INF 367 A V20 EXERCISE 4

NAPHAT AMUNDSEN

1. Know: $P(X=t|C=C_1)=\theta_1$, $P(X=t|C=C_2)=\theta_2$, $X=\{h,t\}$, $C=\{C_1,C_2\}$,
$P(C=C_1)=\pi$, t= tails, h=heads, X and C are binary vars.

1a) $$P(C=C_1|X=h)=\frac{\pi(1-\theta_1)}{\pi(1-\theta_1)+(1-\pi)(1-\theta_2)}$$

$$P(C=C_1|X=t)=\frac{\pi\,\theta_1}{\pi\theta_1+(1-\pi)\theta_2}$$

1b) Let $P(C_1)=P(C=C_1)=\pi$

independent given C_1

$$P(C_1|x_N,x_{N-1},\dots,x_1)=\frac{P(x_N,x_{N-1},\dots,x_1|C_1)\,P(C=C_1)}{P(x_N,x_{N-1},\dots,x_1)}$$ ← marginalizing away C

$$=\frac{P(x_N|C_1)P(x_{N-1}|C_1)\dots P(x_1|C_1)P(C_1)}{P(x_N,x_{N-1},\dots)}=\frac{P(C_1)\prod_i^{n_t}P(x_i=t|C_1)\prod_j^{n_h}P(x_j=h|C_1)}{\sum_C P(C)\prod_i^{n_t}P(x_i=t|C)\prod_j^{n_h}P(x_j=h|C)}$$

$$=\frac{P(C_1)P(x=t|C_1)^{n_t}P(x=h|C_1)^{n_h}}{P(C_1)P(x=t|C_1)^{n_t}P(x=h|C_1)^{n_h}+P(C_2)P(x=t|C_2)^{n_t}P(x=h|C_2)^{n_h}}$$

$$=\frac{\pi\,\theta_1^{n_t}(1-\theta_1)^{n_h}}{\pi\,\theta_1^{n_t}(1-\theta_1)^{n_t}+(1-\pi)\,\theta_2^{n_t}(1-\theta_2)^{n_h}}$$

2a) Choose $P(C_1) = \pi = 0.69 \quad P(C_2) = \pi^c = 1 - 0.69 = 0.31$

2b) $P(C=C_1 | \underbrace{htthh}_{D_1}) = \frac{0.69 \cdot 0.7^2 \cdot (1-0.7)^3}{0.69 \cdot 0.7^2 \cdot (1-0.7)^3 + 0.31 \cdot 0.4^2 \cdot (1-0.4)^3}$

$= 0.46$

$P(C=C_2 | htthh) = 1 - 0.46 = 0.54$

2c) $P(C | D_1, D_2) = \frac{P(D_2|C) P(C|D_1)}{P(D_2|D_1)} = \frac{P(D_2|C) P(C|D_1)}{\sum_C P(D_2|C) P(C|D_1)} = \frac{P(D_2|C) P(C|D_1)}{P(D_2|C=C_1) P(\underbrace{C=C_1}_{C=C_1}|D_1) + P(D_2|C=C_2) P(C=C_2|D_1)}$

$P(x=t|C_1) = \theta_1 = 0.7 \quad P(D_2|C=C_1) = P(ttht|C=C_1) = 0.7^4 (1-0.7)^1 = 0.07203$

$P(x=t|C_2) = \theta_2 = 0.4 \quad P(D_2|C=C_2) = P(ttht|C=C_2) = 0.4^4 (1-0.4)^1 = 0.01536$

$P(C=C_1 | D_1, D_2) = \frac{0.07203 \cdot 0.46}{0.07203 \cdot 0.46 + 0.01536 \cdot 0.54} = 0.7998$

$P(C=C_2 | D_1, D_2) = 1 - 0.7998 = 0.2002$

2d) $D = D_1 + D_2 = htthh\ ttht$

$P(C=C_1 | D) = \frac{0.69 \cdot 0.7^6 \cdot (1-0.7)^4}{0.69 \cdot 0.7^6 \cdot (1-0.4)^4 + 0.31 \cdot 0.4^6 \cdot (1-0.4)^4} = 0.7998$

They are the same

3) "Today's posterior is tomorrows prior" - Pekka Parviainen 2020

Use $P(C|D_1)$ as prior:

$P_{D_1}(x=t) = P(C=C_1|D_1) P(x=t|C_1) + P(C=C_2|D_1) P(x=t|C_2)$

$= 0.46 \cdot 0.7 + 0.54 \cdot 0.4 = 0.538$

4 Let A = Alice, B = Bob

$P(C=C_1|A) = \pi_A = 0.99$

$P(C=C_1|B) = \pi_B = 0.01$

My hypothesis: Their posteriors will converge to the same value when increasing amount of data

4a) $D_3 = \{3 \text{ tails}, 7 \text{ heads}\}$

$$P_A(C=C_1|D_3) = \frac{0.99 \cdot 0.7^3 \cdot 0.3^7}{0.99 \cdot 0.7^3 \cdot 0.3^7 + 0.01 \cdot 0.4^3 \cdot 0.6^7} = \underline{\underline{0.8056}}$$

$$P_B(C=C_1|D_3) = \frac{0.01 \cdot 0.7^3 \cdot 0.3^7}{0.01 \cdot 0.7^3 \cdot 0.3^7 + 0.99 \cdot 0.4^3 \cdot 0.6^7} = \underline{\underline{0.0004}}$$

4b) $D_4 = \{36 \text{ tails}, 64 \text{ heads}\}$

$$P_A(C=C_1|D_4) = \frac{0.99 \cdot 0.7^{36} \cdot 0.3^{64}}{0.99 \cdot 0.7^{36} \cdot 0.3^{64} + 0.01 \cdot 0.4^{36} \cdot 0.6^{64}} = \underline{\underline{3.014 \cdot 10^{-9} \approx 0}}$$

$$P_B(C=C_1|D_4) = \frac{0.01 \cdot 0.7^{36} \cdot 0.3^{64}}{0.01 \cdot 0.7^{36} \cdot 0.3^{64} + 0.99 + 0.4^{36} \cdot 0.6^{64}} = \underline{\underline{3.074 \cdot 10^{-13} \approx 0}}$$

4c) Their posteriors converged to be approximately the same. <u>My hypothesis was correct.</u>