הסקה סטטיסטית שיעור 5

הקדמה חומר קריאה א5, חומר קריאה ב5, חומר קריאה ג5, חומר קריאה ד5

1 עדכון בייסיאני: חיזוי הסתברותי (חומר קריאה א5)

מילים של הערכת הסתברות

- "מחר ירד גשם"
- "מחר כנראה ירד גשם"
- "מחר ירד גשם בהסתברות 60%".

יש הרבה דרכים לתאר משהו הסתברותי, והרבה מהן גורמות לבלבול.

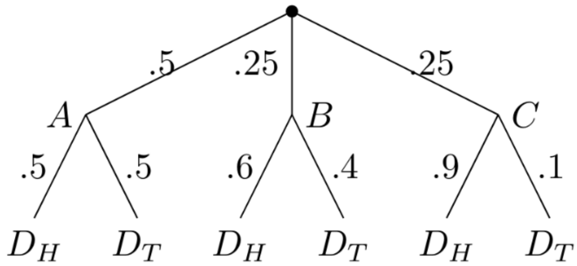
הסתברויות חיזוי

חיזוי הסתברותי זה השמה של הסתברות לכל תוצאה אפשרית של ניסוי. ניזכר בדוגמה של 3 סוגי המטבעות מהשיעור הקודם:

נניח שיש 3 סוגי מטבעות, כל אחד עם הסתברות שונה לעץ. סוג A הוגן, סוג B עם הסתברות 0.6, סוג C עם הסתברות 0.9. נניח שיש קופסה עם 4 מטבעות: 2 מסוג A, 1 מסוג B, ואחת מסוג C. הוצאנו מטבע אחד באקראי.

הסתברויות חיזוי קודמות

נוכל לחשב את ההסתברות שייצא עץ או פלי אם נטיל את המטבע: $D_H =$ עץ.



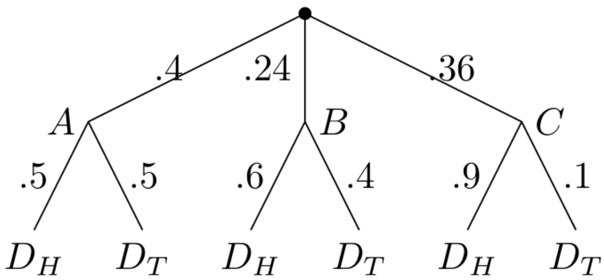
$$\begin{aligned} P(D_H) &= P(D_H|A)P(A) + P(D_H|B)P(B) + P(D_H|C)P(C) \\ &= 0.5 \cdot 0.5 + 0.6 \cdot 0.25 + 0.9 \cdot 0.25 \\ &= 0.625 \end{aligned}$$

הסתברויות חיזוי אוחרות

נניח שהטלנו את המטבע ויצא עץ. עכשיו יש דאטא D, ונוכל לעדכן את ההשערה שלנו להסתברות אוחרת.

אם קיבלנו עץ, נוכל לחשב את ההסתברות שאותו מטבע ייצא עץ בהטלה שנייה. הפעם, במקום הסתברויות קודמות נשים הסתברויות אוחרות:

hypothesis	prior	likelihood	Bayes numerator	posterior
H	$P(H)$	$P(D\|H)$	$P(D\|H)P(H)$	$P(H\|D)$
A	0.5	0.5	0.25	0.4
B	0.25	0.6	0.15	0.24
C	0.25	0.9	0.225	0.36
total	1		0.625	1



$$\begin{aligned} P(D_H|D) &= P(D_H|A)P(A|D) + P(D_H|B)P(B|D) \\ &\quad + P(D_H|C)P(C|D) \\ &= 0.5 \cdot 0.4 + 0.6 \cdot 0.24 + 0.9 \cdot 0.36 \\ &= 0.668 \end{aligned}$$

סיכום

כל השערה נותנת הסתברות אחרת לעץ, אז ההסתברות השלמה היא ממוצע משוקלל. עבור הסתברות החיזוי הקודם, המשקלים מתקבלים מההסתברויות הקודמות של ההשערות. עבור הסתברות החיזוי האוחר, המשקלים מתקבלים מההסתברויות האוחרות של ההשערות.

2 עדכון בייסיאני: סיכויים (חומר קריאה 5ב)

בהשוואה בין שני מאורעות, בדרך כלל מנסחים את הטענה ההסתברותית על ידי **סיכויים.** הגדרה: הסיכויים של מאורע E מול מאורע E' הם היחס בין ההסתברויות שלהם $P(E)/P(E')$. אם לא מוגדר במפורש, נניח שהמאורע השני הוא המשלים של E. כלומר:

$$O(E) = \frac{P(E)}{P(E^c)}$$

לדוגמה, $O(rain) = 2$ אומר שההסתברות שירד גשם גדולה פי 2 מההסתברות שלא ירד. 2/3 מול 1/3. אפשר גם לומר שהסיכוי לגשם הוא 2 ל-1. הסיכוי לקבל 4 בקובייה הוגנת הוא 1 ל-5. באופן כללי, אפשר להמיר:

$$P(E) = p \quad \Longrightarrow \quad O(E) = \frac{p}{1-p}, \qquad O(E) = q \quad \Longrightarrow \quad P(E) = \frac{q}{1+q}$$

הסתברויות הן בין 0 ל-1. סיכויים הם בין 0 ל-∞.

התכונה $P(E^c) = 1 - P(E)$ שקולה לתכונה $O(E^c) = 1/O(E)$.

עדכון סיכויים

כמו שהיו לנו הסתברויות קודמות ואוחרות, יש סיכויים קודמים וסיכויים אוחרים.

דוגמה: תסמונת מרפן

מחלה שמופיעה 1 לכל 15,000 אנשים. חלק מהתסמינים הנראים לעין הם A, B, C. ל 70% מהאנשים עם התסמונת יש לפחות אחד מהתסמינים, ורק ל 7% מהאנשים **בלי** התסמונת יש לפחות אחד מהם.

אם לאדם יש לפחות אחד מהתסמינים, מה ה**סיכוי** שיש לו את התסמונת?

נגדיר: M=לאדם יש את התסמונת, F=לאדם יש לפחות אחד מהתסמינים.

אנחנו יודעים את ההסתברות הקודמת של M, ואת ה*likelihoods* של F בהינתן M או M^c:

$$P(M) = 1/15000\,, \qquad P(F|M) = 0.7, \qquad P(F|M^c) = 0.07$$

hypothesis	prior	likelihood	Bayes numerator	posterior
H	$P(H)$	$P(F\|H)$	$P(F\|H)P(H)$	$P(H\|F)$
M	0.000067	0.7	0.0000467	0.00066
M^c	0.999933	0.07	0.069995	0.99933
total	1		0.07004	1

נחשב את הסיכויים הקודמים:

$$O(M) = \frac{P(M)}{P(M^c)} = \frac{1/15000}{14999/15000} = \frac{1}{14999} \approx 0.000067$$

$$O(M|F) = \frac{P(M|F)}{P(M^c|F)} = \frac{P(F|M)P(M)}{P(F|M^c)P(M^c)} = 0.000667$$

נשים לב שבמעבר השני דילגנו על החלוקה ב-$P(F)$, כי זה יצטמצם.

הסיכויים האוחרים גדולים פי 10 מהקודמים. כלומר, התסמינים הם ראיות חזקות לכיוון זה שיש את התסמונת. ועדיין, לא סביר שיש לאדם את התסמונת.

3 מקדם בייס

המקדם 10 מהדוגמה הקודמת נקרא מקדם בייס. נגדיר פורמלית:

עבור השערה H ודאטא D, המקדם בייס הוא היחס של ה-*likelihoods*:

$$\text{Bayes factor} = \frac{P(D|H)}{P(D|H^c)}$$

$$O(H|D)=\frac{P(H|D)}{P(H^c|D)}=\frac{P(D|H)P(H)}{P(D|H^c)P(H^c)}=\frac{P(D|H)}{P(D|H^c)}\cdot\frac{P(H)}{P(H^c)}=\frac{P(D|H)}{P(D|H^c)}\cdot O(H)=BF(D,H)\cdot O(H)$$

$$\text{Posterior odds} = \text{Bayes factor} \times \text{Prior odds}$$

מכאן אפשר לראות שמקדם בייס אומר לנו האם הדאטא תומכת בהשערה או סותרת אותה:

- אם $BF>1$, אז הסיכוי האוחר גדול מהקודם. כלומר, אחרי שקיבלנו את הדאטא, הסיכוי עלה. הדאטא תומכת בהשערה.
- אם $BF<1$, הסיכוי האוחר קטן מהקודם. הדאטא הקטינה את הסיכוי, כלומר לא תומכת בהשערה.
- אם $BF=1$, הסיכויים שווים לפני ואחרי הדאטא. כלומר היא לא משפיעה.

עדכונים חוזרים

המקרה שהדאטא מגיעה בשני שלבים, ראינו בדוגמאות שאפשר לחשב בפעם אחת או בשלבים, שקודם מעדכנים את ה*prior* לפי ה*likelihood* של D_1 ואז מעדכנים את התוצאה לפי ה*likelihood* של D_2. השיטה השנייה עובדת אם ה*likelihoods* הם בת"ל ואז מתקיים:

$$P(D_1,D_2|H)=P(D_1|H)P(D_2|H)$$

מכיוון שה*likelihoods* תלויים בהשערה, נאמר שD_1,D_2 הם **בלתי תלויים באופן מותנה** אם המשווה נכונה לכל השערה H.

ניזכר, שאי תלות מותנית לא בהכרח גוררת אי תלות.

4 סיכויי לוג

אפשר לעבוד עם לוג לשם נוחות. כלומר במקום:

$$O(H|D_1,D_2)=BF_2\cdot BF_1\cdot O(H)$$

נכתוב:

$$\ln\big(O(H|D_1,D_2)\big)=\ln(BF_2)+\ln(BF_1)+\ln\big(O(H)\big)$$

כלומר, הסיכוי לוג האוחר הוא הסכום של הסיכוי לוג הקודם וכל הממצאים $\ln(BF_i)$. נשים לב שאם עשינו לוג, אז ממצאים תומכים $(BF_i>1)$ יהיה חיובי, וממצאים סותרים $(BF_i<1)$ יהיה שלילי.

5 עדכון בייסיאני עם קודמים רציפים (חומר קריאה 5ג)

נרחיב את הרעיון של עדכון בייסיאני למצבים שיש לנו טווח רציף של השערות. במקום PMF יהיה PDF, ובמקום סכומים יהיו אינטגרלים. דוגמאות:

1) מערכת שיכולה להיכשל בהסתברות $p\in[0,1]$. בדרך כל נמדל את זה ע"י מטבע מוטה.
2) זמן החיים של איזוטופ ממודל ע"י התפלגות מעריכית $Exp(\lambda)$. הזמן הממוצע $1/\lambda$ יכול להיות כל מספר ב $(0,\infty)$.

מוסכמות סימון

כמו בדוגמאות, ההשערה היא פרמטר כלשהו θ. האות θ בדרך כלל תייצג היפותזה כלשהי. האותיות p,f,x ימשיכו לייצג את הPMF, PDF, ודאטא, כרגיל. במקום לומר "ההשערה עם הפרמטר שמעניין אותנו, θ", נאמר "ההשערה θ".

נוסחת ההסתברות השלמה

עבור פרמטר רציף θ בתחום $[a,b]$, ודאטא מקרי בדיד x. נניח ש θ אקראית עם PDF $f(\theta)$. אזי ההסתברות של x היא:

$$p(x) = \int_a^b p(x|\theta) f(\theta)\, d\theta$$

דוגמה: נניח שיש מטבע עם הסתברות θ לעץ, ושהPDF היא $f(\theta) = 2\theta$ בתחום [0.1]. אז ההסתברות לעץ היא:

$$p(x=1) = \int_0^1 p(x=1|\theta) f(\theta)\, d\theta = \int_0^1 \theta \cdot 2\theta\, d\theta = \int_0^1 2\theta^2\, d\theta = \frac{2}{3}$$

נוסחת בייס עבור פונקציות צפיפות רציפות

עבור פרמטר רציף θ בתחום $[a, b]$, ודאטא מקרי בדיד x. נניח ש θ אקראית עם PDF $f(\theta)$. אזי מתקיים:

$$f(\theta|x)\, d\theta = \frac{p(x|\theta) f(\theta)\, d\theta}{p(x)} = \frac{p(x|\theta) f(\theta)\, d\theta}{\int_a^b p(x|\theta) f(\theta)\, d\theta}$$

נסביר: יהי Θ מ"מ שמקבל את הערך θ. נשקול את המאורעות:

$$H = \text{`}\Theta\ is\ in\ an\ interval\ of\ width\ d\theta\ around\ the\ value\ \theta\text{`}, \qquad D = \text{`}the\ value\ of\ the\ data\ is\ x\text{`}$$

כלומר: תטא הוא אינטרוול ברוחב $d\theta$ מסביב הערך θ, ערך הדאטא הוא x. אזי:

$$P(H) = f(\theta) d\theta, \qquad P(D) = p(x), \qquad P(D|H) = p(x|\theta)$$

עכשיו, נוסחת בייס הרגילה הופכת להיות:

$$f(\theta|x) d\theta = P(H|D) = \frac{P(D|H)P(H)}{P(D)} = \frac{p(x|\theta) f(\theta)\, d\theta}{p(x)}$$

ובגלל ש $d\theta$ מופיע במונה בשני הצדדים, אפשר לצמצם אותו ולקבל:

$$f(\theta|x) = \frac{p(x|\theta) f(\theta)}{p(x)} = \frac{p(x|\theta) f(\theta)}{\int_a^b p(x|\theta) f(\theta)\, d\theta}$$

עדכון בייסיאני עם קודמים רציפים

הטבלה של קודם רציף: בגלל שאי אפשר לעשות אינסוף (לא בן מניה) שורות עבור כל השערה, תהיה שורה אחת שמייצגת את כל ההשערות θ. בכך שנכלול את $d\theta$, כל התאים הופכים להסתברויות וכל החוקים של הסתברויות חלים.

בהמשך לדוגמה עם המטבע, נניח שהטלנו פעם אחת וקיבלנו עץ. נחשב את הPDF האוחרת:

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$f(\theta)\, d\theta$	$p(x=1\|\theta)$	$p(x=1\|\theta) f(\theta)\, d\theta$	$f(\theta\|x=1)\, d\theta$
θ	$2\theta\, d\theta$	θ	$2\theta^2\, d\theta$	$3\theta^2\, d\theta$
total	$\int_a^b f(\theta)\, d\theta = 1$		$p(x=1) = \int_0^1 2\theta^2\, d\theta = 2/3$	1

כלומר הPDF האוחרת היא $f(\theta|x) = 3\theta^2$. כמה הערות:

- מכיוון שהשתמשנו בהסתברות הקודמת $f(\theta)\, d\theta$, ההשערה אמורה להיות:
"הפרמטר הלא ידוע הוא אינטרוול ברוחב $d\theta$ בסביבת θ"
אבל אין לנו כוח להגיד או לכתוב את זה כל פעם, אז פשוט נחשוב את זה כל פעם שאומרים שההשערה היא θ.

- כרגיל, $p(x)$ היא הסתברות שלמה. במקום סכום עושים אינטגרל.

הסתברות קודמת אחידה

הסתברות שמניחה שכל תוצאה אפשרית באותה מידה. אם ל-θ יש תחום $[0,1]$, אז $f(\theta)=1$ היא קודמת אחידה. לדוגמה:

מטבע מוטה בהסתברות θ לעץ, הטלנו פעם אחת וקיבלנו פלי. נניח שההסתברות הקודמת אחידה ונמצא את ההסתברות האוחרת ל-θ.

hypothesis θ	prior $f(\theta)\,d\theta$	likelihood $p(x=0\mid\theta)$	Bayes numerator	posterior $f(\theta\mid x=0)\,d\theta$
θ	$1\cdot d\theta$	$1-\theta$	$(1-\theta)\,d\theta$	$2(1-\theta)\,d\theta$
total	$\int_a^b f(\theta)\,d\theta=1$		$p(x=0)=\int_0^1(1-\theta)\,d\theta=1/2$	1

ההסתברות האוחרת שהמטבע מוטה לכיוון עץ היא:

$$P(\theta>0.5|x=1)=\int_{0.5}^{1}f(\theta|x=1)\,d\theta=\int_{0.5}^{1}2\theta\,d\theta=[\theta^2]_{0.5}^{1}=\frac{3}{4}$$

הסתברויות חיזוי

בהמשך לדוגמה, יש לנו מטבע עם הטיה לא ידועה θ לעץ. ל-θ יש PDF קודמת 2θ. הסתברות החיזוי הקודמת היא:

$$p(x_1=1)=\int_0^1 p(x_1=1|\theta)f(\theta)\,d\theta=\int_0^1 2\theta^2\,d\theta=\frac{2}{3}$$

אנחנו יודעים שהPDF האוחרת היא $3\theta^2$, אז הסתברות החיזוי האוחרת היא:

$$p(x_2=1)=\int_0^1 p(x_2=1|\theta,x_1=1)f(\theta|x_1=1)\,d\theta=\int_0^1\theta\cdot 3\theta^2\,d\theta=\frac{3}{4}$$

מעדכון בייסיאני בדיד לרציף

אינטואיציה למעבר מבדיד לרציף:

1) נשערך את טווח ההשערות הרציף ע"י מספר סופי.
2) נייצר טבלת עדכון בדידה למספר הסופי של ההשערות.
3) נשקול איך הטבלה משתנה כאשר מספר ההשערות שואף לאינסוף.

נמשיך עם הדוגמה של המטבע, עם קודמת אחידה $f(\theta)=1$.

נתחיל עם 4 השערות: נחלק את $[0,1]$ ל-4 אינטרוולים $\left[0,\frac{1}{4}\right]\left[\frac{1}{4},\frac{1}{2}\right]\left[\frac{1}{2},\frac{3}{4}\right]\left[\frac{3}{4},1\right]$. לכל אחד יש רוחב $\Delta\theta=1/4$. נשים כל השערה במרכז של אינטרוול:

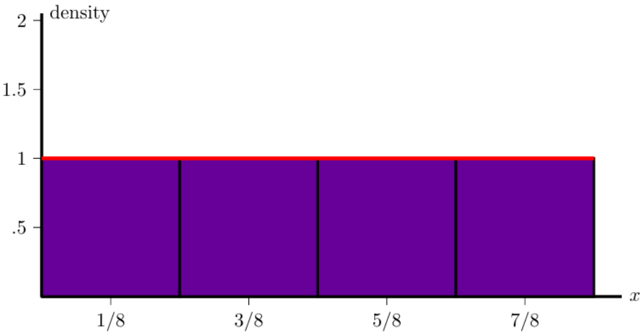
$$\theta_1=\frac{1}{8},\qquad\theta_2=\frac{3}{8},\qquad\theta_3=\frac{5}{8},\qquad\theta_4=\frac{7}{8}$$

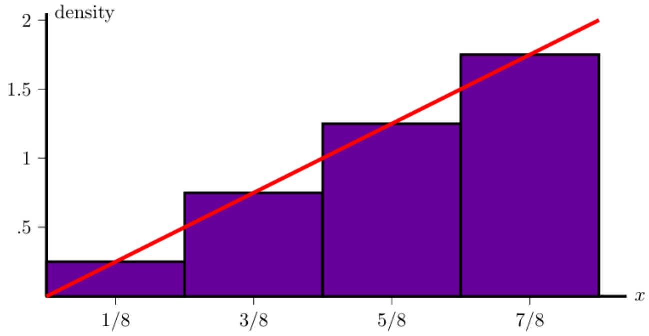
הקודמת האחידה נותנת לכל השערה הסתברות $1/4=1\cdot\Delta\theta$.

hypothesis	prior	likelihood	Bayes num.	posterior
$\theta=1/8$	$1/4$	$1/8$	$(1/4)\times(1/8)$	$1/16$
$\theta=3/8$	$1/4$	$3/8$	$(1/4)\times(3/8)$	$3/16$
$\theta=5/8$	$1/4$	$5/8$	$(1/4)\times(5/8)$	$5/16$
$\theta=7/8$	$1/4$	$7/8$	$(1/4)\times(7/8)$	$7/16$
Total	1	–	$\sum_{i=1}^{n}\theta_i\,\Delta\theta$	1

להלן היסטוגרמות הצפיפות של הPDF הקודמת והאוחרת.

הקודמת והאוחרת הרציפות מהדוגמה הקודמת זה הקווים האדומים:



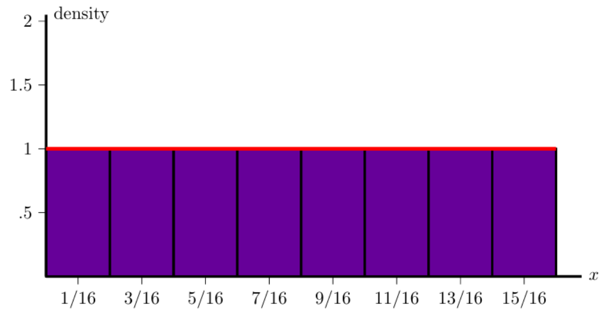


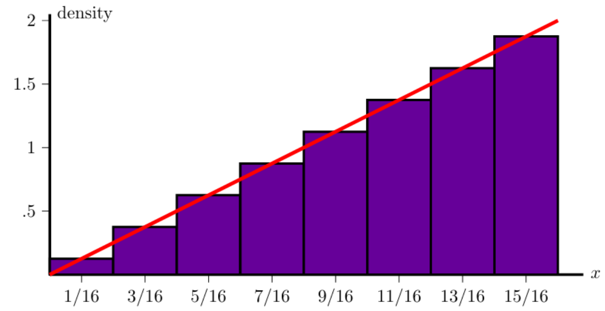
וכנ"ל עבור 8:

θ_1: '$\theta = 1/16$', θ_2: '$\theta = 3/16$', θ_3: '$\theta = 5/16$', θ_4: '$\theta = 7/16$'
θ_5: '$\theta = 9/16$', θ_6: '$\theta = 11/16$', θ_7: '$\theta = 13/16$', θ_8: '$\theta = 15/16$'

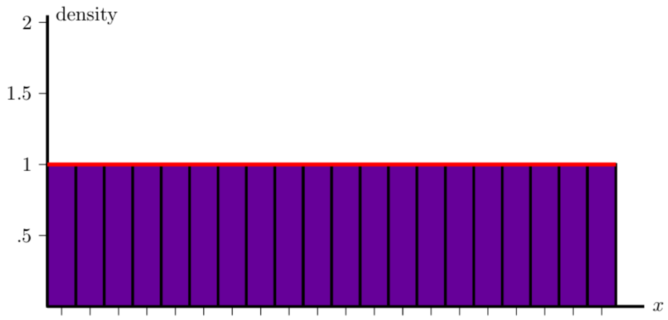
הקודמת האחידה נותנת לכל השערה הסתברות $1/8 = 1 \cdot \Delta\theta$.

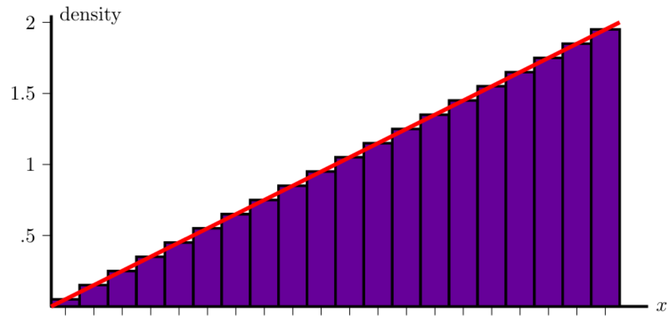
hypothesis	prior	likelihood	Bayes num.	posterior
$\theta = 1/16$	1/8	1/16	$(1/8) \times (1/16)$	1/64
$\theta = 3/16$	1/8	3/16	$(1/8) \times (3/16)$	3/64
$\theta = 5/16$	1/8	5/16	$(1/8) \times (5/16)$	5/64
$\theta = 7/16$	1/8	7/16	$(1/8) \times (7/16)$	7/64
$\theta = 9/16$	1/8	9/16	$(1/8) \times (9/16)$	9/64
$\theta = 11/16$	1/8	11/16	$(1/8) \times (11/16)$	11/64
$\theta = 13/16$	1/8	13/16	$(1/8) \times (13/16)$	13/64
$\theta = 15/16$	1/8	15/16	$(1/8) \times (15/16)$	15/64
Total	1	–	$\sum_{i=1}^{n} \theta_i \, \Delta\theta$	1





ועבור 20, נראה רק את ההיסטוגרמות:





לסיכום:

	hypothesis	prior	likelihood	Bayes numerator	posterior
	$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
Discrete θ:	θ	$p(\theta)$	$p(x\|\theta)$	$p(x\|\theta)p(\theta)$	$p(\theta\|x)$
Continuous θ:	θ	$f(\theta)\,d\theta$	$p(x\|\theta)$	$p(x\|\theta)f(\theta)\,d\theta$	$f(\theta\|x)\,d\theta$

הרצאה מצגת שיעור 5א, מצגת שיעור 5ב, מצגת שיעור 5ג,

1 עדכון בייסיאני (מצגת 5א, שקופית 11 והלאה)

תרגיל כיתה: קוביות

יש 5 קוביות: עם 4, 6, 8, 12, ו-20 פאות. נבחר אחת באקראי ובלי להסתכל נזרוק. נגיד שיצא 13. טבלת ה-*likelihood*:

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	1/5	0	0	0
$\mathcal{H}_6$	1/5	0	0	0
$\mathcal{H}_8$	1/5	0	0	0
$\mathcal{H}_{12}$	1/5	0	0	0
$\mathcal{H}_{20}$	1/5	1/20	1/100	1
total	1		1/100	1

ואם יצא 5:

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	1/5	0	0	0
$\mathcal{H}_6$	1/5	1/6	1/30	0.392
$\mathcal{H}_8$	1/5	1/8	1/40	0.294
$\mathcal{H}_{12}$	1/5	1/12	1/60	0.196
$\mathcal{H}_{20}$	1/5	1/20	1/100	0.118
total	1		0.085	1

או 9:

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	1/5	0	0	0
$\mathcal{H}_6$	1/5	0	0	0
$\mathcal{H}_8$	1/5	0	0	0
$\mathcal{H}_{12}$	1/5	1/12	1/60	0.625
$\mathcal{H}_{20}$	1/5	1/20	1/100	0.375
total	1		.0267	1

עדכונים חוזרים

נניח שזרקנו 5 ואז 9. יהיו שני עדכונים:

$$Bayes\ numerator_1 = \\ = likelihood_1 \times prior$$

$$Bayes\ numerator_2 = \\ = likelihood_2 \times Bayes\ numertor_1$$

hyp.	prior	likel. 1	Bayes num. 1	likel. 2	Bayes num. 2	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1\mid\mathcal{H})$	$***$	$P(\mathcal{D}_2\mid\mathcal{H})$	$***$	$P(\mathcal{H}\mid\mathcal{D}_1,\mathcal{D}_2)$
$\mathcal{H}_4$	1/5	0	0	0	0	0
$\mathcal{H}_6$	1/5	1/6	1/30	0	0	0
$\mathcal{H}_8$	1/5	1/8	1/40	0	0	0
$\mathcal{H}_{12}$	1/5	1/12	1/60	1/12	1/720	0.735
$\mathcal{H}_{20}$	1/5	1/20	1/100	1/20	1/2000	0.265
total	1				0.0019	1

ואם זרקנו 9 ואז 5:

ההסתברות בסוף יוצאת אותו דבר.

hyp.	prior	likel. 1	Bayes num. 1	likel. 2	Bayes num. 2	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1\mid\mathcal{H})$	$***$	$P(\mathcal{D}_2\mid\mathcal{H})$	$***$	$P(\mathcal{H}\mid\mathcal{D}_1,\mathcal{D}_2)$
$\mathcal{H}_4$	1/5	0	0	0	0	0
$\mathcal{H}_6$	1/5	0	0	1/6	0	0
$\mathcal{H}_8$	1/5	0	0	1/8	0	0
$\mathcal{H}_{12}$	1/5	1/12	1/60	1/12	1/720	0.735
$\mathcal{H}_{20}$	1/5	1/20	1/100	1/20	1/2000	0.265
total	1				0.0019	1

אפשר גם לבצע את החישוב בשלב אחד:

כאן $\mathcal{D} =$ \`$rolled\ 9\ then\ 5$\`

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})$	$P(\mathcal{D}\mid\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\mid\mathcal{D})$
$\mathcal{H}_4$	1/5	0	0	0
$\mathcal{H}_6$	1/5	0	0	0
$\mathcal{H}_8$	1/5	0	0	0
$\mathcal{H}_{12}$	1/5	1/144	1/720	0.735
$\mathcal{H}_{20}$	1/5	1/400	1/2000	0.265
total	1		0.0019	1

תרגיל כיתה: חיזוי הסתברותי

מצאנו הסתברויות אוחרות של השערות. נרצה גם לבצע חיזוי אוחר (*posterior predictions*) לגבי הזריקה הבאה.

לדוגמה, נמצא את $P(D_2 = 4|D_1 = 5)$, $P(D_1 = 5)$:

hyp.	prior	likel. 1	Bayes num. 1	post. 1	likel. 2	post. 1 × likel. 2
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}_1\|\mathcal{H})$	$***$	$P(\mathcal{H}\|\mathcal{D}_1)$	$P(\mathcal{D}_2\|\mathcal{H},\mathcal{D}_1)$	$P(\mathcal{D}_2\|\mathcal{H},\mathcal{D}_1)P(\mathcal{H}\|\mathcal{D}_1)$
$\mathcal{H}_4$	1/5	0	0	0	$*$	0
$\mathcal{H}_6$	1/5	1/6	1/30	0.392	1/6	$0.392 \cdot 1/6$
$\mathcal{H}_8$	1/5	1/8	1/40	0.294	1/8	$0.294 \cdot 1/40$
$\mathcal{H}_{12}$	1/5	1/12	1/60	0.196	1/12	$0.196 \cdot 1/12$
$\mathcal{H}_{20}$	1/5	1/20	1/100	0.118	1/20	$0.118 \cdot 1/20$
total	1		0.085	1		0.124

לפי נוסחת ההסתברות השלמה, $P(D_1)$ הוא הסכום של עמודת $Bayes\ numerator\ 1$. $P(D_1) = 0.085$. $P(D_2|D_1)$ הוא הסכום של העמודה האחרונה: $P(D_2|D_1) = 0.124$.

2 חיזוי וסיכויים (מצגת 5ב)

דוגמה: 3 סוגי מטבעות

נניח שיש 3 סוגי מטבעות, כל אחד עם הסתברות שונה לעץ. סוג A הוגן, סוג B עם הסתברות 0.6, סוג C עם הסתברות 0.9. נניח שיש קופסה עם מטבע אחד מכל סוג. נניח שבחרנו מטבע אחד באופן מקרי ואחיד.

הסתברות נחזית קודמת (prior predictive probability): לפני שנאסוף דאטא, מה ההסתברות שייצא עץ?

איסוף דאטא: נניח שההטלה הראשונה יצאה עץ.

הסתברות נחזית אוחרת (posterior predictive probability): מה ההסתברות שההטלה הבאה תהיה עץ?

לפי נוסחת ההסתברות השלמה:

$$P(D_{1H}) = P(D_{1H}|A)P(A) + P(D_{1H}|B)P(B) + P(D_{1H}|C)P(C) = 0.5 \cdot \frac{1}{3} + 0.6 \cdot \frac{1}{3} + 0.9 \cdot \frac{1}{3} = 0.6667$$

נניח שקיבלנו את הדאטא D_{1H}. נעדכן את ההסתברויות לכל סוג מטבע:

hypothesis	prior	likelihood	Bayes numerator	posterior
H	$P(H)$	$P(D_{1,H}\|H)$	$P(D_{1,H}\|H)P(H)$	$P(H\|D_{1,H})$
A	1/3	0.5	0.1667	0.25
B	1/3	0.6	0.2	0.3
C	1/3	0.9	0.3	0.45
total	1		0.6667	1

$$\begin{aligned} P(D_{2H}|D_{1H}) &= P(D_{2H}|A)P(A|D_{1H}) \\ &+ P(D_{2H}|B)P(B|D_{1H}) \\ &+ P(D_{2H}|C)P(C|D_{1H}) = 0.71 \end{aligned}$$

עכשיו, נניח שבחרנו מטבע אחד, הטלנו 5 פעמים, וקיבלנו עץ 1. מה ההסתברות לעץ בהטלה הבאה?

hypothesis	prior	likelihood	Bayes numerator	posterior
H	$P(H)$	$P(D\|H)$	$P(D\|H)P(H)$	$P(H\|D)$
A	1/3	$\binom{5}{1}0.5^5$	0.0521	0.669
B	1/3	$\binom{5}{1}0.6 \cdot .4^4$	0.0256	0.329
C	1/3	$\binom{5}{1}0.9 \cdot .1^4$	0.00015	0.002
total	1		0.0778	1

$$\begin{aligned} P(heads|D) &= (0.669 \cdot 0.5) + (0.329 \cdot 0.6) \\ &+ (0.002 \cdot 0.9) = 0.53366 \end{aligned}$$

נשים לב: הסדר שבו מתקבלים עץ אחד ו-4 פלי, לא משפיע על ההתפלגות האוחרת של סוג המטבע, או על ההתפלגות הנחזית האוחרת של תוצאת ההטלה הבאה.

סיכויים

אם מאורע הוא נדיר, אז $P(E) \approx O(E)$

תרגיל כיתה

מחלה נפוצה ב 0.005 מהאוכלוסייה. יש בדיקה עם false positive 0.05%, false negative 0.02%.

הסיכוי הקודם שיש לאדם מסויים את המחלה, הוא:

$$O(H_+) = \frac{P(H_+)}{P(H_-)} = \frac{0.005}{0.995} = 0.00503$$

נניח שהאדם נבדק ויצא חיובי. טבלת הlikelihoods היא:

	$\mathcal{T}_+$	$\mathcal{T}_-$
$\mathcal{H}_+$	0.98	0.02
$\mathcal{H}_-$	0.05	0.95

מקדם בייס של הדאטא הוא:

$$\text{Bayes factor} = \text{Ratio of likelihoods} = \frac{P(T_+|H_+)}{P(T_+|H_-)} = \frac{0.98}{0.05} = 19.6$$

הסיכוי האוחר הוא:

$$\text{Posterior odds} = \text{Bayes factor} \times \text{Prior odds} = 19.6 \cdot 0.00504 = 0.0985$$

מקדם בייס 19.6 מהווה עדות חזקה לכך שלאדם יש את המחלה. אבל, הסיכוי האוחר עדיין קטן כי הסיכוי הקודם מאד קטן.

אפשר לחשב את הסיכוי האוחר על ידי טבלת עדכון בייסיאנית:

$$O(H_+|T_+) = \frac{0.0897}{0.9103} = 0.0985$$

hypothesis	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{T}_+|\mathcal{H})$	$P(\mathcal{T}_+|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}|\mathcal{T}_+)$
$\mathcal{H}_+$	0.005	0.98	0.00490	0.0897
$\mathcal{H}_-$	0.995	0.05	0.04975	0.9103
total	1		0.05474	1

3 עדכון בייסיאני: קודמים רציפים (מצגת 5ג)

בניית האינטואיציה

יש 3 סוגי מטבעות עם הסתברות 1/4, 1/2, 3/4 לעץ. נניח שהיחס בין הסוגים הוא 1 ל-2 ל-1. נניח שבחרנו מטבע, הטלנו פעמיים וקיבלנו TT. נחשב את ההסתברות האוחרת שלמצטבע יש סיכוי 1/4:

לפי נוסחת בייס וההסתברות השלמה:

$$P(0.25|D) = \frac{P(D|0.25)P(0.25)}{P(D)} = \frac{P(D|0.25)P(0.25)}{P(D|0.25)P(0.25) + P(D|0.5)P(0.5) + P(D|0.75)P(0.75)} = \frac{(0.75)^2(0.25)}{(0.75)^2(0.25) + (0.5)^2(0.5) + (0.25)^2(0.25)} = 0.5$$

בטבלת עדכון:

hypotheses	prior	likelihood	Bayes numerator	posterior
$\mathcal{H}$	$P(\mathcal{H})$	$P(\text{data}|\mathcal{H})$	$P(\text{data}|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}|\text{data})$
$C_{0.25}$	$1/4$	$(0.75)^2$	0.141	0.500
$C_{0.5}$	$1/2$	$(0.5)^2$	0.125	0.444
$C_{0.75}$	$1/4$	$(0.25)^2$	0.016	0.056
Total	1		$P(\text{data}) = 0.281$	1

ואם יש 5 סוגי מטבעות? או 9? או 99...?

hypotheses $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\|\mathcal{H})$	Bayes numerator $P(\text{data}\|\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\|\text{data})$
$C_{0.1}$	1/9	$(0.9)^2$	0.090	0.297
$C_{0.3}$	2/9	$(0.7)^2$	0.109	0.359
$C_{0.5}$	3/9	$(0.5)^2$	0.083	0.275
$C_{0.7}$	2/9	$(0.3)^2$	0.020	0.066
$C_{0.9}$	1/9	$(0.1)^2$	0.001	0.004
Total	1		$P(\text{data}) = 0.303$	1

(k מחושב כך שהסכום הוא 1)

hypotheses $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\|\mathcal{H})$	Bayes numerator $P(\text{data}\|\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\|\text{data})$
$C_{0.1}$	$k(0.1\cdot 0.9)$	$(0.9)^2$	0.0442	0.1483
$C_{0.2}$	$k(0.2\cdot 0.8)$	$(0.8)^2$	0.0621	0.2083
$C_{0.3}$	$k(0.3\cdot 0.7)$	$(0.7)^2$	0.0624	0.2093
$C_{0.4}$	$k(0.4\cdot 0.6)$	$(0.6)^2$	0.0524	0.1757
$C_{0.5}$	$k(0.5\cdot 0.5)$	$(0.5)^2$	0.0379	0.1271
$C_{0.6}$	$k(0.6\cdot 0.4)$	$(0.4)^2$	0.0233	0.0781
$C_{0.7}$	$k(0.7\cdot 0.3)$	$(0.3)^2$	0.0115	0.0384
$C_{0.8}$	$k(0.8\cdot 0.2)$	$(0.2)^2$	0.0039	0.0130
$C_{0.9}$	$k(0.9\cdot 0.1)$	$(0.1)^2$	0.0005	0.0018
Total	1		$P(\text{data}) = 0.298$	1

ברור שלא נעשה טבלה עם 99 שורות. ננסה למצוא שיטה אחרת:

יהי θ ההסתברות לעץ. $\theta = 0.01, 0.02, \ldots, 0.99$. θ גם יסמן את ההשערה שההטיה היא θ. יש לנו נוסחה להסתברות הקודמת: $p(\theta) = k\theta(1-\theta)$. ה-likelihood: $p(D|\theta) = P(TT|\theta) = (1-\theta)^2$. אז הטבלה:

hyp. $\mathcal{H}$	prior $P(\mathcal{H})$	likelihood $P(\text{data}\|\mathcal{H})$	Bayes numerator $P(\text{data}\|\mathcal{H})P(\mathcal{H})$	posterior $P(\mathcal{H}\|\text{data})$
θ	$k\theta(1-\theta)$	$(1-\theta)^2$	$k\theta(1-\theta)^3$	$0.006\cdot\theta(1-\theta)^3$
Total	1		$P(\text{data}) = 0.300$	1

אז המקרה הכללי של המטבע: יש מטבע עם הסתברות לא ידועה θ לעץ, וPDF קודמת $f(\theta) = 3\theta^2$. נוסחת ההסתברות השלמה:

$$p(x=0) = \int_0^1 p(x=0|\theta)f(\theta)\,d\theta = \int_0^1 (1-\theta)3\theta^2\,d\theta = \frac{1}{4}$$

דוגמה: ברנולי

אם $X\sim Ber(\theta)$, אז המ"מ שלנו **בדיד** (מקבל ערכים 0 או 1) אבל מרחב ההשערות הוא **רציף** (θ יכול להיות כל ערך בין 0 ל-1).

hypothesis	prior	likelihood	Bayes numerator	posterior
θ	$f(\theta)\,d\theta$	$p(x\|\theta)$	$p(x\|\theta)f(\theta)\,d\theta$	$\dfrac{p(x\|\theta)f(\theta)\,d\theta}{p(x)}$
Total	1		$p(x) = \int_0^1 p(x\|\theta)f(\theta)\,d\theta$	1

תרגיל כיתה

נתון מטבע עם הסתברות לא ידועה θ לעץ. קודמת: $f(\theta) = 2\theta$ בתחום $[0,1]$. הטלנו וקיבלנו עץ. אז הPDF האוחרת היא:

$$f(\theta|x=1) = \frac{p(x=1|\theta)f(\theta)\,d\theta}{p(x=1)} = \frac{\theta\cdot 2\theta\,d\theta}{\int_0^1 \theta\, f(\theta)\,d\theta} = \frac{2\theta^2 d\theta}{\int_0^1 2\theta^2\,d\theta} = \frac{2\theta^2}{\theta^3/3} = 3\theta^2$$

hypoth.	prior	likelihood	Bayes numerator	posterior
θ	$2\theta\,d\theta$	θ	$2\theta^2\,d\theta$	$3\theta^2\,d\theta$
Total	1		$T = \int_0^1 2\theta^2\,d\theta = 2/3$	1

ואם הטלנו שוב וקיבלנו פלי:

hypoth.	prior	likelihood	Bayes numerator	posterior
θ	$3\theta^2\,d\theta$	$1-\theta$	$3\theta^2(1-\theta),d\theta$	$12\theta^2(1-\theta)\,d\theta$
Total	1		$\int_0^1 3\theta^2(1-\theta)\,d\theta = 1/4$	1

$$f(\theta|x_2=0)=\frac{p(x_2=0|x_1=1)f(\theta)\,d\theta}{p(x_2=0)}=\frac{(1-\theta)3\theta^2 d\theta}{\int_0^1(1-\theta)3\theta^2\,d\theta}=\frac{(1-\theta)3\theta^2 d\theta}{\int_0^1 3\theta^2-3\theta^3\,d\theta}=\frac{(1-\theta)3\theta^2 d\theta}{\left[\theta^3-\frac{3\theta^4}{4}\right]_0^1}$$

$$=\frac{(1-\theta)3\theta^2 d\theta}{1/4}=12\theta^2(1-\theta)$$

תרגיל כיתה: ברנולי

מטבע עם הסתברות θ. PDF קודמת $f(\theta)=1$ בתחום $[0,1]$. קיבלנו 15 עץ ו-12 פלי. נמצא את האוחרת: לא נמצא את האינטגרל המנרמל מפורשות – נשאיר אותו כפרמטר T:

$$f(\theta|D)=\frac{p(D|\theta)f(\theta)\,d\theta}{p(D)}=\frac{\theta^{15}(1-\theta)^{12}d\theta}{\int_0^1\theta^{15}(1-\theta)^{12}\,d\theta}=\frac{\theta^{15}(1-\theta)^{12}d\theta}{T}=\frac{1}{T}\theta^{15}(1-\theta)^{12}$$

(השיטה היא באינטגרציה בחלקים, כל פעם חלק הופך להיות מקדם וזה נהיה סוג של עצרת).

קוראים לזה **התפלגות בטא.** ל- $Beta(a,b)$ יש צפיפות:

$$f(\theta)=\frac{(a+b-1)!}{(a-1)!\,(b-1)!}\theta^{a-1}(1-\theta)^{b-1}$$

המקדם הוא גורם מנרמל אז אם:

$$f(\theta)=c\cdot\theta^{a-1}(1-\theta)^{b-1}$$

היא PDF, אז:

$$c=\frac{(a+b-1)!}{(a-1)!\,(b-1)!}$$

ו-$f(\theta)$ היא הPDF של התפלגות $Beta(a,b)$.

תרגול

תרגיל 1

יש רובוט שלא יכול לזהות אותיות אבל כן מזהה צבעים. הוא צודק בצבע רק 70% מהזמן.

שלב א: ידוע שהרובוט מזהה שחור. מה ההסתברות שהרובוט מול לוח B?

H	Prior	likelihood	Bayes numerator	Posterior
H_A	0.25	0.7	7/40	7/20
H_B	0.25	0.7	7/40	7/20
H_C	0.25	0.3	3/4	3/20
H_D	0.25	0.3	3/4	3/20
			1/2	

שלב ב: הרובוט עושה רבע סיבוב עם כיוון השעון. מה ההסתברות שהוא יראה ירוק?

הposterior הופך לprior: נשים לב שהיה רבע סיבוב.

H	prior	likelihood	Bayes numerator	Posterior
H_A	3/20	0.3	9/200	
H_B	7/20	0.3	21/200	
H_C	7/20	0.7	49/200	
H_D	3/20	0.7	21/200	
			1/2	

בעצם לא קיבלנו דאטא חדשה, אז ההסתברות נשארת אותו דבר.

תרגיל 2

ניזכר בנוסחת בייס למשתנה רציף, ונציג את התפלגות $Beta$:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int_{-\infty}^{\infty} p(x|\theta)P(\theta)\, d\theta} \propto \mathcal{L}(\theta)p(\theta), \qquad Beta(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$$

יהיו $X_1, \dots, X_n \sim Ber(\theta)$, ונניח ש $\theta \sim Beta(\alpha,\beta)$

$$p(\theta|x) \propto \mathcal{L}(\theta)P(\theta) \propto$$

עבור s הוצאות:

$$\propto \theta^s(1-\theta)^{n-s}\theta^{\alpha-1}(1-\theta)^{\beta-1} = C_n \cdot \theta^{s+\alpha-1} \cdot (1-\theta)^{n-s+\beta-1} \sim Beta(s+\alpha, n-s+\beta)$$

תרגיל 3

יהיו $X_1, \dots, X_n \sim U(0,\theta)$, ונניח $p(\theta) \propto 1/\theta$.

$$p(\theta|x) \propto \mathcal{L}(\theta)p(\theta) \ \propto \ \frac{1}{\theta}\prod_{i=1}^{n}\frac{1}{\theta}\cdot I\{x_i \leq \theta\} = \frac{1}{\theta}\cdot\frac{1}{\theta^n}\cdot I\{\max\{x_1, \dots, x_n\} \leq \theta\}$$

$$1 = \int_{-\infty}^{\infty} p(\theta)\, d\theta = \int_{m}^{\alpha} p(\theta)\, d\theta = \int_{m}^{\alpha} C\,\frac{1}{\theta^{n+1}}\, d\theta = \frac{C}{m^n n}$$

$$p(\theta) \sim \begin{cases} 0, & \theta < m \\ \dfrac{n \cdot m^2}{\theta^{n+1}}, & else \end{cases}$$