תרגול מספר 0 - חזרה על הסתברות

תקציר תאוריה

מושגים

- Ω - מרחב המדגם (אוסף תוצאות אפשריות בניסוי). בהטלת קובייה: $\Omega = \{1,2,3,4,5,6\}$
- $\mathcal{F}$ - מרחב המאורעות (אוסף תת-קבוצות של Ω). בהטלת קובייה, אפשרות אחת היא: $\mathcal{F} = \{\{1,3,5\},\{2,4,6\},\varnothing,\Omega\}$
- הסתברות - פונקציה $\mathbb{P}:\mathcal{F}\to[0,1]$. ההסתברות למאורע E היא $\mathbb{P}(E)$.
- איחוד מאורעות (union) – מאורע הכולל את כל תוצאות הניסוי שכלולות במאורע A או במאורע B או בשניהם, נקרא האיחוד של A ו B ומסומן $A\cup B$
- חיתוך מאורעות (intersection) – מאורע הכולל את כל תוצאות הניסוי שכלולות גם במאורע A וגם במאורע B, נקרא האיחוד של A ו B ומסומן $A\cap B$

משתנה אקראי

משתנה אקראי הוא פונקציה $X:\Omega\to\mathbb{R}$.

עבור משתנה אקראי בדיד, מוגדרת פונקציית ההסתברות (probability mass function):

$$p_X(x)=\mathbb{P}(X=x)$$

עבור משתנה אקראי רציף, מוגדרת פונקציית צפיפות ההסתברות (probability density function):

$$p_X(x)dx=\mathbb{P}(x\le X\le x+dx)$$

הערה: לשם הפשטות, ההגדרות מנקודה זו ואילך מתייחסות למשתנה בדיד, כאשר במקרה הרציף סכומים מוחלפים באינטגרלים, פונקציית ההסתברות מוחלפת בצפיפות ההסתברות וכו'. כמו-כן מעתה נשמיט את ה subscript שמציין לאיזה משתנה אקראי מתייחסת פונקציית ההסתברות, במקרים בהם זה ברור מההקשר. למשל, נכתוב $p(x)$ במקום $p_X(x)$.

תוחלת (mean/expectation) של משתנה אקראי:

$$E[X]=\sum_i x_i p(x_i)\equiv\mu$$

שונות (variance) של משתנה אקראי:

$$Var(X)=E[(X-\mu)^2]=E[X^2]-\mu^2\equiv\sigma^2$$

שונות משותפת (covariance) של שני משתנים אקראיים X ו Y:

$$Cov(X,Y)=E[(X-\mu_X)(Y-\mu_Y)]=E[XY]-\mu_X\mu_Y$$

הסתברות משותפת (joint), הסתברות שולית (marginal), והסתברות מותנית (conditional)

עבור שני משתנים אקראיים X ו Y מוגדרת פונקציית ההסתברות המשותפת:

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

ההסתברות השולית מתקבלת ע"י סכימה (או אינטגרציה במקרה הרציף) על אחד המשתנים:

$$p(x)=\sum_i p(x,y_i)$$

ההסתברות המותנית של Y בהינתן X מוגדרת ע"י:

$$p(y\,|\,x)=\frac{p(x,y)}{p(x)}$$

שימו לב כי באופן כללי $p(y\,|\,x)\neq p(y)$, כלומר ידיעת ערכו של X מוסיפה לנו מידע לגבי ההסתברות של Y. מצב זה אינו קורה אם X ו Y בלתי-תלויים, ואז מתקיים $p(x,y)=p(x)p(y)$ כלומר ידיעת X לא הוסיפה מידע על Y ו $p(y\,|\,x)=p(y)$.

התוחלת המותנית של X בהינתן Y מוגדרת ע"י:

$$E[X\,|\,Y]=\sum_i x_i\, p(x_i\,|\,y)$$

כמו-כן מתקיים:

$$E[X]=E_Y[E_X[X\,|\,Y]]=\sum_i E[X\,|\,Y=y_i]\,p(y_i)$$

נוסחת ההסתברות השלמה וכלל Bayes

מהגדרת ההסתברות המותנית נקבל:

$$p(x,y)=p(y\,|\,x)p(x)=p(x\,|\,y)p(y)$$

אם נסכום נוסחה זו על כל ערכי y האפשריים ונשתמש בהתפלגות השולית של x, נקבל את נוסחת ההסתברות השלמה:

$$p(x)=\sum_i p(x,y_i)=\sum_i p(x\,|\,y_i)\,p(y_i)$$

משתי נוסחאות אלה נקבל את כלל Bayes:

$$p(y|x)=\frac{p(x|y)p(y)}{p(x)}=\frac{p(x|y)p(y)}{\sum_i p(x|y_i)p(y_i)}$$

כאשר אפשר לראות את המכנה כגורם נירמול. את כלל Bayes אפשר לרשום במלים בדרך הבאה:

$$posterior=\frac{likelihood\cdot prior}{evidence}$$

תרגיל 1

בבדיקת דם לגילוי מחלה מסויימת יש סיכוי של 95% לקבל תוצאה חיובית אם האדם הנבדק חולה, ויש סיכוי של 1% לקבל תוצאה חיובית אם הנבדק בריא. ידוע כי המחלה קיימת אצל 0.5% מהאוכלוסיה. מה הסיכוי שנבדק מסוים אכן חולה, אם התקבלה אצלו תוצאה חיובית בבדיקה?

תרגיל 2

במשחק רולטה יש 38 תאים, מתוכם 18 שחורים, 18 אדומים, ו 2 ירוקים (עליהם אי-אפשר להמר). מהמר זוכה אם הצליח לנחש את צבע התא בו נופל הכדור. סכום הזכייה הוא כסכום ההימור, כלומר הימור מוצלח בגובה שקל מזכה את המהמר בשקל נוסף.

א. חשבו את תוחלת הרווח (סכום הזכיות) והשונות, עבור מהמר המשחק n פעמים.
ב. חשבו את ההסתברות שעבור $n=100$ המהמר יהיה ברווח. השוו את התוצאה להערכה המתקבלת ע"י שימוש במשפט הגבול המרכזי.

תרגיל 3

א. הוכיחו את אי-שויון מרקוב: אם X הוא משתנה אקראי המקבל ערכים אי-שליליים עם תוחלת $E[X]=\mu$, אז עבור $a>0$ מתקיים:

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

ב. השתמשו באי-שויון מרקוב כדי להוכיח את אי-שויון צ'בישב: אם X הוא משתנה אקראי עם תוחלת μ ושונות σ^2, אז לכל $k>0$ מתקיים:

$$\mathbb{P}(|X-\mu|\geq k)\leq\frac{\sigma^2}{k^2}$$

ג. נתונים שני מטבעות, אחד הוגן והשני עם סיכוי של ¾ ל heads. כמה פעמים צריך להטיל את המטבע כדי לקבוע איזה משני המטבעות הוא, בוודאות של לפחות 95%?