סיכום מתומצת להסתברות

15 ביוני 2025

מבוא

* **הסתברות מותנית:** נסמן כך - $P(A|B)$ והמשמעות היא "מה הסיכוי של A בהינתן ש־B קרה".
הנוסחה - $P(A|B) = \frac{P(A\bigcap B)}{P(B)}$.
* **כלל ההכלה והפרדה:** $P(A\bigcup B) = P(A) + P(B) - P(A\bigcap B)$
משפט ההסתברות השלמה: תהי חלוקה של מרחב המדגם Ω למאורעות $A_1, A_2, ...A_n$. כמו כן, לכל i נתון $P(A_i)$ וכן $P(B|A_i)$. אזי - $P(B) = \sum_{i=1}^{n} P(A_i) * P(B|A_i)$.
כלל בייס: $P(A|B) = \frac{P(B|A)*P(A)}{P(B)}$
אי תלות: יהי Ω מרחב מדגם ויהיו A, B מאורעות. נאמר כי אלו מאורעות בלתי תלויים אם מתקיים $P(A\bigcap B) = P(A) * P(B)$.
נוסחת ברנולי: ההסתברות ל־k הצלחות מתוך n נסיונות בהתפלגות בינומית כאשר ההסתברות להצלחה היא p היא - $P_n(k) = \binom{n}{k} * p^k * (1-p)^{n-k}$.

משתנה רנדומי בדיד

פונקציית מסת ההסתברות: נקראת גם PMF. משייכת כל ערך להסתברות שמתאימה לו. מסומנת f לרוב ברציף ו־p_x בבדיד. כמובן $\sum_x p_X = 1$.
פונקציית ההצטברות / התפלגות: נקראת CDF ומוגדרת להיות $F_X(t) = P(x \leq t)$. כלומר, ההסתברות לקבלת כל הערכים שקטנים מ־t. כמובן $P(x > t) = \overline{F_X(t)} = 1 - P(x \leq t)$. וכן $P(a < x \leq b) = F_x(b) - F_x(a)$.

משתנה ברנולי:

יסומן $x \sim Ber(p)$ משתנה עם פרמטר p אם מתקיים שיש שתי אפשרויות בלבד - הצלחה או כשלון. במצב זה p סיכוי להצלחה ו־$1-p$ לכשלון.
* התוחלת $E[X] = p$
* השונות $E[X] = p(1-p)$

משתנה אחיד:

יסומן $x \sim U[a,b]$ משתנה עם פרמטרים $a \leq b$ כאשר המשתנה יכול לקבל את הערכים מ־$[a,b]$ בהסתברות שווה (למשל הטלת קוביה). ההסתברות למאורע תהיה $\frac{1}{b-a+1}$.
* פונקציית מסת ההסתברות תהיה קו מקביל לציר האיקס - $P(x) = \frac{1}{b-a+1}$
* התוחלת תהיה $E[X] = \frac{a+b}{2}$
* השונות תהיה $Var[X] = \frac{(b-a+1)^2-1}{12}$

משתנה בינומי:

יסומן גם $x\sim Bin(n,p)$. n הטלות בלתי תלויות כאשר הסיכוי להצלחה הוא p. במקרה כזה הסיכוי לk הצלחות יהיה $P_x(k)=\binom{n}{k}*p^k*(1-p)^{n-k}$.

* **התוחלת** $E[X]=np$.
* **השונות** - $Var[X]=np(1-p)$

משתנה גאומטרי:

יסומן גם $x\sim Geo(p)$. הטלת אנסוף מטבעות בלתי תלויות, אם הסיכוי לקבלת תוצאה היא p אזי המשתנה בודק את הסיכוי לקבלת תוצאה לאחר k פעמים.

* תוחלת משתנה גאומטרי (כולל ההצלחה) היא $E[X]=\frac{1}{p}$. התוחלת משתנה גאומטרי כאשר סופרים מס' כשלונות עד להצלחה (ללא ההצלחה) היא $E[X]=\frac{1-p}{p}$
* **טענה:** משתנה גאומטרי הוא חסר זכרון. כלומר - $P_{X-n|x>n}(k)=P_x(k)$.
* פונקציית מסת ההסתברות - $P_x(k)=(1-p)^{k-1}*p$
* השונות - $Var[X]=\frac{1-p}{p^2}$

על תוחלת ושונות

הגדרה של תוחלת: $E[X]=\sum_x x*P_x(X=x)$. אם c קבוע אזי $E[c]=c$. לינאריות התוחלת - $E[aX+b]=aE[X]+b$.

הגדרה של שונות: כמה המשתנים המקריים רחוקים זה מזה. $Var[x]=E[(X-E[X])^2]$.

טענה: $Var[aX+B]=a^2Var[x]$

טענה: $Var[x]=E[X^2]-(E[X])^2$

סטיית תקן - $\sigma_x=\sqrt{Var[X]}$.

התנייה של משתנה רנדומי במאורע - $P_{x|A}(x)=P(X=x|A)=\frac{P(\{X=x\}\bigcap A)}{P(A)}$

משפט התוחלת השלמה - נחלק את מרחב המדגם לקבוצות $A_1,...A_n$ ואז מתקיים כי $E[X]=P(A_1)*E[X|A_1]+.....+P(A_N)*E[X|A_N]$.

פונקציית מסת ההסתברות המשותפת: $P_{x,y}(x,y)=P(X=x\&Y=y)$. כלומר, ההסתברות ששני המשתנים יקבלו כל אחד את הערך שלו.

תוחלת לשני משתנים: $E[X+Y]=E[X]+E[Y]$.

טענה חשובה: $E[E[X|Y]]=E[X]$.

תוחלת מותנית: $E[X|A]=\sum_x xp_{X|A}(x)$

אי תלות בין מאורע ומשתנה מקרי: $\forall x,P(X=x\wedge A)=p(X=x)*p(A)=p_x(x)*p(A)$

* בכדי להראות שמשתנים הם תלויים יש להראות דוגמה אחת בה השוויון לא מתקיים. בשביל להראות אי תלות צריך לעבור על כל הזוגות האפשריים ולהראות שמתקיים השוויון.
* אם X,Y בלתי תלויים אזי $E[XY]=E[X]*E[Y]$ וכן $Var[X+Y]=Var[X]+Var[Y]$.

שונות מותנית: $Var[X|Y=y]=E[(X-E[X|Y=y])^2|Y=y]$

משפט השונות השלמה: $Var[x]=E[Var(X|Y)]+Var[E[X|Y]]$

מניפולציה על משתנים מקריים: בהינתן שני משתנים מקריים, נניח והגדרנו $Z=X+Y$. אזי $p_Z(z)=p(X+Y=z)=\sum_{(x,y)|x+y=z}P(X=x,Y=y)=\sum_x p(X=x,Y=z-x)=\sum_x p_X(x)*p_Y(z-x)$.

שונות משותפת:

תסומן $Cov(X,Y)$ ותתאר את הקשר בין המשתנים. הערך עצמו פחות רלוונטי. אם קיבלנו כי הוא חיובי אז ככל שאחד עולה השני עולה ואם שלילי אז ככל שאחד עולה השני יורד. נגדיר כך - $Cov(X,Y)=E[(X-E[X])(Y-E[Y])]$.

הגדרה שקולה ונוחה יותר - $Cov(X,Y)=E[XY]-E[X]E[Y]$. תכונות מעניינות -

* $Cov(X,X)=E[X^2]-(E[X])^2=Var(X)$
* $Cov(aX+b,Y)=aCov(X,Y)$
* $Cov(X,Y+Z)=Cov(X,Y)+Cov(X,Z)$
* $Var[\sum_{i=1}^{n}X_i]=\sum_{i=1}^{n}Var[X_i]+\sum_{(i,j):i\neq j}Cov(X_i,X_j)$.
* **חשוב מאוד** - $Var[X+Y]=VAR[X]+VAR[Y]+2Cov(X,Y)$

מקדם המתאם: נגדירו כך $\rho(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$. עבורו מתקיים - $\rho(X,-X)=-1$ וכן $\rho(X,X)=1$, לכן תמיד ערכו חסום בין 1 למינוס אחד. ככל שמקדם המתאם קרוב יותר לאחד, כך הקשר בין שני המשתנים קרוב ללינארי.

משתנה מקרי רציף

פונקציית צפיפות ההסתברות: תקרא כאן PDF ותסומן כאן f_X.

פונקציית ההצטברות: CDF. תוגדר כאן להיות $P_X(t)=P(X<t)=\int_{-\infty}^{t}f_x(x)dx$. הסתברות לקבלת ערך בודד הינו 0.

חשוב - מתקיים $P_X(x)'=f_x(x)$ (כלומר אם נגזור את הCDF נקבל את הPMF). וכן $x\to-\infty\Rightarrow F_X(X)=0$ וכן $x\to\infty\Rightarrow F_X(X)=1$.

תוחלת: $E[X]=\int_{-\infty}^{\infty}xf_x(x)$. כל התכונות הלינאריות נשמרות. כמובן גם שכלל התוחלת - $E[g(x)]=\int_{-\infty}^{\infty}g(x)f_X(x)$.

שונות: $Var[X]=\int_{-\infty}^{\infty}(x-E[x])^2f_X(x)dx$

פונקציית הסתברות בהינתן התנייה:

$$f_{X|X\in A}(x):=\left\{\begin{array}{cc}\frac{f_x(x)}{P(A)} & x\in A\\ 0 & x\notin A\end{array}\right\}$$

תוחלת מותנית: $E[X|A]=\int_{-\infty}^{\infty}xf_{X|A}(x)$

משפט ההסתברות השלמה למשתנה רציף: $f_x(x)=P(A_1)*f_{X|A_1}+....P(A_n)*f_{X|A_n}$

משפט התוחלת השלמה למשתנה רציף: $E[X]=P(A_1)E[X|A_1]+...+P(A_n)E[X|A_n]$

משתנה רציף אחיד

אם $x\sim U[a,b]$, אזי הPDF יהיה -

$$f_x(x):=\left\{\begin{array}{cc}\frac{1}{b-a} & a\leq x\leq b\\ 0 & x>b\vee x<a\end{array}\right\}$$

וכן הCDF יהיה:

$$F(x):=\left\{\begin{array}{cc}0 & x\leq a\\ \frac{x-a}{b-a} & a<x<b\\ 1 & x\geq b\end{array}\right\}$$

* **התוחלת** תהיה $E[X]=\frac{a+b}{2}$
* **השונות** תהיה $Var[x]=\frac{(b-a)^2}{12}$, **וסטיית התקן** $\sigma(x)=\frac{b-a}{\sqrt{12}}$.

משתנה מקרי מעריכי

יסומן $x \sim exp(\lambda)$. מייצג את הזמן עד למאורע מסויים. כלומר - עד המאורע הראשון של אותו אירוע.

פונקציית PDF :

$$f_X(x) := \left\{ \begin{array}{cc} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{array} \right\}$$

פונקציית CDF:

$$F_X(x) := \left\{ \begin{array}{cc} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{array} \right\}$$

* **התוחלת** תהיה $E[X] = \frac{1}{\lambda}$ וכן **השונות** $Var[X] = \frac{1}{\lambda^2}$.

טענה: X חסר זכרון $\Longleftrightarrow$ $\left\{ \begin{array}{c} x \sim G(p) \\ x \sim exp(\lambda) \end{array} \right\}$, חוסר זכרון הכוונה $P(x > t + s | x > t) = P(x > s)$

משתנה מקרי פואסון

נסמן $x \sim Poi(\lambda)$. ישנו מאורע שיכול לקרות בכל רגע, אך יש את הנתון לכמה ממנו קורים ביחידת זמן מסויימת. המשתנה יגדיר את ההסתברות למופע אחד לאורך יחידת הזמן. חשוב לשים לב ולהגדיר היטב את יחידת הזמן ולפעול לפיה.

* פונקציית מסת ההסתברות - ההסתברות לk מופעים בזמן הדגימה $f_x(k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

* תוחלת ושונות: $E[X] = Var[X] = \lambda$.

* הקשר בין פואסון למעריכי: אם נגדיר $X \sim Poi(\lambda)$, מס' מאורעות ביחידת זמן כאשר הצפי הוא λ מאורעות ביחידת זמן, נגדיר Y- זמן ההמתנה עד להתרחשות המאורע הראשון, אזי $Y \sim exp(\lambda)$.

משתנה מקרי נורמלי

לרוב בשאלות יהיה נתון לנו שמשתנה מתפלג נורמלית.

* **סטנדרטי:** התוחלת היא 0 והשונות היא 1 אזי $N(0,1) : f_x(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ (זו פונקציית הפעמון)

* **גאוסייני:** משתנה נורמלי כללי. $N(\mu, \sigma^2) : f_x(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$.

* **טענה:** בהינתן $X \sim N(\mu, \sigma^2)$ אם נגדיר $Y = aX + b$ אזי $E[Y] = a\mu + b, Var[Y] = a^2\sigma^2$, וכן $Y \sim N(a\mu + b, a^2\sigma^2)$, כלומר הזזה לינארית של משתנה נורמלי הוא משתנה נורמלי.

* לא ניתן לבצע אינטגרל על פונקציית ההסתברות - איננו ניתן לתאר באמצעות פונקציות אלמנטריות. לכן יש טבלה שממפה ערכים. כלומר את $\Phi(Y) = F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$

* **טענה:** תמיד מתקיים $\phi(X) + \phi(-x) = 1$

* כיצד עוברים ממשתנה נורמלי כלשהו לסטנדרטי על מנת להשתמש בטבלה? נגדיר תמיד $Y = \frac{X - E[X]}{\sigma[X]}$ שיקיים $Y \sim N(0,1)$ וכן את אי השוויון שרצינו נתרגם במונחים של Y.

משתנה מקרי מעורב

כאשר מתקיים

$$X := \left\{ \begin{array}{cc} Y & p \\ Z & 1-p \end{array} \right\}$$

למשל, וכן Y בדיד וZ רציף, נקרא לX משתנה מעורב. במקרה זה הCDF יהיה

$$P_X(x) = pP(Y \leq x) + (1-p)P(Z \leq x) = pF_Y(y) + (1-p)F_Z(z)$$

משתנים מקריים מרובים

פונקצייה - $P((X,Y) \in B) = \int\int_{(x,y)\in B} f(x,y)dxdy$, תמיד יתקיים $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = 1$

מעבר מPDF משותף לשולי: $f_x(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$,$f_y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

פונקציית מסת ההסתברות המשותפת: $P_X(x) = P(X < x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(u,v)du$ (מדוע השתנה לu וv? יש משתנה באינטגרל x. את y יכלנו להשאיר אך בחרנו אחידות מחלקתית ושינינו גם)

חשוב - תמיד מתקיים $f_x(x) = \frac{dF_x(x)}{dx} = \int_{-\infty}^{\infty} f_{X,Y}(x,v)dv$

פונקציית הCDF: $F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{x,y}(u,v)dudv$

חשוב - מתקיים $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$, כלומר אם נגזור פעמיים פעם אחת לפי איקס ופעם אחת לפי y נקבל את $f_{X,Y}$.

התנייה: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$, $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$

תוחלת בהתנייה: $E[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$

אי תלות: אם X וY בלתי תלויים כלומר $f_{X,Y}(x,y) = f_x(x) * f_Y(y)$, אזי $E[XY] = E[X]E[Y]$ וכן $VAR[X+Y] = VAR[X] + VAR[Y]$.

מניפולציה על משתנים מקריים רציפים: אם $Z = X+Y$ אזי $f_Z(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)dx$

אי שוויונות

אי שוויון מרקוב: אם $X \geq 0$ וכן $a > 0$ אזי $P(X \geq a) \leq \frac{E[X]}{a}$.

אי שוויון צ'בישב: יהי X משתנה מקרי כך ש$C > 0$, אזי $P(|X - E[x]| \geq c) \leq \frac{Var[X]}{c^2}$

החוק החלש של המספרים הגדולים: נתונים $X_1, ..., X_n$ משתנים בלתי תלויים זהים (כלומר עם אותה התפלגות) עם תוחלת μ ושונות σ^2. כלומר לכל $1 \leq i \leq n$ מתקיים $E[X_i] = \mu$. נסתכל על ממוצע המדגם: $M_n = \frac{X_1 + ... + X_n}{n}$, נשים לב $E[M_n] = \mu$ וכן $Var[M_n] = \frac{\sigma^2}{n}$. אזי, אזי, לכל $\epsilon > 0$ כאשר $n \to \infty$ מתקיים

$$P(|M_n - \mu| \geq \epsilon) = P(|\frac{X_1 + ... + X_n}{n} - \mu| \geq \epsilon) \to 0$$

משמעות החוק החלש: חוזרים מס' רב של פעמים על אותו הניסוי. בכל שלב מקבלים תוצאה $X_i = \mu + W_i$כך שW_i בלתי תלויים, אזי לא סביר שממוצע המדגם יהיה גדול מהתוחלת באמת. לא סביר - הסתברות אפס.

משפט הגבול המרכזי: נתונים $X_1, ..., X_n$ משתנים בלתי תלויים זהים (כלומר עם אותה התפלגות) עם תוחלת μ ושונות σ^2. כלומר לכל $1 \leq i \leq n$ מתקיים $E[X_i] = \mu$. כך ש$S_n = X_1 + X_2 + ... + X_n$.נגדיר $Z_n = \frac{S_n - \mu n}{\sigma\sqrt{n}}$. מתקיים $E[Z_n] = 0$ וכן $Var[Z_n] = 1$. אזי, יהא $Z \sim N(0,1)$.

$$\forall z : lim_{n\to\infty} P(Z_n \leq z) = P(Z \leq z) = \Phi(z)$$

מה החידוש כאן? לא משנה כיצד $X_1, ..., X_n$ התפלגו קודם לכן. Z_n תמיד יהיה משתנה מקרי נורמלי. לרוב נרצה לפתור $P(S_n \le a) \approx b$. כאשר שניים יהיו נתונים ונרצה למצוא השלישי.
קירוב דה מוויאיר לפלס: $S_n \sim Bin(n,p)$ וכן k, l שלמים חיוביים. אזי,

$$P(k \le S_n \le l) = \Phi(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}) - \Phi(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}})$$