdot 2^{1-\binom{n}{t}}<1$$

מסקנה:

קיבלנו שההסתברות שיהיה קליקה או קבוצה בלתי תלויה בגרף קטנה ממש מ-1 וזה אומר שקיימת הסתברות שלא יהיה קליקה וגם לא קבוצה בלתי תלויה ולכן יש גרף עם n קודקודים שלא מכיל קליקה או קבוצה בלתי תלויה בגודל t ולכן רמזי צריך יותר מ-n.

שיטת המומנט הראשון (מקרה פרטי של אי-שוויון מרקוב, משתמשים כדי להראות קיום):

יהי X משתנה מקרי אי-שלילי, נניח ש-X תלוי באיזה שהוא פרמטר $n\in\mathbb{N}$, בגדול נרצה לדעת האם $X\neq 0$ או $X=0$, נשים לב כי:

$$\lim_{n\to\infty}P(X\neq 0)=\lim_{n\to\infty}P(X>0)\underbrace{=}_{\substack{\text{מקבל ערכים}\\ \text{אי}-\text{שליליים}}}\lim_{n\to\infty}(P\ge 1)\underbrace{\le}_{\text{מרקוב}}\lim_{n\to\infty}\frac{E(X)}{1}=\lim_{n\to\infty}E(X)$$

מסקנה:

אם $\lim_{n\to\infty}E(X)=0$ אז $\lim_{n\to\infty}P(X=0)=1$

שיטת המומנט השני (מקרה פרטי של אי-שוויון צ'בישב, משתמשים כדי להראות שמשהו לא קיום):

מתי נשתמש במומנט השני? כאשר התוחלת שואפת ל-1 ומעלה ואז לא נותנת לנו שום חסם.

$$P(X=0)=P(X-E(X)=-E(X))\leq P\big(|X-E(X)|=E(X)\big)$$
$$\leq P\big(|X-E(X)|\geq E(X)\big)\underbrace{\leq}_{\text{צבישב}}\frac{Var(X)}{E(X)^2}$$

המטרה הסופית היא להראות של $P(X=0)\leq\cdots\to 0$ וכן $P(X=0)\to 0$.

גרפים אקראיים:

גרפים אקראיים הם מודל בשיטה ההסתברות (הוכחות קיום או אי-קיום דרך הסתברות) על מנת להוכיח תכונות בגרפים, קיימים 2 מודלים של גרפים אקראיים:

1. $\boldsymbol{G(n,m)}$ – מרחב ההסתברות של כל הגרפים עם n קודקודים ו-m צלעות.
   ההסתברות לכל גרף ספציפי לבחור m מקומות מתוך ה-$\binom{n}{2}$ הפוטנציאליים לצלע.
   לכן זה 1 חלקי כמות הגרפים השונים שיש עם n קודקודים ו-m צלעות: $\binom{\binom{n}{2}}{m}$.
   כלומר: $P\big(G\in G(n,m)\big)=\frac{1}{\binom{\binom{n}{2}}{m}}$ (התפלגות אחידה).
2. $\boldsymbol{G(n,p)}$ – מרחב הסתברות של כל הגרפים עם n קודקודים כאשר כל צלע בגרף נמצאת בהסתברות של p, כלומר: $P\big(G\in G(n,p)\big)=p^{|E|}\cdot(1-p)^{\binom{n}{2}-|E|}$ (לא התפלגות אחידה).

תכונות:

תכונות של גרפים (כמו גרף קשיר, קודקוד בודד וכו') תסומן ב-Q ו-G גרף המקיים אותה יסומן כ-$G\in Q$.

תכונה מונוטונית היא עולה/יורדת אם נוסיף/נוריד צלעות מהגרף התכונה תשמר בגרף G.

פורמאלית:

$P(G(n,p_1)\in Q)\leq P(G(n,p_2)\in Q)$ אם $p_1\leq p_2$ או $P(G(n,m)\in Q)\leq P(G(n,m+1)\in Q)$.

דוגמאות:

1. קשירות – תכונה מונוטונית עולה.
2. קבוצה בלתי תלויה – תכונה מונוטונית יורדת.
3. מעגל בגרף – תכונה מונוטונית עולה.

פונקציית סף:

פונקציית סף היא פונקציה של הסתברות, לכל גרף עם n קודקודים יש הסתברות לצלע התלויה ב-n : $p(n)$, שואלים מהו הגבול בו עוברים מלא לקיים איזה שהיא תכונה לכן לקיים איזה שהיא תכונה, אם לא מקיימים את תכונה Q ההסתברות שואפת ל-0, אם מקיימים את התכונה ההסתברות שואפת ל-1, בצורה פורמאלית:

פונקציה $p(n)$ היא סף אם:

1. $p=\omega(p(n))$ אז $P(G(n,p)\in Q)\to 1$.
2. $p=o(p(n))$ אז $P(G(n,p)\in Q)\to 0$.

במילים:

1. אם לא מקיימים את התכונה אז ההסתברות שואפת ל-1.
2. אם כן מקיימים את התכונה אז ההסתברות שואפת ל-0.

דוגמה:

מציאת סף עבור קיום קודקוד בודד בגרף $G(n,p)$:

נסמן ב-X_i – אינדיקטור ש-v_i הוא קודקוד בודד, ומתקיים $X=\sum_{i=1}^{n} X_i$,

אם $X=0$ אז אין קודקודים בודדים.

נרצה לדעת מתי אין קודקודים בודדים, ההסתברות שלא יהיה קודקוד בודד:

$$P(X=0)=1-P(X\geq 1)$$

לפי מומנט ראשון (מרקוב):

$$P(X\geq 1)\leq E(X)$$

נחשב את התוחלת:

$$E(X)=E\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} E(X_i)=\sum_{i=1}^{n} P(\text{בודד } v_i)=\sum_{i=1}^{n}(1-p)^{n-1}=n\cdot(1-p)^{n-1}$$

סה"כ:

$$P(X=0)=1-P(X\geq 1)\geq 1-E(X)=1-n\cdot(1-p)^{n-1}$$

המטרה היא שההסתברות תשאף ל-1 ולכן צריך לדרוש שהביטוי $n\cdot(1-p)^{n-1}$ ישאף ל-0.

טענת עזר (לפי $\boldsymbol{Euler}$): $(1-x)^n\leq e^{-xn}$, $x=e^{\ln(x)}$

מכאן $n(1-p)^{n-1}\leq e^{\ln(n)}\cdot e^{-pn}\cdot e^{p}=e^{ln(n)-pn+p}$

נדרוש: $e^{\ln(n)-pn+p}\ll 0$ או במילים אחרות: $e^{\ln(n)-np+p}=O(1)$, ולכן: $\ln(n)-np+p\ll 0$,

נעביר אגפים: $\ln(n)\ll p(n-1)\leftarrow \frac{\ln(n)}{n-1}\ll p$

ולכן: $p=\omega\left(\frac{\ln(n)}{n-1}\right)$, נשמיט את ה-1 כי n קבוע: $p=\omega\left(\frac{\ln(n)}{n}\right)$ ואז פונקציית הסף היא $p(n)=\frac{\ln(n)}{n}$.

מסקנה: אם $p=\omega\left(\frac{\ln(n)}{n}\right)$ אז ההסתברות שלא יהיה קודקוד שואפת ל-1.

דוגמה:

יהי $G\sim G(n,p)$ כאשר $p=\frac{10\ln(n)}{n}$, יהי δ מספר המינימלי של דרגות ה-G, ויהי Δ מספר מקסימלי של צלעות ב-G,הוכח את הטענות הבאות:

1. $\lim_{n\to\infty} P(\delta\geq \ln(n))=1$
2. $\lim_{n\to\infty} P(\Delta\leq 20\ln(n))=1$

פתרון 1:

כאשר רוצים להוכיח גבול שואף ל-1, צריך להראות שהמשלים שואף ל-0, המשלים: $1-P(\delta<\ln(n))$

נגדיר $\deg i$ להיות הדרגה של קודקוד $v_i\in G$, כעת נחשב:

$$P(\delta < \ln(n)) = P\left(\bigcup_{i=1}^{n} \deg i < \ln(n)\right) \le \sum_{i=1}^{n} P(\deg i < \ln(n))$$

נחשב את התוחלת של $\deg i$, נשים לב כי הערכים ש-$\deg i$ יכול להיות הוא בין 0 לבין $n-1$, והוא מתפלג בינומית $Bin(n-1,p)$, ולכן $E(\deg i) = (n-1)p$, נציב $p = \frac{10\ln(n)}{n}$, ונקבל:

$$E(\deg i) = \frac{(n-1)10\ln(n)}{n}$$

כאשר n שואף ל-∞ ניתן לומר:

$$\sum_{i=1}^{n} P(\deg i < \ln(n)) \le \sum_{i=1}^{n} P\left(\deg i \le \frac{1}{2} \cdot E(\deg i)\right)$$

לפי צ'רנוף (2) עבור $\varepsilon = \frac{1}{2}$ נקבל:

$$\sum_{i=1}^{n} P\left(\deg i \le \frac{1}{2} \cdot E(\deg i)\right) \le \sum_{i=1}^{n} e^{-\frac{\frac{1}{4}E(\deg i)}{2}} = ne^{\ln\left(n^{-\frac{5}{4}}\right)} \to 0$$

ולכן קיבלנו:

$$\lim_{n\to\infty} P(\delta \ge \ln(n)) = 1 - \lim_{n\to\infty} P(\delta < \ln(n)) = 1$$

פתרון 2:

כאשר רוצים להוכיח גבול שואף ל-1, צריך להראות שהמשלים שואף ל-0, המשלים: $1 - P(\Delta > 20\ln(n))$

נגדיר $\deg i$ הדרגה של קודקוד $v_i \in G$, כעת נחשב:

$$P(\Delta > 20\ln(n)) = P\left(\bigcup_{i=1}^{n} \deg i > 20\ln(n)\right) \le \sum_{i=1}^{n} P(\deg i > 20\ln(n)) \le \sum_{i=1}^{n} P(\deg i \ge 20\ln(n))$$
$$\le \sum_{i=1}^{n} P\big(\deg i \ge 2E(\deg i)\big)$$

לפי חסם צ'רנוף (3) עובר $\varepsilon = 1$ נקבל:

$$\sum_{i=1}^{n} P\big(\deg i \ge 2E(\deg i)\big) \le \sum_{i=1}^{n} e^{-1\cdot\frac{E(\deg i)}{3}} = \sum_{i=1}^{n} e^{-\frac{10(n-1)\ln(n)}{3n}} \to ne^{-\frac{10\ln(n)}{3}} = n \cdot n^{-\frac{10}{3}} \to 0$$

ולכן קיבלנו:

$$\lim_{n\to\infty} P(\Delta \le 20\ln(n)) = 1 - \lim_{n\to\infty} P(\Delta > 20\ln(n)) = 1$$

ראינו שסף הוא עם חישוב של סדרי גודל התלוי ב-n למשל $p(n)=\frac{1}{n}$, אבל מה קורה כאשר יש הסתברות לתכונה מסוימת בהסתברות של $c\cdot\frac{1}{n}$?

סף חד:

p_0 ייקרא סף חד של תכונה Q מונוטונית עולה אם:

$P(G(n,p)\in Q)\to 1$ אז $p\geq(1+\varepsilon)\cdot p_0$.

$P(G(n,p)\in Q)\to 0$ אז $p=(1-\varepsilon)\cdot p_0$.

בעוד שלכל תכונה מונוטונית עולה יש סף, לא בהכרח שיש לה סף חד.

אלגוריתמים אקראיים:

יש 3 סוגים של אלגוריתמים אקראיים שאיתם נשתמש:

1. לאס וגאס – אלגוריתם שתמיד צודק בתשובה אבל זמן הריצה שלו תלוי במקרה המקרי שלו.
2. מונטה קרלו (אלגוריתם שטועה בהסתברות מסוימת (בדרך כלל נמוכה)):
   a. טעות חד-צדדית – אם צריך להחזיר $true$ אז האלגוריתם תמיד צודק, אבל אם צריך להחזיר $false$ האלגוריתם יכול להחזיר $true$ בהסתברות נמוכה (או להיפך).
   b. טעות דו-צדדית – בכל מקרה יש הסתברות נמוכה של טעות.

דוגמה:

האם שני פולינומים שווים?

קלט: $F(x),G(x)$ פונקציות פולינומיות מסדר d (החזקה הגבוהה ביותר).

פלט: $true$ אם $F(x)\equiv G(x)$ לכל x בתחום, $false$ אם זו לו אותה פונקציה.

אלגוריתם 1:

1. בחר מספר מהתחום $k\in S$, נגדיר $S=\{1,2,3,\dots,100d\}$
2. הצב את k ב-$F(x)$ ו-$G(x)$ וחשב.
3. אם $F(k)=G(k)$ החזר $true$, אחרת החזר $false$.

סיבוכיות זמן ריצה: $O(d)$ (עבור מכפלות, חיבורים וחזקות).

אם $F(x)\equiv G(x)$ אז אין הסתברות לטעות,

אם $F(x)\not\equiv G(x)$ ההסתברות לטעות היא שנבחר את אחת מלכל היותר d נקודות החיתוך לכן ההסתברות היא לכל היותר $\frac{1}{100}=\frac{d}{100d}=\frac{d}{|A|}$

אלגוריתם 2 (אלגוריתם 1 אבל עם חזרות):

1. בצע עבור i מ-1 עד k:
   a. בחר מספר $k\in\{1,2,\dots,2d\}$
   b. הצב את k ב-$F(x)$ ו-$G(x)$ וחשב.
   c. אם $F(k)\neq G(k)$ החזר $false$.
2. החזר $true$

סיבוכיות זמן: $O(kd)$ (אם k הוא קבוע אז $O(d)$).

אם $F(x)\equiv G(x)$ אז אין הסתברות לטעות,

אם $F(x) \not\equiv G(x)$ ההסתברות לטעות היא שנבר את אחת מלכל היותר d נקודות חיתוך בכל איטרציה:

$$P(Error) = P(\text{נבחרה נקודה חיתוך})^k = \frac{1}{2^k}$$

דוגמה:

יהי $2 \le k \le n$ וגם $1 \le m \le 2^{k-2}$ מספרים שלמים, יהי $\{A_1, \dots, A_m\}$ משפחה של תתי קבוצות של $\{1, \dots, n\}$ כל קבוצה כזו $(1 \le i \le m)$ A_i היא בגודל k.

יש לכתוב אלגוריתם המקיים את התנאים הבאים:

1. הקלט הוא מספר n וקבוצה $\{A_1, \dots, A_m\}$.
2. הפלט הוא הקבוצה $\{1, \dots, n\}$ צבועים בכחול ואדום.
3. עם הסתברות של לפחות $1 - 2^{-100}$ שהאלגוריתם צביעה יוציא לכל קבוצה A_i צביעה של לפחות אלמנט אחד בכחול ולפחות אלמנט אחד באדום.
4. זמן הריצה של האלגוריתם יהיה $O(n + km)$

פתרון:

אלגוריתם:

1. עבור k מ-1 עד 100:
   a. בחר צביעה אקראית עבור $\{1, \dots, n\}$ באדום או כחול.
   b. עבור j מ-1 עד m:
      i. בדוק שיש ב-A_i שני איברים לפחות הצבועים שונה.
   c. אם בכל קבוצה A_i היה שני צבעים שונים החזר את הצביעה
2. החזר את הצביעה האחרונה.

ניתוח הסתברות:

נגדיר X_i אינדיקטור עבור המאורע: האם A_i כולו באותו צבע כלומר:

$$\begin{cases} X_i = 0, & \text{יש 2 צבעים שונים} \\ X_i = 1, & \text{כולם אותו צבע} \end{cases}$$

נגדיר $X = \sum_{i=1}^{m} X_i$,

נחשב:

$$P(X_i = 1) = \underbrace{\frac{1}{2^k}}_{\text{הסתברות לכחול}} + \underbrace{\frac{1}{2^k}}_{\text{הסתברות לאדום}} = \frac{2}{2^k}$$

$$P(Error) = P(X > 0) = P\left(\bigcup_{i=1}^{m} X_i = 1\right) \le \sum_{i=1}^{m} P(X_i = 1) = \frac{2m}{2^k}$$

נזכר כי נתון: $m \le 2^{k-2}$ ולכן $\frac{2m}{2^k} \le \frac{2^{k-1}}{2^k} = 2^{k-1-k} = 2^{-1} = \frac{1}{2}$

אבל מכיוון שאנחנו מבצעים באלגוריתם 100 איטרציות נקבל:

$$\frac{1}{2^{100}} = 2^{-100}$$

סיבוכיות: מעבר על כל n איברים: $O(n)$ לצביעה, בנוסף מעבר על m ו-k כדי לבדוק תקינות צביעה: $O(mk)$,

סך הכל: $100 \cdot O(n + mk) = O(n + mk)$.

מרחב הסתברות רציף:

משתנה X על מרחב הסתברות $(\Omega\ or\ \mathbb{R}, \mathcal{F}, \mathrm{P})$

[כאשר Ω או $\mathbb{R}$ זה מרחב הדגימה, $\mathcal{F}$ סיגמא אלגברה, P פונקציית הסתברות]

ייקרא משתנה מקרי רציף אם קיימת פונקציה: $f_X: \mathbb{R} \to \mathbb{R}$ כך ש:

$$P(X \in A) = \int f_{X(x)dx}$$

לפונקציה f_X קוראים פונקציית צפיפות.

תכונות:

1. $P(X = a) = \int_a^a f_{X(x)dx} = 0$
2. $P(X \in \mathbb{R}) = P(\mathbb{R}) = \int_{-\infty}^{\infty} f_{X(x)dx} = 1$
3. $P(a < X < b) = P(a \le X \le b) = \int_a^b f_{X(x)dx}$

מסקנה – כדי להוכיח שמדובר במחרה הסתברות צריך להוכיח שהאינטגרל על הכל שווה ל-1.

פונקציית צפיפות מצטברת:

$$F_X(a) = P(X \le a) = \int_{-\infty}^{a} f_X(x)dx$$

דוגמה:

עבור אלו ערכים של פרמטר c הפונקציות הבאות הן פונקציות צפיפות?:

a. $f: \mathbb{R} \to \mathbb{R}, f(x) = \begin{cases} c(4x - 2x^2), 0 < x < 2 \\ 0 \quad\quad , else \end{cases}$

b. $g: \mathbb{R} \to \mathbb{R}, g(x) = \begin{cases} c \cdot e^{-\frac{x}{100}}, x \ge 0 \\ 0 \quad\quad , else \end{cases}$

פתרון א':

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 c(4x - 2x^2)dx = c\left(2x^2 - \frac{2}{3}x^3\right)\Big|_0^2 = 1$$

נמצא את c:

$$c\left(8 - \frac{16}{3}\right) - c \cdot 0 = c\left(\frac{24}{3} - \frac{16}{3}\right) = \frac{1}{\frac{8}{3}} = \frac{3}{8} = c$$

פתרון ב':

$$\int_{-\infty}^{\infty} g(x)dx = \int_0^{\infty} ce^{-\frac{x}{100}}dx = c\left(\frac{e^{-\frac{x}{100}}}{-\frac{1}{100}}\right)\Big|_0^{\infty} = c\left(-100^{-\frac{x}{100}}\right)\Big|_0^{\infty} = -100ce^{-\infty} - (-100ce^0)$$
$$= 0 + 100c = 1 \to c = \frac{1}{100}$$